CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Tropical AgriSciences

Genetic diversity of indigenous fruit tree species in West African region

BACHELOR'S THESIS

Prague 2022

Author: Alžběta Kubíková

Supervisor: prof. Ing. Bohdan Lojka, Ph.D.

Co-supervisor: Ing. Anna Maňourová, Ing. Marie Kalousová

Declaration

I hereby declare that I have done this thesis entitled Genetic diversity of indigenous fruit tree species in West African region independently, all texts in this thesis are original, and all the sources have been quoted and acknowledged by means of complete references and according to Citation rules of the FTA.

In 15.4.2022

.....

Alžběta Kubíková

Acknowledgements

I would like to thank prof. Ing. Bohdan Lojka Ph.D. for his trust and help during processing this thesis and his patience with my stubbornness. I would like to thank Ing. Anna Maňourová for her help during processing this thesis, her kind approach while working in the lab and her friendly nature which made every bad day a little more bearable. I would like to thank Ing. Marie Kalousová for her help in the lab, for processing data analyses and for a solution to a small but significant problem that occurred during the writing of this bachelor thesis. I would also like to thank myself for my strength to start writing the thesis and also, with the help of all of the above, to finish it, and last but not least, for my courage to start fulfilling my dreams as a woman in science.

Abstract

Genetic diversity is one of the most important steps in domestication strategies. By a board literature review, this thesis summarized information on the genetic diversity of six selected species *Allanblackia* spp., *Cola* spp., *Dacryodes edulis, Garcinia kola, Irvingia spp*. and *Ricinodendron heudelotii* native to West Africa. To practically examinate the knowledge of genetic diversity *G. kola* was evaluated by AFLP markers. Based on the literature review, low genetic diversity was detected in *D. edulis*, *G. kola* and *R. heudelotii*, whereases high diversity occurred in the other species. Additionally, genetic diversity research methods used in the studies seem to be outdated, therefore the results should be supplemented by newer methods such as sequencing. Followingly, the low genetic diversity of *G. kola* was also confirmed by our laboratory experiment. The genetic diversity of selected species is not sufficient because not much research about their genetic diversity has been done. It is necessary to conserve and protect the genetic diversity of indigenous species for future research and domestication because due to environmental changes these genetic resources may be lost.

Key words: agroforestry, neglected species, molecular markers, domestication, *Garcinia kola*

Table of content

List of tables

List of figures

List of abbreviations

AFLP	–	Amplified Fragment Length Polymorphism
CTAB	–	Cetyl trimethylammonium bromide
ddH_2O	–	Double distilled water
DNA	–	Deoxyribonucleic acid
dNTPs	–	Deoxynucleotide triphosphate
ddNTPs	–	Dideoxynucleotide triphosphate
DS	–	Domestication syndrome
*Eco*RI-a	–	*Eco*RI-adaptor
EtBr	–	Ethidium Bromide
H_2O	–	Water, dihydrogen oxide
IAA	–	Isoamyl alcohol
ICRAF	–	International Centre of Research in Agroforestry
ISSR	–	Inter Simple Sequence Repeat
IUCN	–	International Union for Conservation of Nature
MseI-a	–	*MseI*-adaptor
PCR	–	Polymerase chain reaction
PIC	–	Polymorphic Information Content
RAPD	–	Random amplified polymorphic DNA
RFLP	–	Restriction fragment length polymorphism
R-L	–	Restriction-ligation
RNA	–	Ribonucleic acid
S.E.	–	Standard error
SNPs	–	Single Nucleotide Polymorphisms
SSR	–	Singled Sequence Repeat

1. Introduction

Underutilized tree species are little explored but extremely important all around the world. Their significance lies in their benefit to humankind. Exceedingly important are mainly in developing countries where they play part in many areas for local food security and wellbeing. The trees provide shade for crops and produce resins, fibres, or dyes (Waruhiu et al. 2004), but their most important benefit is in poverty due to their high nutrition value, provision of food and medicine as well as additional income in rural poor communities. Increasing incomes may also help to achieve other development goals, such as improving health care and education and protecting the environment. These species are also part of traditional diet and culture. Knowledge of the management and usage of these species exists thanks to indigenous people (Schreckenberg et al. 2006). Many of these species also have health benefits and are locally used as medicine. There is a big potential for some of these species to be domesticated but because there is only little known about them there is a lack of information for proper domestication.

The first step of successful domestication is a description of genetic diversity (Rao & Hodgkin 2002). Genetic diversity is crucial in many ways such as evaluation of disturbance, primary productivity and population recovery (Hughes et al. 2008) or helping to understand the evolutionary process (Ellegren & Galtier 2016) and adaptive potential (Holderegger et al. 2006) of populations. A wide genetic base that can adapt to changes in the environment or to farmers' requirements is necessary to maintain the value of on-farm tree stands, their conservation and long-term viability. The wide genetic base also increases the populations' ability to withstand possible inbreeding depression through future generations of farmer propagation (Hollingsworth et al. 2005).

Therefore, this thesis focused on summarizing available information on the genetic diversity of selected species: *Allanblackia* spp., *Irvingia* spp., *Cola* spp. *Darcyodes edulit*, *Garcinia kola* and *Ricinodendron heudelotii* which are important neglected fruit tree species in West and Central Africa and were recommended by ICRAF for domestication programmes (Franzel & Kindt 2012).

2. Literature review

2.1 Tree domestication

Plant domestication is an evolutionary process initiated by a human. Humankind uses plant species (or also animal species) and changes them in morphological and physiological ways (Purugganan & Fuller 2009). The process consists of several steps; selection, breeding and adaptation to the quality of the product during cultivation (Dawson et al. 2012). There are more than 2,500 plant species in different stages of domestication (Clement et al. 2010; Meyer et al. 2012; Purugganan 2019).

There is a general consensus that domestication represents the culmination of a process that includes the exploitation of wild plants, the cultivation of wild plants that have not yet acquired genetic differentiation from wild plants, and finally, a human selection that results in differences in morphology and genetics that distinguish domestication from its wild progenitor (Pickersgill 2007). These changes can be identified as domestication syndrome (Ekué et al. 2011).

Domestication syndrome (DS) distinguishes a domesticated plant from its wild ancestor. DS includes several differences such as fruit size, taste, nutrition qualities, change in reproductive strategy or change in secondary metabolites. Changes in the content of secondary metabolites are also the most common DS trait (66 %) (Pickersgill 2007; Meyer et al. 2012). It can evolve for many generations until the desirable traits become fixed in the crop genome (Fuller 2007; Meyer et al. 2012). Domestication traits arise from a selection. It can reduce fitness or can cause the inability to survive in wild nature. It is because of these differences that domesticated plants are less likely to survive in the wild, thus they depend on humans for growth and reproduction (Pickersgill 2007). Moving plants from wild nature to cultivating environment changes selection pressures which leads to a decreased fitness of plants in subsequent generations. Changes in selection pressures can be also caused by cultivating management as planting or harvesting techniques. In conscious selection, desirable phenotypes are selected, and less desirable phenotypes are displaced or removed (Meyer et al. 2012). This can have a negative effect on the genetic diversity of a species (Wiehle et al. 2014). Typical characterization of many wild plants species is that those

have more than one reproductive strategy such as sexual breeding system and asexual strategies, but only one strategy is usually exploited as a propagation method under cultivation for given species (Meyer et al. 2012).

There are two types of domestication: landscape domestication and plant population domestication. Plant population domestication consists of genetic techniques, historical ecology, and landscape management (cultivation). It is a co-evolution process based on a selection of plants that leads to genotype and phenotype changes in subsequent populations and makes them more useful for human use, cultivation management and landscape (Clement et al. 2010). Landscape domestication does not include terms like evolution, selection, or genetic diversity, it is defined as a process of intervention and manipulation resulting in a change in landscape ecology and plant population demographics. It leads to a more productive and affable landscape for humans (Smith 2014).

There are two main tree domestication strategies: farmer-driven and market-driven. A farmer-driven strategy is based on farmers' preferences (Leakey & Asaah 2013). They choose plants they want to cultivate by themselves or through programs with the goal of genetic improvement (Leakey et al. 2017). Farmers mostly select familiar and locally marketed indigenous fruit and nut species as top of their priority. It is happening because traditionally important products are not available in abundance in the wild nature, and it is important to domesticate them because of their high cultural and nutritional value (Leakey & Asaah 2013). According to Leakey et al. (2017) research, market-driven domestication is based on these strategies: (1) priorities for domestication and the need of the farmers; (2) sustainable production of agroforestry tree products; (3) reducing deforestation and restoring degraded land; and (4) wise conservation and use of genetic resources.

By increasing knowledge, the participatory programs help farmers to develop skills to found village nurseries and apply basic approaches to its management. The participatory method has been developed to encourage farmers to reduce their poverty level, hunger and malnutrition, and to support food and nutritional security in the communities (Leakey & Asaah 2013). Adoption of participatory domestication has led to improvements in income, diet, and rural business development, enabling producers to

diversify their farms, make them more resilient, increase their productivity and had a positive impact on the social wellbeing of local communities (Ofori et al. 2014). The basis of our understanding of domestication processes comes from a small group of well-studied crops, including major economic crops and model crops (crops with a genome analysed and suitable for transformation) such as the Poaceae family (Meyer et al. 2012).

Many species have been domesticated and are already part of the global trade market but there is a neglected group of species still waiting to be fully discovered and described. The domestication of these trees is supported by ICRAF, which has also recommended the domestication of trees from West and Central Africa, which also includes selected species on which this thesis focuses (Franzel & Kindt 2012).

2.2 Indigenous fruit tree species in Africa

There are not many domesticated indigenous fruit trees in Africa and those that are domesticated are often overlooked in national or international poverty reduction strategies though they provide a source of income for many farmers (Schreckenberg et al. 2006). In tropical countries, there are a lot of indigenous fruit tree species that still have the potential to be domesticated and thus to increase the economic and livelihood level of local smallholder farmers. Many of these trees are a valuable source of nutrition with many health benefits (Leakey & Schreckenberg 2003). In Sub-Saharan African countries indigenous species play a crucial role in human diets. Whereas staple crops fail, indigenous trees bear fruits even during drought because they are well adapted to the local environment (Ræbild et al. 2011; Stadlmayr et al. 2013). Lots of these trees provide oil for cooking or pharmaceutic utilization. Some trees are used for their health benefit or timber products such as *Garcinia kola* or *Allanblackia* spp. (Maňourová et al. 2019; Kyereh et al. 2021).

2.3 Genetic improvement of indigenous fruit trees

Compared to annual crops and temperate fruit trees, tropical fruit trees have made limited progress in genetic improvement. Breeding programs for tropical fruit

trees require considerable resources to create new cultivars that can be shipped and stored in a modern manner (Arias et al. 2012). To domesticate these species, we have to deal with low knowledge about genetic diversity (Leakey & Simons 1998). Despite farmers being interested in indigenous fruit species and their nutritional as well as medical benefits, these species do not receive attention from foresters or agriculturalists. The reason is probably the lack of success of many planting programmes in the second half of the last century. However, many of these species may have a local and international impact on market income and a number of them meet the needs of farmers (Schreckenberg et al. 2006).

To solve these problems and start new domesticating programmes is necessary to explore and describe the genetic diversity of these species. Nowadays we know that genetic diversity plays a fundamental role and is widely used in domesticating programmes.

2.3.1 The importance of genetic diversity

The World Conservation Union (IUCN) recognizes genetic diversity as one of three types of biodiversity that deserves protection (Reed & Frankham 2003). It is essential to analyse genetic diversity in any tree domestication programme in order to identify superior genotypes and to properly deploy clones (Kyereh et al. 2021). Variations in genetic polymorphism among species and within genomes have important implications for the evolution of the species and its conservation (Ellegren & Galtier 2016). There has been a huge loss of genetic diversity during the “Green revolution”. Even though we cannot deny the success of this process in Asia, but spreading and planting monocultures of dwarf hybrids of wheat or rice led to the loss of genetic diversity and extinction of adaptive and primitive genes (Govindaraj et al. 2015).

Genetic diversity is important for the adaptability of populations and affects their fitness. In order to make predictions about a population's genetic diversity, it is necessary to study how genetic diversity is related to fitness (Booy et al. 2000). Low heterozygosity levels are directly associated with a reduction in population fitness as a result of inbreeding depression, so genetic diversity is necessary for populations to adapt to environmental changes (Reed & Frankham 2003) and (a)biotic stresses.

Adaptation to global changes such as new climates, pests or diseases, or pollution depends on the diversity of genetic material in species, but this survival mechanism is critically threatened by the effects of anthropogenic disturbance such as climate change, land-use changes, overexploitation or deforestation (Rimlinger et al. 2021).

