

## Article

# Genetic diversity and population structure in a regional collection of Kersting's groundnut (*Macrotyloma geocarpum* (Harms) Maréchal & Baudet)

Konoutan M. Kafoutchoni <sup>1,2,\*</sup>, Eric E. Agoyi <sup>1</sup>, Symphorien Agbahoungba <sup>1</sup>, Achille E. Assogbadjo <sup>1</sup> and Clément Agbangla <sup>2</sup>

<sup>1</sup> Non-Timber Forest Products and Orphan Crop Species Unit, Laboratory of Applied Ecology (LEA), Faculty of Agronomic Sciences, University of Abomey-Calavi; 05 BP 1752 Cotonou, Benin

<sup>2</sup> Laboratoire de Génétique Moléculaire et d'Analyse des Génomes (LGMAG), Faculty of Sciences and Techniques, University of Abomey-Calavi; BP 142, Abomey-Calavi, Benin

\* Correspondence: k.medard@gmail.com; Tel.: +229-94-58-0569

Received: date; Accepted: date; Published: date

**Abstract:** Kersting's groundnut is an important source of protein and essential nutrients that contribute to food security in West Africa. However, the crop is still underexploited by the populations and under-researched by the scientific community. This study aimed to investigate the genetic diversity and population structure of 217 Kersting's groundnut accessions from five origins using 886 DArTseq markers. Gene diversity was low and ranged from 0.049 to 0.064. The number of private alleles greatly varied among populations (42–192) and morphotypes (40–339). Moderate to very high levels of selfing and inbreeding were observed among populations ( $s=56\text{--}85\%$ ,  $F_{IS}=0.389\text{--}0.736$ ) and morphotypes ( $s=57\text{--}82\%$ ,  $F_{IS}=0.400\text{--}0.691$ ). Moreover, little to very high genetic differentiations were observed among populations ( $0.006 \leq F_{IS} \leq 0.371$ ) and morphotypes ( $0.029 \leq F_{IS} \leq 0.307$ ). Analysis of molecular variance partitioned 38.5% of the genetic variation among and 48.7% within populations ( $P < 0.001$ ). Significant isolations by distance were detected between populations ( $R^2=0.612$ ,  $P=0.011$ ) and accessions ( $R^2=0.499$ ,  $P < 0.001$ ). Discriminant analysis of principal components and neighbour joining consistently distinguished eight distinct clusters. These data provide a global picture of the existing genetic diversity for Kersting's groundnut and will guide the choice of breeding strategies to increase production.

**Keywords:** orphan crop; genotyping-by-sequencing; inbreeding; pre-breeding; population genetics; DArTseq; isolation by distance

## 1. Introduction

Kersting's groundnut (*Macrotyloma geocarpum* (Harms) Maréchal & Baudet) is a geocarpic orphan legume crop grown in the semi-arid regions of West Africa. It is a diploid ( $2n = 22$ ) species and likely originated from an area extending between central Benin and northern Togo [1], though a putative wild form *M. geocarpum* var. *tisserantii* ( $2n = 20$ ) was found in northern Cameroon [2]. Kersting's groundnut is mainly grown for its grains which contain significant amounts of protein, essential amino acids, minerals (Ca, Mg, K, Zn, Na, P) and vitamins (A, B1, B2) [3]. It is also an excellent source of carbohydrate. Besides, the crop has the potential to restore soil fertility through atmospheric  $N_2$ -fixation in root nodules [4]. This ranks Kersting's groundnut among the crops that improve productivity in agroecosystems [5]. In Benin, Kersting's groundnut is ranked as the most priority and the most valuable orphan legume crop due to its substantial contribution to food security and income for resource-poor rural households [6, 7]. Despite these attributes, cultivation of Kersting's groundnut is gradually being abandoned in the growing countries [8-12], especially in

Benin [13]. Low yield, small grain size and lack of improved cultivars are among the commonly reported factors responsible for its reduced cultivation [13-15] and its anticipated extinction [4, 9]. Thus, it is urgent to develop new Kersting's groundnut varieties with improved agronomic performances for boosted production.

Successful development of improved cultivars requires variations within germplasm [16, 17]. Assessment of genetic diversity and relatedness among accessions and populations is therefore fundamental to the success of any breeding program. Knowledge of this is crucial in the selection of parental lines and the identification of genotypes that harbour valuable alleles for breeding targets [18]. Moreover, characterisation of genetic diversity is important for effective management of germplasm and development of core collection of genotypes representing maximum diversity in a minimum number of accessions [19]. However, genetic diversity in Kersting's groundnut is poorly documented.

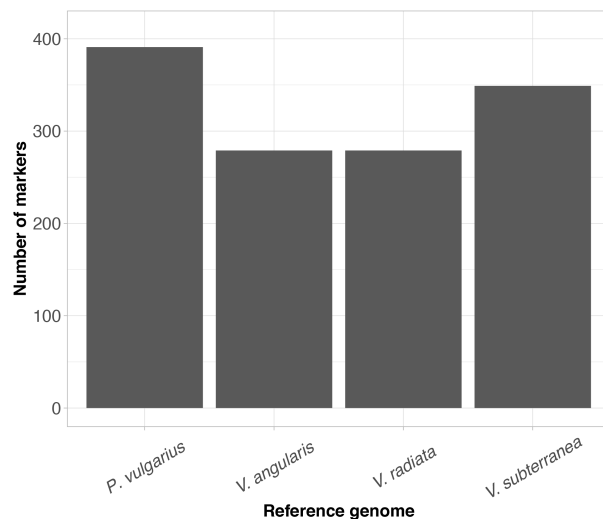
To date, most of the investigations on genetic diversity in Kersting's groundnut were based on agromorphological traits [13, 20-24]. Although germplasm characterisation using agromorphological traits is frequently used by breeders in sub-Saharan Africa [17], this approach also has many limitations. It is time consuming and prone to environmental effects compared to molecular characterisation [25]. The first attempt to investigate the genetic diversity in Kersting's groundnut using molecular markers used only 19 isozymes and was based on 20 accessions [26], thus limiting its power to depict the actual diversity in the crop. Isozyme markers are less in number, detect less polymorphism, have little reproducibility and are known to be affected by the extraction method, plant tissues and plant growth stages, and can therefore produce unreliable diversity estimates [16, 27]. To understand the real genetic variation within the crop, a marker system that produces a lot more markers should be used in a large panel of accessions. The high throughput genotyping approach Genotyping-by-Sequencing (GBS), that reduces complexity in genome to partially sequence crop genome, has shown significant success in many annual legume crops including soybean [28-30] in cowpea [31, 32] Common beans [33-35]. DArTseq technology (Diversity Array Technology), bases on GBS principle to generate thousands of SNP markers. It works well with many crops including neglected crops without reference genome [36]. DArTseq based markers have been successfully used to unfold genetic diversity patterns in various crops including watermelon [37], garlic [38], rice [39], wheat [40], and Bambara groundnut [41]. Recently, Akohoue et al. [42] used 493 DArTseq SNP markers on a collection of 281 Kersting's groundnut accessions and found a low genetic diversity influenced by the geographic origins of the accessions. Nevertheless, the study did not include accessions from all known geographical origins of the crop, which may affect the level of reported genetic diversity. The present study was undertaken to investigate the level of genetic diversity among a regional collection of 227 Kersting's groundnut accessions from five west African countries and to determine population structure within the germplasm.

## 2. Results

### 2.1. Marker quality

The DArTseq genotyping of the panel of 227 Kersting's groundnut generated a total of 1938 Single Nucleotide Polymorphism (SNP) markers of which 391 (20%) were aligned with the reference genome of Common bean (*Phaseolus vulgaris*), 279 (14%) with Adzuki bean (*Vigna angularis*), 279 (14%) with Mung bean (*Vigna radiata*), and 349 (18%) with Bambara groundnut (*Vigna subterranean*) (Figure 1). The markers contained 20.23 % missing data. The probabilistic principal components analysis (PPCA) imputation method that had the highest performance, with an accuracy of 0.97 SMC (Simple matching coefficient), was used to impute the missing SNPs. After applying the various filtering criteria, 1052 SNPs (54.28%) were filtered out and the remaining 886 markers were used for further analyses (Table S1). Ten accessions had over 50% missing SNPs and were removed from the final dataset. All the filtered markers had 100% call rate and an average of 99.8% reproducibility (Table 1). Read depth ranged between 5 and 140 with a mean across loci of 21.71. The distribution of read depths showed that most markers had read depth between 5 and 15 (Figure 2a). The mean PIC

of the reference alleles (0.064) was greater than that of the SNP alleles (0.054) (Table 1). The average PIC ranged from 0.004 to 0.499 with a mean value of 0.059. About 83% of the SNP markers (738) had PIC below 0.05 (Figure 2b). Only 16.7% (148 SNPs) scored greater than 5% PIC. The minor allele frequency (MAF) varied from 0.002 to 0.495 with a mean of 0.044 (Table 1). About 12% (123) of markers had minor allele frequency greater than 0.05 (Figure 3c). Furthermore, mean expected and observed heterozygosity were estimated at 0.06 (0.010-0.500) and 0.008 (0-0.793) respectively. The total gene diversity ( $H_T$ ) across SNPs ranged from 0.002 to 0.501, with a mean  $H_T$  of 0.057 (Table S2).

