



# Morphological and genetic diversity of cacao (*Theobroma cacao* L.) in Uganda

David Gopaulchan<sup>1</sup> · Lambert A. Motilal<sup>1</sup>  · Frances L. Bekele<sup>1</sup> · Séverine Clause<sup>2</sup> · James O. Ariko<sup>2</sup> · Harriet P. Ejang<sup>2</sup> · Pathmanathan Umaharan<sup>1</sup>

Received: 15 August 2018/Revised: 7 December 2018/Accepted: 11 December 2018/Published online: 1 January 2019  
© Prof. H.S. Srivastava Foundation for Science and Society 2019

**Abstract** Cocoa is among the top foreign exchange earners in Uganda's agriculture sector and has benefitted the livelihood of farmers involved in production. Although cacao cultivation was adopted in the early 1900s, little is known about the on-farm diversity of the crop. A total of 125 cacao landraces were surveyed from eight districts in the Central and Western Regions to evaluate the morphological and genetic diversity of cacao in Uganda. Passport data included site, tree, fruit and seed information. Trees were genotyped using 96 single nucleotide polymorphism markers on a Fluidigm platform. Low heterozygosity was detected in the germplasm in both the Central [observed heterozygosity ( $H_o$ ) = 0.295, expected heterozygosity ( $H_e$ ) = 0.334] and Western Regions ( $H_o$  = 0.317,  $H_e$  = 0.322). Genetic variation in both regions was generally comparable but the regions could be differentiated from each other. Inbreeding was noted in the Central Region while a greater sharing of genetic material was observed in the Western Region. The morphological and genetic data indicated that the Ugandan collection was an interspersed group with low to moderate variation with some separation of the Central from Western regions. Ancestry analysis indicated that the majority of the accessions were hybrids

of Marañon lineage but also had Amelonado and Iquitos genetic backgrounds. These findings are consistent with the history of the movement of cacao into Uganda. A core collection of 18 individuals to represent the genetic diversity as well as 12 additional trees with possible advantageous traits is proposed.

**Keywords** Africa · DNA fingerprinting · Single nucleotide polymorphism · Cacao landraces · Passport data · Population structure

## Introduction

Cacao (*Theobroma cacao* L.) belongs to the Malvaceae family and genus *Theobroma*, which contains 22 species (Cuatrecasas 1964; Alverson et al. 1999; Bayer et al. 1999) with *T. cacao* being the main economic species. Cacao originated in the South American Amazon basin and occurs as an understory tropical tree (van Hall 1914; Cheesman 1944; Cuatrecasas 1964; Toxopeus 1985a; Coe and Coe 1996; Bartley 2005). The industry commonly refers to the soft commodity as cocoa which is among the top ten agricultural commodities in the world (Utro et al. 2012). The fermented and dried cotyledons of the seeds (beans) are the main economic raw products. Beans are used in the production of chocolate and other confectionery products in the multibillion dollar candy industry. Cocoa butter from the beans is also used in the pharmaceutical and cosmetic industries. A variety of by-products can be made from the fruit (pod) walls or the mucilaginous pulp of the seeds (Oddoye et al. 2013). The cocoa flavanols are beneficial to human health leading to the current interest in the development of nutraceuticals (Cooper et al. 2008; Corti et al. 2009; Araujo et al. 2016).

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12298-018-0632-2>) contains supplementary material, which is available to authorized users.

✉ Lambert A. Motilal  
lamotilal@yahoo.com

<sup>1</sup> Cocoa Research Centre, The University of the West Indies, Sir Frank Stockdale Bldg., St. Augustine 330912, Republic of Trinidad and Tobago

<sup>2</sup> Sunshine Agro Products Ltd, 24, Avenue de la Toison d'Or, 1050 Brussels, Belgium

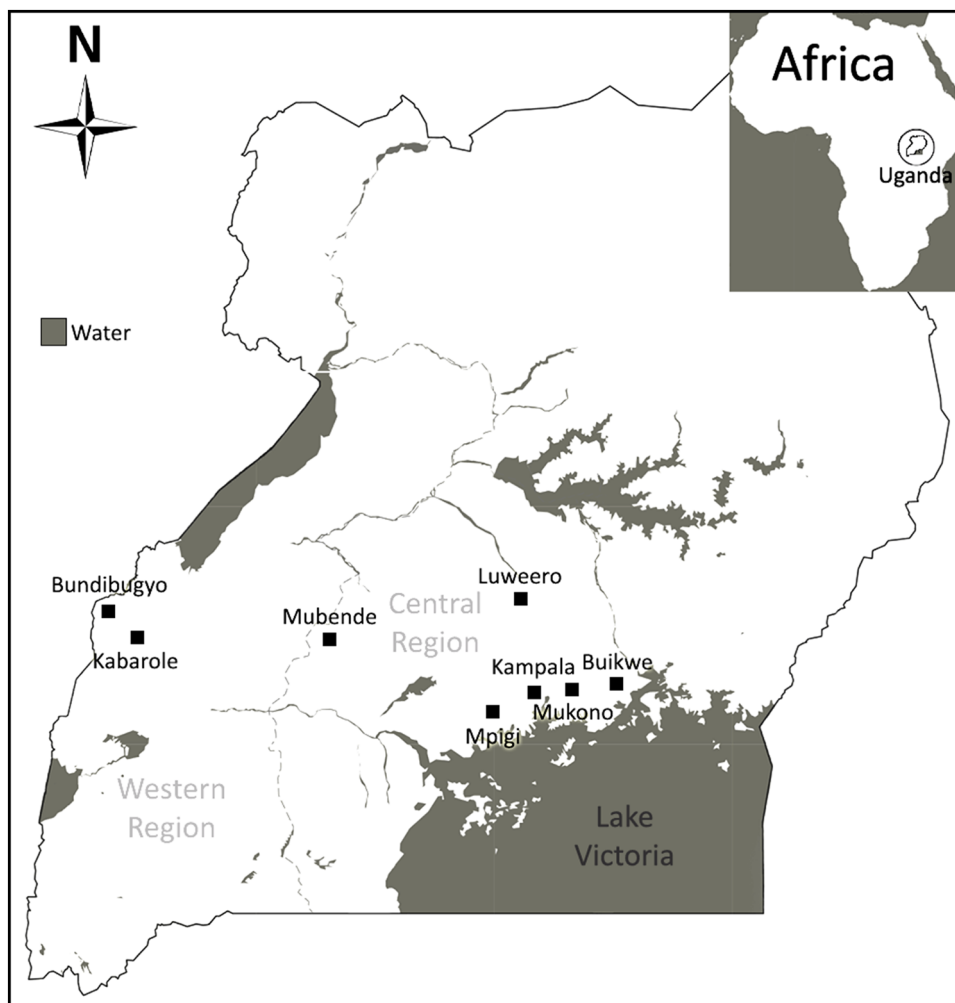
Generally, the crop is grown by smallholder, subsistence farmers in producing countries throughout the tropics (International Cocoa Organization 2014; Dalberg 2015). Cocoa is a major export in African, Asian, Central and South American countries that lie 15° to 20° north and south of the equator. The west coast of Africa contributes more than 70% of the world's cocoa supply with Côte d'Ivoire, Ghana, Nigeria and Cameroon being the leading producers (International Cocoa Organization 2017). Yet, Uganda, a landlocked state that borders Lake Victoria (Fig. 1) in Eastern Africa is not to be discounted. Recent figures indicated that 25,915 tonnes of bulk cocoa worth more than US\$56.6 million were exported in 2015 (Uganda Bureau of Statistics 2016). The major cocoa producing areas are Bundibugyo in the Western Region and Mukono in the Central Region. Several other districts also contribute to cocoa cultivation in the country.

Traditionally the crop has been considered to be divided into three agromorphological groups: Forastero, Criollo and Trinitario (Cheesman 1944; Toxopeus 1985a). Since cacao originated in the Amazon, there was a tendency to

separate Lower Amazon Forastero from Upper Amazon Forastero. While there is some confusion as to the meaning of these names in the historical record, the industry recognises Forasteros as vigorous varieties with dark purple beans with a relatively bitter flavour and often an acidic taste (Toxopeus 1985a). In contrast, Criollo varieties are less vigorous and produce beans that are white or very slightly pigmented with a strong desirable aroma and only slight bitterness (Toxopeus 1985a). Trinitario varieties are traditionally thought to be hybrids derived from open pollination between Lower Amazon Forastero and Criollo varieties (Toxopeus 1985a; Motamayor et al. 2003). Recently, Trinitarios were suggested to be an admixed group of various Forastero lineages among themselves and with Criollo (Motilal and Sreenivasan 2012; Motilal et al. 2011; Yang et al. 2013).

Unbiased groupings were achieved using a variety of molecular markers. In a seminal study using microsatellite markers, ten distinct genetic groups of cacao were recognised (Motamayor et al. 2008). Recent studies suggested that additional populations from Bolivia (Zhang et al.