The problem with modern crops is that they were developed for high yield but only under favourable growing conditions. These plants are therefore not very suitable for small farmers in developing countries practising low input agriculture in an environment with a number of constraints (Govindaraj et al. 2015).

2.3.2 Why do we need to study genetic diversity?

By monitoring and analysing genetic diversity through the probability of gene origin we can discover how populations have been managed in the past and can be used to plan future management to prevent the loss of diversity (Sargolzaei et al. 2006). The use of population-level molecular analyses can clarify the demographic histories of the domestication process itself, which, in conjunction with expanded archaeological studies, can help shed light on the origin of crops (Clement et al. 2010; Smýkal et al. 2018).

Genetic diversity as well as the specific distribution of genetic variation within species is crucial for producing suitable conservation and utilization strategies (Ude et al. 2006). Tree breeding and improvement is usually the prerogative of research institutes. Nowadays domestication is based on the wild gene pool, and it is up to the scientists to understand and develop the potential of the species while preserving the diversity of the genetic resource (Leakey & Schreckenberg 2003). Genetic resources are stored in the platforms such as DNA libraries or gene banks and are used for crop improvement because of global challenges like food and nutritional security (Govindaraj et al. 2015). Genomic analysis of tropical fruit trees has been largely limited to the development of dominant PCR markers for the analysis of germplasm diversity and clonal fingerprinting (Arias et al. 2012).

2.4 Genetic diversity of selected fruit tree species from sub-Saharan Africa

2.4.*1 Allanblackia* spp.

It is a dioecious fruit tree from the Clusiaceae family. The main product is its fruits where the seed is mostly used. (Russell et al. 2009). *Allanblackia* can be also used as medicine and timber species (Ofori et al. 2013). The species has a great potential to improve nutrition, food security and can be an alternative income source for the rural population (Kyereh et al. 2021). The most common utilization of the seed is oil production. These seeds contain oleic and stearic acids which reduce plasma cholesterol and help prevent heart attacks (Atangana et al. 2010). *Allanblackia* oil can be used for margarine production, cooking, soap making or for ointment manufacture.

There are three *Allanblackia* species that have significant economic importance, *Allanblackia parviflora*, *Allanblackia floribunda* and *Allanblackia stuhlmanii* (Kyereh et al. 2021). The first one, *A. parviflora* is an evergreen tree and grows to about 40 m in height (Peprah et al. 2009). Based on morphological and molecular differences there is significant genetic diversity within the species (Kyereh et al. 2021). In a study conducted by Peprah et al. (2009), there have been demonstrated some differences in fruit parameters but the results were not significant and the variation is among individual trees and not between populations or ecological zones. The significant variety was only observed among fruit volume where the average volume was 1,545 cm^3 (528-2,793 cm^3). Trees with higher fruit volume have high heritability (h = 0,822) and also genetic gain (G = 20.1 %).

Another significant result was a positive correlation found between seed weight and fruit volume. To ensure the availability of *A. parviflora* genetic resources, there was established a gene bank in 2007 at Benso village in the Western Region of Ghana to capture the adequate genetic diversity of eight different populations across three ecological zones (Ofori et al. 2016).

The high diversity was determined also by AFLP and SSR analysis of the *Allanblackia* genus in Ghana, Cameroon, and Tanzania. Its revealed differences between individual species and between countries. AFLP analysis detects higher genetic diversity

in samples from Cameroon than from Tanzania, but SSR analysis detects higher genetic diversity in Tanzania than in Cameroon. The lowest genetic diversity was in Ghana. The study also determined taxonomic misidentification during field collection (Russell et al. 2009). In summary, there is a big potential for *Allanblackia* species for domestication (Russell et al. 2009; Peprah et al. 2009; Kyereh et al. 2021).

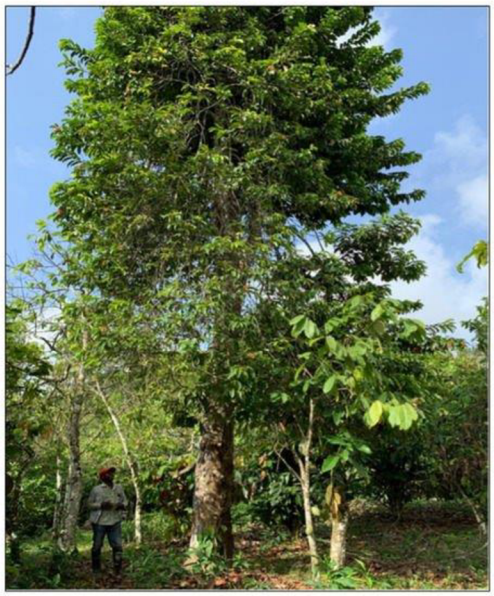

Figure 1. *A. parviflora* tree.

(source: Kyereh et al. 2021)

2.4.3 *Cola spp.*

The Cola genus plays a crucial role in many basic needs of Sub-Saharan communities. These trees belong to the Sterculiaceae family but in the broad sense they are incorporated into Malvaceae (Dah-Nouvlessounon et al. 2016). There are about 140 species of the *Cola* genus, but commonly used are mostly *Cola acuminata*, *Cola nitida* and *Cola anomala* (Onomo et al. 2006). In the study based on secondary metabolites content of *Cola* spp. conducted by Niemenak et al. (2008) there is a significant difference between *C. anomala* and the other two species. The analyses categorized *Cola* spp. into two groups. The first chemotype included *C. anomala* with high theobromine and epicatechin content. The second chemotype encompassed *C. nitida* and *C acuminita* with high caffeine and catechin contents. The results showed a high genetic distance of *C. anomala* from the other two *Cola* species. One of the factors influencing the differences might be distinct habitats. *C. anomala* is found in high altitudes, whereas

C. acuminata and *C. nitida* grow in low altitudes. In other research based on RAPD primes where genetic diversity of *C. acuminata* and *C. nitida* has been studied, was proven that *C. acuminata* has a higher variation level (71.5 %) than *C. nitida* (58.3 %) but the interpopulation differentiation index was higher at *C. nitida* (46.7 %) than *C. acuminata* (42.5 %), both *Cola* species showed high genetic variation (Akinro et al. 2019). The wide polymorphism among *Cola* spp. also supports isoenzyme variety (Onomo et al. 2006).

For *Cola nitida,* there is a significant difference in the coefficient of variation in bioactive compounds content that suggests that there are variations among *Cola* genotypes which provides a chance to select promising genotypes to optimize nutraceutical contents (Nyadanu et al. 2020). Resulting of the cutting of trees for timber and firewood production and land-use changes the population of *Cola nitida* decreased and the genetic resources are increasingly vulnerable to genetic erosion (Nyadanu et al. 2021).

Figure 2. Morphology of *Cola* trees in home gardens. a) *C. acuminata*; b) *C. nitida* *(source: Dah-Nouvlessounon et al. 2016)*

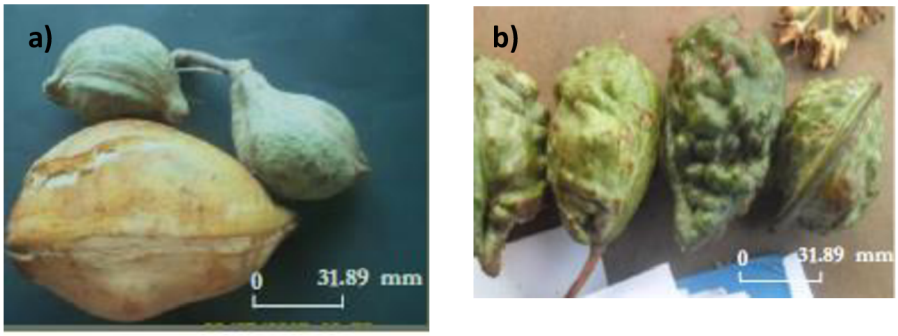

Figure 3. The fruit of *Cola* spp. a) C. acuminata b) C. nititda

(source: Dah-Nouvlessounon et al. 2016)

2.4.3 *Dacryodes edulis*

Also known as African plum or prune belongs to the *Burseraceae* family. It is a tall tree cultivated especially within cocoa/coffee agroforests (Rimlinger et al. 2019). It also provides fruit locally named safou (Ondo-Azi et al. 2009). The genetic diversity of *Dacryodes edulis* is based on fruit variations (Kengue et al. 2002).

High genetic diversity was observed also in the study of 200 trees from four populations in Cameroon. Where the differences were found in fruit width (23.3 to 53.5 mm), length (33.5 to 122.4 mm), mass (10.0 to 114.0 g) and flesh mass (12.5 to 106.0 g) and depth (0.6 to 11.1 mm). The significant differences were detected in taste, oiliness and shin and flesh colour (Waruhiu et al. 2004). Similar results were published by Anegbeh et al. (2005) in their research on 100 trees in southeast Nigeria based on fruit variation. Significant differences were shown in fruit width (21.82 to 43.75 mm), fruit length (39.0 to 95.1 mm), fruit mass (10.2 to 71.4 g), flesh mass (6.8 to 62.2 g), kernel mass (1.3 to 15.1 g) and flesh thickness (1.82 to 6.39 mm).

A study conducted in Cameroon that compares urban and rural populations based on microsatellites markers has demonstrated that the allelic richness of *Dacryodes edulis* was high (rarefied for k = 300 gene copies) and in both populations was similar despite a wide range of geographic coverage. The genetic composition between the two populations was also similar, the genetic differentiation was weak but significant (F_{ST} = 0.0057) and different from 0 (Rimlinger et al. 2021). In a study of three *D. edulis* provenance at ICRAF breeding trials in Cameroon, there was also measured low genetic diversity among populations based on the SSR marker technique. The AMOVA

analysis revealed that there existed only 2 % of genetic variation between populations. The heterozygosity was 0.022 – 0.5385 and the F_{ST} = 0.018 was low. The PCoA analysis did not reveal significant separation between populations (Makueti et al. 2015). The other research base on microsatellite analysis detected high polymorphism (PIC from 0.15 to 0.78) and heterozygosity (0.53) but the F_{ST} was low (0.012). The variation among populations was only 1 % but the variation among individuals within populations was 93 %. Wild individuals possessed 45 % of the loci with private alleles (Tchinda et al. 2016).

Figure 4. Morphology of *Dacryodes edulis*. a) mature tree; b) fruit variety

(source: Rimlinger et al. 2019)

2.4.4 *Garcinia kola*

Garcinia kola is an indigenous tree commonly found in West and Central Africa. The tree is known as bitter kola, and it is a multipurpose tree that can be used as a medicine. The most valuable product are seeds which are consumed for prevention and also for curing gastric disorders (Maňourová et al. 2019) such as the fight against indigestion, impotence, as antiinflammatory and treatment of stomach aches (Kamga et al. 2019). The bark is also used for treatment, which is peeled off along the entire length of the tree for this purpose, which can lead to the death of the tree (Yogom et al. 2020) In eastern Cameroon, these practices exploit and destroy a very high percentage of individuals (62 % of standing trees) (Kamga et al. 2019).

In a study from seven states in Nigeria, there was measured high phenotypic (98.63 %) variance at crown width and the highest heritability was a stem girth (29.98 % and shoot wet weight (17.55 %) (Azeez et al. 2020).

The study of genetic diversity in Benin implies little genetic structure of *G. kola*. Low diversity of heterozygosity per population (0.196 to 0.228), while homozygosity per population was higher (0.223 to 0.248), was found by SNP analysis using ultra-high-throughput diversity array technology (DArT). The results showed that the genetic differentiation between populations was low, while the analysis of molecular variance revealed that most of the variation was between populations (97.86 %). Additionally, the results of this study show very high levels of inbreeding (F_{IS} = 0.781 – 0.848). Based on clustering and discriminant analysis of principal components, there were two admixed clusters in population structure, indicating little genetic diversity. Based on the average PIC of 0.3, the markers appear to be moderately informative (Dadjo et al. 2020).

In contrast in the study conducted by Olawuyi & Azeez (2019) made on plants from Nigeria that used RAPD markers, there was high genetic diversity (PIC 0.93). Using six decamer primers revealed high and very close polymorphisms, indicating they are all suitable and reliable as genetic diversity markers for *G. kola.* It is possible that the differences in PIC values observed between the two marker systems are due to the bi-allelic nature of DART-SNP markers, for which the maximum value for PIC is 0.5, compared with multi-allelic RAPD markers, which have a maximum value for PIC of 1 (Dadjo et al. 2020). Genetic diversity in Central Cameroon populations by AFLP also showed low genetic diversity (Irikidzai 2021).