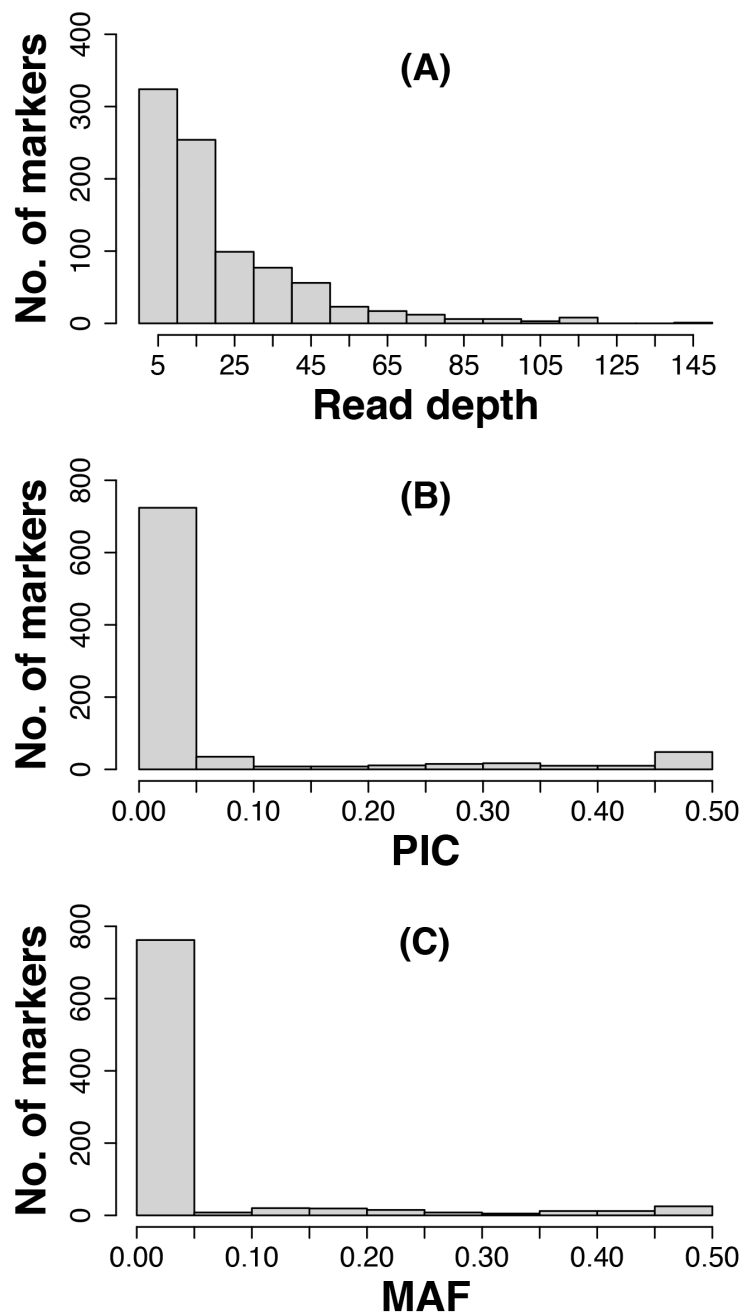


**Figure 1.** Alignment of DArTseq generated SNP markers of Kersting's groundnut with the reference genome of related species.

**Table 1.** Quality control metrics for the 886 SNP markers

Marker Quality Parameter	Mean	Min	Max
Call rate	1.000	1.000	1.000
Read depth	21.714	5.000	140.100
Reproducibility	0.998	0.950	1.000
PIC - reference allele	0.064	0.000	0.499
PIC - SNP allele	0.054	0.008	0.500
Average PIC	0.059	0.004	0.499
Minor allele frequency	0.044	0.002	0.495

PIC Polymorphic information content



**Figure 2.** Distribution of (a) read depth, (b) Polymorphism information content (PIC), and (c) minor allele frequency (MAF) for 886 DArTseq SNP markers in Kersting's groundnut.

## 2.2. Genetic diversity analysis based on geographical populations and morphotypes

The genetic diversity parameters estimated for each geographical population are summarised in Table 2. Number of different alleles ( $N_a$ ), number of effective alleles ( $N_e$ ) and allelic richness ( $A_r$ ) varied slightly among the populations.  $N_a$  ranged from 1.19 (Nigeria) to 1.54 (Zou). The number of effective alleles was nearly the same for all populations. The mean  $N_e$  across populations (1.10) was lower than mean  $N_a$  (1.30) suggesting that only a few alleles contributed to the variation.  $A_r$  ranged between 1.12 (Nigeria) and 1.20 (Ghana) with a mean value of 1.17. Private alleles were found in all populations with Zou possessing the highest number of private alleles ( $N_p = 192$ ) whereas Nigeria had the lowest number ( $N_p = 42$ ). The accessions with the greater number of private alleles were Gbo5, Lili\*Sie, and Itk from Zou, Ghana and Collines populations, respectively (Table S3). In general, the observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) and unbiased expected heterozygosity ( $uH_e$ ) significantly varied ( $P < 0.001$ ) among populations.  $H_o$  values were approximately 6-7 times

lower than  $H_e$  and  $uH_e$ .  $H_o$  ranged between  $0.69 \times 10^{-2}$  and  $1.02 \times 10^{-2}$ , with an average of  $0.79 \times 10^{-2}$  across populations. The more diverse population was from Atacora ( $H_e = 6.12 \times 10^{-2}$ ,  $uH_e = 6.24 \times 10^{-2}$ ) followed by Zou population ( $H_e = 6.07 \times 10^{-2}$ ,  $uH_e = 6.12 \times 10^{-2}$ ), whereas Ghana's population was the least diverse ( $H_e = 5.32 \times 10^{-2}$ ,  $uH_e = 5.40 \times 10^{-2}$ ). The inbreeding coefficient within populations varied from moderate in Zou ( $F_{IS} = 0.39$ ) to very high in Collines ( $F_{IS} = 0.74$ ) and Nigeria ( $F_{IS} = 0.74$ ). The selfing rate greatly varied among populations. It ranged between  $s = 0.56$  for Zou population and  $s = 0.85$  for Collines and Nigeria populations. The across populations mean was  $s = 0.72$ , indicating a considerable level of selfing. Selfing rates was proportional to inbreeding coefficients and inversely proportional to polymorphism within populations (Table 2). Percentage of polymorphism ranged from 18.51 % (Nigeria) to 53.50 % (Zou) with an average of 39.39 % indicating a moderate level of polymorphism and variation within geographical populations.

**Table 2.** Genetic diversity parameters for Kersting's groundnut populations using 886 SNP markers for geographical populations of Kersting's groundnut

Population	N	N <sub>a</sub>	N <sub>e</sub>	A <sub>r</sub>	N <sub>p</sub>	H <sub>o</sub> *	H <sub>e</sub> *	uH <sub>e</sub> *	F <sub>IS</sub>	s	%P
Atacora	25	1.25	1.10	1.19	75	0.89	6.12	6.24	0.65	0.79	25.17
Burkina	48	1.34	1.09	1.19	141	0.69	5.61	5.67	0.54	0.70	33.52
Collines	13	1.21	1.10	1.14	49	0.83	5.93	6.17	0.74	0.85	21.11
Ghana	34	1.36	1.08	1.20	84	0.70	5.32	5.40	0.55	0.71	35.55
Nigeria	12	1.19	1.10	1.12	42	0.70	5.66	5.91	0.74	0.85	18.51
Togo	19	1.25	1.10	1.17	51	0.71	5.92	6.08	0.64	0.78	25.40
Zou	66	1.54	1.10	1.18	192	1.02	6.07	6.12	0.39	0.56	53.50
Overall	217	1.30	1.10	1.17	-	0.79	5.80	5.94	0.57	0.72	39.39

\*  $\times 10^{-2}$ , N = Sample Size, N<sub>a</sub> = Number of different alleles, N<sub>e</sub> = Number of effective alleles, A<sub>r</sub> = Allelic richness, N<sub>p</sub> = Number of private alleles, H<sub>o</sub> = Observed Heterozygosity, H<sub>e</sub> = Expected Heterozygosity, uH<sub>e</sub> = Unbiased Expected Heterozygosity, F<sub>IS</sub> = Inbreeding coefficient, s = Selfing rate, %P = Percentage of Polymorphic Loci.

