agronomy

MDPI

Article

Genetic Diversity and Population Structure of Cacao (*Theobroma cacao* L.) Germplasm from Sierra Leone and Togo Based on KASP–SNP Genotyping

Ranjana Bhattacharjee [1,*], Mohamed Mambu Luseni [2,3], Komivi Ametefe [4], Paterne A. Agre [1], P. Lava Kumar [1] and Laura J. Grenville-Briggs [2]

1 International Institute of Tropical Agriculture, PMB 5320, Ibadan 200001, Nigeria; p.agre@cgiar.org (P.A.A.); l.kumar@cgiar.org (P.L.K.)
2 Department of Plant Protection Biology, Swedish University of Agricultural Sciences, 750 07 Lomma, Sweden; mmluseni@gmail.com (M.M.L.); laura.grenville.briggs@slu.se (L.J.G.-B.)
3 Sierra Leone Agricultural Research Institute, Kenema Forestry and Tree Crops Research Center, PMB 1313, Freetown, Sierra Leone
4 Institut Togolais de Recherche Agronomique, Centre de Recherche Agronomique-Zone Forestière (CRA-F), Kpalime, BP 90, Togo; amexkom@gmail.com
* Correspondence: r.bhattacharjee@cgiar.org

Abstract: Cacao (*Theobroma cacao* L.) is a tropical tree species belonging to the Malvaceae, which originated in the lowland rainforests of the Amazon. It is a major agricultural commodity, which contributes towards the Gross Domestic Product of West African countries, where it accounts for about 70% of the world's production. Understanding the genetic diversity of genetic resources in a country, especially for an introduced crop such as cacao, is crucial to their management and effective utilization. However, very little is known about the genetic structure of the cacao germplasm from Sierra Leone and Togo based on molecular information. We assembled cacao germplasm accessions (235 from Sierra Leone and 141 from Togo) from different seed gardens and farmers' fields across the cacao-producing states/regions of these countries for genetic diversity and population structure studies based on single nucleotide polymorphism (SNP) markers using 20 highly informative and reproducible KASP–SNPs markers. Genetic diversity among these accessions was assessed with three complementary clustering methods, including model-based population structure, discriminant analysis of principal components (DAPC), and phylogenetic trees. STRUCTURE and DAPC exhibited some consistency in the allocation of accessions into subpopulations or groups, although some discrepancies in their groupings were noted. Hierarchical clustering analysis grouped all the individuals into two major groups, as well as several sub-clusters. We also conducted a network analysis to elucidate genetic relationships among cacao accessions from Sierra Leone and Togo. Analysis of molecular variance (AMOVA) revealed high genetic diversity (86%) within accessions. A high rate of mislabeling/duplicate genotype names was revealed in both countries, which may be attributed to errors from the sources of introduction, labeling errors, and lost labels. This preliminary study demonstrates the use of KASP–SNPs for fingerprinting that can help identify duplicate/mislabeled accessions and provide strong evidence for improving accuracy and efficiency in cacao germplasm management as well as the distribution of correct materials to farmers.

Keywords: cacao; STRUCTURE; DAPC; hierarchical clustering; genetic diversity

Citation: Bhattacharjee, R.; Luseni, M.M.; Ametefe, K.; Agre, P.A.; Kumar, P.L.; Grenville-Briggs, L.J. Genetic Diversity and Population Structure of Cacao (*Theobroma cacao* L.) Germplasm from Sierra Leone and Togo Based on KASP–SNP Genotyping. *Agronomy* **2024**, *14*, 2458. https://doi.org/10.3390/agronomy14112458

Academic Editor: Mark P. Widrlechner

Received: 2 September 2024
Revised: 17 October 2024
Accepted: 18 October 2024
Published: 22 October 2024

1. Introduction

Cacao, *Theobroma cacao* L., is a tropical tree native to the humid tropics of the central and northern parts of South America [1]. It is a diploid (2n = 2x = 20) vegetatively propagated species domesticated approximately 3000 years ago [2,3]. It is the major ingredient used in the multi-billion-dollar chocolate and confectionary industry, as well as for other

intermediary products such as cacao butter, cacao powder, cacao cake, and cacao liquor. The coastal countries in West Africa, known as the West Africa cacao belt [4], from Sierra Leone, Guinea, and Liberia to southern Cameroon, apart from the Benin Republic, are responsible for the production of about 70% of the world's cacao (http://faostat.fao.org) (accessed on 23 September 2023). The world's chocolate and confectionary industry is heavily dependent on cacao beans from West African countries, both due to the high production and high quality of beans (bulk cacao, may not be specialty cacao) compared to those produced in other cacao-producing regions such as Asia or Central and Southern America [5]. In 2020, cacao was the primary agricultural export commodity for Cote d'Ivoire, Ghana, Nigeria, Cameroon, and Sierra Leone, and the second most important for Guinea, Togo, and Liberia, thus contributing significantly towards the Gross Domestic Products of these countries (www.statista.com) (accessed on 23 September 2023).

It is estimated that Brazilian cacao (of the Amelonado type, which is also known as the Lower Amazon Forastero type) was first introduced into West Africa in the 19th and early 20th centuries by the Portuguese [6], and since then, it has been cultivated by smallholder farmers in this region. Cote d'Ivoire and Ghana remain the highest producers of cacao in the world, accounting for over 60% of global world production of around 4.9 million tons in the 2021/2022 cacao season [7,8]. Among cacao-producing countries in the world, Togo and Sierra Leone rank 15th and 17th with a production of 22,522 and 14,670 metric tons, respectively [9]. In 2017, the World Bank Trade Statistics recorded export earnings of about USD 14,461 million from cacao beans in Sierra Leone [10], even with low productivity. Over the years, cacao production has increased in both Sierra Leone and Togo, which corresponds to an increase in the area under cultivation. However, future yields are expected to be adversely affected by changing climatic conditions. Similar to other West African countries, cacao cultivation in Sierra Leone and Togo faces the challenges of old trees, aged farmers, black pod disease, mirids, poor access to improved planting materials, and other challenges (such as cacao swollen shoot virus) related to its cultivation and management [11,12].

The cacao germplasm introduction in Sierra Leone and Togo followed the same trend as introductions in other West African countries, which is from a common source—Fernando Po [13]. However, in Sierra Leone, there may also have been introductions from other sources (the West Indies), raising questions as to the contributions of various sources [14]. There is, therefore, an interesting possibility that present-day cacao in West Africa (apart from recent introductions through the University of Reading) and germplasm exchanges between West African countries such as introductions from Ghana to both Sierra Leone and Togo at experimental stations of national institutes is of dual origin. In Togo, cacao germplasm is conserved in a gene bank consisting of clones introduced from countries in the sub-region (Ghana, Côte d'Ivoire, Cameroon, and Nigeria) and international collections (University of Reading, United Kingdom). In both of these countries, historical phenotypic data or any other type of data were unavailable to understand the nature of these genotypes (personal communications). It is, therefore, crucial to understand and assess genetic relationships and genetic diversity among cacao germplasms within and between these two countries to effectively conserve and efficiently use them for further breeding and crop improvement. Several efforts have been put forward to understand and assess the genetic diversity currently available in major cacao-growing countries in the world, including several West African countries, using both molecular markers and morphological traits [15–19]. However, there are no studies to date that have targeted germplasms collected from different cacao-growing regions in Sierra Leone and Togo to assess genetic diversity and understand their population structure.

There is a general agreement that cultivated cacao in West Africa has a narrow genetic base and faces issues of mislabeling [15–18]. A 2019 review that compared both modern and historical introductions did not detect significant genetic diversity or improvements in yield or pest and disease resistance during the last 20 years [20]. Any improvements in yield resulted from better management practices, although some of the recent varieties or hybrids developed in major West African countries produced significantly higher yields with better