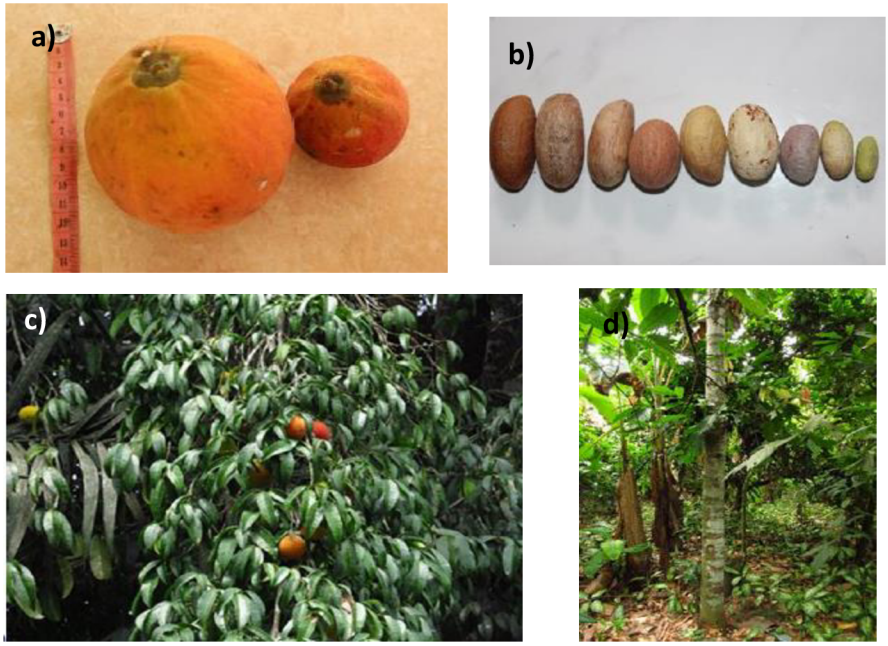

Figure 5. Morphology of *Garcinia kola*. a) ripe fruits; b) seeds at varying stages of Maturity; c) tree crown; d) tree in the home garden
(source: Irikidzai 2021)

2.4.2 *Irvingia spp.*

An economically important tree from this genus is the *Irvingia gabonensis* from the Irvingiaceae family originating in the rainforests of West and Central Africa. This fruit is known as sweet bush mango or Dika nut and is used as a snack or fresh fruit and also for soup thickener or thickening agents. There is also a bitter bush mango, *Irvingia wombolu*. These two trees are genetically distinct based on RAPD analysis (Lowe et al. 2000) but hard to distinguish them based on morphology (Vihotogbé et al. 2013). However, genetic differentiation is significant. A study carried out in Cameroon, Nigeria and Gabon demonstrated that the population of *I. gabonensis* and *I. wombolu* is genetically differentiated among countries within both species (Lowe et al. 2000; Ude et al. 2006). The Gabon and Nigeria clusters were clearly separate, while the Cameroonian population overlapped them. There is also a closer relationship between Cameroon and Gabon populations. The range and average of the genetic distance within each country was Cameroon 62 % (53 – 76 %) Nigeria 52 % (32.3 – 84.8 %) and Gabon 50 % (45 – 53 %) so there is also higher genetic diversity in Cameroon (Ude et al. 2006). The genetic similarities of each population declined according to the geographical distance of each individual. “Hot spot” of genetic diversity for *I. gabonensis* was found in southern

Cameroon, southern Nigeria and central Gabon, and for *I. wombolu* in southern Nigeria and southern Cameroon (Lowe et al. 2000).

Figure 6. Young trees. a) *I. gabonensis*; b) *I. wombolu*

(source: Onyekwelu & Stimm 2006)

Figure 7. Fruit on the branches. a) *I. gabonensis*; b) *I. wombulu*

(source: Onyekwelu & Stimm 2006)

2.4.5 *Ricinodendron heudelotii*

It is an oilseed tree from the *Euphorbiaceae* family. *R. heudelotii* is used by local communities to improve nutrition and income (Mpeck et al. 2003) and has medical benefits (Boko-Haya et al. 2021). The kernel is known as “ndjanssang” or “njanssang” is one of the most traded products in Cameroon (Mpeck et al. 2003).

There is very little information known about this tree and much less is known about its genetic diversity but according to Mpeck et al. (2003) who conducted a study in three villages in Cameroon of 64 trees, he found highly significant variation between trees in fruit mass, weight, and fruit length. A significant difference exists also between parameters of fruit mass, fruit length, fruit width, nuts, shells, and kernels. Which suggests a possible high genetic diversity. Compared to the research carried out in five states (Oyo, Osun, Ondo, Edo, and Cross River) in Southern Nigeria in which it was found that the genetic diversity in populations was from 0.24 to 0.12 and the genetic similarities were from 60.03 % to 57.6 % between Oyo and Osun (Onefeli 2021).

Figure 8. *Ricinodendron heudelotii*.

(source: Onefeli 2021)

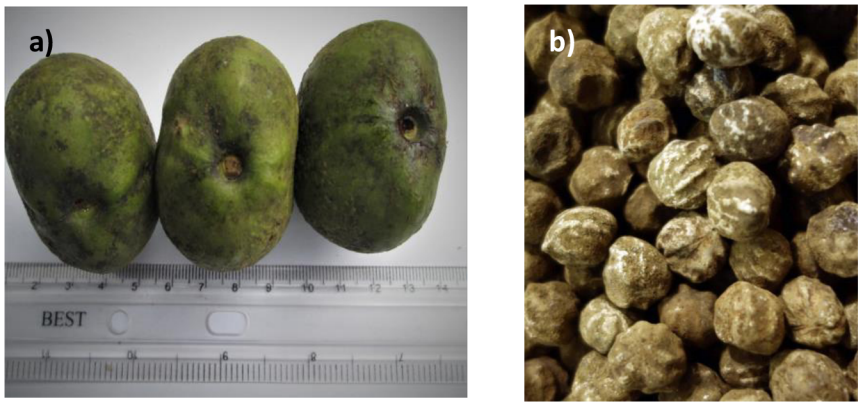

Figure 9. Indehiscent fruits. a) Fruit of *Ricinodendron heudelotii*; b) Seeds of *Ricinodendron heudelotii*

(source: Onefeli 2021)

2.5 Molecular markers

The general definition of genetic markers is (1) chromosomal tags or alleles that allow the identification of specific regions of DNA; (2) specific DNA fragments with a specific location in the genome; and (3) a gene that easily distinguishes phenotypic expression. Markers can be utilized to identify cells or individuals or as labelling probes in organelles, chromosomes, or chromosomal loci. The established terminology divides genetic markers into three main groups: morphological or phenotypic markers (determined by the organism's phenotype) molecular markers (defined at the level of nucleic acid) and biochemical markers (Chesnokov et al. 2020). The molecular method for detecting DNA sequence variations is usually based on the use of restriction enzymes that recognize and cut short DNA sequences and polymerase chain reactions (PCR), which involve the amplification of DNA sequences based on oligonucleotide primers (Rao 2003).

Molecular markers are greatly assisting breeders in solving several challenging issues associated with tropical fruit tree breeding such as the development of complex family structures for recombination mapping and recurrent selection (Arias et al. 2012). In plant genetics and breeding, the use of various types of markers not only helped to accelerate scientific progress but also made it possible to apply the knowledge acquired in theory such as in the implementation of breeding programs with the least amount of time and resource consumption (Chesnokov et al. 2020). Molecular markers have proven to be effective and authentic in plant breeding. Next-Generation Sequencing (NGS) facilities and other bioinformatics tools have assisted in identifying various putative micro RNAs (miRNAs) in plants. Different organisms express miRNAs, which are noncoding RNAs with a length of 21 to 24 nucleotides. As miRNA sequences are highly conserved, especially in the stem-loop (hairpin loop) region, there is a potential for developing novel genetic markers. Although miRNA-based molecular markers have overwhelmingly been used in animal systems, they are currently gaining attention among plant breeders for the development of new markers in plant systems (Adhikari et al. 2017).

Based on technical principles, there are three categories of molecular markers: Polymerase Chain Reaction based on DNA amplification (RAPD, AFLP, SSRs), single

nucleotide polymorphism (SNPs) and nucleic acid hybridization based on complementary bases (RFLPs) (Akemi et al. 2012; Adhikari et al. 2017).

2.5.1 PCR

The polymerase chain reaction is a method for detecting nucleic acids in biological samples even at very low levels (Albayrak & Arican 2004). The technique was developed by Kary Mullins in 1983 (Mullis 1990). Nowadays, the PCR technique is used to diagnose diseases by clinicians and researchers, for clone and genes sequence, fingerprinting of DNA, and also for quantitative and genomic studies (Welsh et al. 1992).

PCR is based on thermostable DNA polymerase. First isolated enzyme *Taq* polymerase from *Thermus aquaticus* (Saiki et al. 1987; Pavlov et al. 2004). It is a simple enzymatic assay where the specific DNA fragments are amplificated from a complex pool of DNA. For the process is necessary to prepare the mixture composed of dNTPs (adenine, cytosine, guanine and thymine) specific primers, DNA polymerase and DNA template (Garibyan & Avashia 2013). PCR consists of three steps which are repeated several times during the whole process. Denaturation, primer annealing and extension (Weier & Gray 1988; McPherson & Moller 2005).

In the denaturation step, the strands of the DNA template are separate usually at 94 - 95 °C. In annealing, the temperature gets lower usually about 37 °C, depending on the used primer. The primers hybridised complementary sequences on the template strands. In the final step, the DNA polymerase is ligated to primers and the synthesis of new strands begins. All these steps are repeated about 25 - 40 times. The temperature of the last step depends also on the used enzymes but usually, the temperature is about 72 - 75 °C (McPherson & Moller 2005).

The extension complementary products are capable of bending primers. Each cycle doubles the amount of DNA from the preceding cycle. The result of this is an exponential accumulation of the specific target fragment, roughly 2'n, where n is the number of cycles (Saiki et al. 1987).

2.5.2 RFLP

Restriction Fragment Length Polymorphism is commonly used for genotyping in almost any organism, including plants, animals, and humans. It was the first DNA profiling method and has been widely used for genome mapping and variation analysis, such as genetic fingerprinting, building genetic maps, identifying candidate genes for different traits, hereditary disease diagnosis, and paternity testing (Dai & Long 2015). Originally the technique was used to recognize point mutation within PCR (Hashim & Al-Shuhaib 2019). It is characterized as a detection of differences in homologous DNA sequences by the presence of different length fragments after digestion of DNA samples with specific restriction enzymes (NCBI 2017). A restriction endonuclease digestion method identifies DNA polymorphisms in genes or regions of DNA of interest using restriction endonuclease digestion. Most commonly, RFLPs are used to identify candidate genes by analysing a polymorphism that is known to be near or in a specific gene. When there is no known polymorphism associated with the candidate gene, we search for polymorphisms to identify a suitable RFLP (Jarcho 1994). RFLP is powerful and simple but the use of this technology is limited to restriction enzyme recognition sites and other sequences are disregarded (Hashim & Al-Shuhaib 2019). Because of this, the technique is not the best for analysing a large number of different SNPs (Rasmussen 2012; Hashim & Al-Shuhaib 2019).

Known-sequence chemically/radioactively labelled DNA probes (from cDNA or genomic libraries) are hybridized with restriction digested DNA samples to create differential DNA fragment profiles. Point mutations at the restriction enzyme recognition site or DNA rearrangements such as insertions and deletions, translocations, inversions, and transpositions are the cause of differential DNA profiles. Markers are potential candidates for linkage analysis and can be used to determine whether a recessive trait occurs in a homozygous or heterozygous state in a particular individual (Winter & Kahl 1995; Adhikari et al. 2017). In PCR-RFLP analysis, the first step is an amplification of a fragment containing the variation. Following this, a restriction enzyme is applied to the amplified fragment. By electrophoresis of the fragments, it is possible to analyse alleles based on their sizes due to the presence or absence of the restriction enzyme recognition site (Rasmussen 2012).

A cloned piece of eukaryotic DNA was first used as a genetic marker by Botstein et al. (1980). According to these authors, it is expected that organisms with differing base sequences would yield fragments of varying size when digested with a restriction enzyme (Tanksley et al. 1989; Kochert 1991). RFLP have a few advantages compared to the other techniques. First, RFLP loci are located throughout the entire genome and the markers are relatively highly polymorphic. The second benefit is that they are reproducible and mostly inherited. As a third point, RFLP can reliably detect polymorphic loci across multiple varieties, without regard to environmental or genetic interactions. Because of these characteristics, the method provides an opportunity to simultaneously screen numerous samples. In addition, DNA blots can be analysed repeatedly with different RFLP probes by stripping and reprobing (Dai & Long 2015). In recent years, the use of RFLPs in whole genome searches for linkage has become less common, as newer and more powerful markers have been developed, like simple sequence repeats (Jarcho 1994).