The estimated genetic diversity parameters also varied among morphotypes as shown in Table 3. The number of alleles (N<sub>a</sub>) varied between 1.17 in brown seed to 1.59 in cream seed, with an average of 1.35 across morphotypes. The average number of effective alleles across morphotypes was 1.09. The highest allelic richness (1.22) was found in black morphotype. The number of private alleles varied greatly among morphotypes with cream seed possessing the highest number of private alleles (N<sub>p</sub> = 339) and brown seed the lowest (N<sub>p</sub> = 40). Genetic diversity was low but significantly ( $P < 0.001$ ) varied among morphotypes (Table 3). The more diverse morphotype were cream seed with black eye ( $H_e = 6.29 \times 10^{-2}$ ,  $uH_e = 6.40 \times 10^{-2}$ ) and black ( $H_e = 6.18 \times 10^{-2}$ ,  $uH_e = 6.25 \times 10^{-2}$ ), whereas the least diverse one was brown seed ( $H_e = 4.69 \times 10^{-2}$ ,  $uH_e = 4.88 \times 10^{-2}$ ). Consistent with that, high levels of inbreeding ( $F_{IS} = 0.69$ ) and selfing ( $s = 0.82$ ) were observed within brown seed, while the lowest levels were found in cream with black eye seed ( $F_{IS} = 0.40$ ,  $s = 0.57$ ). The average percentage of polymorphism was 35.40% suggesting a moderate level of variation within morphotypes. The percentage of polymorphism was lowest within brown seed (17.27%) and highest in cream seed (59.37%).

**Table 3.** Genetic diversity parameters for Kersting's groundnut morphotypes using 886 SNP markers

Morphotype	N	N <sub>a</sub>	N <sub>e</sub>	A <sub>r</sub>	N <sub>p</sub>	H <sub>o</sub> *	H <sub>e</sub> *	uH <sub>e</sub> *	F <sub>IS</sub>	s	%P
Black	49	1.36	1.10	1.22	147	0.79	6.18	6.25	0.55	0.71	36.12
Brown	13	1.17	1.08	1.18	40	0.75	4.69	4.88	0.69	0.82	17.27
Brown Mottled	23	1.24	1.09	1.18	66	0.73	5.45	5.57	0.68	0.81	23.81
Cream	102	1.59	1.10	1.21	339	0.67	5.82	5.85	0.41	0.58	59.37
Cream BE	30	1.40	1.10	1.21	76	1.50	6.29	6.40	0.40	0.57	40.41
Overall	217	1.35	1.09	1.20	-	0.89	5.69	5.79	0.50	0.67	35.40

\*  $\times 10^{-2}$ , BE = Black eye, N = Sample Size, Na = Number of different alleles, Ne = Number of effective alleles, Ar = Allelic richness, Np = Number of private alleles,  $H_o$  = Observed Heterozygosity,  $H_e$  = Expected Heterozygosity,  $uH_e$  = Unbiased Expected Heterozygosity,  $F_{is}$  = Inbreeding coefficient,  $s$  = Selfing rate, %P = Percentage of Polymorphic Loci.

### 2.3. Genetic differentiation across geographic populations and morphotypes

Genetic differentiation of geographic populations (Table 4) and morphotypes (Table 5) were examined based on  $F_{ST}$ . The  $F_{ST}$  coefficients among populations varied from 0.006 to 0.371, indicating little to very great differentiation. The highest level of differentiation was observed between Burkina and Zou populations. The lowest  $F_{ST}$  was observed between Collines and Zou. A very low genetic differentiation was also found between Atacora and Zou ( $F_{ST} = 0.008$ ), and between Atacora and Collines ( $F_{ST} = 0.017$ ). Therefore, individuals between populations from Benin (Atacora, Collines, Zou) were closely related ( $F_{ST} < 0.05$ ). Equally low genetic differentiation was also observed between populations from Benin and Nigeria ( $0.033 < F_{ST} < 0.065$ ) (Table 4). Low to very high differentiation was also observed between morphotypes (Table 5). The lowest differentiation was found between cream and brown ( $F_{ST} = 0.029$ ) while differentiation was the highest between brown mottled and brown morphotypes ( $F_{ST} = 0.307$ ). A very high differentiation was also found between brown mottled and cream morphotypes ( $F_{ST} = 0.302$ ). The genetic distance between cream and cream with black eye morphotypes was also relatively high ( $F_{ST} = 0.265$ ).

**Table 4.** Pairwise genetic differentiation ( $F_{ST}$ ) between populations of Kersting's groundnut

Population	Atacora	Burkina	Collines	Ghana	Nigeria	Togo
Burkina	0.301					
Collines	0.017	0.275				
Ghana	0.295	0.099	0.285			
Nigeria	0.056	0.243	0.065	0.248		
Togo	0.192	0.111	0.226	0.151	0.192	
Zou	0.008	0.371	0.006	0.315	0.033	0.183

**Table 5.** Pairwise genetic differentiation ( $F_{ST}$ ) between morphotypes of Kersting's groundnut

Morphotype	Black	Brown	BM	Cream
Brown	0.091			
BM	0.128	0.307		
Cream	0.231	0.029	0.302	
CBE	0.077	0.183	0.202	0.265

BM= Brown Mottled, CBE= Cream with black eye

The AMOVA of 217 accessions from seven populations revealed that 38.53% of the total variation was among populations ( $F_{ST} = 0.385$ ) and 48.72 % within populations ( $F_{is} = 0.793$ ), while the rest of variation (12.74%) was within accessions ( $F_{IT} = 0.873$ ; Table 6). All values were statistically significant ( $P < 0.001$ ). The two-level AMOVA across morphotypes showed significant levels of genetic variation within (49.13%) and among morphotypes (38.73%), with 12.14% of the variation retained within accessions (Table 7). The overall fixation ( $F_{ST}$ ) value was equal to 0.387 among morphotypes. There was homozygote excess as revealed by the high overall inbreeding coefficient ( $F_{is} = 0.802$ ). The three-level AMOVA (Table 7) partitioned 22.10% of the genetic variation among the seven geographic populations and about 25% of variation among morphotypes. In addition, 40.39% of the total variation was observed within morphotypes, whereas the remaining 12.72% was found within individuals. All the values were statistically significant ( $P < 0.001$ ) except at population level ( $P = 0.097$ ).



**Table 6.** Partitioning of genetic variance across geographic populations of Kersting's groundnut

Source	Df	SS	MS	Est. Var.	%	F-statistics	P*
Among pops	6	8113.87	1352.31	22.1	38.53	$F_{ST} = 0.385$	0.001
Among accessions/within pops	210	13271.54	63.2	27.94	48.72	$F_{IS} = 0.793$	0.001
Within accessions	217	1586	7.31	7.31	12.74	$F_{IT} = 0.873$	0.001
Total	433	22971.41	53.05	57.35	100		

Df = Degree of freedom, SS= Sum of squares, MS= Mean square, Est.Var.= Estimated variance, %= Percentage of variation, \*= *P*-value based on 999 permutations test,  $F_{ST}$  = Fixation index,  $F_{IS}$ = Inbreeding coefficient of an individual relative to subpopulations,  $F_{IT}$ = Inbreeding coefficient of an individual relative to the total population