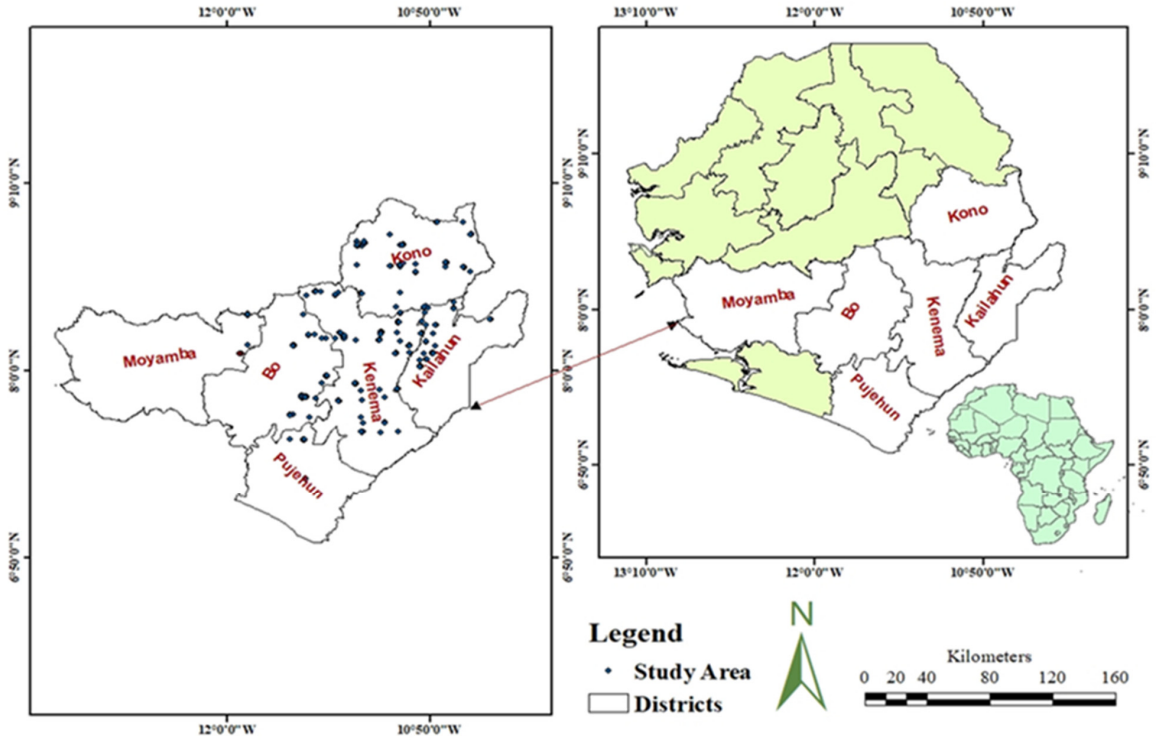
pest and disease resistance under intensive cultivation. Similar improvements in yield as well as other traits are needed in other countries so that there will be enough supply for the world's increasing demand for dry cacao beans. Therefore, it is critical that cacao germplasms available in other West African cacao-producing countries be characterized, documented, identified, conserved, and utilized. This will allow searches for promising unique genetic materials that can be used to develop strong, local breeding programs in areas that currently depend on Cote d'Ivoire, Ghana, or Nigeria for improved varieties or planting materials.

Assessments of genetic diversity and identification of mislabeled accessions in germplasm collections have been conducted by using single nucleotide polymorphism (SNP) markers in several crops, a fast, high-throughput, and affordable tool for whole-genome genetic diversity analysis. SNPs have been successfully used to characterize crops such as maize [21] and soybean [22]. Similarly, SNP markers have been used for fingerprinting cacao germplasm collections in several studies [23–27]. In West Africa, Takrama et al. [28] used 54 SNPs to fingerprint 160 cacao trees from the germplasm collection at the Cocoa Research Institute of Ghana (CRIG) for accurate identification of individual genotypes. However, high-throughput sequence-based KASP (kompetitive allele-specific PCR)-SNPs have been used in only two studies for cacao germplasm from West Africa [17,18]. KASP–SNP is an effective method compared to traditional SNP genotyping using electrophoresis systems because of a low genotyping error rate, cost-effectiveness, and flexibility to automation [29]. In this present study, a subset of 20 KASP–SNPs from a set of 100 KASP–SNPS used in genotyping cacao genotypes from Nigeria and Ghana was used for genotypic characterization of cacao from Sierra Leone and Togo. This is the first study to use a subset of 20 selected KASP–SNPs to understand the extent of genetic diversity and population structure within and among cacao germplasms.

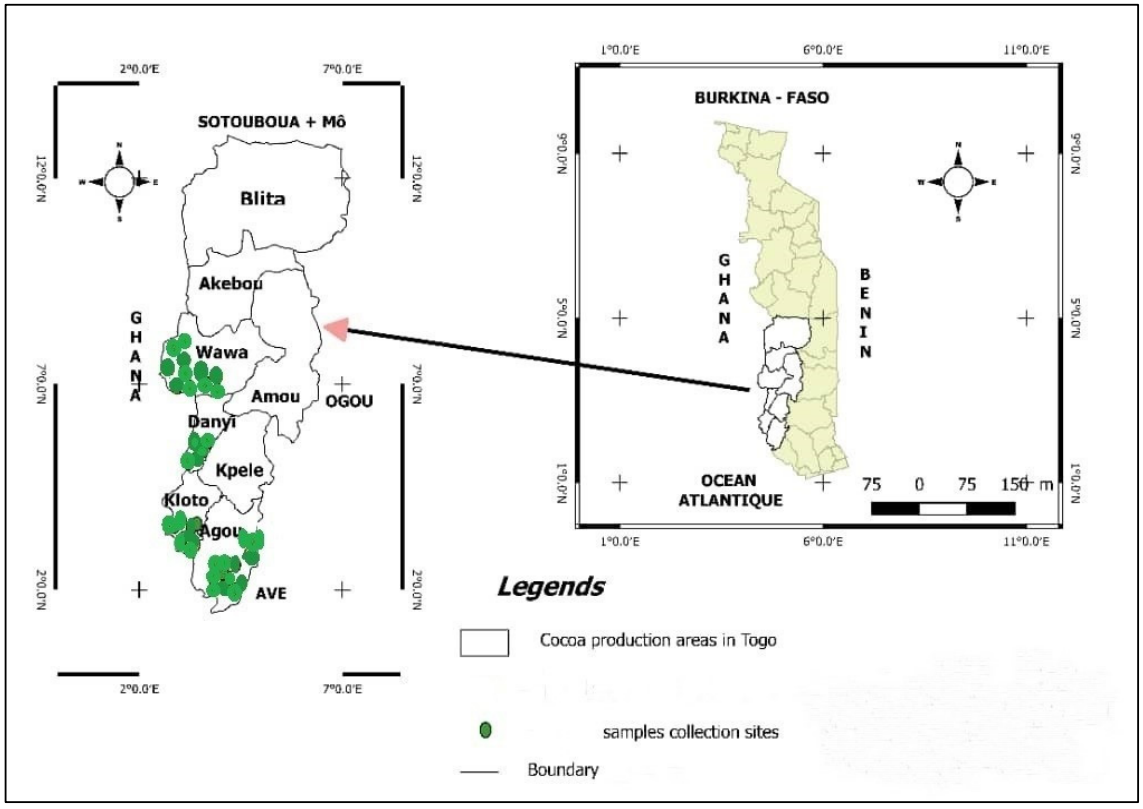
2. Materials and Methods

2.1. Cacao Sampling

In Sierra Leone and Togo, we sampled 235 and 141 cacao accessions, respectively, from different cacao growing regions (Table S1a,b). In Sierra Leone, leaf samples were collected from cacao farms operating across the three major cacao-producing districts: Kenema (N: 8°02.805′, W: 11°02.032′), Kailahun (N: 8°21.570′, W: 10°23.477′), and Kono (N: 8°36.196′, W: 10°56.458′). In addition, leaf samples were collected from minor cacao-producing districts such as Bo (N: 8°10.653′, W: 11°41.720′) and Pujehun (N: 7°08.684′, W: 11°22.652′). Leaf samples were also collected from trees in the Njala University research garden in the Moyamba district (N: 8°06.493′, W: 12°04.950′) (Figure 1a). Similarly, in Togo, leaf samples were collected from 15 cacao plantations located in the areas of Agou (N: 6°49′18″, E: 0°52′04″), Kloto (N: 7°04′, E: 0°44′), Danyi (N: 6°49′40.2″, E: 0°43′07″), and Litimé (N: 8°06′; E: 1°00′) (Figure 1b). Individual trees were tagged and geo-referenced from which fresh young cacao leaf samples were collected and dried using silica gel (note: the individual tagged trees are maintained in the seed gardens or farmers' farms; no trees were destroyed during this study). The dried leaf samples were then shipped to the Bioscience Center of the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, where six leaf discs of approximately 5 mm in diameter from dried leaf samples of each genotype were punched into labeled 8-strip 1.1 mL propylene tubes with strip caps, with up to 12 strips placed in 96-well boxes (Thistle Scientific, Warwickshire, UK). A total of four labeled 96-well boxes were then shipped to a genotyping service provider (Intertek, Kista, Sweden) for automated DNA extraction and genotyping with 20 SNP markers using the KASP assay. Two blank controls were included in each box during genotyping.



(**a**)



(**b**)

Figure 1. (**a**). Map of Sierra Leone showing the districts and locations where leaf samples were collected (note: map generated using ArcMap 10.8.2 (Esri ArcGIS Desktop) and GPS information collected during sampling). (**b**). Map of Togo showing the districts and locations where leaf samples were collected (note: map generated using QGIS 3.32.3 and the GPS information collected during sampling).