2.5.3 RAPD

Random amplified polymorphic DNA is one of the PCR-based techniques. As a molecular technique for developing DNA markers, RAPD has been widely used (Kumar & Gurusubramanian 2011). Markers detect polymorphisms in nucleotide sequences among individuals by using a single, long (decamer), random oligonucleotide primer. DNA polymorphisms are detected as changes in nucleotide sequence at or between oligonucleotide primer binding sites (Adhikari et al. 2017). In the case of short primers, there is a high probability that the genome will contain several inverted priming sites close to one another. By scanning a genome for such small inverted repeats, the technique amplifies intervening DNA segments of variable lengths (Hadrys et al. 1992).

Dominant markers can detect multiple loci in a single assay ranging from 0.5 to 5 kb. Electrophoresis of agarose gels fractionates DNA fragments and visualizes them with ethidium bromide. RAPD technique has been shown to be a promising tool for the identification of markers related to agronomically important traits due to the simplicity of the method, the absence of sequence information for designing primers, and the low cost. This method is not preferential for genome wide studies due to their inherent

problems of reproducibility, sensitivity to experimental conditions, the presence of artefactual bands (false positives), and the inability to distinguish homozygous from heterozygous individuals (Adhikari et al. 2017).

Several applications of RAPD can be used in molecular ecology, such as determining taxonomic identity, determining kinship relationships, analysing mixed genome samples, and developing specific probes. In addition to its suitability for working with anonymous genomes, RAPD technology can also be used in situations where only limited amounts of DNA are available (Hadrys et al. 1992).

2.5.4 AFLP

Amplified fragment length polymorphism is another technique based on PCR markers and is used for rapid screening and identification of genetic diversity. By generating hundreds of highly replicable markers from any organism's DNA, it allows for high-quality genotyping of fingerprinting. AFLPs are more efficient, replicable, and have a higher resolution than other markers, but they primarily generate dominant rather than codominant markers. AFLP markers have emerged as important genetic markers with widespread applications, in population genetics, pathotyping, DNA fingerprinting and quantitative trait loci mapping due to their high replicability and ease of use (Mueller & Wolfenbarger 1999).

A three-step procedure is utilized. In the first step, DNA is restricted, oligonucleotide adapters are ligated, and then sets of restriction fragments are amplified selectively, followed by gel analysis of the amplified fragments. To amplify restriction fragments by PCR, the adapter sequence and restriction site sequence serve as the target sites for primer annealing. A selective amplification process can be achieved by using primers that extend into restriction fragments, resulting in only amplifying fragments whose primer extensions match with the flanking nucleotides of restriction sites. PCR can visualize restriction fragment sets without knowing the nucleotide sequences of the fragments. A specific co-amplification of many restriction fragments can be achieved using this method (single steps of AFLP are described below). However, the number of fragments that can be analysed simultaneously depends on the detection system's resolution (Vos et al. 1995).

2.5.5 DNA sequencing

Sequencing was firstly introduced in 1968 (Hutchison 2007), and next-generation sequencing was available during the 21st century. The first known sequencing method was Sangler sequencing which caused a revolution in biological research (Mardis 2017). This method is based on 2′,3′-dideoxynucleotides (Dey 2018). DNA dependent polymerase generates a copy on the basis of complementarity (Valencia et al. 2013) anneals to the primer which is complementary to the known sequence flanking the region of interest. The detecting takes place in a cycle sequencing reaction. Reactions contain template strand, dNTPs, primer, and DNA polymerase. In addition, there are four dideoxynucleotides phosphates (ddNTPs) such as ddATP, ddTTP, ddCTP and ddGTP are labelled by fluorescent dyes, each has different colour. The cycles include denaturation of the template strand, primer annealing and extension. The new strand is synthesised from dNTPs and the synthesis end by incorporating one of ddNTP. For determining the DNA sequence is used high-resolution electrophoretic separation of the end-labelled and single-stranded products in a capillary based polymer gel. Laser read excitation of the fluorescent label as fragments exit capillary and the data are assessed by software translates to the DNA sequence (Shendure & Ji 2008; Dey 2018).

Although this sequencing method is still used, the most widely used method today is next-generation sequencing.

Next-generation sequencing (NGS)

This type of sequencing is based on three techniques. (1) sequencing by synthesis (SBS) (454 [Roche] pyrosequencing, Solexa/Illumina, Ion Torrent, Qiagen gene reader); (2) single-molecule sequencing (Helicos, Oxford Nanopore Technologies) and (3) sequencing by ligation (SOLiD) (Mardis 2017). Some of these technologies are no longer in use (Helicos, SOLiD). Another division may be based on the time of origin of the technology, there is a second generation (SBS), third generation (Pacific Biosciences [PacBio]) and fourth-generation (Oxford nanopore technologies) (Slatko et al. 2018), but sometimes the third generation is associated with the fourth generation. There are four of the most prevalent technologies Illumina and Ion Torrent as short-read sequencing

techniques and PacBio and Oxfords Nanopore Technologies as long-read sequencing techniques (Hu et al. 2021).

Illumina

This technique was developed by Solexa and Lynx Therapeutics. Before sequencing, there is a step of clonal amplification (cluster generation) by “bridge amplification” where DNA fragments (about 550 bp) with adapters ligated on each end are used for amplification synthesis reaction (Slatko et al. 2018). Before the amplification, the adenine is added to the DNA fragments' ends. Adapters with a single-base 'T' overhang are then ligated to both ends of DNA fragments. On a solid support-flow cell, DNA fragments are immobilized at one end after denaturation. Its surface is densely covered with adapters and complementary adapters (Xiaoguang et al. 2010).

The attached DNA fragments form “bridges”. After fragments 'bend over' and hybridise to complementary adapters on the surface, they start complementary strand synthesis. By repeating the amplification, clusters are created (van Dijk et al. 2014). The steps of this process are visualized in Figure10.

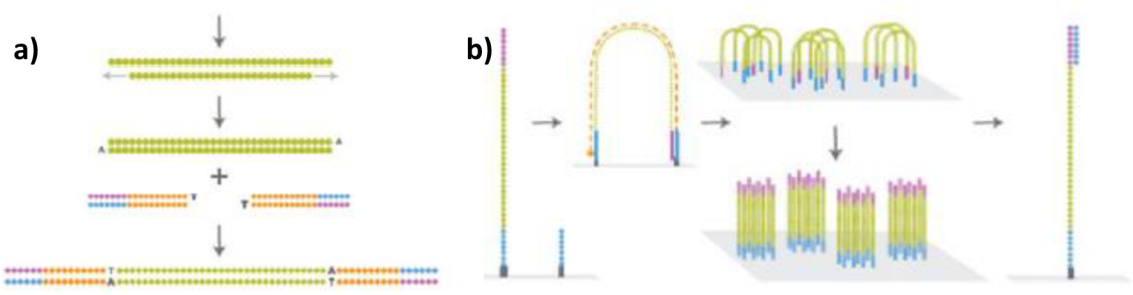

Figure 10. Ilumina sequencing. a) DNA fragments with ligated adaptors; b) bridge amplification and clusters

(source: Ansorge 2009)

Ion Torrent

The technique is based on 454 (Roche) pyrosequencing. But Ion sensors detect proton release during nucleotide incorporation instead of pyrophosphate; imaging is not required (van Dijk et al. 2014). By using Ion Sphere particles in a micro-well, emulsion PCR on the Ion Torrent platform enables clonal amplification. An adapter sequence is attached to a DNA fragment, followed by a droplet of water-in-oil emulsion (micelle),

which contains a bead of complementary adapters, deoxynucleotides (dNTPs), primers, and DNA polymerase. A micelle acts as an independent micro-PCR reactor, allowing PCR to be amplified independently. A semiconductor chip contains a flow chamber and complementary metal-oxide semiconductors to measure the pH level. A flow chamber and a complementary metal-oxide semiconductor sensor constitute the Ion torrent semiconductor chip. In the sequencing process, microwells on a semiconductor chip are loaded with micelles, and the chips are flooded sequentially with unmodified A, T, G, or C nucleotides. When incorporated single nucleotide, it releases a hydrogen ion which is detected by the pH sensor. Ion Torrent is the first to perform semiconductor sequencing without using a light source (Hu et al. 2021).

Pacific Biosciences (PacBio)

Is current technology leader. Two sequencing systems were commercialised by them, RSII and Sequel. Pac Bio is referred also to as SMRT (Single Molecule Real Time). This method enables the sequencing of long fragments (up to 30 -50 kb or longer) (Slatko et al. 2018). The sequencing is made on zero-mode waveguides (ZMW) in an SMRT cell. This cell consists of millions of sub-wavelength holes. ZMW is like a small chamber leading illuminating light into a small area whose dimensions are small relative to the wavelength. Due to this, the detection takes place at the bottom of the well where the DNA polymerase incorporates single fluorescent labelled nucleotide into the growing chain. After incorporation, the fluorescent moiety bound to phosphate is released and creates a flash of light, where each nucleotide has a different colour (Xiaoguang et al. 2010; Slatko et al. 2018).

Oxfords Nanopore Technique (ONT)

This technique detects the DNA composition of native ssDNA molecules without secondary signals like light or colour. The DNA passed through a protein pore while a current flows through this pore (Goodwin et al. 2016). Processive enzymes are first bound to long dsDNA molecules. When the complex encounters a nanopore only one DNA strand enters it. DNA polymerase synthesis and translocation regulate the translocation rate through the pore. DNA polymerase synthesis and translocation

regulate the translocation rate through the pore. A nucleotide passing through a nanopore disrupts an electrical current. There is a characteristic electronic signal associated with each nucleotide that is recorded as a disruption of current. A nanopore can be used by another DNA molecule once DNA leaves the pore (Goodwin et al. 2016; Slatko et al. 2018).

3. Aims of the thesis

The main objective of this thesis was to describe the current knowledge of the genetic diversity of selected African indigenous fruit trees. The specific objectives were the following:

1) To review and summarize all available information about the genetic diversity of *Allanblackia* spp., *Cola* spp., *Dacryodes edulis, Garcinia kola, Irvingia* spp., and *Ricinodendron heudelotii*
2) To identify knowledge gaps in the genetic diversity of the selected species and to propose their potential solutions
3) To analyse the genetic diversity and population structure of *Garcinia kola* in the South region of Cameroon with AFLP markers

4. Material and methods

4.1 Study site and sample collection

The samples of *G. kola* leaves were collected in Cameroon near cities Ebolowa, Kye-Ossi, Sangmelima and Zoétélé in August and September 2019 by Anna Maňourová. There were 80 samples from four geographically separated populations: Ebolowa (26), Kye-Ossi (26), Sangmelima (16) and Zoétélé (12). Trees come from different environments such as cocoa agroforestry (planted by farmers or their ancestors), farm (planted by farmers), wild nature (naturally occurring), or home gardens (planted by owners) and the minimum distance of individuals was 100 m. Most trees were found in agroforestry systems (38) and forests (22). The origin of each tree is shown in Appendix A.

In this region, the climate is very hot and humid. The average temperature is 27 – 33 °C during the day and 20 – 21 °C at night and there are about 20 to 30 rainy days each month. The average total precipitation is about 1700 mm per year, 155 to up to 400 mm per month during wet seasons (March - November) but only 48 to 97 mm during dry seasons (December – February) The average humidity is 70 to 90 %.