**Fig. 1** Uganda: the study was conducted in the districts of Buikwe, Kampala, Luweero, Mpigi, Mubende, Mukono in the Central Region and Bundibugyo, Kabarole in the Western Region (Google Maps 2017). Districts selected were areas where cacao was grown or is still being cultivated. The major cacao producing areas are Bundibugyo and Mukono



2012), Peru (Motamayor et al. 2010) and Colombia (Osorio-Guarín et al. 2017) may also exist. It is likely that other unique cacao populations will be found as more wild areas of South America are explored. The current means of recording genetic diversity is with single nucleotide polymorphisms (SNPs) which have the highest potential to uncover differences in the genomes of plant varieties. The biallelic state of the SNPs, their high throughput assays and low error rate contribute to their being the markers of choice for genetic diversity studies. Understanding the genetic diversity of the cacao germplasm in the country helps in repositioning the industry. Informed decisions can be made on the importation of germplasm to complement the landrace background. Further, since cacao is currently conserved as field gene banks, the identification of a minimal set of trees for conservation will reduce germplasm maintenance costs. A core collection is a subset of the total collection that could be considered as optimal to retain the bulk of the genetic diversity. In addition, some ancestral lineages like Criollo have highly desirable and valued flavour notes (Frauendorfer and Schieberle 2008; Kongor et al. 2016) and can be used to obtain higher farm gate prices and produce luxury chocolates.

Genetic diversity of cacao on a local scale is affected by germplasm movement. Cacao has been repeatedly moved into, around, and among various countries. Long distance movement of germplasm is described in Kennedy and Mooleedhar (1993) and Lockwood and End (1993). In the early days, long distance transport was by sea and the vagaries of the journey at sea, meant that the limited varieties were exported and few fruits or live plants reached their final destination. The global cocoa industry is therefore based on a narrow genetic origin (Toxopeus 1985b; Zhang and Motilal 2016). Cacao was first introduced to Uganda in 1901 and was established at the Botanic Gardens in Entebbe in the Central Region (Brown and Hunter 1913). The plants, likely seedlings, were procured from the Royal Botanic Gardens in England, which were likely descendants of material collected from Trinidad between 1880 and 1881 (Urquhart 1958; Bartley 2005). Germplasm was subsequently imported from the West Indies in 1903, from Ceylon (presently known as Sri Lanka) in 1910 mainly as Forastero seedlings and from Zanzibar in 1913 as seeds (Urquhart 1958). The established plants provided planting material for estates, and by 1917, cocoa was being exported. Nevertheless, by 1924, the early plantings were abandoned due to low yields; fall in cocoa prices and the build-up of pests and diseases to the non-native crop (Krug and Quartey-Papafio 1964; ADC/IDEA 1998).

Upper Amazon hybrid seeds were also imported from Ghana in 1956 (Krug and Quartey-Papafio 1964). The crop was subsequently reintroduced with plantings from surviving plots, most likely of Amelonado and Upper Amazon

material from Ghana (Krug and Quartey-Papafio 1964; ADC/IDEA 1998). However, the political and economic turmoil of the 70s and early 80s in Uganda negatively affected the cocoa industry and there was an abandonment of farms and subsequent reduction in cocoa acreage. In 1986, under a Food and Agriculture Organization of the United Nations and United Nations Development Programme project, cocoa nurseries were developed in cocoa growing areas at Damba Island on Lake Victoria for the quarantining of imported materials (ADC/IDEA 1998). High yielding varieties of Criollo, Forastero and Trinitario hybrid groups were acquired from Ghana and South America in 1987 (ADC/IDEA 1998). Hybrid seeds resulting from open pollination mainly between Upper Amazon and Trinitario, were secured from Costa Rica and Trinidad in 1988, and were established on the island to broaden the genetic base for cocoa breeding (Petithuguenin 2000).

However, apart from these prior historical records, there are little, if any empirical studies on the genetic diversity and population structure of the cacao germplasm currently existing in Uganda. Hence, the objective of this study was to characterise the morphological and genetic diversity of cacao landraces on farmer fields in eight districts in the Central and Western Regions in Uganda. This will give information on the conservation and sustainable management of these landraces towards the rejuvenation of the cocoa industry in Uganda.

## Materials and methods

### Passport data and plant material

Passport data and plant material were collected between 28th May to 6th September 2011, in the districts of Buikwe, Kampala, Luweero, Mpigi, Mubende and Mukono in the Central Region and the districts of Bundibugyo and Kabarole in the Western Region (Fig. 1). The districts selected were areas where cacao was grown and/or where cacao is still cultivated. The location sites varied between 1074 and 1280 m in altitude in the Central Region and 662–960 m in the Western Region. Most plantations were surrounded by forest vegetation. On-site visits of farms were carried out and a total of 125 cacao trees were sampled from the Central (76 trees; 60.8%) and Western (49 trees; 39.2%) Regions of Uganda (Fig. 1). The collection of passport data included GPS co-ordinates, site description, tree habit, tree height and canopy diameter and number of mature fruits (Table 1). Qualitative and quantitative traits from fruits and seeds were also recorded as described in Bekele and Butler (2000) and Bekele et al. (2006) with the following exceptions: fruit quantitative traits were measured from two fruits per tree and seed

**Table 1** Location and description of Ugandan cacao (*Theobroma cacao* L.) accessions

Region	District	Sample code	Number of trees	Tree height (m)		Canopy height (m)		Canopy diameter (cm)		Trunk girth (cm)	
				Mean	SE	Mean	SE	Mean	SE	Mean	SE
Central	Buikwe	U001 to U013, U038 to U061	37	5.28	0.21	4.37	0.18	522.70	22.02	69.88	4.61
	Kampala	U016 to U018	3	5.29	1.18	3.91	1.27	608.00	82.71	49.00	7.00
	Luweero	U019 to U026	8	5.44	0.40	4.40	0.34	594.50	69.84	55.13	7.23
	Mpigi	U062 to U077	16	6.73	0.28	5.22	0.39	740.94	42.71	123.63	55.08
	Mubende	U014 and U015	2	5.35	0.85	4.68	0.73	605.00	45.00	59.00	1.00
	Mukono	U028 to U037	10	5.82	0.46	4.98	0.44	587.70	45.02	44.80	2.80
	All districts		76	5.67	0.16	4.62	0.15	590.29	19.28	75.23	11.95
Western	Bundibugyo	U078 to U092, U111 to U126	31	4.91	0.24	3.90	0.24	468.71	26.94	39.32	1.91
	Kabarole	U093 to U110	18	5.49	0.35	3.65	0.28	476.33	27.94	51.50	4.59
	All districts		49	5.12	0.20	3.81	0.18	471.51	19.72	43.80	2.21
Total			125	5.46	0.13	4.30	0.12	543.73	14.92	62.91	7.43

quantitative traits were measured from twelve seeds from each fruit. Two young leaves per tree were collected for SNP genotyping.

### DNA extraction

DNA was extracted from young leaf samples using a ZR Plant/Seed DNA MiniPrep kit (Zymo Research Corp, Irvine, CA). In brief, 0.8 mL of lysis solution with 5% PVPP (w/v) and 7% thioglycerol (v/v) was added to 0.04 g dried leaf tissue in a CKMix tube (Bertin Corp, Rockville, MD) and the sample was homogenised at 6000 rpm for 30 s (three sets with 5 s intervals between sets) in a Pre-cellys 24 homogeniser (Bertin Corp, Rockville, MD). The homogenate was incubated at 4 °C for 1 h and centrifuged at 18,000g for 2 min. DNA was then extracted from the supernatant according to the manufacturer's instructions with the exception that 3% thioglycerol (v/v) was added to the plant/seed DNA binding buffer. The DNA was quantified using a NanoDrop 2000c UV–Vis (Thermo Fisher Scientific Inc., Wilmington, DE) and purity was assessed from the absorbance ratios of 260/280 nm and 260/230 nm. The integrity of the DNA was confirmed with agarose gel electrophoresis and DNA extracts were stored at – 80 °C.

### SNP markers and genotyping

The 96 SNP assays (Online Resource 1) were selected from the global reference SNP panel recommended by Motilal et al. (2017). SNP type genotyping assays were manufactured by Fluidigm Corporation (San Francisco, CA). Genotyping was performed on a Fluidigm Juno System using the Juno SNP Type Genotyping Reagent Kit and a

Juno 96.96 Genotyping IFC according to the manufacturer's instructions. The Juno 96.96 thermal cycling protocol script was used for amplification as recommended by the manufacturer. In brief, a multiplex specific target amplification (STA) step was performed prior to the SNP genotyping to allow the enrichment of template molecules. STA thermal cycling conditions were 95 °C for 2 min, followed by 14 cycles at 95 °C for 15 s and 60 °C for 4 min. For SNP genotyping, reactions were initially heated to 95 °C for 10 min, followed by 4 cycles at 95 °C for 15 s, 64–61 °C (1 °C decrease with each cycle) for 45 s, and 72 °C for 15 s, followed by 39 cycles at 95 °C for 15 s, 60 °C for 45 s and 72 °C for 15 s. Fluorescent intensity was quantified using the Fluidigm EP1 and genotypic calls were made automatically using Fluidigm SNP Genotyping Analysis software v4.1.3.

Young leaves from 109 reference accessions were collected from the International Cocoa Genebank, Trinidad (ICGT). The reference accessions incorporated the ten cacao genetic groups identified by Motamayor et al. (2008) with four (Purús), six (Criollo), seven (Amelonado, Contamana, Nacional), eight (Guiana), nine (Marañón, Nanay), ten (Iquitos) and 18 (Curaray) pure high ancestry members. Two additional groups (Refractario Group1 and Refractario Group 2), proposed by Motilal et al. (2013), each with 12 members were also used. Each leaf sample was washed and dried and four to five leaf discs (six mm diameter) were prepared using a punch (World Precision Instruments Ltd., Hertfordshire, UK) from the interveinal areas. The discs were stored in a 96-well tube storage rack from LGC's leaf sampling kit (LGC Genomics, UK) and shipped to LGC Genomics, UK. DNA extraction and SNP genotyping using KASP chemistry was performed by LGC Genomics at the 96 SNP sites previously described. Selected samples of the

reference accessions were also genotyped with the Fluidigm Juno System to verify that both technologies produced comparable SNP calls.