2.2. DNA Extraction, Preparation and Genotyping

DNA quality and quantity were checked on a 0.8% agarose gel. We used 20 high-quality SNPs to genotype the cacao leaf samples. These 20 SNPs were chosen from a set of 1536 SNPs identified from previous studies that used expressed sequence tags (ESTs) of a wide range of cacao tissue and organs displaying transcriptomic differences [30,31]. A subset of 100 SNPs was then selected from the initial set based on call rate, representativeness across the ten chromosomes and heterozygosity, and their use by cacao researchers in past studies [25,32,33]. The selected 100 SNPs were then converted using KASP™ assays at LGC Genomics (http://www.biosearchtech.com/support/education/kasp-genotyping-reagents) (accessed on 20 August 2022) for earlier studies on cacao genetic diversity in West Africa [17,18]. For this present study, 20 highly polymorphic, high-quality KASP–SNPs (Table S2) were carefully selected from the original 100, while genotyping was carried out at Intertek, Sweden. We performed simulations on 100 KASP–SNPs to select these 20, which are routinely used in cacao breeding for low-density genotyping at Intertek. The KASP assay protocol followed the KASP manual [34] in which genotyping was carried out with high-throughput PCR SNPline workflow by using 1 µL reaction volume in 384-well PCR plates. The KASP genotyping reaction mix consisted of three components, including sample DNA (10 ng), marker assay mix comprising target-specific primers, and KASP-TF™ master mix containing two universal fluorescence resonant energy transfer cassettes (FAM and HEX), passive reference dye (ROX™), Taq polymerase, free nucleotides, and $MgCl_2$ in an optimized buffer solution. The SNP assay mix is specific to each marker and consists of kompetitive allele-specific forward and reverse primer. After PCR, the plates were fluorescently read, and allele calls were made by using KRAKEN™ software (LGC Biosearch Technologies, Hoddesdon, UK) and scored on a Cartesian plot, (cluster plot), assigning each DNA sample to a class: homozygous for either allele 1 or 2 or heterozygous in the case of biallelic SNPs.

2.3. Data Analyses

For data analyses, all cacao genotypes across both target countries were combined with an assumption that genotypes with the same name may be genetically different and genotypes with different names may be genetically similar. Genetic diversity analyses based on minor allele frequency (MAF), polymorphism information content (PIC), expected heterozygosity (He), and observed heterozygosity (Ho) parameters were conducted with vcftools and plink 1.9 [35]. A genetic distance matrix-based identity by state (IBS) generated with plink 1.9 [35] served as the basis for hierarchical clustering. An unrooted phylogenetic tree was constructed to visualize how closely accessions were related within each country by using the ape (analyses of phylogenetics and evolution) library package [36] and phangorn, an R package [37]. The dissimilarity matrix was then used to construct network relationships among cacao accessions from both Sierra Leone and Togo with QGRAPH [38] implemented in R. In addition, as a complementary analysis, Discriminant Analysis of Principal Component (DAPC) was carried out using 'genind object' and the find.clusters function in the adegenet package [39]. In order to properly assign the accessions to groups, the Bayesian information criterion (BIC) was used to assign accessions to groups and determine the optimal number of clusters to be retained. A binary file was generated from the filtered VCF file and was then subjected to cross-validation for population structure analysis [35]. A cut-off value of 50% ancestry suggested through the Admixture analysis was used to estimate membership probabilities of all accessions for the groups identified [36]. The model-based clustering approach implemented in ADMIXTURE assumes linkage equilibrium among loci and Hardy–Weinberg equilibrium within ancestral populations [40]. However, such assumptions may not apply in clonally propagated tree species like cacao due to the presence of clonal duplicates in germplasm collections. To validate the clustering pattern obtained from ADMIXTURE and hierarchical clustering algorithms, DAPC, an assumption-free multivariate clustering method, was used. Genetic differentiation among and within groups (individual country level and combined) was

estimated via an analysis of molecular variance (AMOVA), and its significance was tested with a non-parametric approach with 999 permutations by using GenAlex v. 6.503 [41]. Coefficients of genetic differentiation among populations (Sierra Leone, Togo, and combined) were calculated based on pairwise F_{ST} (fixation index) to estimate genetic distances and relationships among populations used in this study.

3. Results

3.1. Cacao Accessions from Sierra Leone and Togo

The cacao accessions collected from both Sierra Leone and Togo represented several clones with the same name (Table S3) even when they were collected from different trees either at the same farm or from different farms. The GPS coordinates as well as the village/farm names and district names for each genotype are provided in Tables S1a and S1b. Most of the cacao genotypes from Sierra Leone with similar names were collected from trees located at different farms, indicating that these trees either belong to the same clone or there are issues of mislabeling. The situation is the same in Togo, although genotypes with the same names were present at the same farm as well as at different farms.

The 235 cacao genotypes from Sierra Leone represented 144 accessions with unique names. These 144 accessions were sourced from Ghana (66 accessions) or the University of Reading (77 accessions), while the source of one genotype was unknown (Table S3). Similarly, the genotypes from Togo represented 77 accessions with unique names. These seventy-seven accessions were sourced from Cameroon (four accessions), Cote d'Ivoire (eleven accessions), Ghana (seventeen accessions), Nigeria (twenty-eight accessions), and Togo (nine accessions), with a few clones representing both Cameroon and Ghana (ICS 6 × NA33; ICS16), Cote d'Ivoire and Cameroon (IFC1 × SNK13), Ghana and Cote d'Ivoire (NA2), or Ghana and Nigeria (SCA 12 × NA32; C23; C26 × SCA 6), and five genotypes whose accession names were lost (Table S3). For the accessions with lost names, it is assumed that these may represent PA7/A19 clones (Table S3). There are eleven accessions/accession names common between both Sierra Leone and Togo, of which PA7, ICS60, IMC47, Na34, and Pa35 were sourced from the University of Reading and Nigeria in Sierra Leone and Togo, respectively. C20 and C70 listed Ghana as the source country for both Sierra Leone and Togo; C26 and C42 listed Ghana and Nigeria as the source country for Sierra Leone and Togo, respectively; C23 listed Ghana for Sierra Leone and Nigeria/Ghana for Togo as the source country; Pound 7 listed the Gene Bank of SLARI and Ghana as the source for Sierra Leone and Togo, respectively; and PA7 listed both the Gene Bank of SLARI and Nigeria as the source (Table S3).

3.2. Genetic Diversity Parameters

The 20 KASP–SNP markers used herein were distributed across the ten chromosomes of cacao, with two SNPs on each chromosome (Table S2). The average PIC (polymorphic information content), He (expected heterozygosity), Ho (observed heterozygosity), and MAF (minor allele frequency) values for 235 cacao accessions collected in Sierra Leone were 0.22, 0.30, 0.24, and 0.21, respectively, while for 141 accessions collected in Togo, these values were 0.21, 0.29, 0.22, and 0.19, respectively. For the combined population of 376 cacao accessions, the average PIC, He, Ho, and MAF values were 0.24, 0.30, 0.26, and 0.23, respectively (Table 1). Table S4 and Figure S1 present summary statistics and the distribution of He, Ho, MAF, and PIC values for the combined population, respectively. Of the 20 SNPs used in this study, many had a minor allele frequency (MAF) above 0.2 and showed high PIC values with a peak distribution above 0.2. Similarly, low observed and expected heterozygosity was recorded. The observed heterozygosity for the combined population was higher than individual country-level observed heterozygosity, which was also represented by higher PIC values (Table 1).

Table 1. Descriptive statistics based on 20 KASP–SNP markers across cacao accessions.

Cacao Population	N	He	Ho	MAF	PIC
Sierra Leone	235	0.30	0.24	0.21	0.22
Togo	141	0.29	0.22	0.19	0.21
Combined	376	0.30	0.26	0.23	0.24

N: number of accessions.