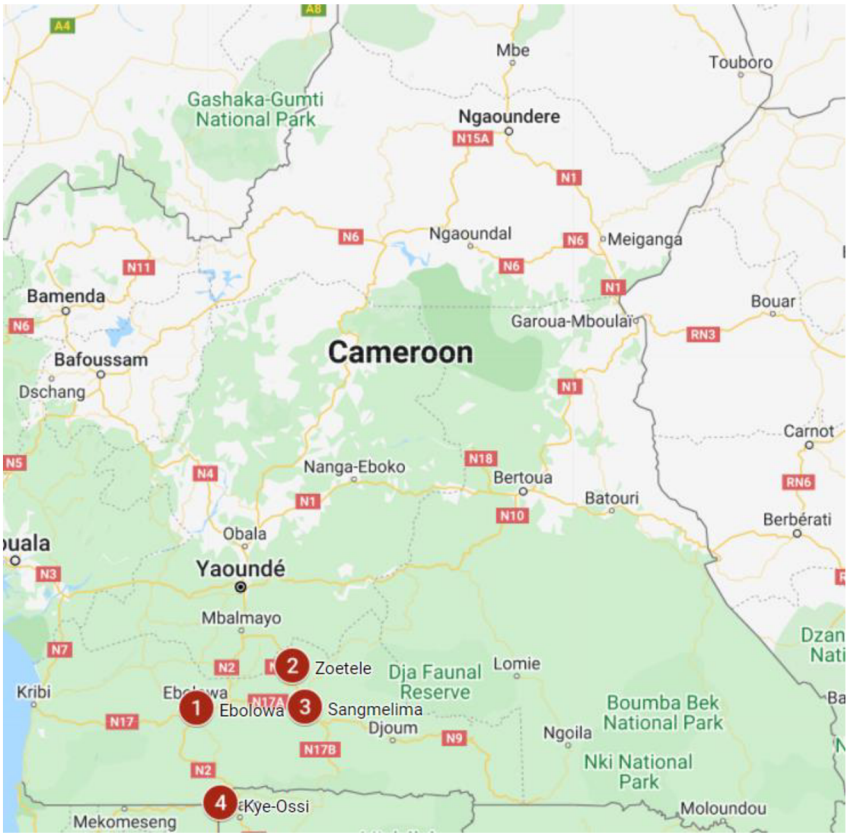

Figure 11. Map of Cameroon with highlighted study site locations. 1 – Ebolowa; 2 – Zoetele; 3 – Sangmelima; 4 – Kye-Ossi

(source: Google Maps 2022)

4.2.1 DNA extraction

DNA extraction was carried out at the Laboratory of Molecular Genetics, Faculty of Tropical AgriSciences, CZU Prague. The extraction was done by CTAB (Cetyl trimethylammonium bromide) method because of the high quantity and satisfactory quality of the final extracted DNA. The method was developed by Doyle and Doyle (1987) and was modified by Faleiro (2002). DNA was extracted from small pieces of the dried leaves (about 2 x 2 cm) which were grinded by using mortar and pestle with fine sand. The homogenized mass was transferred into 2 ml tubes, 800 µl of CTAB extraction buffer was added and the mixture was vortexed. 5 µl of Proteinase K was added into the mixture and kept for 1 hour at 65 °C and agitate every 10 minutes. After that, 700 µl of chloroform: IAA (24:1) was added and mixed for 10 minutes and centrifugated at 14,000 RPM at 4 °C for another 10 minutes. After centrifugation, the solid part was sedimented and the liquid supernatant was transferred into new 2 ml test tubes, while 55 µl of 7 % CTAB was added and mixed for 10 minutes. The chloroform and centrifugation steps

were repeated due to the increasing purity of the final DNA samplings. The supernatant was transferred into new 1.5 ml tubes and washed by adding 700 µl of isopropanol, mixed for 5 minutes and was stored at -20 °C for an hour or at 4 °C overnight. Then the solution was centrifuged again at 14,000 RPM for 10 minutes. After this step, the DNA was separated, and the pellets were settled at the bottom of the tubes. The liquid solution was released, and the pellets were washed twice with ethanol. The first washing was done with 96 % ethanol for 3 minutes at 37 °C, then the alcohol was discarded. For the second washing was used 70 % ethanol for 5 minutes at 21 °C (room temperature). After this time, the solution was centrifuged for 3 minutes at 14,000 RPM and the ethanol was discarded. The pellets were dried at room temperature overnight or at 37 °C for about two hours. Finally, the pellets were dissolved by adding 50 µl of ddH_2O and 5 µl of RNase at 37 °C. When the pellets were dissolved, spectrophotometry was performed.

4.2.2 Spectrophotometric quantification of DNA

The concentration of the final DNA was determined by NanoDrop One Microvolume UV-Vis spectrophotometer (Thermo Scientific). Primarily, the concentration of the DNA was observed, the clarity was a secondary criterium. For the analysis, 1 µl of the sample was used. First, the blank sample, pure ddH_2O, was performed, followed by the measurement of the individual DNA samples. The low limit of concentration was set at 500 ng/µl.

4.2.3 AFLP

First, the genomic DNA is digested with two restriction enzymes. The next steps are preamplification (pre-selective amplification) and selective amplification using specific oligonucleotide primes combinations.

Restriction – ligation

This step is usually done in a single reaction. In the R-L two restriction endonucleases, *Mse*I and *Eco*RI were used, also including their adaptors *Mse*I-a and *Eco*RI-a. *Mse*I

recognised four bases pair 5’-TTAA and *Eco*RI recognised five bases pair 5’-GAATTC. The adaptor sequences are described in Table 1. These enzymes split the DNA in a particular place. Another final master mix components were DNA samples, H_2O, T4 ligase with T4 ligase buffer and CutSmart buffer. T4 ligase catalysed the formation of a phosphodiester bond between 5’-phosphate and 3’-hydroxyl in DNA fragments during the ligation and the buffer created an optimal environment for the reaction. The CutSmart buffer was also important, it facilitated the course of the reaction. The content of the master mix for one sample was compiled in Table 2. The final mixture was vortexed and 19 µl of its content was added into the single cell of 96 well plates subsequently the DNA was added. The plate was pasted with a plastic film to prevent evaporation during the process. The plate with samples was inserted into a thermocycler and incubated at 37 °C for four hours and at 65 °C for 20 minutes afterwards (Table 3.). The product was checked by gel electrophoresis on a 2 % agarose gel stained by EtBr, run at 90 V for 1 hour. The product was diluted 10x for further processing.

Table 1. Adaptor sequences.

Adaptor	Sequences
MseI A 1	GACGATGAGTCCTGAG
MseI A 2	TACTCAGGACTCAT
EcoRI A 1	CTCGTAGACTGCGTACC
EcoRI A 2	AATTGGTACGCAGTCTAC

Table 2. Content of the master mix for one sample.

Reagent	Volume (µl)	Concentration
DNA template	1	500 ng
T4 Ligase	0.17	67 U
T4 Ligase buffer	2	
EcoRI	0.25	5 U
EcoRI adaptor	1	5 pmol/µl
MseI	0.1	1 U
MseI adaptor	1	50 pmol/µl
Cut smart buffer	1	
H2O	10.48	

Table 3. Steps if cycles of restriction-ligation.

Temperature	Time
37 °C	4 hours
65 °C	20 min.
4 °C	∞

Preselective amplification

Large amounts of DNA fragments that were formed by R-L had to be reduced. Preamplification works similarly to ligation. The main difference is that the enzymes *Mse*I+C and *Eco*RI+A (Table 4.) are able to recognize one base pair more than in R-L. So only 1 ⁄ 16 of DNA fragments were amplificated. The mixture was mixed with Q-solution, Qiagen Master mix, R-L PCR product and of course with mentioned enzymes. The PCR products were tenfold diluted before use. Table 5. shows the final form of the mixture. 15 µl of the mixture and 5 µl of R-L product were added into the individual well and vortexed. The plate was put in the thermocycler. Single steps of cycling were summarized in Table 6. At the end of the process, the samples were left at 4 °C until they were removed and stocked at -20 °C.

Table 4. Primer sequences of preselective amplification.

Primer	Sequences
MseI+C	GATGAGTCCTGAGTAA**C**
EcoRI+A	GACTGCGTACCAATTC**A**

Table 5. The final form of the mixture.

Reagent	Volume (µl)
R-L product	5
Eco+1 primer	1.5
Mse+1 primer	1.5
Q-solution	2
Qiagen Master mix	10

Table 6. Steps of cycles of preselective amplification.

Temperature	Time	Touchdown	Repeats
95 °C	15 min		
95 °C	30 sec		
62 °C	30 sec	-1 °C/cycle	10x
72 °C	2 min		
95 °C	30 sec		
52 °C	30 sec		20x
72 °C	1 min		
72 °C	10 min		
4 °C	∞		

Selective amplification

The selective amplification was performed with four mixtures, each had a different combination of enzymes. The total content of the mixture was summarized in Table 7. and primer sequences in Table 8. 9 µl of the mixture was added to a single well followed by 5 µl of preamplification product. The whole plate was vortexed and put in the thermal cycler. The profile of cycling was compiled in Table 9. The final products were stored at -20 °C. After an initial screening of 24 combinations of selective primers, the following four primer combinations were selected based on amplification results. The Eco selective primer was fluorescently labelled at 5‘ end with 6-FAM dye (Selective nucleotides are in bold).

The selective amplification product was separated by capillary electrophoresis on a 3500 Series Genetic Analyzer (Applied Biosystems, USA). The results of fragment analysis were visualized using Geneious Prime software

Table 7. Final mixture of selective amplification.

Reagent	Volume (µl)
preamplification product	5
Eco+3 primes	1
Mse+3 primer	1
Qiagen Master mix	7

Table 8. Primer sequencing of selective amplification.

Primer	Eco+3 (labelled with 6-FAM)	Mse+3
Sequences	GACTGCGTACCAATTC**ATT**	GATGAGTCCTGAGTAA**CCT**
		GATGAGTCCTGAGTAA**CTA**
	GACTGCGTACCAATTC**AAT**	GATGAGTCCTGAGTAA**CGA**
		GATGAGTCCTGAGTAA**CAT**

Table 9. Steps of cycles of selective amplification.

Temperature	Time	Touchdown	Repeats
95 °C	15 min		
95 °C	30 sec		
62 °C	30 sec	-1 °C/cycle	10x
72 °C	2 min		
95 °C	30 sec		
52 °C	30 sec		20x
72 °C	1 min		
72 °C	10 min		
4 °C	∞		

4.2.4 Gel electrophoresis

To assess the concentration of the DNA gel electrophoresis method was performed. Electrophoresis works on the principle of negative DNA charge. The DNA fragments flow through agar gel from the negative acquired part to the positive acquired part. The agarose gel was 1 % and was prepared with 0,8 g of agarose and 80 ml TBE buffer (Tris/Boric Acid/EDTA). The Ethidium Bromide (EtBr) was added after heating. EtBr binds to DNA and caused fluorescence under UV light. Then the gel was left to solidify with a comb. 5 µl of DNA sample was mixed with loading dye and put into small holes which were made by com. Then the gel was put into a bath with TBE buffer and covered. The electrophoresis was run at 120 V for 30 minutes. Finally, the gel was placed at the transilluminator with UV light and the bands fluorescent.

4.3. Data analysis

A binary matrix with all alleles was created, where they were scored as present (1) or absent (0). The binary matrix was then subjected to data analysis. Basic diversity indices were calculated in AFLP-SURV (Vekemans et al. 2002), including the number and percentage of polymorphic loci and expected heterozygosity, along with population structure indicators such as total gene diversity (Ht); mean gene diversity within populations (Hw); average gene diversity among populations (Hb); Wright's fixation index (F_{ST}) based on 1,000 permutations and Nei's genetic distance between populations supported by 1,000 bootstraps.

The genetic structure of sampled populations was assessed by model-based clustering implemented in Structure (Pritchard et al. 2000). The program was run for K=1-6 with 5 iterations for each K. The settings included a burn-in period of 10,000 followed by 50,000 MCMC steps. Because weak structuring was expected, the analysis was run with the locprior parameter. The optimal number of K was determined by the Evanno method (Evanno et al. 2005) in Structure Harvester (Earl & von Holdt 2012). The results of the repeated runs were aligned using Clumpp (Jakobsson & Rosenberg 2007) and the bar plots were visualized in the district (Rosenberg 2003).

5. Results

Genetic diversity

The total number of loci was 1,299 and mean the number of fragments per individual was 99.0, and the total number of segregating fragments was 1,116 (85.9 %). Observed heterozygosity in populations (Hj) ranged from 0.061 (Ebolowa) to 0.10 (Zoétélé), the percentage of polymorphic loci was highest in Sangmelima (34.0 %) and lowest in Ebolowa (21.6 %) (Table 10.). The highest diversity was observed in the population from Zoétélé and the lowest in Ebolowa.

Table 10. Population data (Lynch & Milligan method).

Population	n	#loc.	#loc_P	PLP	Hj
Ebolowa	26	1299	281	21.6	0.061
Kye-Ossi	26	1299	358	27.6	0.072
Sangmelima	16	1299	442	34.0	0.072
Zoétélé	12	1299	410	31.6	0.102

n = average number of scored individuals; **#loc** = number of loci scored; **#loc_P** = number of polymorphic loci at the 5 % level, i.e. loci with allelic frequencies lying within the range 0.05 to 0.95; **PLP** = proportion of polymorphic loci at the 5 % level, expressed as a percentage; **Hj** = expected heterozygosity under Hardy-Weinberg genotypic proportions.

Population genetic

Total gene diversity (Ht) showed almost the same value as mean gene diversity (Hw) (0.0804 and 0.0806, respectively). This means that the main source of variation comes from differences among individuals, not among populations. The genetic differentiation among populations (Hb) is also low. Wright's fixation index (F_{ST}) also shows there is no genetic structure between populations. Data are summarised in Table 11.