## Data analysis

### Phenotypic dataset

The dataset was analysed in XLSTAT version 2014.5.03 (Addinsoft 2014) statistical software and data analysis add-on for Excel. Principal component analysis (PCA) was performed on fruit and seed dimensions [heritable traits based on the findings of Toxopeus (1972) and Cilas et al. (2010)]. Pearson correlation adjustment was implemented and observations with missing data were removed. Means were compared using the *t* test on two independent samples.

### SNP dataset

Pairwise multilocus matching of the Ugandan samples, the probability of identity among siblings (PID<sub>SIB</sub>) analysis, which is the probability that two sibling individuals drawn at random from a population have identical genotypes (Evelt and Weir 1998; Waits et al. 2001) and summary statistics for the SNP assays were determined from the routines implemented in GenAlEx v6.502 (Peakall and Smouse 2006, 2012). Summary statistics included the effective number of alleles [ $N_e$ ; Brown and Weir (1983)]; the Shannon Information Index [ $I$ ; Brown and Weir (1983)]; observed and expected heterozygosity [ $H_o$ ,  $H_e$  respectively; Hartl and Clark (1997)], unbiased heterozygosity [ $uH_e$ ; Peakall and Smouse (2012)], Fixation index [ $F$ ; Hartl and Clark (1997)] and deviation from Hardy–Weinberg equilibrium (HWE). Unique Ugandan samples were subsequently analysed along with the 109 cacao reference accessions. A dendrogram was constructed based on the default simple matching dissimilarity index using the Neighbour-Joining method with 1000 bootstrap replicates in DARwin v6 (Perrier and Jacquemoud-Collet 2006) and the figure was rendered with FigTree v1.4.2 (Rambaut 2012). A principal coordinate analysis (PCoA) based on standardised pairwise genetic distances was performed in GenAlEx v6.502 (Peakall and Smouse 2006, 2012).

Spatial effects in the Ugandan dataset was assessed via Mantel tests and spatial autocorrelation tests implemented in GenAlEx v6.502 (Peakall and Smouse 2006, 2012). A global Mantel test was applied using log transformed GPS data and linearized genetic distance matrix and 999 permutations. A global spatial autocorrelation analysis was conducted using the multipop option to test for population effect, 999 permutations, 999 bootstraps, three distance classes (max distance from GenAlEx was three and was

therefore used as the default neighbour size) to estimate the spatial correlogram wide test statistic, Omega (Smouse et al. 2008). A two-dimensional local spatial autocorrelation analyses (2D LSA) was conducted using 999 permutations, one-tail testing, a class size of one and three distance classes. Significance at  $P < 0.01$  for the Heterogeneity test was used as recommended by Banks and Peakall (2012).

Population structure of the Ugandan germplasm was determined using Structure v2.3.4 (Pritchard et al. 2000). An admixture model with alpha inferred, independent allele frequency with 200,000 burn-ins and 500,000 Monte Carlo Markov Chain repetitions was used. The number of clusters ( $K$ ) was set from 6 to 18 and thirty iterations were run at each  $K$  value. CLUMPAK (Cluster Markov Packager Across  $K$ ) (Kopelman et al. 2015) was used to estimate the degree of congruence between independent runs from STRUCTURE for each  $K$  value. The optimum  $K$  value was determined using the ad hoc  $\Delta K$  method described by Evanno et al. (2005). Ugandan samples with high membership and their related populations were identified from the STRUCTURE output. A refined STRUCTURE analysis was then conducted on this dataset to corroborate the ancestral estimates of these Ugandan samples. The number of populations was set from 2 to 7 with thirty iterations for each  $K$  value. Graphical rendering of STRUCTURE results was created with Distruct v1.1 (Rosenberg 2004).

BOTTLENECK v1.2.02 (Cornuet and Luikart 1996) was used to analyse the population to determine recent effective size reductions from allele data frequencies by testing mode shift and heterozygosity excess (Cornuet and Luikart 1996; Luikart and Cornuet 1998). The analysis was performed using 30 different combinations of the two phase model (TPM) and proportion (%) of stepwise mutation model (SMM) in TPM for 1000 iterations. Significance was determined by the Sign test and the Standardized differences test as implemented in the program following Campoy et al. (2016) in their SNP study of *Prunus* diversity.

Core selection was conducted in PowerCore v.1.0 (National Institute of Agricultural Biotechnology 2006) based on an M (maximization) strategy from three randomised datasets of the unique Ugandan SNP multilocus profiles. A common set of individuals over the three cores was retained as the core collection. The suitability of the selected core accessions was verified by comparing descriptive statistics and estimator of actual differentiation ( $D_{est}$ ) values (Jost 2008) of the core accessions with (a) the entire dataset and (b) the rest of the collection after the core accessions were removed.

## Results

### Farm site description

The majority of the sampled trees were from Buikwe (29.6%) in the Central Region and Bundibugyo (24.8%) in the Western Region. Plantations in the Central Region were small to intermediate in size ranging between 2 and 70 acres with an average of 11.3 acres. In the Western Region, small-scale farms were predominant and ranged between 1 and 10 acres with an average of 4.8 acres. Cacao trees were grown in mixed farming systems along with timber trees, fruit trees, vegetable crops, vining crops and tuber crops among others.

### Tree and fruit morphology

Overall, Mpigi had the tallest and thickest trees with the largest canopy dimensions while Bundibugyo had the shortest and thinnest trees and the smallest diameter of the canopy (Table 1). The number of fruits per tree at the time of assessment ranged from 0 to 96 fruits, with 13 trees from the Central Region in the districts of Buikwe and Mpigi bearing 37 or more fruits (data not shown). Mean fruit length ranged from 248.3 mm to 157.6 mm while mean fruit diameter ranged from 108.8 mm to 75.9 mm (Table 2). The longest fruits (> 250 mm) were found in Kampala district of the Central Region (“U017” and “U018”) and in the Bundibugyo district of the Western Region (“U118” and “U122”). The widest fruits (diameters > 100 mm) were found in Kampala (“U016”, “U017” and “U018”) of the Central Region and in Kabarole (“U095”, “U099” and “U106”) of the Western Region. Mean seed length, width and thickness ranged from 22.63,

13.4 and 6.8 mm to 25.48, 16.44 and 9.28 mm respectively (Table 2). Cotyledons exhibited a range of purple shades and notably, there was an absence of cream or white coloured cotyledons.

The PCA of fruit and seed dimensions revealed a single group with indication of separation between Central and Western Regions (Fig. 2). Two axes explained 71.5% of the variation with seed dimensions accounting for 70.4% of the loadings in the first vector while fruit dimensions contributed 67.8% to the second vector. Seed width and seed thickness appeared to be separated based on region (Fig. 2). A comparison of seed width and seed thickness means between the regions indicated that both seed width and thickness were larger ( $P < 0.001$ ) in the Central Region compared to the Western Region.

### Genetic diversity statistics

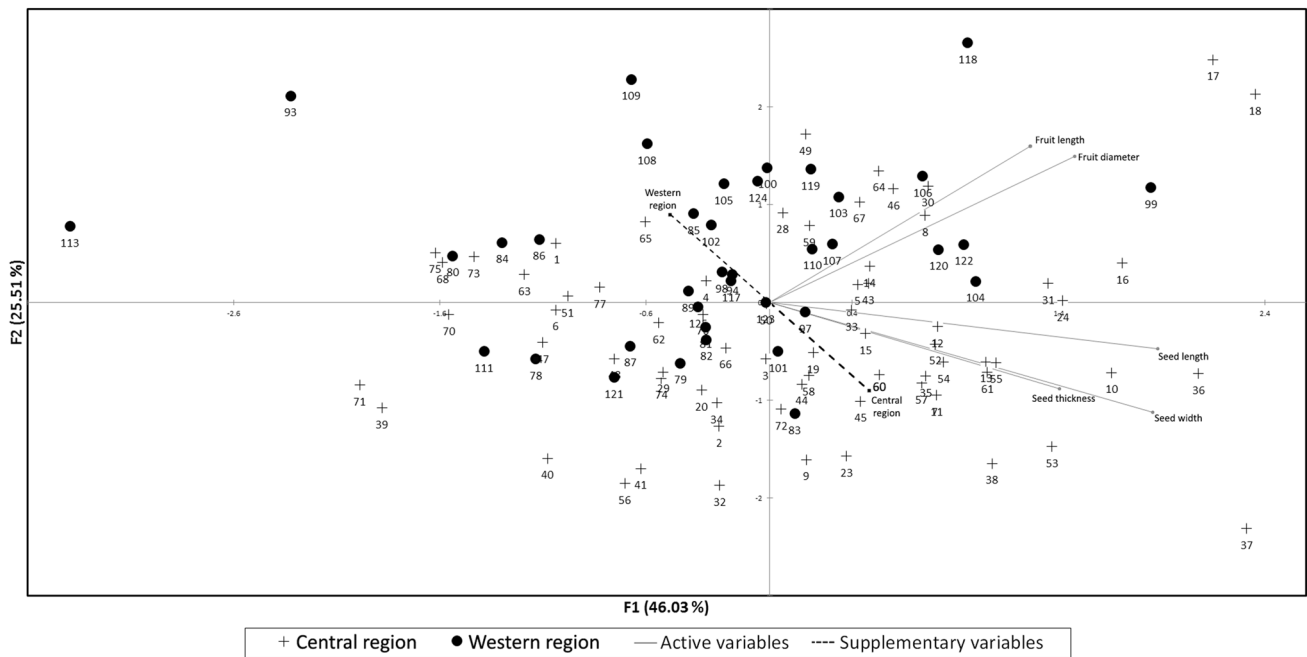
Suitable DNA was obtained from 109 trees as leaf samples from 16 trees were degraded. Nine assays (TcSNP: 198, 326, 397, 529, 703, 886, 945, 1201, 1442) of the 96 SNP markers, had more than 10% missing allelic data and were excluded. Missing allelic data in the culled dataset was at most 5.75% per sampled tree. Identity matching gave a combined  $PID_{SIB}$  of  $9.8 \times 10^{-14}$  and all samples were unique. Interestingly, sample “U026” matched reference accession “REDAMEL 1/31” (Amelonado) at 86 loci (99%), while both “U032” and “U033” matched reference accessions “IMC 67” (Iquitos) and “SCA 6” (Contamana) respectively at 85 loci (98%).