3.3. Population Structure and Genetic Relationships

The model-based population structure analysis of the combined cacao population (235 accessions from Sierra Leone and 141 from Togo) showed that the delta K values from the mean log-likelihood probabilities stagnated at K = 4 (Figure 2a), although the optimal number of clusters obtained initially was K = 2 (Figure S2). The 376 cacao accessions were divided into four subpopulations at K = 4 (Figure 1b). Based on an 80% membership probability threshold, 270 accessions (73.37%) were successfully assigned to the four sub-populations. In comparison, 106 accessions with a probability of <80% were designated as an admixed population (Table S3). Sub-population 1 consisted of 123 accessions (Sierra Leone: 94 accessions; and Togo: 29 accessions). Sub-populations 2, 3, and 4 constituted 3.72%, 22.07%, and 13.30% of the accessions, respectively, with ten accessions from Sierra Leone and four accessions from Togo in sub-population 2, thirty-six accessions from Sierra Leone and forty-seven accessions from Togo in sub-population 3, and twenty-six accessions from Sierra Leone and twenty-four accessions from Togo in sub-population 4 (Table S3, Figure 1b). The admixed group consisted of 69 accessions from Sierra Leone and 37 accessions from Togo. Few additional smaller peaks observed (Figure 2b) implied the presence of subgroups within the four major subpopulations. Therefore, individual STRUCTURE analysis was performed for accessions representing Sierra Leone and Togo. Sub-clustering of cacao germplasms from Sierra Leone and Togo showed that delta K values stagnated at K = 3 and K = 4, respectively (Figure 2a,b). A higher degree of admixture was observed in the cacao germplasm from Togo than in the cacao germplasm from Sierra Leone. The 235 cacao accessions from Sierra Leone were divided into three subpopulations with 102 accessions in subpopulation 1, 29 in subpopulation 2, and 93 in subpopulation 3; whereas, only 11 accessions (4.7%) were in an admixed group (Table S3; Figure 3a). In contrast, one hundred and forty-one cacao accessions from Togo were grouped into four subpopulations, with four accessions in subpopulation 1, thirty-six in subpopulation 2, fifty in subpopulation 3, and twenty-five in subpopulation 4. The admixed group consisted of 26 accessions, representing 18.4% admixture among the accessions (Table S3; Figure 3b).

Using the Bayesian information criterion (BIC) implemented in DAPC, a maximum of K = 4 was obtained, which corresponded to the four groups obtained for the combined population (Figure 4) and for the germplasm collection in Togo (Table S3). Designation of cluster membership showed that cluster 3 included the largest number of accessions (158), followed by cluster 2 with 116 accessions and cluster 1 with 86 accessions, with cluster 4 having the fewest accessions (16). Of the 158 accessions in cluster 3, 105 accessions (66.5%) were from Sierra Leone and 53 (33.5%) from Togo (Table S3). Cluster 2 included 51 accessions from Sierra Leone and 65 accessions from Togo, whereas cluster 1 had 68 accessions from Sierra Leone and 18 accessions from Togo. The smallest cluster, cluster 4, represented eleven accessions from Sierra Leone and five from Togo.

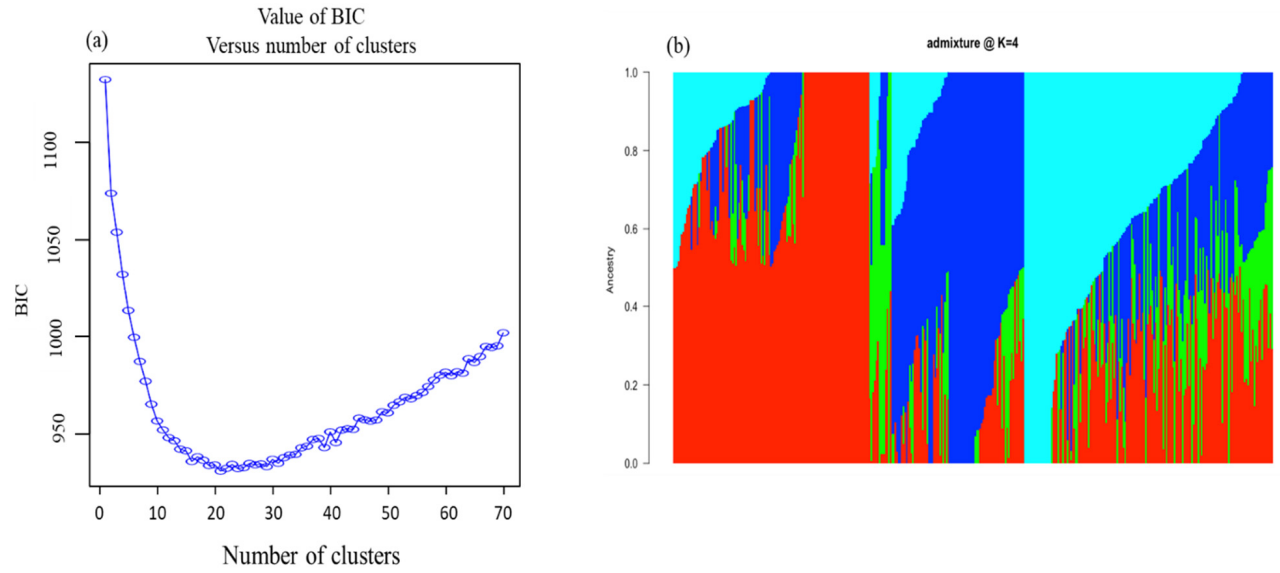


Figure 2. Graphical representation of the population structure of 376 cacao accessions. (**a**) Plot of mean likelihood of delta K against the number of K groups. (**b**) Subpopulations at K = 4. The colors represent four subpopulations of 376 accessions: red: subpopulation 1, blue: subpopulation 2, light blue: subpopulation 3, and green: subpopulation 4.

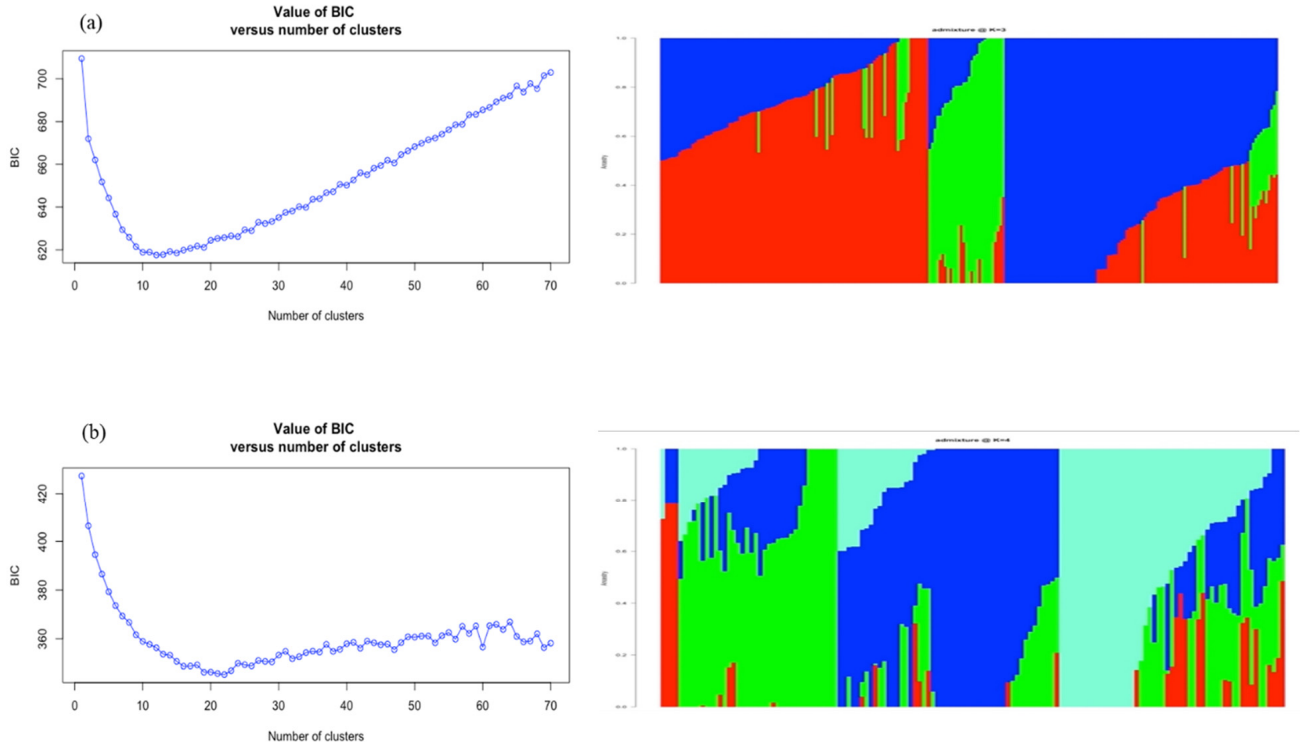


Figure 3. Population structure (**a**) Sierra Leone: plot of mean likelihood of delta K against the number of K groups; subpopulations at K = 3. The colors represent three subpopulations of 235 cacao accessions: red—subpopulation 1, green—subpopulation 2, blue—subpopulation 3. (**b**) Togo: plot of mean likelihood of delta K against the number of K groups; subpopulations at K = 4. The colors represent four subpopulations of 141 cacao accessions: red—subpopulation 1, green—subpopulation 2, blue—subpopulation 3, turquoise—subpopulation 4.