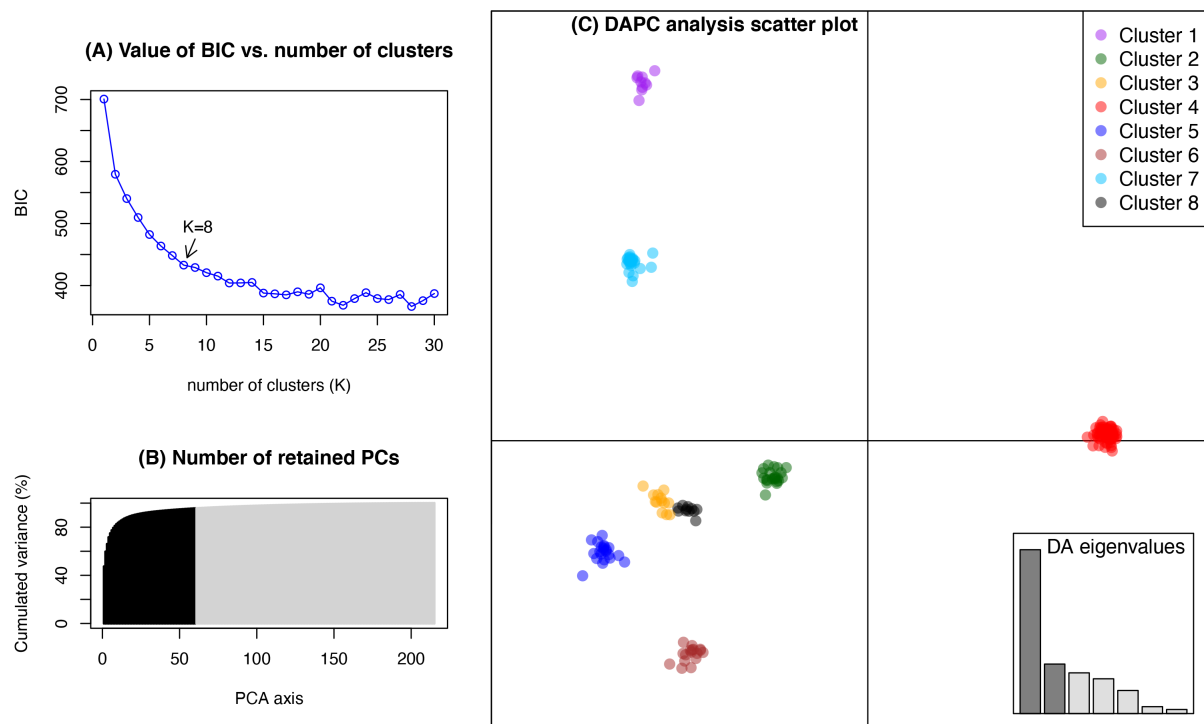
**Table 7.** Partitioning of genetic variance across morphotypes of Kersting's groundnut

Source	Df	SS	MS	Est. Var.	%	F-statistics	P*
<i>Two-level</i>							
Among morphotypes	4	7291.48	1822.87	23.32	38.73	$F_{ST} = 0.387$	0.001
Among individuals/within morphotypes	212	14093.94	66.48	29.59	49.13	$F_{IS} = 0.802$	0.001
Within individuals	217	1586	7.31	7.31	12.14	$F_{IT} = 0.879$	0.001
Total	433	22971.41	53.05	60.22	100		
<i>Three-level</i>							
Among populations	6	8113.87	1352.31	12.69	22.10		0.097
Among morphotypes/Within populations	9	2476.27	275.14	14.24	24.79		0.001
Among individuals/within morphotypes	201	10795.27	53.71	23.2	40.39		0.001
Within individuals	217	1586	7.31	7.31	12.72		0.001
Total	433	22971.41	53.05	57.44	100		

Df= Degree of freedom, SS= Sum of squares, MS= Mean square, Est. Var.= Estimated variance, %= Percentage of variation, \*= *P*-value based on 999 permutations test,  $F_{ST}$ = Fixation index,  $F_{IS}$ = Inbreeding coefficient of an individual relative to subpopulations,  $F_{IT}$ = Inbreeding coefficient of an individual relative to the total population

#### 2.4. Population structure

A discriminant analysis of principal components (DAPC) was used to infer population structure of Kersting's groundnut. The optimum number of clusters inferred using the *find.clusters* function was eight (Figure 3a, c). DAPC analysis was performed using the inferred number of clusters. Based on DAPC cross-validation analysis, the 60 first PCs (95% of variance conserved) of PCA (Figure 3b) and 7 discriminant functions were retained. Clusters 1 to 8 consisted of 10, 22, 13, 98, 22, 17, 24 and 11 accessions, respectively. Cluster 1 was predominantly composed of accessions from Burkina Faso, with one accession from Atacora. Cluster 2 contained the majority (52.6%) of accessions from Togo, about 7.5% from Zou, with a few accessions from Nigeria, Burkina Faso, Collines and one accession from Atacora. Almost all the remaining accessions from Atacora possessed 80% membership in Cluster 4 which also included the majority of accessions from Zou, Nigeria and Collines (80.3%, 83.3% and 84.6% respectively). Cluster 3 contained mainly accessions from Ghana and Burkina Faso, whereas Cluster 5 included one third of the accessions from Burkina Faso and about 15% from Ghana. Clusters 6 and 7 were predominantly composed of accessions from Burkina Faso and the majority (~60%) of the accessions from Ghana, respectively (Table 8). Cluster 8 contained almost 37% of the accessions from Togo, and the remaining accessions from Atacora and Zou.



**Figure 3.** Discriminant analysis of principal components (DAPC) for 217 Kersting's groundnut accessions using 886 SNPs. (a) Statistical determination of the optimum number of clusters, (b) Number of retained DAPC principal components, (c) Scatter plot from the DAPC analysis.

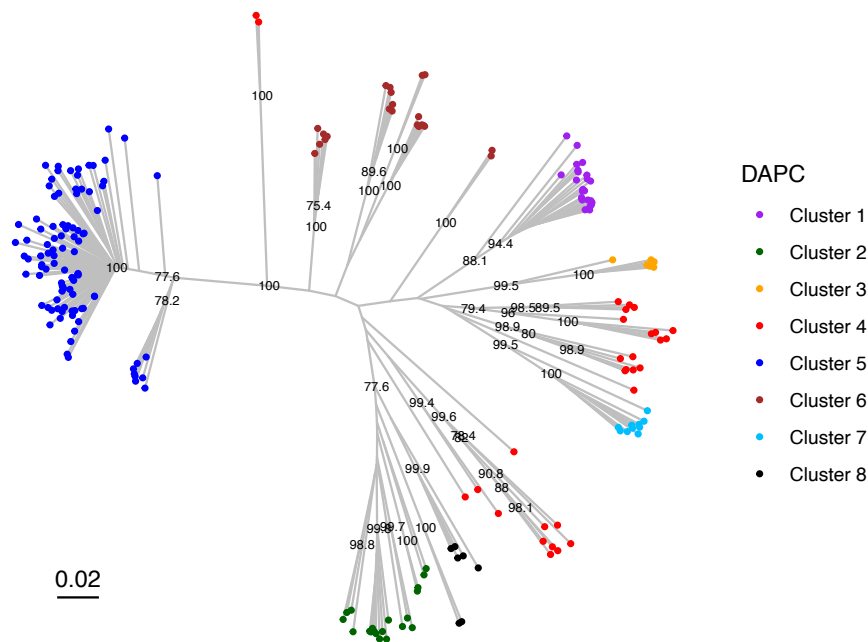
**Table 8.** Cluster assignment based on DAPC results.

Population	DAPC inferred clusters							
	C1	C2	C3	C4	C5	C6	C7	C8
Atacora	1	1	1	20	-	-	-	2
Burkina	9	2	5	-	16	14	2	-
Collines	-	2	-	11	-	-	-	-
Ghana	-	-	6	3	5	1	19	-
Nigeria	-	2	-	10	-	-	-	-
Togo	-	10	-	1	-	1	-	7
Zou	-	5	1	53	1	1	3	2

For each population, the values represent the number of accessions assigned to a cluster.

To obtain a more complete picture of the relationships among the 217 Kersting's groundnut accessions a neighbour joining (NJ) tree was constructed based on Edward's distance (Figure 4). The NJ tree separated all inferred clusters from DAPC, in highly supported groups (>75% bootstrap support (BS) value), though there was a branch of cluster 3 for which the BS value was slightly lower (66%; Figure 4).