Within the Ugandan germplasm, one of 87 SNP loci was monomorphic and significant deviations from HWE were found for 18, 19 and nine loci in the full dataset, Central region and Western Region respectively. Of the nine loci in

**Table 2** Fruit and seed dimensions in eight districts in the Central and Western Regions of Uganda

Region	District	Fruit dimension (mm)				Seed dimension (mm)					
		Length		Diameter		Length		Width		Thickness	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Central	Buikwe	174.55	2.71	79.46	1.30	24.56	0.15	15.04	0.09	8.22	0.08
	Kampala	248.33	16.80	108.83	3.59	24.94	0.31	16.44	0.26	9.28	0.27
	Luweero	157.56	4.78	75.88	2.80	25.48	0.44	15.77	0.22	8.02	0.15
	Mpigi	162.10	4.24	80.94	2.01	22.63	0.26	13.40	0.12	6.83	0.12
	Mubende	170.50	5.42	91.50	5.25	24.71	0.36	15.08	0.25	8.08	0.22
	Mukono	182.30	8.09	82.70	2.29	25.21	0.27	15.78	0.14	8.76	0.13
	All districts	174.05	2.48	81.30	1.01	24.31	0.11	14.90	0.06	8.03	0.06
Western	Bundibugyo	177.48	5.09	80.46	1.36	23.54	0.17	13.61	0.11	6.80	0.11
	Kabarole	195.82	5.20	88.35	1.88	23.96	0.24	13.80	0.14	6.95	0.10
	All districts	184.57	3.81	83.51	1.17	23.72	0.14	13.69	0.09	6.86	0.08
Total		177.92	2.12	82.11	0.77	24.10	0.09	14.47	0.05	7.62	0.05

Seed width and thickness differed significantly ( $P < 0.001$ ) between the Central and Western Regions



**Fig. 2** Principal component analysis plot of fruit and seed dimensions of Central and Western Ugandan cacao (*Theobroma cacao* L.). The analysis was performed using Pearson (n) correlation adjustment and observations with missing data removed

the Western Region, three and four of these were also found to deviate from HWE in the full Ugandan dataset and Central Region respectively. The  $H_o$  ranged from 0.289 to 0.326 and was similar in both regions (Table 3). The  $H_e$  and I was highest in Mukono (0.351 and 0.521 respectively) and lowest in Kampala (0.264 and 0.389 respectively) (Table 3). The average  $H_e$  in the Central (0.334) and

Western (0.322) Regions were similar to each other (Table 3). Jost’s estimate of differentiation ( $D_{est}$ ) between the Central and Western Region was significant but low ( $D_{est} = 0.003$ ;  $P = 0.008$ ). The fixation index was notably lower in the Western Region (0.016) in comparison to the Central Region (0.123).

**Table 3** Descriptive statistics of the Ugandan cacao (*Theobroma cacao* L.) germplasm

Region	District	N	$N_e$		I		$H_o$		$H_e$		$uH_e$		F	
			Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Central	Buikwe	31	1.564	0.037	0.487	0.022	0.296	0.018	0.325	0.018	0.330	0.018	0.081	0.025
	Kampala	3	1.460	0.041	0.389	0.030	0.307	0.033	0.264	0.021	0.317	0.026	-0.163	0.056
	Luweero	8	1.492	0.037	0.442	0.024	0.294	0.022	0.292	0.018	0.311	0.019	-0.034	0.035
	Mpigi	14	1.514	0.036	0.464	0.022	0.295	0.020	0.305	0.017	0.317	0.018	0.040	0.035
	Mukono	10	1.607	0.034	0.521	0.019	0.289	0.019	0.351	0.015	0.369	0.016	0.166	0.041
	All districts	66	1.577	0.036	0.502	0.020	0.295	0.016	0.334	0.016	0.337	0.016	0.123	0.021
Western	Bundibugyo	27	1.543	0.037	0.480	0.021	0.312	0.019	0.318	0.017	0.324	0.017	0.016	0.024
	Kabarole	15	1.546	0.038	0.475	0.023	0.326	0.022	0.316	0.018	0.327	0.018	-0.033	0.029
	All districts	42	1.553	0.037	0.486	0.021	0.317	0.018	0.322	0.017	0.326	0.017	0.016	0.021
Total		108	1.573	0.036	0.499	0.020	0.304	0.016	0.332	0.016	0.334	0.016	0.087	0.017

Mean and standard error presented. N, number of samples;  $N_e$ , number of effective alleles (Brown and Weir 1983); I, Shannon’s information Index (Brown and Weir 1983);  $H_o$ , observed heterozygosity (Hartl and Clark 1997);  $H_e$ , expected heterozygosity (Hartl and Clark 1997);  $uH_e$ , unbiased expected heterozygosity (Peakall and Smouse 2012); F, fixation index (Hartl and Clark 1997); were obtained from GenAlEx v6.502 (Peakall and Smouse 2006, 2012)

The district of Mubende was excluded as data was available for only a single tree (U015)

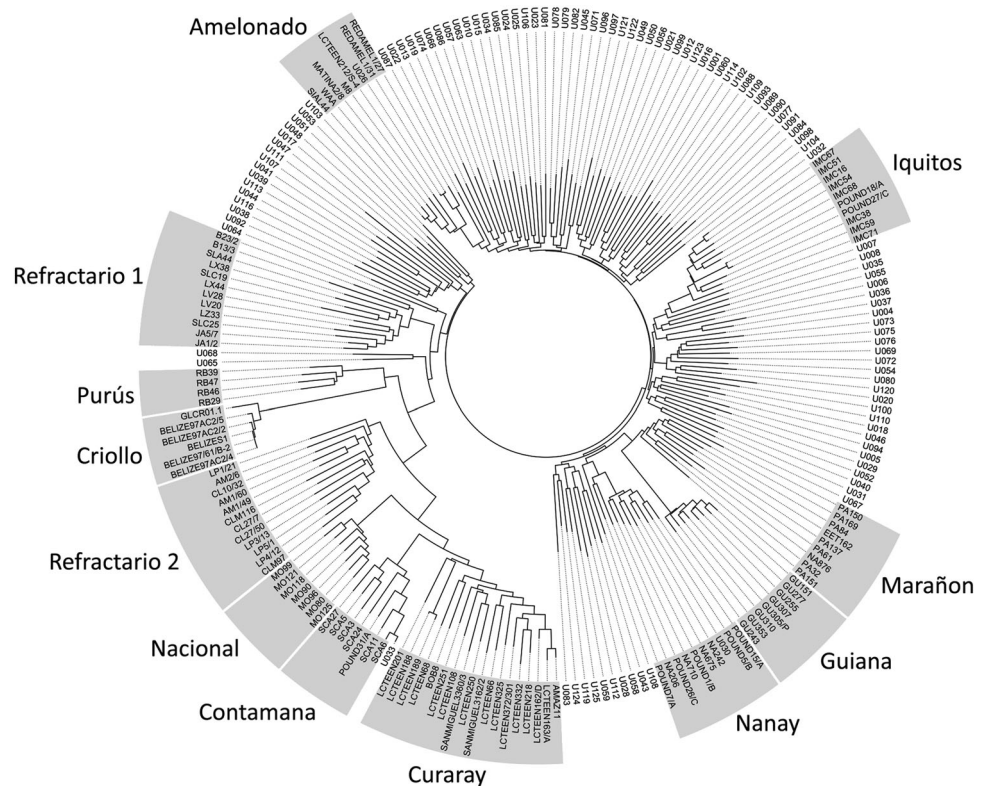
### Spatial genetic analysis

The Mantel test returned a low but significant correlation coefficient ( $R_{xy} = 0.037$ ;  $P = 0.010$ ). Global spatial analysis revealed a significant spatial genetic correlation coefficient ( $\Omega = 40.060$ ;  $P = 0.001$ ). However, non-significant tests of the correlation between the linearized genetic matrices of the Central and Western regions, even at all size classes were obtained. The correlogram coefficient for the Western Region by itself was just non-significant ( $\Omega = 21.193$ ;  $P = 0.011$ ) whereas the coefficient was markedly significant ( $\Omega = 37.288$ ;  $P = 0.001$ ) in the Central Region. Using two nearest neighbours, seven samples (four from Central “U031”, “U036”, “U037” and “U076”; and three from Western “U089”, “U111” and “U122”) were found to have significant ( $P = 0.001$  to  $0.005$ ) correlation ranging from 0.117 to 0.239 in value.