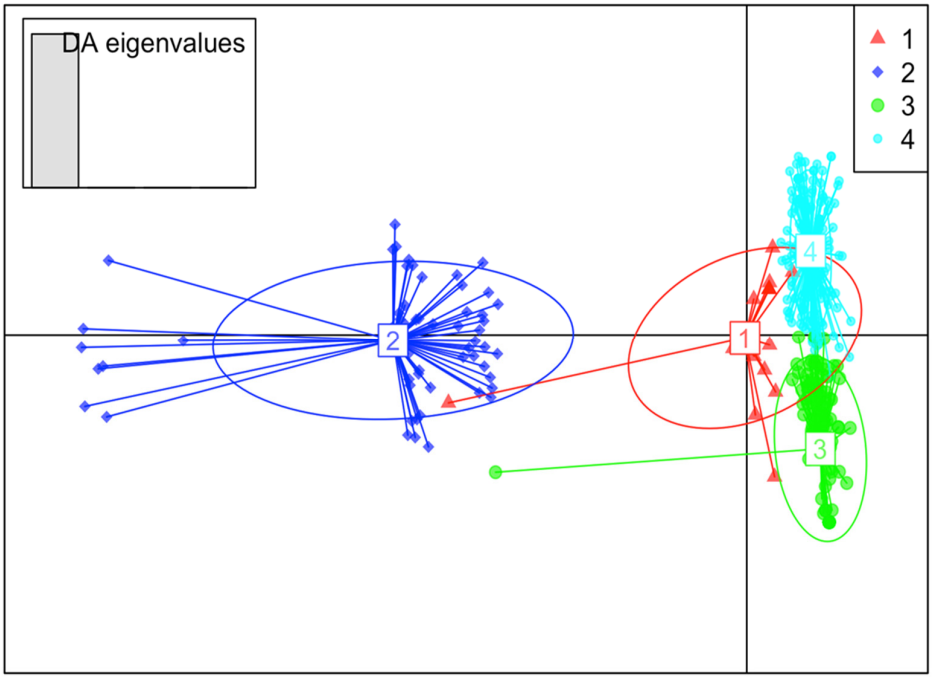


Figure 4. Discriminant analysis of principal components (DAPC) using 20 KASP–SNP markers. The axes represent the first two linear discriminants (LD). Each color represents a cluster, while each dot represents an individual. Numbers represent different subpopulations identified by DAPC analysis.

Contrary to the results of STRUCTURE and DAPC, the hierarchical clustering assigned all 376 cacao accessions to two major clusters with several sub-clusters representing a higher degree of admixture among accessions from both countries (Figure 5a). The hierarchical clustering performed for cacao accessions from Sierra Leone and Togo independently is displayed in Figure 5b. The cacao accessions from Sierra Leone split into three main clusters (Figure 5b(i)), while those from Togo divided into four main clusters (Figure 5b(ii)) (Table S3), consistent with results obtained from our STRUCTURE and DAPC analyses. For Sierra Leone, cluster 1 was the largest, with 119 cacao accessions, while 72 accessions and 44 accessions were grouped into clusters 2 and 3, respectively. For Togo, cluster 3 was the largest, with 52 cacao accessions followed by cluster 1 (46 accessions), cluster 2 (29 accessions), and cluster 4 (14 accessions) (Table S3). A network analysis between cacao accessions from Sierra Leone and Togo (Figure 6) showed strong genetic relationships, indicating that these accessions share a similar genetic background across the two countries. The central core of the QGRAPH (Figure 6) represents a set of accessions collected from Sierra Leone and Togo that are genetically similar to each other. While some peripheral genotypes depict more genetic divergence, this was more evident for accessions from Sierra Leone than for those from Togo.

A comparison of all three methods (STRUCTURE, DAPC, and hierarchical clustering) did not reveal any identical clustering patterns among these 376 accessions, except for two accessions that were both from Sierra Leone with the same accession name (C77) (Figure S2; Table S2). However, the DAPC and hierarchical clustering showed similar patterns of clustering among 135 out of 376 cacao accessions, in which one hundred and twenty-seven accessions were from Sierra Leone and eight accessions were from Togo. Meanwhile, the DAPC and STRUCTURE analyses showed similar patterns of grouping among 61 out of 376 cacao accessions (30 from Sierra Leone and 31 from Togo) (Figure S3).

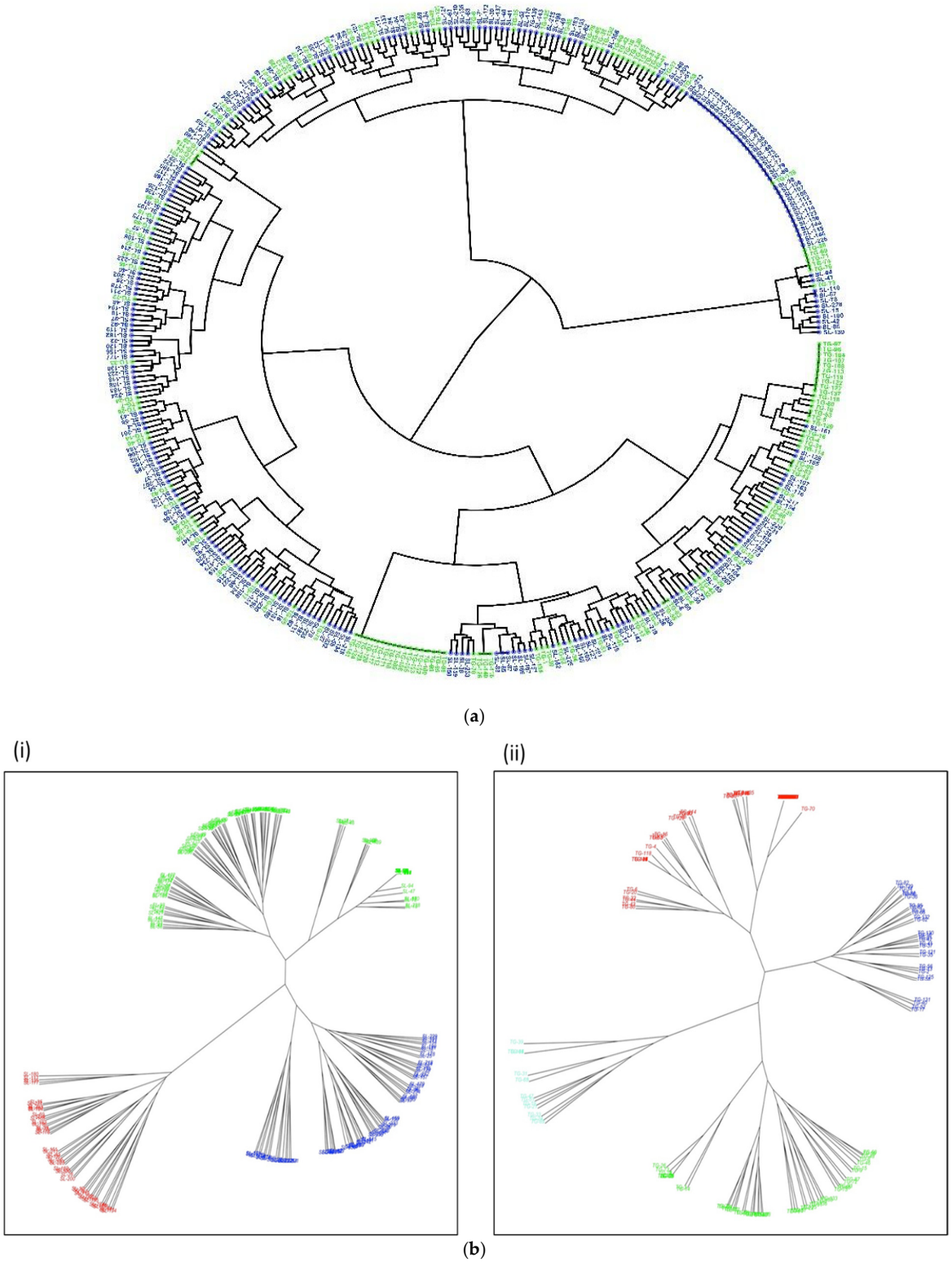


Figure 5. (**a**) Phylogenetic tree for 376 cacao accessions from Sierra Leone and Togo. The blue and green colors represent accessions from Sierra Leone and Togo, respectively. (**b**) Phylogenetic tree depicting genetic relationships among cacao accessions from (i) Sierra Leone and (ii) Togo. Colors of each tree represent genetic groups within each country.