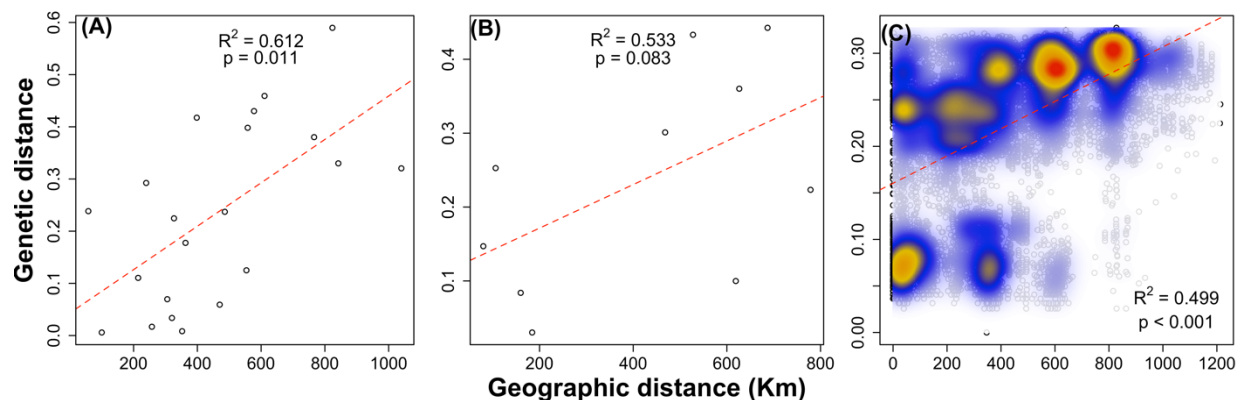




**Figure 4.** Unrooted neighbour joining tree based on Edward's distance calculated for 217 accessions of Kersting's groundnut, coloured by cluster calls DAPC. Numbers associated with branches indicate bootstrap values based on 999 replications.

### 2.5. Isolation by distance

The Mantel test comparing genetic distance [ $F_{ST}/(1-F_{ST})$ ] and geographic distance per population indicated the presence of a significant isolation by distance (IBD) ( $R^2 = 0.612$ ,  $P = 0.011$ ; Figure 5a). In contrast, no significant IBD was detected among morphotypes ( $R^2 = 0.533$ ,  $P = 0.083$ ; Figure 5b). Figure 6c showed that genetic distance between pairs of Kersting's groundnut accessions increased significantly with geographic distance, with a Mantel test correlation of 0.499 and a probability of  $P < 0.001$ . This IBD among accessions did not result from a single continuous cline across the range of the crop, but of eight distant populations that are genetically distinct (Figure 5c).



**Figure 5.** Scatter plots illustrating the pairwise relationship between genetic distance ( $F_{ST}/(1-F_{ST})$ ) and geographic distances (Km) between populations (a) morphotypes (b) and samples (c). Warmer colors indicate higher densities of points.

## 3. Discussion

Genetic diversity is particularly important to breeders as a source of novel alleles for improving diverse complex traits of agronomic interest, particularly with regards to exacerbated climate challenges and sophisticated end-user demands. In the present study, genetic diversity and population structure were assessed in a panel composed of 217 Kersting's groundnut accessions from different origins, using 886 DArTseq-generated SNP markers. The information provided is useful to

breeders and geneticists. It will help for efficient management and conservation of Kersting's groundnut genetic resources, as the crop is almost on the verge of extinction in many growing countries [8, 9, 26].

### 3.1. SNP discovery and marker quality

DArTseq has been successfully applied in many recent plant breeding studies for quantifying genetic diversity and deciphering population structure in major [43-46] as well as minor or orphan crops [40, 41, 47]. The number of SNP markers detected in Kersting's groundnut (1938 SNPs) is almost four-fold than that found by Akohoue et al. [42] in the same crop (493 SNPs). This can be explained by the fact that the 281 accessions used by Akohoue et al. [42] were all from Benin and Togo and may not have included important sources of genetic variation from the other Kersting's groundnut growing countries. The number of detected marker is however lower than in other orphan legumes such as Bambara groundnut (3148 SNPs) [41], pigeon pea (3 713 723 SNPs) [48], chickpea (3339 SNPs) [49] and cowpea (2276 SNPs) [32, 50]. Kersting's groundnut is therefore not as polymorphic as other self-pollinated legume crops including faba bean [51] and cowpea [32, 50].

Our results showed a relatively low alignment (14-20%) of markers with reference genomes of related crops such as common bean, mung bean, Bambara groundnut, and adzuki bean. This is in line with the findings of Akohoue et al. [42], though we obtained better alignment here. It is therefore necessary to assemble a reference genome for Kersting's groundnut to enhance the accuracy of SNP calling and thereby increase their use in the pre-breeding and future breeding pipelines of the crop.

The markers used in the study were highly reproducible (0.99) and scored 100% for call rate, denoting their high quality. However, they exhibited low polymorphism information contents (average PIC=0.059) suggesting that most of them would not be very informative for linkage analysis [52]. These low PIC values also deviate from the average reported for other underutilised legume crops including Bambara groundnut (PIC=0.58) [41], mung bean (PIC=0.25) [53] and pigeon pea (PIC=0.25) [54]. Nevertheless, our results showed that SNPs with lower average PIC have greater density, which would increase their effectiveness in quantifying the genetic diversity and discriminating the Kersting's groundnut accessions. Furthermore, the high sequencing depth observed in this study suggested that null alleles might be responsible for most missing data. In addition, low mean heterozygosity ( $H_o$ =0.008) was found across markers. Low mean heterozygosity value (0.04) was also recorded in Kersting's groundnut by Akohoue et al. [42], indicating a low genetic variation within Kersting's groundnut accessions.

### 3.2. Genetic diversity across populations and morphotypes

The overall genetic variability measured as expected heterozygosity  $H_e$  (Nei's gene diversity) and unbiased expected heterozygosity  $uH_e$  across geographical populations ( $H_e$ =0.058;  $uH_e$ =0.059) and morphotypes ( $H_e$ =0.057;  $uH_e$ =0.058) of Kersting's groundnut is lower compared to the value of 0.22 reported by Akohoue et al. [42]. The seemingly higher average heterozygosity found by these authors could, however, be misleading regarding the real level of genetic diversity in the crop as it was estimated from 113 markers with minor allele frequency  $MAF \geq 0.05$ . As demonstrated by O'Leary et al. [55], the common practice to assume that loci with a  $MAF < 0.05$  at population level are not useful and to exclude them from datasets may, unfortunately, lead to excluding true rare alleles from the datasets, especially in orphan crops with little or no genomic resources. However, consistently with Akohoue et al. [42], we found genetic variation in Kersting's groundnut, which is incongruent with Pasquet et al. [26] who reported a total absence of genetic variation in the crop. The total absence of diversity reported by Pasquet et al. [26] is due to the marker system (allozyme) used and the relatively low number (20) of accessions analysed. This further is an indication that genotyping-by-sequencing (GBS) has power to depict diversity in species with very narrow genetic base.

In any case, the present study showed that genetic variability within Kersting's groundnut is undeniably limited, and this may be due to the selfing nature of the crop which resulted in long generations of inbreeding. Indeed, from our results, selfing rates were estimated as 79% across

populations, with values as high as 85% in some populations (Table 2). Similarly, very high selfing rates were observed across morphotypes with values reaching up to 82% within brown seed morphotypes. Moreover, there was a high predominance of homozygous individuals indicating that Kersting's groundnut is primarily a selfing crop just like horsegram (*Macrotyloma uniflorum*) its closely related species [56, 57]. In fact, Kersting's groundnut bears 'pre-anthesis cleistogamous' flowers in which pollination and fertilisation occur in closed flower buds [58], leading to very high levels of homozygosity as reflected here by the low observed heterozygosity among populations and morphotypes (Tables 2 and 3). Therefore, induced mutagenesis seems to be the most promising way to broaden the genetic base of Kersting's groundnut.

The higher genetic diversity found in Atacora and in Cream seeds with black eye can be favourably used in Kersting's groundnut genetic improvement. However, the greatest numbers of private alleles found respectively in Zou and in Cream seed morphotypes suggest that there are potential sources of novel alleles and traits of qualitative or agronomic importance. The first three accessions with greater number of private alleles were Gbo5, Lili\*Sie and Itk and possessed respectively 18, 16 and 12 private alleles (Table S3). These accessions can thus be good candidate parents for the development of improved Kersting's groundnut cultivars with desirable traits for farmers and consumers.

### 3.3. Genetic differentiation and partitioning of genetic variation across populations and morphotypes

The ranges of  $F_{ST}$  values among populations (Table 4) and among morphotypes (Table 5) of Kersting's groundnut revealed substantial genetic differentiation. As expected, little differentiation is observed between the three populations from Benin. This suggests a high level of gene flow in the form of seed exchange among farmers, especially between Zou and Collines, two neighbouring districts. In general, considerable genetic differentiation was observed between populations from different countries, indicating some degree of genetic isolation among countries. Indeed, a significant correlation was detected between genetic and geographic distances between individuals and populations suggesting that seed exchange between very distant localities may not be common. These are in agreement with many previous studies that reported cultural, social, and natural barriers may strengthen reproductive isolation, thereby limiting or promoting gene flow among cultivars [59, 60].