### Dendrogram construction and principal coordinate analysis

The dendrogram showed that the majority of the Ugandan samples had a close relationship to the Amelonado, Iquitos, Marañon/Guiana and Nanay clades (Fig. 3). Sixteen samples also showed a stronger relationship with the Refractario Group 1 samples. In contrast, only two samples

**Fig. 3** Phylogenetic tree of 109 Ugandan samples and 109 cacao (*Theobroma cacao* L.) reference accessions from the International Cocoa Genebank, Trinidad. Amelonado, Contamana, Criollo, Curaray, Guiana, Iquitos, Marañon, Nacional, Nanay and Purús genetic groups described by Motamayor et al. (2008) and Refractario Groups 1 and 2 proposed by Motilal et al. (2013). Phylogenetic tree was constructed with 87 SNPs using the Neighbor-Joining method in DARwin v6 (Perrier and Jacquemoud-Collet 2006). Ugandan samples are denoted by “U” followed by a unique 3 digit number

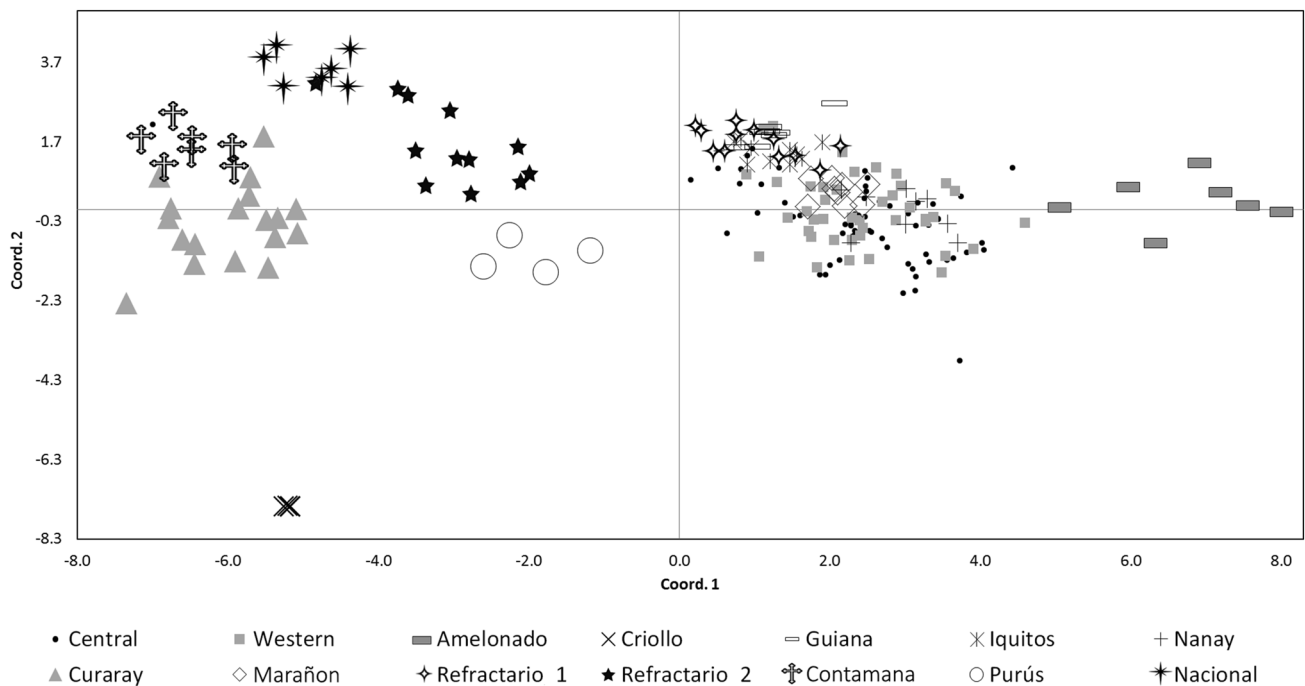


“U065” and “U068”, which were collected from Mpigi, were derived from plants that gave rise to the Criollo, Curaray, Contamana, Purús, Nacional and Refractario 2 accessions. Notably, four samples from the Central Region (Luweero district: “U026”; Mukono district: “U030”, “U032” and “U033”) grouped tightly with the Amelonado, Nanay, Iquitos and Contamana genetic clusters respectively. The PCoA showed that the diversity of the Ugandan cacao population represented a subset of that observed at the ICGT (Fig. 4). The plane of the first two main axes, which accounted for 45.7% of the total variation expressed, showed that the Ugandan samples were grouped into one main cluster which included the Marañon, Nanay, Iquitos, Amelonado, Guiana and Refractario Group 1 accessions.

### Ancestry analysis and population bottleneck

At  $K = 8$ , the majority of the Ugandan samples were placed in a cluster comprised predominantly of individuals of Marañon ancestry (61.5% of the trees had  $\geq 80\%$  Marañon lineage) (Fig. 5a). However, at  $K = 9$  to 11, the samples separated from the grouping and formed two distinct and undefined sub-clusters (data not shown). The 218 samples differentiated best into thirteen genetic clusters ( $K = 13$ ), the optimal value determined by the ad hoc  $\Delta K$  statistic (Evanno et al. 2005) (Fig. 5a). At  $K = 13$ ,





**Fig. 4** Principal coordinate plot of the Central and Western Ugandan samples and 109 reference cacao (*Theobroma cacao* L.) accessions. Reference accessions are from the International Cocoa Genebank

Trinidad and they represent the genetic backgrounds of the genetic groups described by Motamayor et al. (2008) and Motilal et al. (2013). Multivariate analysis was conducted with 87 SNPs

22.0% of the Ugandan accessions (24 individuals) fitted into one of the undefined cluster and 17.4% (19 individuals) fitted into the second undefined cluster, using a threshold of 80% for group allocation. Accessions “U026”, “U030”, “U032” and “U033” had high genetic ancestry (> 95% lineage) allied to Amelonado, Nanay, Iquitos and Contamana genetic clusters respectively, which complemented the findings of the phylogenetic analysis. Only five trees had appreciable Criollo ancestry. Samples “U019”, “U037”, “U036” had moderate levels of 31.4%, 24.3% and 19.7% Criollo ancestry respectively, while “U088” and “U041” had lower levels of 12.3% and 10.6% respectively. The remaining trees had mixed ancestries with appreciable levels of Amelonado, Marañon and Iquitos lineages (Fig. 5a). A refined STRUCTURE analysis, using the 43 Ugandan cacao samples with more than 80% membership to the two undefined clusters (24 individuals from undefined cluster 1 and 19 individuals from undefined cluster 2), revealed that the Ugandan accessions were predominantly admixed Marañon hybrids with Amelonado and Iquitos lineages (Fig. 5b;  $K = 3$ ).

Bottleneck analysis showed a deviation of allelic diversity and heterozygosity from mutation-drift equilibrium and revealed significant mode-shift and excess heterozygotes ( $P < 0.05$ ) for each of the 30 different combinations of the TPM and proportion of SMM in TPM (%) (data not shown).

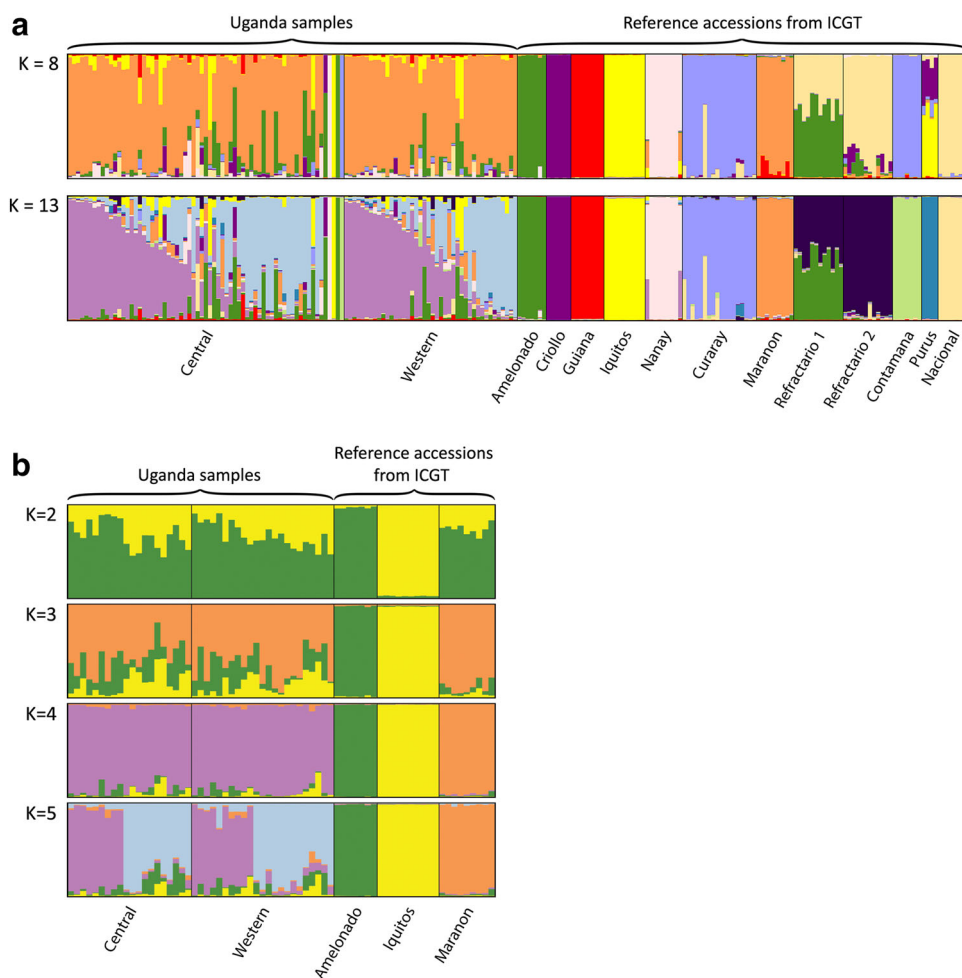
**Core collection**

A set of 18 individuals (Table 4) was returned every time with full matching among the runs. The core of 18 captured similar genetic diversity as the entire set of 109 sampled accessions from 87 SNPs in Uganda. Differentiation was absent between the core selected individuals and the entire collection ( $D_{est} = -0.004$ ,  $P > 0.05$ ) or the collection after removal of the core individuals ( $D_{est} = -0.001$ ;  $P > 0.05$ ).