Table 11. Population genetic structure (Lynch & Milligan method).

n	Ht	Hw	Hb	F_{ST}
4	0.0804	0.0806	0.0	0.0
S.E.		0.008841	0.000137	0.789474
Var		0.000078	0.000000	0.623269

Ht = the total gene diversity; **Hw** = the mean gene diversity within populations (analogous to Nei's Hs); **Hb** = the average gene diversity among populations in excess of that observed within populations; **F_{ST}** = Wright's fixation index; **n** = number of populations; **Var** = variance

Nei's genetic distance confirms the trend shown by the low values of Hb and F_{ST} in Table 11. It proves that there is almost no differentiation between populations.

Table 12. Nei's genetic distance.

Ebolowa	0.0000	0.0000	0.0000	0.0001
Kye-Ossi	0.0000	0.0000	0.0000	0.0000
Sangmelima	0.0000	0.0000	0.0000	0.0000
Zoétélé	0.0001	0.0000	0.0000	0.0000

Genetic structure

Results show that the populations are divided into two clusters. Graph (Figure 12) presents with what probability each population belongs to each cluster. Each column represents one population. There are not many differences between populations. No special structure is apparent. The biggest difference is between Ebolowa and Zoétélé.

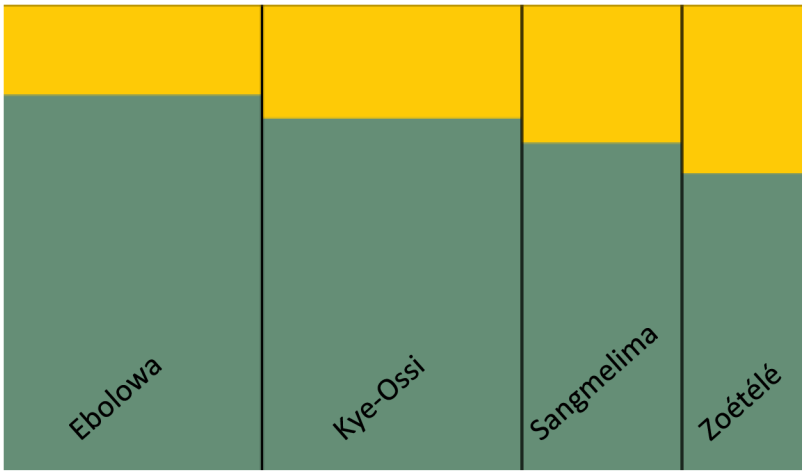

Figure 12. Structure graph of populations. Green – cluster one; yellow – cluster two

The structure graph for each individual (Figure 13.) is similar to the population but there are more marked differences between them.

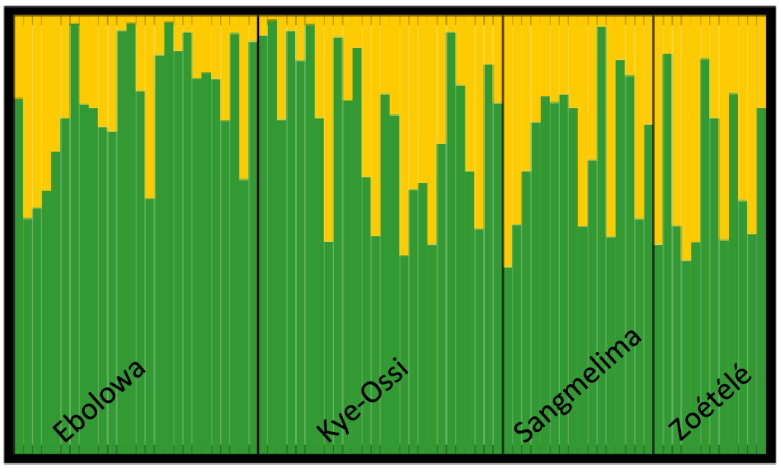

Figure 13. Structure graph of individual trees. Green – cluster one; yellow – cluster two

6. Discussion

6.1. Genetic diversity of African fruit trees with high potential for domestication

As was shown in the literature review the genetic diversity of six selected species, with high potential for domestication, is often unclear because not much research focusing on their genetic diversity was done yet. It can be also caused by different analytical methods used for individual species or by relatively small sample size, in some cases, which did not correspond to the natural distribution of the species (Table 13.).

Table 13. Information on genetic diversity research of the selected species.

Name of the species	Analytical methods	Genetic diversity	Areas of species research	Area of species distribution	Number of genetic research	References
Allanblackia spp.	AFLP, SSR	high	Ghana, Cameroon, Tanzania	West / Central Africa	2	Russell et al. 2009; Kyereh et al. 2021
Cola spp.	RAPD	high	Cameroon, Nigeria	West / Central Africa	1	Akinro et al. 2019
Darcyodes edulis	SSR	low	Cameroon, Nigeria	West / Central Africa	3	Makueti et al. 2015; Tchinda et al. 2016; Rimlinger et al. 2021
Garcinia kola	RAPD, SNP, AFLP	low	Benin, Cameroon, Nigeria	West / Central Africa	4	Olawuyi & Azeez 2019 Azeez et al. 2020; Dadjo et al. 2020; Irikidzai 2021
Irvingia spp.	AFLP, RAPD	high	Cameroon, Nigeria, Gabon	Central Africa	2	Lowe et al. 2000; Ude et al. 2006
Ricinodendron heudelotii	ISSR	low	Cameroon, Nigeria	West / Central Africa	1	Onefeli 2021

In *Allanblackia* spp. there was found high genetic diversity based on AFLP and SSR markers (Ofori et al. 2016), which is supported by morphological variations which showed also a high diversity. A study in Ghana indicated that a large proportion of the fruit's phenotype might be genetically controlled, adding that the selection should be individual and based on fruit size (Peprah et al. 2009). The genetic data were obtained from both wild populations and farm remnants (trees retained during forest clearance. To properly domesticate *Allanblackia*, it is essential to address key knowledge gaps, such as the lack of genetic information within the genus and the absence of information on gene flow within and between populations (Russell et al. 2009).

In *Cola* spp., low diversity was measured based on secondary metabolites and bioactive compounds (Onomo et al. 2006; Niemenak et al. 2008). But there is a lack of information to unequivocally describe the level of genetic diversity. The RAPD analysis of trees grown on farms has shown that the genetic diversity is high (Akinro et al. 2019). There was only one study that described genetic diversity carried out at farms in Nigeria, so the knowledge gap is the lack of genetic data in wild populations of *Cola* spp. from wide areas of its distribution.

Populations of *Dacryodes edulis* have a high diversity mainly based on fruit variation in Nigeria and Cameroon (Waruhiu et al. 2004; Anegbeh et al. 2005). However, in genetic studies from Cameroon, the SSR analysis measured low genetic diversity with only 2 % of genetic variation (Makueti et al. 2015; Tchinda et al. 2016; Rimlinger et al. 2021). This could be a result of high gene flow, that the populations of *D. edulis* could be maintained with low genetic differentiation and their genetic relationships closed. The introgression of new alleles may broaden the genetic basis of *D. edulis*. It is also needed to discover and characterize new markers in *D. edulis* research such as gene discovery or gene mapping (Makueti et al. 2015). Because most private alleles have been found in wild individuals, these alleles could be the source for crossing new cultivars (Tchinda et al. 2016). *D. edulis* is also the most studied species out of the six selected species.

Concerning *Irvingia spp*. there was a significant genetic differentiation between the two species *I. gabonensis* and *I. wombulu* (Lowe et al. 2000). However, in Cameroon,

the populations of these two species have a close relationship. These facts may indicate that Cameroon is the centre of the diversity of *Irvingia gabonensis* and could be a source of original materials in other countries because the population of Cameroon genetically overlaps with the population of Ghana and Nigeria. To ensure optimum protection of the existing genetic diversity in *I. gabonensis*, the collection of germplasm from Cameroon should be prioritized. The agreement between Lowe et al.'s (2000) RAPD results and AFLP results implies that AFLP is as effective as RAPD in identifying intraspecific variation within *I. gabonensis* (Ude et al. 2006).

There is high diversity based on fruit variation in *Ricinodendron heudelotii*, but the genetic diversity of Nigerian populations was low (Onefeli 2021), and the populations were genetically similar to each other. Studies also indicated that the earliest population of *R. heudelotii* originated in Benin (Onefeli 2021). There is a lack of information about this species especially about genetic diversity so further research is strongly recommended.

6.2 Genetic diversity of *Garcinia kola*

The information about the genetic diversity of *Garcinia kola* is remarkably diverse. The level of genetic diversity depends on the analysis used. The RAPD analysis (multi-allelic markers) proved high genetic diversity compared (Olawuyi & Azeez 2019) to SNP analysis (bi-allelic markers) (Dadjo et al. 2020) which proved low genetic diversity. However, as mentioned in the literature review the use of multi-allelic markers is better because their maximum PIC value is higher than the bi-allelic markers (Olawuyi & Azeez 2019).

Results of our study confirmed low genetic diversity and population structure. The main differences were between the Ebolowa and Zoétélé populations, which do not correspond to the geographical distribution of the populations. It was assumed that the largest differences would be between the southernmost population (Kye-Ossi) and the northernmost population (Zoétélé) because of the larger geographical distance. However, this did not happen. This may be due to the close proximity of individual populations or the small range of sampled areas. Another reason may be the high flow

of genes due to trade in products and the movement of people or seed spreading by mammals.

To improve the gene pool of *G. kola* and to discover new qualities at the morphological and chemical level it needs to enlarge the scope and size of collection during the fruiting season throughout the distribution areas (Azeez et al. 2020). The genetic information on *G. kola* is still insufficient and further research should be carried out. It is also recommended to protect the tree through the adaptation of interventions (Dadjo et al. 2020) such as managing trees in their natural habitats within their environmental range, dynamically evolving artificial populations elsewhere or sustainable cultivation for medical purposes and last but not least, domestication.

Conclusion

Based on the literature review the high genetic diversity was observed in *Allanblackia* spp., *Cola* spp., and *Irvingia* spp., whereas low genetic diversity occurred in other species. Most studies were performed on *G. kola* (4) and *D. edulis* (3). In most of the studies relatively outdated methods were used such as RAPD or AFLP. In future research, it is necessary to supplement the results with more modern methods such as SNP or sequencing. This data could then be used for genome mapping and analysis. The genetic diversity of *Garcinia kola* in South Cameroon is low so this population is not very suitable for domestication, but it would be appropriate to continue genetic research in other areas. It is also recommended to conduct research in a wider area of the natural distribution of the trees and with a larger number of examined samples. The genetic diversity of selected species should be protected and conserved.

7. References

Adhikari S, Saha S, Biswas A, Rana TS, Bandyopadhyay TK, Ghosh P. 2017. Application of molecular markers in plant genome analysis: a review. The Nucleus **60**:283–297.

Akemi A, Pereira J, Macedo P, Alessandra K. 2012. Genetic Diversity in Microorganisms. Page 149 – 170 in Caliskan M, editor. Microsatellites as Tools for Genetic Diversity Analysis. In Tech. Available from: http://www.intechopen.com/books/geneticdiversity-in-microorganisms/microsatellites-as-tools-for-genetic-diversity-analysis (accessed March 2022).

Akinro L, Adesoye A, Fasola T. 2019. Genetic diversity in *Cola acuminata* and *Cola nitida* using RAPD primers. Genetika **51**:227–236.

Albayrak G, Arican E. 2004. Amplification of Specific Genes by using RT-PCR Technique in Plants. Biotechnology & Biotechnological Equipment **18**:15–18.

Anegbeh PO, Ukafor V, Usoro C, Tchoundjeu Z, Leakey RRB, Schreckenberg K. 2005. Domestication of *Dacryodes edulis*: 1. Phenotypic variation of fruit traits from 100 trees in southeast Nigeria. New Forests **29**:149–160.

Ansorge WJ, 2009. Next-generation DNA sequencing techniques. New Biotechnology **25**:195–203.

Arias RS, Borrone JW, Tondo CL, Kuhn DN, Irish BM, Schnell RJ. 2012. Genomics of Tropical Fruit Tree Crops. Pages 209–239 in Schnell RJ, Priyadarshan PM, editors. Genomics of Tree Crops. Springer New York, New York, NY. Available from http://link.springer.com/10.1007/978-1-4614-0920-5_7 (accessed February 16, 2022).

Atangana AR, Beaulieu J, Khasa DP. 2010. Wild genetic diversity preservation in a small-sized first generation breeding population of *Allanblackia floribunda* (Clusiaceae). Tree Genetics & Genomes **6**:127–136.