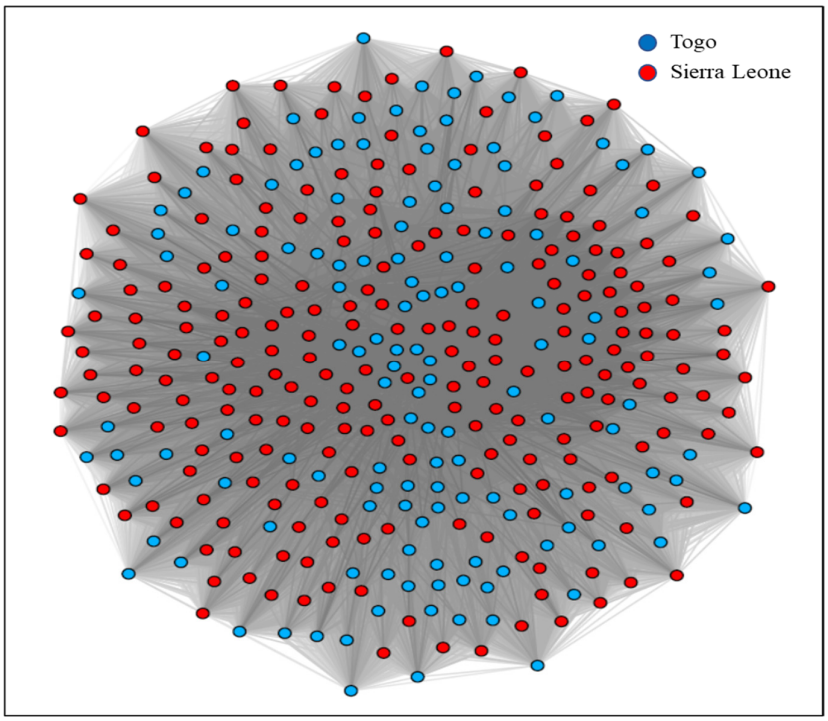


Figure 6. Genetic networks obtained by using QGRAPH. This diagram visualizes networks among cacao accessions from Sierra Leone and Togo, with the node size depicting genetic relationships among different accessions based on observed heterozygosity and allelic richness.

3.4. Analyses of Molecular Variance and Genetic Differentiation

The AMOVA analysis revealed a variability of 86% within accessions and 14% among populations (Combined data) (Table 2). The overall F_{ST} value was 0.601. A significant level of population divergence based on pairwise F_{ST} ($p < 0.0001$) was also observed among different populations, while strong genetic relationships with much less divergence were observed within populations (Table 3). The average F_{ST}-based population differentiation was highest for the combined population (0.096) and lowest for cacao accessions from Sierra Leone (0.049). The pairwise F_{ST} values ranged from 0.010 (combined population vs. Togo) to 0.045 (Sierra Leone vs. Togo).

Table 2. Analysis of molecular variance (AMOVA) among and within three different genetic populations (Sierra Leone, Togo, and combined).

Source	d.f.	SS	MS	Est. Var.	% Var.	*p* Value
Among populations	3	189.16	63.05	0.362	14	0.001
Among accessions	372	690.62	1.86	0.000	0	0.001
Within accessions	376	809.50	2.15	2.153	86	0.001
Total	751	1689.28		2.515	100	
Fixation index (F_{ST})	0.601					0.001

d.f.: Degree of freedom, SS: Sum of squares, MS: Mean sum of squares, Est. Var.: Estimated variance, % Var.: Percent Variance.

Table 3. Pairwise fixation index (F_{ST}) values within and among cacao accessions collected from Sierra Leone and Togo.

	Sierra Leone	Togo	Combined Population
Sierra Leone	−0.0024		
Togo	0.0451	−0.0040	
Combined Population	0.0201	0.0101	−0.0015
Average F_{ST}	0.049	0.073	0.096

4. Discussion

The results revealed that genotyping cacao germplasms collected from Sierra Leone and Togo by using 20 KASP–SNPs was efficient in assessing the genetic diversity and population structure. These KASP–SNPs represented a subset of 100 KASP–SNPs carefully selected based on their distribution across twenty cacao chromosomes (two SNPs per chromosomes), polymorphic nature, reproducibility, and higher efficiency from earlier studies used in fingerprinting cacao germplasms from Ghana [17], Nigeria [18], and other West African countries (unpublished). The biallelic nature of these selected SNPs has a lower error rate in allele calling, with higher accuracy and efficiency [17].

The germplasms studied represent collections maintained in different seed gardens as well as farmers' fields in both countries. Currently, cacao cultivation is unstructured in Sierra Leone, with cacao trees labeled as 'Forestero' being the main variety cultivated there [11]. However, other varieties known as 'Amazon cacao', 'Ghanian cacao', and 'Ivorian cacao' are also grown, with a significant exchange of planting materials among the farmers. The situation is similar in Togo. A thorough assessment of the available genetic diversity and population structure is necessary for the enhancement and effective use of well-characterized, diverse germplasms in the cacao improvement programs of both countries.

In this present study, diversity indices revealed the presence of substantial genetic diversity in cacao germplasm from both Sierra Leone and Togo indicated by average He (0.30 and 0.29, respectively) and Ho (0.24 and 0.22, respectively), whereas the PIC value was 0.24. Similar He values were reported for cacao collections from Nigeria [18] and Ghana [17], although those studies recorded higher PIC values and observed heterozygosity than was observed in our study. This difference could be explained by the number or selection of KASP–SNPs used or may reflect real differences among the populations. Nonetheless, average He and Ho values for the combined (Sierra Leone and Togo) population in this current study were comparable to those obtained for each country individually, indicating that the cacao germplasms present in both countries may share the same genetic background. This study did reveal that eleven accession names are common between both countries and that there is a presence of mislabeled cacao accessions in seed gardens and farmers' fields, as has already been reported in most West African countries [15–17,28]. It is also clear that a common set of cacao genotypes is shared by major cacao-producing West African countries, i.e., Cote d'Ivoire, Ghana, Nigeria, and Cameroon, likely extending to other smaller countries in the region, suggesting that mislabeled and duplicated cacao germplasms may be shared more widely than previously reported.

Our assessment of genetic relatedness, based on three different analyses (model-based population structure, IBS-based clustering, and DAPC), revealed that the combined population of accessions from Sierra Leone and Togo and the accessions specifically from Togo were composed of four main subpopulations, whereas the one from Sierra Leone was composed of three main subpopulations. The fact that this clustering was supported by a Bayesian approach, a genetic distance-based method, and a DAPC-based analysis, provides strong support for the observed population structure and genetic relationships among accessions.

In studies where reference genotypes are not used and historical pedigree data are unavailable, ancestry information can still provide a framework for determining admixtures and the contributions of genotypes to open pollination or any other sort of natural hybridization [42]. The fact that no common clustering was observed across the three clustering methods used in our study could be because DAPC revealed more clusters than ADMIXTURE, but the latter method assigned genotypes based on ancestries. The DAPC approach relies on discriminant functions that seek to maximize divergence between clusters while minimizing within-cluster diversity [39]. Similar inconsistency has been observed in other studies, such as cassava [42]. In clonally propagated crops such as cacao, genotypes represent complex inter-generational hybridization or open pollination, resulting in complex genetic relationships and clinal patterns of genetic differentiation [39]. Still, there is a

general agreement in cluster assignment across all three methods, specifically for DAPC and ADMIXTURE approaches, wherein >75% genotypes were assigned to specific clusters.

It is worth noticing that in our combined analysis, most cacao accessions (73.37%) were assigned to one of the four subpopulations with probabilities >0.8, and only 106 of 376 cacao accessions (28.19%) were classified as admixtures. In a recent study of cacao germplasm from Nigeria, the rate of admixtures or off-types ranged from 10% to 73% among the clones in seed gardens [18]. Such admixtures in cacao germplasm from both Sierra Leone (4.7%) and Togo (18.4%) probably reflect genotypes with different allelic patterns and labeling errors present in the introduced germplasms from other West African countries. This is also supported by our findings that mislabeling was less frequent in accessions from Sierra Leone (33.33%) than in those from Togo (45.45%). Furthermore, the presence of a higher level of admixed accessions in cacao genotypes from Togo may also reflect recent breeding advances involving open pollination and bi-parental crossing among accessions coupled with strong selection pressure, as observed in other clonally propagated crops [43], which is not the case in Sierra Leone. Labeling errors pose a higher risk of misidentification of clones, which proliferates when beans/pods from such clones are shared or used to establish new seed gardens, when neighboring farmers share materials with each other, or when farmers import materials from neighboring countries without understanding their genetic potential, which is a common practice in both Sierra Leone and Togo. Olasupo et al. [18] explained how the presence of 58% mislabeled accessions in an old seed garden translated into 100% mislabeling in a newly established seed garden in Nigeria. The presence of mislabeling in germplasms from the University of Reading in Sierra Leone and at the seed gardens of the National Institute in Togo indicates a loss of labels or human errors in recording them, occurrences commonly observed in clonally propagated crops with long life cycles such as tree crops [17,18]. There are serious implications when mislabeled or misidentified germplasm is introduced, as it often leads to poor predictions of their performance or their value in improvement programs. This may be one explanation for relatively low yields among cacao accessions introduced from neighboring countries, in addition to the prevalence of pests and diseases.