The results of AMOVA indicated that majority of genetic variation (~49%) was intra-population, while 38.5% of genetic variation was inter-population, and only ~13% of variation was apportioned within accessions. Partitioning of genetic variation in plant populations can be affected by many factors including the breeding system (e.g. selfing or outcrossing rates) of the species [61-63]. Evidence showed that inbreeding or selfing annual plants predominantly maintain more genetic variation among populations than within populations contrary to long-lived perennial and/or outcrossing species [64, 65]. Our finding is not consistent with this theory. Higher within population variation is also reported in other annual crops such as wheat [66, 67], safflower [68] and rice [69]. Besides, greater genetic variation (49%) was also partitioned within morphotypes while the variation among morphotypes was moderate (about 39%) but significant. Therefore, intensive collection of accessions of various morphotypes from different populations/countries would substantially increase the genetic diversity within Benin Kersting's groundnut germplasm.

### 3.4. Population structure

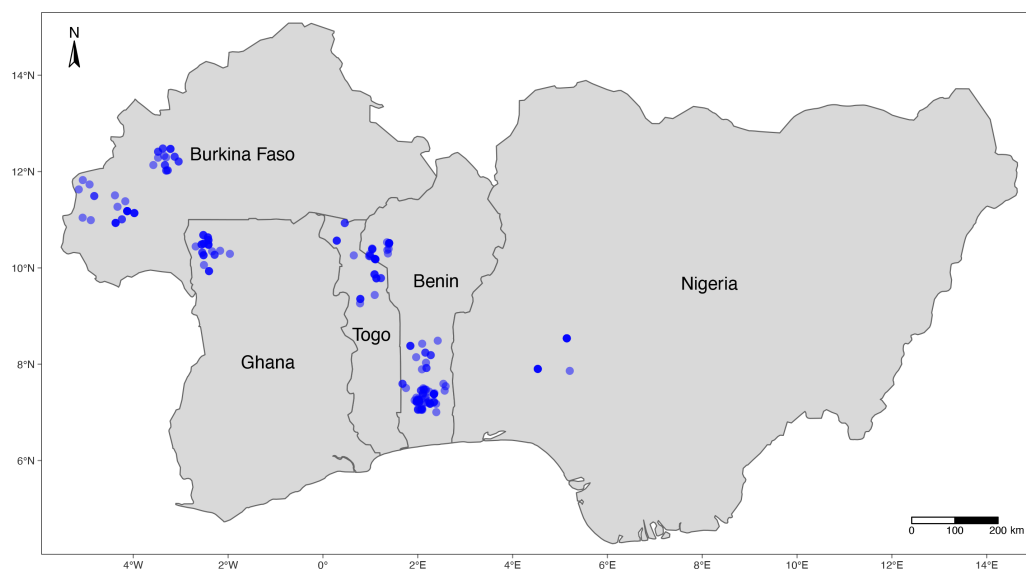
DAPC divided the accessions into eight well-defined clusters which were further confirmed by neighbour joining (NJ) tree. The high number of detected clusters was expected as the accessions included in the study were all landraces from a wide geographical origin. The differences between DAPC and NJ in the assignment of accessions to clusters could be explained by the fact that the NJ method is an agglomerative hierarchical method of clustering, whereas DAPC is based on k-means which is a non-hierarchical clustering method [70]. It is well established that population structure may lead to spurious associations in Genome-wide association studies (GWAS) [71, 72] and should be considered in the development of training populations for Genomic selection (GS) [73]. Therefore,

the clustering of accessions presented in this work provides useful data to inform decision-making for genetic improvement of Kersting's groundnut using molecular breeding tools and techniques.

## 4. Materials and Methods

### 4.1. Plant materials

A total of 227 Kersting's groundnut accessions from five West African countries were used in this study (Table S4). Among them, 112 were from Benin, 48 from Burkina Faso, 36 from Ghana, 12 from Nigeria and 19 from Togo (Figure 6). The accessions were all landraces and composed of 47% (102) cream seed, 22.5% (49) black seed, 13.8% (30) cream seed with black eye, 10.6% (23) brown mottled seed, and 6% (13) brown seed (Table S4). Hereafter, each seed colour was considered as a morphotype.



**Figure 6.** Map showing collection sites of Kersting's groundnut accessions used in this study

### 4.2. Methods

#### 4.2.1. DNA extraction

Each of the 227 accessions was planted in the field in single row plots of 5 m length. Young leaves were collected from 3-week-old plants into 96 deep well sample collection kits and sent to Biosciences Eastern and Central Africa (Beca-ILRI) Hub in Nairobi for Genotyping. DNA was isolated and purified using the NucleoMag® 96 Plant genomic DNA extraction kit (Macherey-Nagel, Düren, Germany), following the manufacturer's instructions. Concentration of the extracted DNA was normalised within the range of 50-100 ng/μl. The quality and quantity of the DNA samples were checked on 0.8% agarose.

#### 4.2.2. Library construction and Genotyping by sequencing

Libraries were constructed following the protocol described in [36]. Two DArTseq™ complexity reduction methods were tested using the rare cutting restriction endonuclease enzyme *Pst*I (5'-CTGCA|G-3') in combination with two different frequently cutting restriction enzymes *Hpa*II (5'-C|CGG-3') and *Mse*I (5'-T|TAA-3'). The *Pst*I/*Mse*I combination was selected as the best performing method. For each sample, 2 μL of DNA was digested with the *Pst*I/*Mse*I restriction enzyme combination. Digestion products were ligated to barcoded adapters pair annealed to the two restriction enzyme overhangs. The *Pst*I-compatible adapter included the partial attachment sequence for the 'Read 1 End' of the Illumina flow cell, a barcode of variable length (4-8 bp) and the *Pst*I-

compatible overhang sequence. The reverse adapter included the partial sequence for the 'Read 2 End' of the Illumina flow cell and *MseI*-compatible overhang sequence. The adapter-ligated fragments were amplified in a Polymerase Chain Reaction (PCR) using optimized settings. After PCR, equimolar amounts of the amplified products from each sample were pooled together using a TECAN robotic liquid handler (TECAN, Männedorf, Switzerland) and loaded on the cBot (Illumina, Inc., San Diego, CA, USA) for clustering on a flow cell. Libraries were sequenced in the Illumina HiSeq 2500 using Single Read sequencing runs for 77 bases.

The IGSS platform uses a high-throughput genotyping-by-sequencing (GBS) DArTseq technology (DArT Pty Ltd, Canberra, Australia). DArTseq markers scoring was achieved using DArTsoft14 a proprietary marker scoring pipeline based on algorithms. Two types of DArTseq markers (SilicoDArT and SNP) were scored as '1' for presence, and '0' for absence and '-' for failure to score. As *Kersting's groundnut* had no available reference genome, both SilicoDArT and SNP markers were aligned to the reference genomes of *Phaseolus vulgaris* ([ftp://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/phaseolus\\_vulgaris/dna/](ftp://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/phaseolus_vulgaris/dna/)), *Vigna angularis* ([ftp://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/vigna\\_angularis/dna/](ftp://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/vigna_angularis/dna/)), *Vigna radiata* ([ftp://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/vigna\\_radiata/dna/](ftp://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/vigna_radiata/dna/)) and *Vigna subterranean* (<http://gigadb.org/dataset/view/id/101055>) which are related species [74].

#### 4.2.3. Quality filtering, Imputation and Markers diversity

SNP markers were imputed using the Diversity Arrays Technology's KDCompute Optimal Imputation plugin (<https://kdcompute.igss-africa.org/kdcompute/login>). Seven imputation methods (MissForest, Random, Nipals, singular value decomposition (SVD), Expectation-Maximization (EM), k-nearest neighbour (kNN), and probabilistic principal component analysis (PPCA)) were tested based on simple matching coefficient (SMC), calculated by excluding an additional 10% of missing values and evaluating the correlation between introduced missing values and the original dataset. Missing calls were imputed using the most accurate imputation method for which the highest SMC was recorded.

The pre-imputation SNP dataset was filtered for various quality control metrics, and markers that passed thresholds were retained for analysis. Various iterative filtering schemes were applied as recommended by O'Leary et al. [55] and the one that resulted in retaining more loci and individuals without compromising the markers quality was retained (data not shown). Consequently, only loci with <50% missing data, >5 read depth, >75 % call rate, >95% reproducibility and >0 polymorphic information content (PIC) were retained. In addition, accessions with >50% missing SNPs were removed from downstream analyses. This narrowed down to 217 accessions that were retained for further analyses. SNP filtering was carried out in R 4.0.0 [75].