**Discussion**

The morphological and SNP genetic diversity of cacao landraces in eight districts of the Central and Western Regions of Uganda were investigated. Overall, the findings indicated that the Ugandan cacao population had limited genetic variation occurring as a common pool of germplasm. Genetic diversity parameters of  $I$ ,  $H_o$ ,  $H_e$  and  $uHe$  were similar between Central and Western regions. However, the estimator of population differentiation  $D_{est}$  (Jost 2008) was low but significant and about twice the number of SNP loci deviated from HWE in the Central Region as compared to the Western Region. Low heterozygosity was detected in the germplasm in both the Central and Western Regions (Table 3). Additionally, while  $H_o$  and  $H_e$  in the Western Region was comparable to each other, the

**Fig. 5** Structure analysis of Ugandan cacao (*Theobroma cacao* L.) samples. Population structure was determined using Structure v2.3.4 and based on 87 SNP markers. Each vertical line represents an individual; **a** 109 Ugandan samples and 109 cacao (*Theobroma cacao* L.) reference accessions from the International Cocoa Genebank, Trinidad representing the different genetic groups described by Motamayor et al. (2008) and Motilal et al. (2013); **b** samples with more than 80% membership to undefined clusters 1 and 2 and the Amelonado, Iquitos and Marañon reference accessions



observed heterozygosity ( $H_o = 0.295$ ) in the Central Region was lower than the expected heterozygosity ( $H_e = 0.334$ ), which indicated that some degree of inbreeding occurred. Interestingly, the fixation index was notably lower in the Western Region (0.016) than in the Central Region (0.123), suggesting that there was a greater sharing of genetic material and therefore more interspersions occurring in the Western Region in comparison to the Central Region. These results were supported by the significant Mantel test and the significant global spatial correlogram. This could have been mainly due to the Central region as unlike the Western Region, it had a significant correlogram. Contrariwise, the linearized genetic distance matrices between Central and Western Regions were non-significant and the 2D LSA detected only a few samples in both Central and Western regions which could have contributed to the significant differences. These findings were further supported by the slight separation of Central from Western samples in the PCA of fruit and seed morphological traits (Fig. 2) and by the difference in mean trait values of seed width and thickness (Table 3). Taken together, these results support the conclusion that the cacao

germplasm is non-randomly distributed and probably had different distribution pathways and sources which were region dependent.

Overall,  $H_o$  was lower than  $H_e$  in Uganda as was observed for farmed cacao in Indonesia (Lukman et al. 2014), Honduras and Nicaragua (Ji et al. 2013) and Ghana (Takrama et al. 2014). However, while the  $H_o$  in Uganda was within the range observed for this parameter in the above farmed samples, the  $H_e$  was lower in Uganda. Likewise, Shannon's Information Index was lower in Uganda than that reported by Ji et al. (2013) and Lukman et al. (2014). This suggests that farmed cacao in Uganda has a more limited genetic diversity than these four countries. The low level of genetic variability observed in the Ugandan cacao population is reflective of cacao germplasm found in other countries where cacao was introduced. For instance, in Cameroon, Efombagn et al. (2006) noted that farmers' planting material was not highly diverse, but genetically close to parental genotypes available in genebanks. Pokou et al. (2009) found that in Cote D'Ivoire, the materials grown on farms were mostly locally selected Amelonado type, while the others were improved hybrids

**Table 4** Location details of 18 trees (core collection) selected to represent the genetic diversity of the Ugandan cacao (*Theobroma cacao* L.) collection and the additional trees to complement the core set

Sample code	Region	District	GPS coordinate (latitude, longitude)
<i>Core collection</i>			
U013	Central	Buikwe	0.198972, 32.920000
U019	Central	Luweero	0.670944, 32.844028
U032	Central	Mukono	0.408944, 33.113806
U033	Central	Mukono	0.409056, 33.113389
U040	Central	Buikwe	0.420361, 33.115111
U041	Central	Buikwe	0.336750, 33.201250
U045	Central	Buikwe	0.336917, 33.200889
U049	Central	Buikwe	0.337194, 33.201472
U055	Central	Buikwe	0.338889, 33.205500
U068	Central	Mpigi	0.160944, 32.237611
U072	Central	Mpigi	0.160944, 32.237611
U073	Central	Mpigi	0.160944, 32.237611
U077	Central	Mpigi	0.131639, 32.226778
U083	Western	Bundibugyo	0.860389, 30.173500
U091	Western	Bundibugyo	0.859056, 30.178583
U094	Western	Kabarole	0.843917, 30.172139
U110	Western	Kabarole	0.841917, 30.165778
U116	Western	Bundibugyo	0.702333, 30.061917
<i>Additional trees to complement the core collection</i>			
U016	Central	Kampala	0.103222, 32.930083
U017	Central	Kampala	0.103611, 32.929972
U018	Central	Kampala	0.103194, 32.930611
U036	Central	Mukono	0.414833, 33.119444
U037	Central	Mukono	0.414750, 33.119444
U093	Western	Kabarole	0.843694, 30.171917
U095	Western	Kabarole	0.844056, 30.172139
U099	Western	Kabarole	0.843333, 30.172056
U106	Western	Kabarole	0.840389, 30.166917
U113	Western	Bundibugyo	0.702528, 30.062472
U118	Western	Bundibugyo	0.702333, 30.061917
U122	Western	Bundibugyo	0.703306, 30.060139

and a mixture of local and improved types. Zhang and Motilal (2016) proposed that the low allele diversity recorded in these cacao producing countries was due to the narrow genetic origin of the crop under domestication as only a relatively small part of the wild cacao gene pool was taken into cultivation. Consequently, breeding efforts to date have been no more than a reshuffling of these minimal genetic variants. Likewise in Uganda, it is probable that a comparatively small set of common alleles were introduced and widely dispersed resulting in the limited allelic diversity recorded in this study.

In addition to examining the genetic diversity, this research provided an insight into the population structure of the Ugandan cacao. The majority of the accessions grouped into two undefined clusters, and it was found that these accessions were in highly admixed states (Fig. 5a).

Further analysis demonstrated that these individuals were mainly of Marañon ancestry; however there were also substantial contributions from Amelonado and Iquitos lineages (Fig. 5b). Moreover, it was found that even though the hybrids were placed into two distinct genetic groups, they were not separated by geographical locations based on their genetic backgrounds. Early introductions of cacao to Uganda were from England's Royal Botanic Gardens, Ceylon and Zanzibar which were obtained directly from the West Indies (Brown and Hunter 1913; Urquhart 1958; Bartley 2005). Estates were established from these accessions, but were abandoned due to low yields; fall in cocoa prices and the incidence of pests and diseases (Krug and Quartey-Papafio 1964; ADC/IDEA 1998). The crop was subsequently reintroduced with plantings from surviving plots, most likely of Amelonado and Upper Amazon

hybrids from Ghana (Krug and Quartey-Papafio 1964; ADC/IDEA 1998). The Upper Amazon germplasm taken into Ghana from Trinidad were hybridised with Amelonado material to generate the WACRI Series II varieties during 1971–1985 (Edwin and Masters 2005). This is congruent with the results of this study in which Amelonado, Iquitos, Marañon and Nanay lineages featured prominently in Ugandan cacao. The Upper Amazon material is represented by the latter three groups. The results suggested that either Parinari accessions or other germplasm with high membership in the Marañon population were brought into Uganda in higher proportions than other germplasm groups or that these Marañon individuals had a better survival rate in Uganda. Farmer varieties in Nigeria, that included WACRI varieties, had the highest gene flow from Marañon population than from the Contamana, Iquitos or Nanay populations (Aikpokpodion et al. 2009). It is likely that the surviving hybrids of Marañon, Amelonado and Iquitos ancestry were distributed as seeds to different parts of Uganda to develop the current plantations. This would account for the backgrounds and diversity of the germplasm in the Central and Western Regions being similar to each other but without any duplicate trees.

Although the genetic diversity as indicated by ancestry from the respective genetic groups described by Motamayor et al. (2008) and Refractario Groups 1 and 2 (Motilal et al. 2013) was limited, the combination of the lineages varied substantially among individuals. Additionally, the majority of the accessions were genetically distant to the reference genotypes and each sampled tree had a unique SNP genotype. Together these findings suggested that a substantial amount of genetic recombination occurred in Uganda. Furthermore, deviation of allelic diversity and heterozygosity from mutation-drift equilibrium suggested that the Ugandan cacao population had recently experienced a bottleneck event. This is in agreement with establishment from a limited introduced diversity with subsequent spread as seedling trees from selected trees. Hybridisation among the limited introduced Ugandan germplasm would also explain the heterozygosity excess. About 23% of the loci deviated from HWE in the sampled Ugandan cocoa. Deviations from HWE can be due to many causes including inbreeding and selection as suggested above.