Network analysis was used to unravel the genetic relationships among the cacao accessions from Sierra Leone and Togo. In the absence of pedigree records or comparisons to international reference clones, the dissection of genetic relationships among cacao accessions from these two countries through network analysis was a worthwhile approach. Network analysis has been successfully used in other clonally propagated food crops such as cassava [42] and white yam [44]. In tree crops such as cacao, which is an outbreeding species with a long-life cycle, open pollination is common, with the source of pollen typically unknown. The extent of genetic diversity observed within cacao accessions in our study can, in part, be attributed to this factor as well as to the presence of a good number of unique accessions in both countries. Hence, these unique cacao accessions could be considered potential parents for local cacao improvement once preliminary agronomic evaluations and trait profiling have been conducted in multiple sites. There is also a need for field conservation of unique cacao accessions in both countries, with long-term support for best management practices and irrigation facilities to mitigate losses associated with biotic and abiotic stresses. It may also be necessary to mainstream DNA fingerprinting of introduced germplasm from neighboring countries and international collections for the regular auditing of cacao accessions for their true-to-typeness and also to check for pollen contamination during hybridization through open or manual pollination.

The low level of genetic diversity of the cacao germplasm historically introduced into West Africa is well known [13,17,18], a situation further compromised by mislabeling and unintentional duplication in both seed gardens and farmers' fields. Mislabeling (the use of the same name for different genotypes or the use of different names for identical genotypes) is common in many clonally propagated crops [42,44,45]. The number of SNPs used herein was relatively low but based on their high polymorphic, informative nature;

our results provide a first look at these limitations within the cacao germplasm of Sierra Leone and Togo.

We suspect that additional studies based on the use of a larger set of SNPs may provide a deeper look into the sub-populations observed in our study, but patterns of genetic diversity and mislabeling may change little. When possible, the original donors should be contacted to determine the correct application of accession names through genotyping original source material.

5. Conclusions

In this study, we made use of 20 KASP–SNP markers to assess the genetic diversity of cacao germplasms from Sierra Leone and Togo. The genetic relationships elucidated among the accessions in each country as well as the identification of mislabeling have provided key information to support future cacao improvement by identifying diverse and correctly named parents. This study also confirmed the reliability and accuracy of KASP–SNPs generated from next-generation sequencing-based genotyping coupled with complementary statistical analyses to generate knowledge on genetic diversity and population structure. In this study, we identified a high degree of mislabeling in most of the introduced materials, which has been attributed to errors from the sources of introduction, labeling errors, and lost labels.

Further detailed research is needed for multi-location phenotyping as well as genotyping with high-density DNA markers or the whole-genome re-sequencing of cacao germplasm in Sierra Leone and Togo to validate and refine our initial findings. The presence of duplicates/mislabeling has serious consequences for improvement programs, and caution should be taken to ensure future accuracy in labeling and consistent identification of clones/accessions before establishing seed gardens or distributing planting materials to farmers. It is recommended that accession naming in both countries follow the International Cocoa Germplasm Database convention [46] and maintain the International Cocoa Quarantine Center, Reading (ICQC, R) names when acquiring germplasms from the University of Reading or other major collections.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy14112458/s1, Table S1. (a) List of cocoa genotypes from Sierra Leone with clone name, GPS information, district name, chiefdom name and village/farm name. Table S1 (b) List of cocoa genotypes from Togo with clone name, GPS information, district name, chiefdom name, and village/farm name; Table S2. Details of 20 KASP–SNPs with their IDs and their position in each linkage group; Table S3. Cacao accessions from Sierra Leone, their grouping based on population structure, DAPC and Hierarchical Clustering (HC), and the source of introductions of cacao germplasm in the two countries; Table S4. Descriptive statistics of He, Ho, MAF, and PIC values for the combined population of 376 cacao accessions. Figure S1. Distribution of He, Ho, MAF, and PIC values for the combined population (376 cacao accessions); Figure S2. Graphical representation of the population structure of 376 cacao accessions. Plot of mean likelihood of delta K against the number of K groups. Figure S3. Comparison of three complementary approaches: STRUCTURE, DAPC, and Hierarchical Clustering for grouping 376 cacao accessions.

Author Contributions: Conceptualization, R.B., P.L.K. and L.J.G.-B.; Methodology, R.B., P.A.A., M.M.L. and K.A.; Data analysis, P.A.A., M.M.L. and K.A.; Writing original draft preparation, R.B.; Writing—review and editing, M.M.L., K.A., P.L.K., P.A.A. and L.J.G.-B.; Supervision, R.B., P.L.K. and L.J.G.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by a research grant from Swedish Research Council Development Vetenskapsrådet (Grant Number: 2017-05621) to L.G.B., R.B., and P.L.K. Additional financial support was received from Partnerskap Alnarp, Sweden for the KASP–SNP genotyping. The first author is grateful for the scholarship granted by the Swedish Research Council Development for his Ph.D. research at the Swedish Agricultural University.

Data Availability Statement: All data supporting the findings in this study are available within this article and supplementary files. The raw SNP data can be made available upon request to the corresponding author.

Acknowledgments: The authors acknowledge a Swedish Research Council Development research grant (2017-05621) to the Swedish University of Agricultural Sciences and the International Institute of Tropical Agriculture for funding this research. The authors acknowledge the services and support of Petra van Roggen and the team at Intertek for the KASP–SNP genotyping. The authors also acknowledge all the cacao farmers in Sierra Leone and Togo for their cooperation during leaf sample collection.

Conflicts of Interest: The authors declare no potential conflicts of interest or personal relationships that could have appeared to influence the research work reported in this paper.

References

1. Motamayor, J.C.; Risterucci, A.M.; Lopez, P.A.; Ortiz, C.F.; Moreno, A.; Lanaud, C. Cacao domestication I: The origin of the cacao cultivated by the Mayas. *Heredity* **2002**, *89*, 380–386. [CrossRef]
2. Davie, J.H. Chromosome studies in the Malvaceae and certain related families. *Genetica* **1935**, *17*, 487–498. [CrossRef]
3. Henderson, J.S.; Joyce, R.A.; Hall, G.R.; Hurst, W.J.; McGovern, P.E. Chemical and archaeological evidence for the earliest cacao beverages. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18937–18940. [CrossRef]
4. International Trade Centre. *Cacao: A Guide to Trade Practices*; International Trade Centre UNCTAD/WTO: Geneva, Switzerland, 2001.
5. Bhattacharjee, R.; Kumar, P.L. Cacao. In *Genome Mapping and Molecular Breeding in Plants*; Kole, C., Ed.; Springer-Verlag: Heidelberg/Berlin, Germany, 2007; pp. 127–142.
6. Bartley, B.G.D. *The Genetic Diversity of Cacao and Its Utilization*; CABI Publishing: Wallingford, UK, 2005.
7. ICCO. ICCO Quarterly Bulletin of Cocoa Statistics, vol XLVII. Cocoa Year 2020/2021. Available online: https://www.icco.org/icco-documentation/quarterly-bulletin-of-cocoa-statistics/ (accessed on 12 September 2023).
8. ICCO. ICCO Quarterly Bulletin of Cocoa Statistics, vol XLVIII. Cocoa Year 2021/2022. Available online: https://www.icco.org/icco-documentation/quarterly-bulletin-of-cocoa-statistics/ (accessed on 12 September 2023).
9. FAOSTAT. Food and Agriculture Organization of the United Nations Database of Agricultural Production. FAO Statistical Databases. 2022. Available online: http://www.fao.org/faostat/ (accessed on 25 September 2023).
10. EU. Ex-Post Evaluation of Agriculture for Development (A4D). 2019. Available online: https://www.eeas.europa.eu/sites/default/files/a4devalfinalreport.pdf (accessed on 14 September 2023).
11. Moinina, A.; Lahlali, R.; Boulf, M. Management practices to improve the cocoa bean value chain in Sierra leone: Cocoa bean production in Sierra Leone. *Moroc. J. Agric. Sci.* **2023**, *4*, 45–52.
12. Chingadu, N.; Kouakou, K.; Aka, R.; Ameyaw, G.; Gutierrez, O.A.; Hermann, H.-W.; Brown, J.K. The proposed new species, cacao red vein virus, and three previously recognized badnavirus species are associated with cacao swollen shoot disease. *Virol. J.* **2017**, *14*, 199. [CrossRef]
13. Aikpokpodion, P.O.; Motamayor, J.C.; Adetimirin, V.O.; Adu-Ampomah, Y.; Ingelbrecht, I.; Eskes, A.B.; Schnell, R.J.; Kolesnikova-Allen, M. Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genet. Genomes* **2009**, *5*, 699–711. [CrossRef]
14. Howes, F.N. The early introduction of cacao to West Africa. *Afr. Aff.* **1969**, *45*, 152. [CrossRef]
15. N'Goran, J.A.K.; Laurent, V.; Risterucci, A.M.; Lanaud, C. The genetic structure of cacao populations (*Theobroma cacao* L.) revealed by RFLP analysis. *Euphytica* **2000**, *115*, 83–90. [CrossRef]
16. Aikpokpodion, P.O.; Kolesnikova-Allen, M.; Adetimirin, V.O.; Guiltinan, M.J.; Eskes, A.B.; Motamayor, J.C.; Schnell, R.J. Population structure and molecular characterization of Nigerian field genebank collections of cacao, *Theobroma cacao* L. *Silvae Genet.* **2010**, *59*, 273–285. [CrossRef]
17. Padi, F.K.; Ofori, A.; Takrama, J.; Djan, E.; Opoku, S.Y.; Dadzie, A.M.; Bhattacharjee, R.; Motamayor, J.C.; Zhang, D. The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. *Tree Genet. Genomes* **2015**, *11*, 1–14. [CrossRef]
18. Olasupo, F.O.; Adewale, D.B.; Aikpokpodion, P.O.; Muyiwa, A.A.; Bhattacharjee, R.; Gutierrez, O.A.; Motamayor, J.C.; Schnell, R.J.; Ebai, S.; Zhang, D. Genetic identity and diversity of Nigerian cacao genebank collections verified by single nucleotide polymorphisms (SNPs): A guide to field genebank management and utilization. *Tree Genet. Genomes* **2018**, *14*, 32. [CrossRef]
19. Adenuga, O.O.; Ariyo, O.J. Diversity analysis of cacao (*Theobroma cacao*) genotypes in Nigeria based on juvenile phenotypic plant traits. *Int. J. Fruit Sci.* **2020**, *20*, S1348–S1359. [CrossRef]
20. Bekele, F.; Phillips-Mora, W. Cocoa Breeding. In *Advances in Plant Breeding: Industrial and Food Crops*; Al-Khayri, J., Jain, S., Johnson, D., Eds.; Springer-Verlag: New York, NY, USA, 2019; pp. 409–487.