Summary statistics (minimum, maximum and mean) were calculated for the remaining SNP markers for the following quality parameters: call rate, read depth, one ratio, PIC, heterozygosity, minor allele frequency and reproducibility. The number of markers successfully aligned were graphed per reference genome.

#### 4.2.4. Analysis of genetic variation and population structure

The average number of alleles ( $N_a$ ), number of effective alleles ( $N_e$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), unbiased expected heterozygosity ( $uH_e$ ), the inbreeding coefficient ( $F_{IS}$ ) and the percentage of polymorphic loci (%P) were calculated by geographic population and by morphotype, using GenAlEx version 6.51b2 [76]. For this purpose, the 112 accessions from Benin were assigned to three putative populations (Atacora, Collines, Zou) while those from the other countries were considered belonging to a single population each. Average allelic richness ( $A_r$ ), and the number of private alleles ( $N_p$ ) were calculated in the 'PopGenReport' [77] R package. Moreover, selfing rate ( $s$ ) was determined using the equation described in [78]:  $s = 2F_{IS} / (1 + F_{IS})$ . Nei's pairwise  $F_{ST}$  was computed between all pairs of populations and morphotypes in the 'hierfstat' R package [79]. The  $F_{ST}$  statistics measures the overall level of genetic divergence among subpopulations. According to Wright [80], a  $F_{ST} < 0.05$  indicates little genetic differentiation, values



comprised between 0.05 and 0.15 indicate moderate genetic differentiation, while  $F_{ST}$  values between 0.15 and 0.25 denotes great differentiation.  $F_{ST} > 0.25$  indicates very great genetic differentiation [80]. Furthermore, analysis of molecular variance (AMOVA) was performed to determine the level of partitioning of diversity within and among populations and morphotypes. AMOVA were performed in the 'poppr' R package [81] using 999 permutations, while significance levels were estimated for variance components based on 9,999 permutations using the *randtest* function in the 'ade4' R package [82].

Population structure was inferred using a Discriminant Analysis of Principal Components (DAPC) performed in the 'adegenet' R package [83]. The most probable number of clusters in the data was inferred using the *find.clusters* function from 'adegenet', running successive K-means and with the Bayesian Information Criterion (BIC) used as the statistical measure of goodness of fit. The number of principal components (PCs) to retain in the DAPC analysis was determined using a cross-validation procedure implemented with the *xvalDapc* function available in the 'adegenet' package, using the default parameters. Furthermore, an unrooted neighbour-joining tree was constructed in the package 'poppr' based on Edward's genetic distance. Node strength in the dendrograms was tested using 1,000 bootstrap analyses in 'poppr' package.

Finally, isolation by distance (IBD) was tested using a Mantel test correlating genetic distance [ $F_{ST}/(1-F_{ST})$ ] and geographic distances among populations, and morphotypes. The correlation between Nei's genetic distance [84] and geographic distance was also examined among accessions using Mantel test. The Mantel test was performed with 999,999 permutations using the function *mantel.randtest* available in 'ade4'. The genetic versus geographic distances were plotted along with a point density of accessions to visualise whether IBD existed as a continuous gradient or resulted from patches of distant and divergent individuals.

## 5. Conclusions

The present study was undertaken to investigate the level of genetic variation and determine population structure among a collection of 217 Kersting's groundnut accessions from different origins using 886 high-quality DArTseq derived SNP markers. Low levels of genetic diversity were found overall. The highest genetic diversity was found in Atacora population while the highest number of private alleles was observed in Zou population. Similarly, the more diverse morphotype was cream seed with black eye whereas the highest number of private alleles was found in cream seed. Furthermore, very high genetic differentiations were observed among populations and morphotypes, with maximal differentiation between populations from different countries. The available level of genetic diversity could be efficiently exploited to develop improved Kersting's groundnut cultivars with desirable characteristics such as high yield and quality traits. However, the use of induced mutagenesis is recommended for fast broadening of the genetic base of the crop. Substantial genetic differentiation was observed among populations and among morphotypes. Likewise, considerable genetic variations were contained within populations as well as within morphotypes. The population structure analysis identified eight clusters containing accessions from different geographic origins. The genomic resource developed in this study can significantly contribute to integrating modern breeding techniques, including Genome-wide association studies and Genomic selection into the Kersting's groundnut breeding pipeline in Benin and West Africa.

**Supplementary Materials:** Supplementary materials can be found at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1).

**Author Contributions:** Conceptualization, Konoutan M. Kafoutchoni and Eric E. Agoyi; Data curation, Konoutan M. Kafoutchoni; Formal analysis, Konoutan M. Kafoutchoni; Funding acquisition, Konoutan M. Kafoutchoni, Symphorien Agbahoungba and Achille E. Assogbadjo; Project administration, Eric E. Agoyi; Resources, Eric E. Agoyi, Achille E. Assogbadjo and Clément Agbangla; Supervision, Achille E. Assogbadjo and Clément Agbangla; Writing – original draft, Konoutan M. Kafoutchoni; Writing – review & editing, Konoutan M. Kafoutchoni, Eric E. Agoyi and Symphorien Agbahoungba.

**Funding:** This research was funded by the World Academy of Sciences (TWAS), grant number 18-238 RG/BIO/AF/AC\_G-FR3240303667.



**Acknowledgments:** We thank Carnegie Cooperation of New York for funding part of the work through the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) grant number RU/2018/TQA/38. We are grateful to the different institutions that provided the accessions used in this study: IITA-Nigeria, INERA-Burkina-Faso, SARI-Ghana. Finally, we thank Dr. Mustapha Mohammed (Tshwane University of Technology, Pretoria, South Africa) for providing additional accessions from Ghana.

**Conflicts of Interest:** The authors declare no conflict of interest

## Abbreviations

%P	Percentage of Polymorphic Loci
AMOVA	Analysis of Molecular Variance
Ar	Allelic richness
BIC	Bayesian Information Criterion
BM	Brown Mottled
BS	Bootstrap support
CBE	Cream with black eye
DAPC	Discriminant Analysis of Principal Components
DArTseq	Diversity array technology sequencing
Df	Degree of freedom
DNA	Deoxyribonucleic acid
EM	Expectation-Maximization
Est.Var.	Estimated variance
F <sub>IS</sub>	Inbreeding coefficient
F <sub>IT</sub>	Inbreeding coefficient of an individual relative to the total population
F <sub>ST</sub>	Fixation index
GBS	Genotyping-by-sequencing
He	Expected Heterozygosity
Ho	Observed Heterozygosity
IBD	Isolation by distance
kNN	k-nearest neighbour
MAF	Minor Allele Frequency
MS	Mean square
Na	Number of different alleles
Ne	Number of effective alleles
NJ	Neighbour joining
Np	Number of private alleles
PC	Principal component
PCA	Principal component analysis
PIC	Polymorphic information content
PPCA	Probabilistic Principal Components Analysis
s	Selfing rate
SMC	Simple matching coefficient
SNP	Single Nucleotide Polymorphism
SS	Sum of squares
SVD	Singular value decomposition
uHe	Unbiased Expected Heterozygosity

## References

1. Achigan Dako, E. G.; Vodouhè, S. R., *Macrotyloma geocarpum* (Harms) Maréchal and Baudet. *Brink M, Belay G (Editeurs). PROTA* **2006**, 1.
2. Hepper, F., Plants of the 1957-58 West African Expedition: II. The Bambara Groundnut (*Voandzeia subterranea*) and Kersting's Groundnut (*Kerstingiella geocarpa*) Wild in West Africa. *Kew Bulletin* **1963**, 16, (3), 395-407.