Dendrogram and ancestry analysis indicated that accessions “U026”, “U032” and “U033” were genetically close to Amelonado, Iquitos and Contamana reference genotypes respectively. In contrast to the admixed hybrids observed, these accessions may represent pure reference accessions as they matched their respective references at 98–99% of the loci studied. Additionally, while the admixed hybrids were widely distributed in both the Central and Western Regions, these accessions were found

only in the Central Region in the Luweero (“U026”) and Mukono (“U032” and “U033”) districts. It is possible that these individuals may have different origins from the admixed hybrids. For instance, they may be introgressed individuals from clonal germplasm taken into Uganda. In addition to possessing unique alleles and being a valuable reservoir for genetic diversity, it is likely they may also have important commercial traits. Therefore, it would be worthwhile to include these accessions in breeding programmes to develop improved hybrids.

The absence of cream or white coloured cotyledons suggested an absence or very low representation of Criollo germplasm based on the observations of Wellensiek (1931) and was confirmed by the STRUCTURE results. Criollo ancestry was negligible except for a few admixed trees such as “U019” from Luweero, and “U037” and “U036” from Mukono that had moderate lineage levels of 31.4%, 24.3% and 19.7% respectively. Sample “U088” from Bundibugyo and “U041” from Buikwe had lower levels of 12.3% and 10.6% respectively. Criollo varieties are characterised by ‘plump’ seeds, while Forastero types such as those of the Nanay, Iquitos, Amelonado and Marañon have flattish seeds (Cheesman 1944). The larger seed width and thickness observed in the Central region coincides with the occurrence of these admixed trees in Luweero and Mukono. Accessions with Criollo background have been linked to the production of chocolate with milder, fine, floral, fruity and nutty flavours as well as a desirable aroma and only slight bitterness (Frauendorfer and Schieberle 2008; Kongor et al. 2016). Such beans are highly sought after to produce quality chocolate products and the development of the fine or flavour cocoa industry. Historical records indicated that Criollo and Trinitario hybrids were acquired by Uganda from Ghana and South America in 1987 (ADC/IDEA 1998). Trinitario varieties from Costa Rica and Trinidad in 1988 were also established at the quarantine station on Damba Island (Petithuguenin 2000). It may be possible that the accessions with the Criollo ancestry were derivatives of these introduced accessions; although it cannot be ruled out that they were from earlier introductions or undocumented sources. Nevertheless, the low incidence of Criollo ancestry of farm varieties suggested that most of the germplasm with appreciable Criollo ancestry and their progenies may have died out. This is in keeping with the low vigour and susceptibility to diseases widely attributed to Criollo germplasm. Natural hybridization with parental trees that had non-Criollo background but which were present in higher numbers would also contribute to a successive decrease of Criollo ancestry in surviving progenies.

To conserve the genetic diversity of cacao, adequate representation of the diversity in germplasms is essential. Identifying a reduced set of accessions to represent the

genetic diversity among individuals in a large population is a cost-effective strategy in the management of germplasm collections. It was found that a set of 18 individuals (Table 4) were able to capture the same genetic diversity as the entire set of 109 landraces sampled in Uganda and therefore these accessions should be prioritised for conservation. The identified core did not include any of those top ranked for fruit dimensions, or the outliers based on the quantitative PCA, and only included one tree with high Criollo ancestry (“U019”). Therefore, the core set could be expanded by including the other samples with high Criollo ancestry (“U036” and “U037”) and outliers from the PCA (“U017”, “U018”, “U093”, “U113” and “U118”). The latter included the top ranked samples for fruit length and two of the top ranked samples for fruit width. The other top ranked for fruit width (“U016”, “U095”, “U099” and “U106”) and length (“U122”) can also be considered. These additional trees (divergent types) may be potentially useful for heterotic combinations in breeding and therefore could be considered to complement the core set based on SNP genetic diversity (Table 4). All the samples should be evaluated for their agronomic performance before being used in any breeding programme. In addition, germplasm from under-represented groups in Uganda can be introduced to broaden the gene pool and increase inter-population variation. The additional germplasm could be selected from Criollo, Curaray, Nacional and Purús groups as germplasm in the Contamana and Guiana groups are known to have small seeds.

To our knowledge, this is the first report to characterise the morphological and genetic diversity of cacao landraces in Uganda and of introduced cacao to East Africa. Superior mother trees needed to implement conservation, breeding and rehabilitation programmes were identified. The findings are consistent with the historical records on the movement of cacao into the country. The study has provided a baseline of genetic data and has improved the understanding of the genetic diversity, population structure and degree of admixture of cacao in Uganda.

**Acknowledgements** The authors are grateful to Saila Ramkissoon and Amrita Mahabir for their assistance in extracting DNA from the collected cacao leaf samples. SNP genotyping was supported by the International Fine Cocoa Innovation Centre (IFCIC), of the Cocoa Research Centre, which is funded by the European Union under the ACP Science & Technology Programme II and The University of the West Indies.

#### Compliance with ethical standards

**Conflict of interest** The authors declare they have no conflict of interest.

## References

- ADC/IDEA (1998) A baseline survey on cocoa production in selected districts of Uganda. Development and Management Consultants International. [http://pdf.usaid.gov/pdf\\_docs/Pnadj521.pdf](http://pdf.usaid.gov/pdf_docs/Pnadj521.pdf)
- Addinsoft (2014) XLSTAT version 2014.5.03: data analysis and statistics software for Microsoft Excel. Paris, France
- Aikpokpodion PO, Motamayor JC, Adetimirin VO, Adu-Ampomah Y, Ingelbrecht I, Eskes AB, Schnell RJ, Kolesnikova-Allen M (2009) Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genet Genomes* 5:699–711. <https://doi.org/10.1007/s11295-009-0221-1>
- Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA (1999) Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *Am J Bot* 86:1474–1486
- Araujo QRD, Gattward JN, Almoosawi S, Silva MdGCPC, Dantas PADS, Araujo Júnior QRD (2016) Cocoa and human health: from head to foot—a review. *Crit Rev Food Sci Nutr* 56:1–12. <https://doi.org/10.1080/10408398.2012.657921>
- Banks SC, Peakall R (2012) Genetic spatial autocorrelation can readily detect sex-biased dispersal. *Mol Ecol* 21:2092–2105. <https://doi.org/10.1111/j.1365-294X.2012.05485.x>
- Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI, Wallingford
- Bayer C, Fay MF, Bruijn AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW (1999) Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcl* DNA sequences. *Bot J Linn Soc* 129:267–303. <https://doi.org/10.1111/j.1095-8339.1999.tb00505.x>
- Bekele F, Butler DR (2000) Proposed short list of cocoa descriptors for characterization. In: Eskes AB, Engels JMM, Lass RA (eds) Working procedures for cocoa germplasm evaluation and selection. Proceedings of the CFC/ICCO/IPGRI project workshop, Montpellier, France, 1998. International Plant Genetic Resources Institute, Rome, Italy, pp 41–48
- Bekele FL, Bekele I, Butler DR, Bidaisee GG (2006) Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. *Genet Resour Crop Evol* 53:933–948. <https://doi.org/10.1007/10722-004-6692-x>
- Brown E, Hunter HH (1913) Planting in Uganda. Longmans, Green & Co., London
- Brown AHD, Weir BS (1983) Measuring genetic variability in plant populations. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, Part A. Elsevier, Amsterdam, pp 219–239
- Campoy JA, Lerigoleur-Balsemin E, Christmann H, Beauvieux R, Girollet N, Quero-García J, Dirlwanger E, Barreneche T (2016) Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Biol* 16:49. <https://doi.org/10.1186/s12870-016-0712-9>
- Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop Agric (Trinidad)* 21:144–159
- Cilas C, Machado R, Motamayor JC (2010) Relations between several traits linked to sexual plant reproduction in *Theobroma cacao* L.: number of ovules per ovary, number of seeds per pod, and seed weight. *Tree Genet Genomes* 6:219–226. <https://doi.org/10.1007/s11295-009-0242-9>
- Coe SD, Coe MD (1996) The true history of chocolate. Thames & Hudson Ltd, London