Azeez AA, Olawuyi OJ, Igata DF. 2020. Genetic diversity *of Garcinia kola* Heckel from selected states in Nigeria. Journal of Research in Forestry, Wildlife & Environment **12:**92-105.

Boko-Haya YY, Ouinsavi CAIN, Ewédjè E-EBK, Akin YY, Zinkpe T, Agbangla C. 2021. Variability of Seed Germination and Seedling Growth Potential of *Ricinodendron heudelotii* (Euphorbiaceae) at Fine Scale in Southern of Benin. East African Journal of Forestry and Agroforestry **3**:1–17.

Booy G, Hendriks RJJ, Smulders MJM, Groenendael JM, Vosman B. 2000. Genetic Diversity and the Survival of Populations. Plant Biology **2**:379–395.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. American Society of Human Genetics. **32**:314–331.

Chesnokov Yu V, Kosolapov VM, Savchenko IV. 2020. Morphological Genetic Markers in Plants. Russian Journal of Genetics **56**:1406–1415.

Clement C, De Cristo-Araújo M, Coppens D'Eeckenbrugge G, Alves Pereira A, Picanço-Rodrigues D. 2010. Origin and Domestication of Native Amazonian Crops. Diversity **2**:72–106.

Dadjo C, Nyende AB, Yao N, Kiplangat N, Assogbadjo AE. 2020. Genome-wide genetic diversity and population structure of *Garcinia kola* (Heckel) in Benin using DArT-Seq technology. PLOS ONE **15** (e0238984) DOI 10.1371/journal.pone.0238984.

Dah-Nouvlessounon D, Adoukonou-Sagbadja H, Nafan D, Adjanohoun A, Noumavo PA, Sina H, Daouda BO, Baba-Moussa. 2016. Morpho-agronomic variability of three kola trees accessions [*Cola nitida* (Vent.) Schott et Endl., *Cola acuminata* (P. Beauv.) Schott et Endl., and *Garcinia kola* Heckel] from Southern Benin. Genetic Resources and Crop Evolution **63**:561–579.

Dai S, Long Y. 2015. Genotyping Analysis Using an RFLP Assay. Pages 91–99 in Batley J, editor. Plant Genotyping. Springer New York, New York, NY. Available from http://link.springer.com/10.1007/978-1-4939-1966-6_7 (accessed February 25, 2022).

Dawson I, Harwood C, Jamnadass R, Beniest J. 2012. Tree domestication for small-scale farmers. In Dawson I, Harwood C, Jamnadass R, Beniest J editors, Agroforestry tree domestication: A primer. World Agroforestry Centre. Kenya. Available from http://apps.worldagroforestry.org/downloads/Publications/PDFS/TM17346.PDF (Access February 2022).

Dey P. 2018. Sanger Sequencing and Next-Generation Gene Sequencing: Basic Principles and Applications in Pathology. Pages 227–231 Basic and Advanced Laboratory Techniques in Histopathology and Cytology. Springer Singapore, Singapore. Available from http://link.springer.com/10.1007/978-981-10-8252-8_23 (accessed March 15, 2022).

Earl DA, von Holdt BM. 2012. Structure harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources **4**:359–361.

Ekué MRM, Gailing O, Vornam B, Finkeldey R. 2011. Assessment of the domestication state of ackee (*Blighia sapida* K.D. Koenig) in Benin based on AFLP and microsatellite markers. Conservation Genetics **12**:475–489.

Ellegren H, Galtier N. 2016. Determinants of genetic diversity. Nature Reviews Genetics **17**:422–433.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology **14**:2611–2620.

Franzel S, Kindt R. 2012. Choosing the right tree: Species priority setting procedures. Pages 36–45. In Dawson I, Harwood C, Jamnadass R, Beniest J editors, Agroforestry Tree Domestication: A Primer. World Agroforestry Centre, Nairobi, Kenya.

Fuller DQ. 2007. Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World. Annals of Botany **100**:903–924.

Garibyan L, Avashia N. 2013. Research Techniques Made Simple: Polymerase Chain Reaction (PCR). Journal of Investigative Dermatology **133**:1–4.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics **17**:333–351.

Google Maps. 2022. Map of Cameroon with highlighted study site locations. Google Maps (online). Available from https://www.google.com/maps/d/u/0/edit?mid=1jGTFfHnTY9rGVp8PfAEp6qJ0DvqnVmfE&hl=cs&ll=5.7556779691584925%2C14.243299889323676&z=7 (Accessed March 2022).

Govindaraj M, Vetriventhan M, Srinivasan M. 2015. Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. Genetics Research International **2015**:1–14.

Hadrys H, Balick M, Schierwater B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular Ecology **1**:55–63.

Hashim HO, Al-Shuhaib MB. 2019. Exploring the Potential and Limitations of PCR-RFLP and PCR-SSCP for SNP Detection: A Review. Journal of Applied Biotechnology Reports **6**:137–144.

Holderegger R, Kamm U, Gugerli F. 2006. Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landscape Ecology **21**:797–807.

Hollingsworth PM, Dawson IK, Goodall-Copestake WP, Richardson JE, Weber JC, Sotelo Montes C, Pennington RT. 2005. Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia: SSR variation in *Inga edulis*. Molecular Ecology **14**:497–501.

Hu T, Chitnis N, Monos D, Dinh A. 2021. Next-generation sequencing technologies: An overview. Human Immunology **82**:801–811.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M. 2008. Ecological consequences of genetic diversity: Ecological effects of genetic diversity. Ecology Letters **11**:609–623.

Hutchison CA. 2007. DNA sequencing: bench to bedside and beyond. Nucleic Acids Research **35**:6227–6237.

Irikidzai PC. 2021. Population structure of *Garcinia kola* Heckel in Central region of Cameroon. Master's thesis. Czech University of Life Sciences, Prague.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23**:1801–1806.

Jarcho J. 1994. Development of Genetic Markers: Restriction Fragment Length Polymorphism Analysis. Current Protocols in Human Genetics. John Wiley & Sons, Inc **1**:1–15.

Kamga YB, Nguetsop VF, Anoumaa M, Kanmegne G, Kanmegne MC, Ngueguim JR. 2019. *Garcinia Kola* (Guttiferae) in Tropical Rain Forests: Exploitation, Income Generation and Traditional Uses, in the East and Central Regions of Cameroon. Journal of Pharmaceutical, Chemical and Biological Sciences **7**:13–27.

Kengue J, Fohouo FNT, Adewusi HG. 2002. Towards the improvement of safou (*Dacryodes edulis*): population variation and reproductive biology. Forests, Trees and Livelihoods **12**:73–84.

Kochert G. 1991. Restriction Fragment Length Polymorphism in Plants and Its Implications. Pages 167–190 in Biswas BB, Harris JR, editors. Plant Genetic

Engineering. Springer US, Boston, MA. Available from http://link.springer.com/10.1007/978-1-4613-9365-8_8 (accessed February 28, 2022).

Kumar NS, Gurusubramanian G. 2011. Random amplified polymorphic DNA (RAPD) markers and its applications. Science Vision **3**:116–124.

Kyereh D, Maňourová A, Hendre PS, Muchugi A, Kalousová M, Choungo Nguekeng PB, Jawo TO, Ofori DA, Lojka B. 2021. Diversity, Chemical Composition, and Domestication Potential of *Allanblackia parviflora* A. Chev. in West Africa. Forests **12**:1758.

Leakey R, Schreckenberg K. 2003. The participatory domestication of West African indigenous fruits. International Forestry Review **5**:338–347.

Leakey RRB, Asaah EK. 2013. Underutilised Species as the Backbone of Multifunctional Agriculture – the Next Wave of Crop Domestication. Acta Horticulture **979**:293–310.

Leakey RRB, Schreckenberg K, Tchoundjeu Z. 2017. Towards a Domestication Strategy for Indigenous Fruit Trees in the Tropics. Pages 165–176 in Akinnifesi, F.K., Leakey, R.R.B., Ajayi, O.C., Sileshi, G., Tchoundjeu, Z., Matakala, P., Kwesiga, F. editors. Multifunctional Agriculture. CAB International, Wallingford, UK Available from https://linkinghub.elsevier.com/retrieve/pii/B9780128053560000167 (accessed December 1, 2021).

Leakey RRB, Simons AJ. 1998. The domestication and commercialization of indigenous trees in agroforestry for the alleviation of poverty. Agroforesry systems **38**:165–176.

Lowe AJ, Gillies ACM, Wilson J, Dawson IK. 2000. Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. Molecular Ecology **9**:831–841.

Makueti JTM, Otieno G, Tchoundjeu Z, Muchugi A, Tsobeng A, Asaah E, Kariba R. 2015. Genetic diversity of *Dacryodes edulis* provenances used in controlled breeding trials. Journal of Plant Breeding and Crop Science **7**:327–339.

Maňourová A, Leuner O, Tchoundjeu Z, Van Damme P, Verner V, Přibyl O, Lojka B. 2019. Medicinal Potential, Utilization and Domestication Status of Bitter Kola (*Garcinia kola* Heckel) in West and Central Africa. Forests **10**:1–18.

Mardis ER. 2017. DNA sequencing technologies: 2006–2016. Nature Protocols **12**:213–218.

McPherson MJ, Moller SG. 2005. PCR: Basics from Background to Bench. Page 1–8 in Bosher A editor. BIOS Scientific. Abingdon; Florence. Available from http://site.ebrary.com/id/5000324 (accessed February 22, 2022).

Meyer RS, DuVal AE, Jensen HR. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytologist **196**:29–48.

Mpeck MLN, Asaah E, Tchoundjeu Z, Atangana AR. 2003. Strategies for the domestication of *Ricinodendron heudelotii*: Evaluation of variability in natural populations from Cameroon. WFL Publisher Science and Technology 1:257–262.

Mueller UG, Wolfenbarger LL. 1999. AFLP genotyping and fingerprinting. Trends in Ecology & Evolution **14**:389–394.

Mullis KB. 1990. The Unusual Origin of the Polymerase Chain Reaction. Scientific American **262**:56–65.

National Centre of Biotechnology Information (NCBI). 2017. Restriction Fragment Length Polymorphism (RFLP). National Centre of Biotechnology Information. U.S. Available from https://www.ncbi.nlm.nih.gov/probe/docs/techrflp/ (accessed April 2022).

Niemenak N, Onomo PE, Fotso, Lieberei R, Ndoumou DO. 2008. Purine alkaloids and phenolic compounds in three *Cola species* and *Garcinia kola* grown in Cameroon. South African Journal of Botany **74**:629–638.

Nyadanu D, Akpertey A, Dadzie AM, Pobee P, Lowor ST, Asirifi ID, Brako-Marfo M, Aboagye LM. 2021. Mapping spatial distribution of genetic resources of kola (*Cola nitida* (Vent.) Schott & Endl.) in Ghana and collection of germplasm for conservation, characterization and development of improved varieties. Genetic Resources and Crop Evolution **68**:69–77.

Nyadanu D, Lowor ST, Akpertey A, Tchokponhoué DA, Pobee P, Dogbatse JA, Okyere D, Amon-Armah F, Brako-Marfo M. 2020. Genetic variability of bioactive compounds and selection for nutraceutical quality in kola [*Cola nitida* (Vent) Schott. and Endl.]. PLOS ONE **15** (e0242972) DOI 10.1371/journal. pone.0242972.

Ofori DA, Gyau A, Dawson IK, Asaah E, Tchoundjeu Z, Jamnadass R. 2014. Developing more productive African agroforestry systems and improving food and nutritional security through tree domestication. Current Opinion in Environmental Sustainability **6**:123–127.

Ofori DA, Kehlenbeck K, Munjuga M, Jamnadass R, Asaah EK, Kattah C, Rutatina F. 2013. *Allanblackia* species: a model for domestication of high potential tree crops in Africa. Acta Horticulture 2:311–317.

Ofori DA, Peprah T, Koech G, Anjarwalla P, Munjuga M, Jamnadass R. 2016. Survival, growth performance and reproductive biology of *Allanblackia parviflora* A. Chev., a high valued indigenous fruit tree species in Ghana. Genetic Resources and Crop Evolution **63**:791–800.

Olawuyi OJ, Azeez AA. 2019. Molecular Evaluation of *Garcinia kola* Heckel Accessions Using RAPD Markers. American Journal of Molecular Biology **09**:41–51.

Ondo -Azi AS, Missang CE, Silou T. 2009. classification of *Dacryodes edulis* (g. don) h.j. lam by using morphological and physical characteristics of the fruits: a statistical approach. Forests, Trees and Livelihoods **19**:99–109.