21. Van Inghelandt, D.; Melchinger, A.E.; Lebreton, C.; Stich, B. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* **2010**, *120*, 1289–1299. [CrossRef] [PubMed]
22. Liu, Z.; Li, H.; Wen, Z.; Fan, X.; Li, Y.; Guan, R.; Guo, Y.; Wang, S.; Wang, D.; Qiu, L. Comparison of genetic diversity between Chinese and American soybean (*Glycine max* L.) accessions revealed by high-density SNPs. *Front. Plant Sci.* **2017**, *8*, 2014. [CrossRef]
23. Motamayor, J.C.; Schnell, R.; Kuhn, D. Applying SNP marker technology in the cacao breeding programme in Ghana. *Afr. Crop Sci. J.* **2012**, *20*, 67–75.
24. Zhang, D.; Susilo, A.W.; Dinarti, D.; Bailey, B.A.; Mischke, S.; Meinhardt, L. Genetic identity, ancestry and parentage in farmer selections of cacao from Aceh, Indonesia revealed by single nucleotide polymorphism (SNP) markers. *Trop. Plant Biol.* **2014**, *7*, 133–143.
25. Livingstone, D.S.; Royaert, S.; Stack, C.; Mockaitis, K.; May, G.; Farmer, A.D.; Saski, C.; Schnell, R.J.; Kuhn, D.; Motamayor, J.C. Making a chocolate chip: Development and evaluation of a 6K SNP array for *Theobroma cacao*. *DNA Res.* **2015**, *22*, 279–291. [CrossRef]
26. Motilal, L.A.; Sankar, A.; Gopaulchan, D.; Umaharan, P. Cocoa. In *Biotechnology of Plantation Crops*; Chowdappa, P., Karun, A., Rajesh, M.K., Ramesh, S.V., Eds.; Daya Publishing House: New Delhi, India, 2017; pp. 313–354.
27. Mahabir, A.; Motilal, L.A.; Gopaulchan, D.; Ramkissoon, S.; Sankar, A.; Umaharan, P. Development of a core SNP panel for cacao (*Theobroma cacao* L.) identity analysis. *Genome* **2020**, *63*, 103–114. [CrossRef]
28. Takrama, J.; Kun, J.; Meinhardt, L.; Mischke, S.; Opoku, S.Y.; Padi, F.K.; Zhang, D. Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers. *Afr. J. Biotech.* **2014**, *13*, 227–2136.
29. Semagn, F.K.; Babu, R.; Hearne, S.; Olsen, M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol. Breed.* **2014**, *33*, 1–14. [CrossRef]
30. Argout, X.; Fouet, O.; Wincker, P.; Gramacho, K.; Legavre, T.; Sabau, X.; Risterucci, A.M.; Da Silva, C.; Cascardo, J.; Allegre, M.; et al. Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. *BMC Genom.* **2008**, *9*, 512. [CrossRef] [PubMed]
31. Allegre, M.; Argout, X.; Boccara, M.; Fouet, O.; Roguet, Y.; Berard, A.; Thevenin, J.M.; Chauveau, A.; Rivallan, R.; Clément, D.; et al. Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. *DNA Res.* **2012**, *19*, 23–35. [CrossRef] [PubMed]
32. Fang, W.; Meinhardt, L.W.; Mischke, B.S.; Bellato, C.; Motilal, L.A.; Zhang, D. Accurate determination of genetic identity for a single cacao bean, using molecular markers with a nanofluidic system, ensures cocoa authenticity and traceability. *J. Agric. Food Chem.* **2014**, *62*, 481–487. [CrossRef] [PubMed]
33. Cosme-Reyes, S.M.; Cuevas, H.E.; Zhang, D.; Oleksyk, T.K.; Irish, B.M. Genetic diversity of naturalized cacao (*Theobroma cacao* L.) in Puerto Rico. *Tree Genet. Genomes* **2016**, *12*, 88. [CrossRef]
34. LGC. Genomics KASPtm Genotyping Chemistry User Guide and Manual. 2013. Available online: http://www.lgcgenomics.com/ (accessed on 20 August 2022).
35. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
36. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **2004**, *20*, 289–290. [CrossRef]
37. Schliep, K.; Potts, A.J.; Morrison, D.A.; Grimm, G.W. Intertwining phylogenetic trees and networks. *Methods Ecol. Evol.* **2017**, *8*, 1212–1220. [CrossRef]
38. Epskamp, S.; Cramer, O.J.C.; Waldorp, L.J.; Schmittmann, V.D.; Borsboom, D. QGRAPH: Network visualizations of relationships in psychometric data. *J. Stat. Software* **2012**, *48*, 1–18. [CrossRef]
39. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 1–15. [CrossRef]
40. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **2011**, *12*, 246. [CrossRef]
41. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Mol. Ecol. Notes* **2012**, *28*, 2537–2539. [CrossRef] [PubMed]
42. Rabbi, I.Y.; Kulakow, P.A.; Manu-Aduening, J.A.; Dankyi, A.A.; Asibuo, J.Y.; Parkes, E.Y.; Abdoulaye, T.; Girma, G.; Gedil, M.A.; Ramu, P.; et al. Tracking crop varieties using genotyping-by-sequencing markers: A case study using cassava (*Manihot esculenta* Crantz). *BMC Genet.* **2015**, *16*, 115. [CrossRef] [PubMed]
43. Darkwa, K.; Olasanmi, B.; Asiedu, R.; Asfaw, A. Review of empirical and emerging breeding methods and tools for yam (*Dioscorea* spp.) improvement: Status and prospects. *Plant Breed.* **2019**, *139*, 474–497. [CrossRef]
44. Bhattacharjee, R.; Agre, P.; Bauchet, G.; De Koeyer, D.; Lopez-Montes, A.; Lava Kumar, P.; Abberton, M.; Adebola, P.; Asfaw, A.; Asiedu, R. Genotyping-by-Sequencing to unlock genetic diversity and population structure in white yam (*Dioscorea rotundata* Poir.). *Agronomy* **2020**, *10*, 1437. [CrossRef]

45. Turnbull, C.S.; Butler, D.; Cryer, N.; Zhang, D.; Lanaud, C.; Daymond, A.; Ford, C.S.; Wilkinson, M.J.; Hadley, P. Tackling mislabeling in cocoa germplasm collections. *Ingenic. Newsl.* **2002**, *9*, 8–11.
46. Wadsworth, R.M.; Harwood, T. *International Cocoa Germplasm Database, ICGD 2000 V4.1*; London International Financial Futures and Options Exchange and the University of Reading: London, UK, 2000.