- Cooper KA, Donovan JL, Waterhouse AL, Williamson G (2008) Cocoa and health: a decade of research. *Br J Nutr* 99:1–11. <https://doi.org/10.1017/S0007114507795296>
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001–2014
- Corti R, Flammer AJ, Hollenberg NK, Lüscher TF (2009) Cocoa and cardiovascular health. *Circulation* 119:1433–1441. <https://doi.org/10.1161/CIRCULATIONAHA.108.827022>
- Cuatrecasas J (1964) Cacao and its allies; a taxonomic revision of the genus *Theobroma* vol 35. Contributions from the U.S. National Herbarium
- Dalberg (2015) Smallholder tree crop renovation and rehabilitation (R&R): a review of the state of the emerging R&R market and opportunities to scale investment. <https://www.idhsustainabletrade.com/uploaded/2017/03/Dalberg-RR-Report.pdf>. Accessed 20 June 2017
- Edwin J, Masters WA (2005) Genetic improvement and cocoa yields in Ghana. *Exp Agric* 41:491–503. <https://doi.org/10.1017/S0014479705002887>
- Efombagn MIB, Sounigo O, Nyassé S, Dauleux MM, Cilas C, Eskes MAB, Allen K (2006) Genetic diversity in cocoa germplasm of southern Cameroon revealed by simple sequences repeat (SSRs) markers. *Afr J Biotechnol* 5:1441–1449. <https://doi.org/10.5897/AJB06.287>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Evetts IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer Associates, Sunderland
- Fraundorfer F, Schieberle P (2008) Changes in key aroma compounds of Criollo cocoa beans during roasting. *J Agric Food Chem* 56:10244–10251. <https://doi.org/10.1021/jf802098f>
- Google Maps (2017). <https://www.google.tt/maps/place/Uganda>. Accessed 23 Dec 2017
- Hartl DL, Clark AG (1997) Principles of population genetics, 3rd edn. Sinauer Associates Inc, Sunderland
- International Cocoa Organization (2014) Cocoa Market Update. <http://www.worldcocoafoundation.org/wp-content/uploads/Cocoa-Market-Update-as-of-4-1-2014.pdf>. Accessed 20 June 2017
- International Cocoa Organization (2017) International Cocoa Organization annual report 2014/2015. [https://www.icco.org/about-us/international-cocoa-agreements/doc\\_download/2647-annual-report-2014-2015-english-french-spanish-russian-full.html](https://www.icco.org/about-us/international-cocoa-agreements/doc_download/2647-annual-report-2014-2015-english-french-spanish-russian-full.html). Accessed 20 June 2017
- Ji K, Zhang D, Motilal LA, Boccara M, Lachenaud P, Meinhardt LW (2013) Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. *Genet Resour Crop Evol* 60:441–453. <https://doi.org/10.1007/s10722-012-9847-1>
- Jost L (2008)  $G_{ST}$  and its relatives do not measure differentiation. *Mol Ecol* 17:4015–4026. <https://doi.org/10.1111/j.1365-294X.2008.03887.x>
- Kennedy AJ, Mooleedhar V (1993) Conservation of cocoa in field genebanks—the International Cocoa Genebank, Trinidad. In: International workshop on conservation, characterisation and utilization of cocoa genetic resources in the 21st century, Port of Spain, Trinidad, 1992. Cocoa Research Unit, The University of the West Indies, pp 21–23
- Kongor JE, Hinneh M, Van de Walle D, Afoakwa EO, Boeckx P, Dewettinck K (2016) Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile—a review. *Food Res Int* 82:44–52. <https://doi.org/10.1016/j.foodres.2016.01.012>
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and population structure inferences across *K*. *Mol Ecol Resour* 15:1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Krug CA, Quartey-Papafio E (1964) World cocoa survey, vol 63. Food and Agriculture Organisation of the United Nations, Rome
- Lockwood G, End MJ (1993) History, technique and future needs for cacao collection. In: International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century, Port-of-Spain, Trinidad, 1992. Cocoa Research Unit, The University of the West Indies, pp 1–14
- Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv Biol* 12:228–237
- Lukman, Zhang D, Susilo AW, Dinarti D, Bailey B, Mischke S, Meinhardt LW (2014) Genetic identity, ancestry and parentage in farmer selections of cacao from Aceh, Indonesia revealed by single nucleotide polymorphism (SNP) markers. *Trop Plant Biol* 7:133–143. <https://doi.org/10.1007/s12042-014-9144-6>
- Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS ONE* 3:e3311. <https://doi.org/10.1371/journal.pone.0003311>
- Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor RG, Martinez WJ, Graham J, Kuhn DN, Brown S, Schnell RJ (2010) No mas forastero: a new protocol for meaningful cacao germplasm classification. In: Proceedings of the 16th international cocoa research conference, 2010. COPAL, Nigeria, pp 179–185
- Motamayor JC, Risterucci AM, Heath M, Lanaud C (2003) Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. *Heredity* 91:322–330
- Motilal LA, Sankar A, Gopaulchan D, Umaharan P (2017) Molecular markers and marker assisted selection: Cocoa. In: Chowdappa P, Karun A, Rajesh MK, Ramesh SV (eds) Biotechnology of plantations crops. Daya Publishing House, New Delhi, pp 313–354
- Motilal LA, Sreenivasan TN (2012) Revisiting 1727: crop failure leads to the birth of Trinitario cacao. *J Crop Improv* 26:599–626. <https://doi.org/10.1080/15427528.2012.663734>
- Motilal LA, Zhang D, Mischke S, Meinhardt LW, Umaharan P (2013) Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa Genebank, Trinidad. *Tree Genet Genomes* 9:1395–1411. <https://doi.org/10.1007/s11295-013-0645-5>
- Motilal LA, Zhang D, Umaharan P, Mischke S, Pinney S, Meinhardt LW (2011) Microsatellite fingerprinting in the International Cocoa Genebank, Trinidad: accession and plot homogeneity information for germplasm management. *Plant Genet Resour* 9:430–438. <https://doi.org/10.1017/S147926211100058X>
- National Institute of Agricultural Biotechnology (2006) PowerCore (v. 1.0). A program applying the advanced M strategy using heuristic search for establishing core or allele mining sets. R. Korea: Rural Development Administration (RDA). <http://www.genebank.go.kr/eng/PowerCore/powercore.jsp>. Accessed 7 Nov 2012
- Oddoye EOK, Agyente-Badu CK, Gyedu-Akoto E (2013) Cocoa and its by-products: identification and utilization. In: Watson RR, Preedy VR, Zibadi S (eds) Chocolate in health and nutrition, vol 7. Humana Press, Totowa, NJ, pp 23–37
- Osorio-Guarín JA, Berdugo-Cely J, Yockeng RA, Zapata YP, Quintero C, Gallego-Sánchez G, Yockeng R (2017) Colombia a source of cacao genetic diversity as revealed by the population structure analysis of germplasm bank of *Theobroma cacao* L. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2017.01994>

- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
- Peakall R, Smouse PE (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. <http://darwin.cirad.fr/darwin>
- Petithuguenin P (2000) The situation of cocoa production in Uganda. First consultancy progress report for the ADC/IDEA project. CIRAD, France
- Pokou ND, N’Goran JAK, Lachenaud PH, Eskes AB, Montamayor JC, Schnell R, Kolesnikova-Allen M, Clément D, Sangaré A (2009) Recurrent selection of cocoa populations in Côte d’Ivoire: comparative genetic diversity between the first and second cycles. *Plant Breed* 128:514–520. <https://doi.org/10.1111/j.1439-0523.2008.01582.x>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rambaut A (2012) FigTree v1. 4.2. <http://tree.bio.ed.ac.uk/>
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Smouse PE, Peakall R, Gonzales E (2008) A heterogeneity test for fine-scale genetic structure. *Mol Ecol* 17:3389–3400. <https://doi.org/10.1111/j.1365-294X.2008.03839.x>
- Takrama J, Kun J, Meinhardt L, Mischke S, Opuku SY, Padi FK, Zhang D (2014) Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers. *Afr J Biotechnol* 13(21):2127–2136. <https://doi.org/10.5897/AJB2013.13331>
- Toxopeus H (1972) Cocoa breeding: a consequence of mating system heterosis and population structure. In: Wastie RL, Earp DA (eds) Proc of the Conference on cocoa and coconuts in Malaysia 25–27 November 1971 Kuala Lumpur. Incorporated Society of Planters, Kuala Lumpur, pp 3–12
- Toxopeus H (1985a) Botany, types and populations. In: Wood GAR, Lass RA (eds) Cocoa, 4th edn. Longman, London, pp 11–37
- Toxopeus H (1985b) Planting material. In: Wood GAR, Lass RA (eds) Cocoa, 4th edn. Longman, London, pp 80–92
- Uganda Bureau of Statistics (2016) Uganda Bureau of Statistics statistical abstract. [http://www.ubos.org/onlinefiles/uploads/ubos/statistical\\_abstracts/2016%20Statistical%20Abstract.pdf](http://www.ubos.org/onlinefiles/uploads/ubos/statistical_abstracts/2016%20Statistical%20Abstract.pdf). Accessed 20 June 2017
- Urquhart DH (1958) Prospects for cocoa-growing in Uganda and Zanzibar. Cadbury Brothers Ltd., Bournville
- Utro F, Cornejo OE, Livingstone D, Motamayor JC, Parida L (2012) ARG-based genome-wide analysis of cacao cultivars. *BMC Bioinformatics* 13(Suppl 19):S17 <http://www.biomedcentral.com/1471-2105/13/S19/S17>
- van Hall CJJ (1914) Cocoa. Macmillan, London
- Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Mol Ecol* 10:249–256. <https://doi.org/10.1046/j.1365-294X.2001.01185.x>
- Wellensiek SJ (1931) The genetics of cotyledon-colour of cocoa as a basis for quality selection. *Archief voor de Koffiecultuur* 5:217–232 (**Translated into English by Toxopeus H, Wessel PC (eds) (1983) In: Archives of Cocoa Research Volume II. Wageningen, ACRI/IOCC**)
- Yang JY, Scascitelli M, Motilal LA, Sveinsson S, Engels JMM, Kane NC, Dempewolf H, Zhang D, Maharaj K, Cronk QCB (2013) Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genet Genomes* 9:829–840. <https://doi.org/10.1007/s11295-013-0601-4>
- Zhang D, Martínez WJ, Johnson ES, Somarriba E, Phillips-Mora W, Astorga C, Mischke S, Meinhardt LW (2012) Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genet Resour Crop Evol* 59:239–252. <https://doi.org/10.1007/s10722-011-9680-y>
- Zhang D, Motilal L (2016) Origin, dispersal, and current global distribution of cacao genetic diversity. In: Bailey BA, Meinhardt LW (eds) Cacao diseases: a history of old enemies and new encounters. Springer, Basel, pp 3–31. [https://doi.org/10.1007/978-3-319-24789-2\\_1](https://doi.org/10.1007/978-3-319-24789-2_1)