Onefeli AO. 2021, August. Morphologic and molecular characterisation of *Ricinodendron heudelotii* (Baill.) Pierre ex Pax in Southern Nigeria. Doctor thesis. University of Ibadan, Nigeria.

Onomo PE, . NN, . DON. 2006. Isoenzyme Variability of Three *Cola* (*Cola acuminata* (Pal. de Beauv, Schott and Endlicher), *Cola nitida* ((Vent) Schott and Endlicher) and *Cola anomala* (Schott and Endlicher)) Germplasm in Cameroon. Pakistan Journal of Biological Sciences **9**:391–397.

Onyekwelu JC, Stimm B. 2006. *Irvingia gabonensis*. Page 1-14 in Schütt P, Weisgerber H, Lang U, Roloff A, Stimm B editos. Enzyklopädie der Holzgewächse. ecomed-Verlag, Landsberg/Lech, München. Available from https://www.researchgate.net/publication/259622863_Irvingia_gabonensis (Accessed February 2022).

Pavlov AR, Pavlova NV, Kozyavkin SA, Slesarev AI. 2004. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. Trends in Biotechnology **22**:253–260.

Peprah T, Ofori DA, Siaw DEKA, Addo-Danso SD, Cobbinah JR, Simons AJ, Jamnadass R. 2009. Reproductive biology and characterization of *Allanblackia parviflora* A. Chev. in Ghana. Genetic Resources and Crop Evolution **56**:1037–1044.

Pickersgill B. 2007. Domestication of Plants in the Americas: Insights from Mendelian and Molecular Genetics. Annals of Botany **100**:925–940.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics **155**:945–959.

Purugganan MD. 2019. Evolutionary Insights into the Nature of Plant Domestication. Current Biology **29**:705–714.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. Nature **457**:843–848.

Ræbild A et al. 2011. Advances in domestication of indigenous fruit trees in the West African Sahel. New Forests **41**:297–315.

Rao KN. 2003. Plant genetic resources: Advancing conservation and use through biotechnology. African Journal of Biotechnology **3**:136–145.

Rao VR, Hodgkin T. 2002. Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell, Tissue and Organ Culture **68**:1–19.

Rasmussen HB. 2012. Restriction Fragment Length Polymorphism Analysis of PCR-Amplified Fragments (PCR-RFLP) and Gel Electrophoresis - Valuable Tool for Genotyping and Genetic Fingerprinting. Page in Magdeldin S, editor. Gel Electrophoresis - Principles and Basics. InTech. Available from http://www.intechopen.com/books/gel-electrophoresis-principles-and-basics/restriction-fragment-length-polymorphism-analysis-of-pcr-amplified-fragments-pcr-rflp-and-related-te (accessed February 25, 2022).

Reed DH, Frankham R. 2003. Correlation between Fitness and Genetic Diversity. Conservation Biology **17**:230–237.

Rimlinger A et al. 2021. Trees and their seed networks: The social dynamics of urban fruit trees and implications for genetic diversity. PLOS ONE **16** (e0243017) DOI 10.1371/journal.pone.0243017.

Rimlinger A, Carrière SM, Avana M-L, Nguegang A, Duminil J. 2019. The Influence of Farmers' Strategies on Local Practices, Knowledge, and Varietal Diversity of the Safou Tree (*Dacryodes edulis*) in Western Cameroon. Economic Botany **73**:249–264.

Rosenberg NA. 2003. Distruct: a program for the graphical display of population structure: PROGRAM NOTE. Molecular Ecology Notes **4**:137–138.

Russell JR, Kadu CAC, Jamnadass R, Booth A, Cordeiro NJ, Woodhead M, Dawson IK. 2009. AFLP and SSR diversity in the African fruit tree Allanblackia: implications for management of a genus newly subject to domestication for the edible oil industry. Tree Genetics & Genomes **5**:517–527.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, HoRN GT, Mullis KB, Erlich HA. 1987. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase **239**:487–491.

Sargolzaei M, Iwaisaki H, Colleau JJ. 2006. CFC: a tool for monitoring genetic diversity. World Congress on Genetics Applied To Livestock Production 8:1–2.

Schreckenberg K, Awono A, Degrande A, Mbosso C, Ndoye O, Tchoundjeu Z. 2006. Domesticating indigenous fruit trees as a contribution to poverty reduction. Forests, Trees and Livelihoods **16**:35–51.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nature Biotechnology **26**:1135–1145.

Slatko BE, Gardner AF, Ausubel FM. 2018. Overview of Next-Generation Sequencing Technologies. Current Protocols in Molecular Biology **122:**1–14.

Smith C, editor. 2014. Encyclopedia of Global Archaeology. Springer New York, New York, NY. Available from http://link.springer.com/10.1007/978-1-4419-0465-2 (accessed November 28, 2021).

Smýkal P, Nelson M, Berger J, Von Wettberg E. 2018. The Impact of Genetic Changes during Crop Domestication. Agronomy **8**:119.

Stadlmayr B, Charrondière UR, Eisenwagen S, Jamnadass R, Kehlenbeck K. 2013. Nutrient composition of selected indigenous fruits from sub-Saharan Africa. Journal of the Science of Food and Agriculture **93**:2627–2636.

Tanksley SD, Young ND, Paterson AH, Bonierbale MW. 1989. RFLP mapping in plant breeding: new tools for an old science. Nat Biotechnol **7**:257–264.

Tchinda ND, Wanjala BW, Muchugi A, Fotso, Nzweundji G, Skilton R. 2016. Genetic diversity and gene flow revealed by microsatellite DNA markers in some accessions of African Plum (*Dacryodes edulis*) in Cameroon. African Journal of Biotechnology **15**:511–517.

Ude GN, Dimkpa CO, Anegbeh PO, Shaibu AA, Tenkouano A, Pillay M, Tchoundjeu Z. 2006. Analysis of genetic diversity in accessions of *Irvingia gabonensis* (Aubry-Lecomte ex O'Rorke) Baill. African Journal of Biotechnol **5**:219–223.

Valencia CA, Pervaiz MA, Husami A, Qian Y, Zhang K. 2013. Sanger Sequencing Principles, History, and Landmarks. Pages 3–11 . Next Generation Sequencing Technologies in Medical Genetics. Springer New York, New York, NY. Available from http://link.springer.com/10.1007/978-1-4614-9032-6_1 (accessed March 15, 2022).

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. Trends in Genetics **30**:418–426.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular Ecology **11**:139–151.

Vihotogbé R, Ronald G, van den Berg RG, Sosef MSM. 2013. Morphological characterization of African bush mango trees (I*rvingia* species) in West Africa. Genetic Resources and Crop Evolution **60**:1597–1614.

Vos P et al. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research **23**:4407–4414.

Waruhiu AN, Kengue J, Atangana AR, Tchoundjeu Z, Leakey RRB. 2004. Domestication of *Dacryodes edulis*. 2. Phenotypic variation of fruit traits in 200 trees from four populations in the humid lowlands of Cameroon. WFL Publisher Science and Technology Food, Agriculture & Environment **1**:340–346.

Weier HU, Gray JW. 1988. A Programmable System to Perform the Polymerase Chain Reaction. DNA **7**:441–447.

Welsh J, Chada K, Dalal SS, Cheng R, Relph D, McClelland M. 1992. Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Research **20**:4965–4970.

Wiehle M, Prinz K, Kehlenbeck K, Goenster S, Mohamed SA, Buerkert A, Gebauer J. 2014. The role of homegardens and forest ecosystems for domestication and conservation of *Ziziphus spina*-christi (L.) Willd. in the Nuba Mountains, Sudan. Genetic Resources and Crop Evolution **61**:1491–1506.

Winter P, Kahl G. 1995. Molecular marker technologies for plant improvement. World Journal of Microbiology & Biotechnology **11**:438–448.

Yogom BT, Avana-Tientcheu M-L, Mboujda MFM, Momo ST, Fonkou T, Tsobeng A, Barnaud A, Duminil J. 2020. Ethnicity Differences in Uses and Management Practices of Bitter Kola Trees (*Garcinia kola*) in Cameroon. Economic Botany **74**:429–444.

Xiaoguang Z, Lufeng R, Qingshu M, Yuntao L, Yude Y, Jun Y. 2010. The next-generation sequencing technology and application. Protein & Cell **1**:520–536.

Appendix A

Table of origin of individual tree.

	ID	Surrounding	Location		ID	Surrounding	Location
1	SKOG2	cocoa agroforestry	Kye - Ossi	41	SEPL1	home garden	Ebolowa
2	SKVO1G	forest	Kye - Ossi	42	SEAX1	cocoa agroforestry	Ebolowa
3	SKAS1G	cocoa agroforestry	Kye - Ossi	43	SEGD2G	cocoa agroforestry	Ebolowa
4	SEDR1	forest	Ebolowa	44	SKFS2G	home garden	Kye – Ossi
5	SKBF3	home garden	Kye - Ossi	45	SKAS3G	cocoa agroforestry	Kye - Ossi
6	SKSD1	cocoa agroforestry	Kye - Ossi	46	SZJY3	forest	Zoétélé
7	SKVO2	forest	Kye - Ossi	47	SKOG4G	cocoa agroforestry	Kye - Ossi
8	SKSD2	cocoa agroforestry	Kye - Ossi	48	SKOG5	forest	Kye - Ossi
9	SSCF1	home garden	Sangmelima	49	SKBF4	home garden	Kye - Ossi
10	SZML1	cocoa agroforestry	Zoétélé	50	SSJM1	forest	Sangmelima
11	SKBF1	home garden	Kye - Ossi	51	SSPP1	cocoa agroforestry	Sangmelima
12	SKFS1	home garden	Kye - Ossi	52	SKLC1G	cocoa agroforestry	Kye - Ossi
13	SKOG3	cocoa agroforestry	Kye - Ossi	53	SZJY2G	forest	Zoétélé
14	SKOG1	cocoa agroforestry	Kye - Ossi	54	SZCL1	cocoa agroforestry	Zoétélé
15	SKVO3	cocoa agroforestry	Kye - Ossi	55	SEGD5G	cocoa agroforestry	Ebolowa
16	SSKD5	forest	Sangmelima	56	SEEL1	cocoa agroforestry	Ebolowa
17	SSKD4	cocoa agroforestry	Sangmelima	57	SKPA1	forest	Kye - Ossi
18	SEGD4	cocoa agroforestry	Ebolowa	58	SSKD1	cocoa agroforestry	Sangmelima
19	SEGD8	forest	Ebolowa	59	SZSV1	cocoa agroforestry	Zoétélé
20	SEDR2G	forest	Ebolowa	60	SKBD1	home garden	Kye - Ossi

21	SEDL1	cocoa agroforestry	Ebolowa	61	SKAS2	cocoa agroforestry	Kye - Ossi
22	SEGD6G	cocoa agroforestry	Ebolowa	62	SKFS3	home garden	Kye - Ossi
23	SSKD6G	forest	Sangmelima	63	SKPA2G	forest	Kye - Ossi
24	SSJP2G	home garden	Sangmelima	64	SETS3G	cocoa agroforestry	Ebolowa
25	SSKD3G	forest	Sangmelima	65	SSJP1	cocoa agroforestry	Sangmelima
26	SZCE1	farm	Zoétélé	66	SSPP2	cocoa agroforestry	Sangmelima
27	SEGD1	cocoa agroforestry	Ebolowa	67	SSKD2	forest	Sangmelima
28	SEEL2G	cocoa agroforestry	Ebolowa	68	SKVO4	cocoa agroforestry	Kye - Ossi
29	SERE1	cocoa agroforestry	Ebolowa	69	SKVO5G	cocoa agroforestry	Kye - Ossi
30	SEGD3	home garden	Ebolowa	70	SZCV1	farm	Zoétélé
31	SELN6	forest	Ebolowa	71	SZGF1	farm	Zoétélé
32	SELNG4G	cocoa agroforestry	Ebolowa	72	SSPP3	cocoa agroforestry	Sangmelima
33	SELN3	cocoa agroforestry	Ebolowa	73	SKBD2	farm	Kye - Ossi
34	SELN5	forest	Ebolowa	74	SSCF2	cocoa agroforestry	Sangmelima
35	SETS2G	cocoa agroforestry	Ebolowa	75	SSSM1	forest	Sangmelima
36	SELN2	forest	Ebolowa	76	SSFG1	forest	Sangmelima
37	SEEM1	farm	Ebolowa	77	SZRN1G	-	Zoétélé
38	SEZE1	home garden	Ebolowa	78	SZBT1	forest	Zoétélé
39	SETS1	cocoa agroforestry	Ebolowa	79	SZJM2	farm	Zoétélé
40	SELN1	home garden	Ebolowa	80	SZJY1	forest	Zoétélé