3. Chikwendu, J. N., Comparative evaluation of chemical composition of fermented ground bean flour (*Kerstingella geocarpa*), cowpea flour (*Vigna unguiculata*) and commercial wheat flour (*Triticum* spp.). *Pak J Nutr* **2015**, 14, 218-224.
4. Jaiswal, S. K.; Mohammed, M.; Dakora, F. D., Microbial community structure in the rhizosphere of the orphan legume Kersting's groundnut [*Macrotyloma geocarpum* (Harms) Marechal & Baudet]. *Molecular biology reports* **2019**, 46, (4), 4471-4481.
5. Varshney, R. K.; Close, T. J.; Singh, N. K.; Hoisington, D. A.; Cook, D. R., Orphan legume crops enter the genomics era! *Current Opinion in Plant Biology* **2009**, 12, (2), 202-210.
6. Dansi, A.; Vodouhe, R.; Azokpota, P.; Yedomonhan, H.; Assogba, P.; Adjatin, A.; Loko, Y. L.; Dossou-Aminon, I.; Akpagana, K., Diversity of the neglected and underutilized crop species of importance in Benin. *ScientificWorldJournal* **2012**, 2012, 932947.
7. Agoyi, E. E.; N'danikou, S.; Kafoutchoni, M.; Ayena, M.; Sodedji, F. A.; Agbahoungba, S.; Sossou, H. S.; Vodouhe, R.; Assogbadjo, A. E., Kersting's Groundnut [*Macrotyloma geocarpum* (Harms) Maréchal & Baudet] Crop Attracts More Field Pests and Diseases than Reported Before. *Agricultural Research & Technology Open Access Journal* **2019**, 21, (5).
8. Amujoyegbe, B.; Obisesan, I.; Ajayi, A.; Aderanti, F., Disappearance of Kersting's groundnut (*Macrotyloma geocarpum* (Harms) Marechal and Baudet) in south-western Nigeria: an indicator of genetic erosion. *Plant Genetic Resources Newsletter (Bioversity International/FAO)* **2007**.
9. Adu-Gyamfi, R.; Fearon, J.; Bayorbor, T.; Dzomeku, I.; Avoronyo, V., The Status of Kersting's Groundnut (*Macrotyloma Geocarpum* [Harms] Marechal and Baudet) An Underexploited Legume in Northern Ghana. *Outlook on AGRICULTURE* **2011**, 40, (3), 259-262.
10. Tamini, Z. Etude ethnobotanique et analyses morphophysiologiques du développement de la lentille de terre (*Macrotyloma geocarpum* (Harms) Marechal et Baudet. Université de Ouagadougou, Burkina-Faso, 1997.
11. Bampuori, A. H., Effect of Traditional Farming Practices on the Yield of Indigenous Kersting's Groundnut (*Macrotyloma geocarpum* Harms) Crop in the Upper West Region of Ghana. *Journal of Developments in Sustainable Agriculture* **2007**, 2, (2), 128-144.
12. Mergeai, G., Influence des facteurs sociologiques sur la conservation des ressources phytogénétiques. Le cas de la lentille de terre (*Macrotyloma geocarpum* (Harms) Marechal & Baudet) au Togo. *Bulletin des Recherches Agronomiques de Gembloux* **1993**, 28, (4), 487-500.
13. Assogba, P.; Ewedje, E. B.; Dansi, A.; Loko, Y.; Adjatin, A.; Dansi, M.; Sanni, A., Indigenous knowledge and agro-morphological evaluation of the minor crop Kersting's groundnut (*Macrotyloma geocarpum* (Harms) Maréchal et Baudet) cultivars of Benin. *Genetic resources and crop evolution* **2015**, 63, (3), 513-529.
14. Akohoue, F.; Sibiya, J.; Achigan-Dako, E. G., On-farm practices, mapping, and uses of genetic resources of Kersting's groundnut [*Macrotyloma geocarpum* (Harms) Maréchal et Baudet] across ecological zones in Benin and Togo. *Genetic Resources and Crop Evolution* **2018**, 66, (1), 195-214.
15. Kafoutchoni, K. M.; Agoyi, E. E.; Sossou, H. S.; Agbahoungba, S.; Agbangla, C.; Assogbadjo, A. E., Impacts of sociodemographic factors on the production system of the orphan legume Kersting's groundnut (*Macrotyloma geocarpum*) and prioritisation of production constraints in Benin. *African Journal of Rural Development* **2019**, (In press).
16. Govindaraj, M.; Vetriventhan, M.; Srinivasan, M., Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics research international* **2015**, 2015.

17. Ajala, S. O.; Olayiwola, M. O.; Ilesanmi, O. J.; Gedil, M.; Job, A. O.; Olaniyan, A. B., Assessment of genetic diversity among low-nitrogen-tolerant early generation maize inbred lines using SNP markers. *South African Journal of Plant and Soil* **2019**, 36, (3), 181-188.
18. Chen, W.; Hou, L.; Zhang, Z.; Pang, X.; Li, Y., Genetic diversity, population structure, and linkage disequilibrium of a core collection of *Ziziphus jujuba* assessed with genome-wide SNPs developed by genotyping-by-sequencing and SSR markers. *Frontiers in plant science* **2017**, 8, 575.
19. Kumar, S.; Ambreen, H.; Variath, M. T.; Rao, A. R.; Agarwal, M.; Kumar, A.; Goel, S.; Jagannath, A., Utilization of molecular, phenotypic, and geographical diversity to develop compact composite core collection in the oilseed crop, safflower (*Carthamus tinctorius* L.) through maximization strategy. *Frontiers in plant science* **2016**, 7, 1554.
20. Akohoue, F.; Achigan-Dako, E. G.; Coulibaly, M.; Sibiya, J., Correlations, path coefficient analysis and phenotypic diversity of a West African germplasm of Kersting's groundnut [*Macrotyloma geocarpum* (Harms) Maréchal & Baudet]. *Genetic Resources and Crop Evolution* **2019**, 66, (8), 1825-1842.
21. Chodaton, G. Y.; Agoyi, E. E.; Houndété, T. A.; Kafoutchoni, K. M.; Sossou, H. S.; Sodédji, F. A. K.; Ayi, S.; Agbahoungba, S.; Chadare, F. J.; Adandonon, A.; Vodouhè, R.; Assogbadjo, A. E.; Sinsin, B. A., Morphological variation and discriminating traits of Kersting's groundnut accessions. *Journal of Crop Improvement* **2020**, (In press).
22. Mensah, A. G.; Gyan, A. D.; Yeboah, A. M., An assessment on variation among genotypes of kersting's groundnut in Northern Ghana. *Advances in Agriculture and Agricultural Sciences* **2016**, 2, (7), 149-155.
23. Adu-Gyamfi, R.; Dzomeku, I. K.; Lardi, J., Evaluation of growth and yield potential of genotypes of Kersting's groundnut (*macrotyloma geocarpum* harms) in Northern Ghana. *International Research Journal of Agricultural Science and Soil Science* **2012**, 2, (12), 509-515.
24. Bayorbor, T. B.; Dzomeku, I. K.; Avoronyo, V. K.; Opoku-Agyeman, M. O., Morphological variation in Kersting's groundnut (*Kerstigiella geocarpa* Harms) landraces from northern Ghana. *Agriculture and Biology Journal of North America* **2010**, 1, (3), 290-295.
25. Singh, S.; Gupta, S.; Thudi, M.; Das, R. R.; Vemula, A.; Garg, V.; Varshney, R.; Rathore, A.; Pahuja, S.; Yadav, D. V., Genetic diversity patterns and heterosis prediction based on SSRs and SNPs in hybrid parents of pearl millet. *Crop Science* **2018**, 58, (6), 2379-2390.
26. Pasquet, R. S.; Mergeai, G.; Baudoin, J.-P., Genetic diversity of the African geocarpic legume Kersting's groundnut, *Macrotyloma geocarpum* (Tribe Phaseoleae: Fabaceae). *Biochemical systematics and ecology* **2002**, 30, (10), 943-952.
27. Mondini, L.; Noorani, A.; Pagnotta, M. A., Assessing plant genetic diversity by molecular tools. *Diversity* **2009**, 1, (1), 19-35.
28. Agoyi, E. E.; Tumuhairwe, J. B.; Chigeza, G.; Tukamuhabwa, P.; Diers, B. W., First Attempt To Identify And Map QTLs Associated With Promiscuous Nodulation In Soybean. *bioRxiv* **2019**, 688028.
29. Iquiria, E.; Humira, S.; François, B., Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC plant biology* **2015**, 15, (1), 5.
30. Jarquín, D.; Kocak, K.; Posadas, L.; Hyma, K.; Jedlicka, J.; Graef, G.; Lorenz, A., Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC genomics* **2014**, 15, (1), 740.
31. Xiong, H.; Shi, A.; Mou, B.; Qin, J.; Motes, D.; Lu, W.; Ma, J.; Weng, Y.; Yang, W.; Wu, D., Genetic diversity and population structure of cowpea (*Vigna unguiculata* L. Walp). *PLoS One* **2016**, 11, (8), e0160941.

32. Fatokun, C.; Girma, G.; Abberton, M.; Gedil, M.; Unachukwu, N.; Oyatomi, O.; Yusuf, M.; Rabbi, I.; Boukar, O., Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. *Scientific reports* **2018**, 8, (1), 1-10.
33. Schröder, S.; Mamidi, S.; Lee, R.; McKain, M. R.; McClean, P. E.; Osorno, J. M., Optimization of genotyping by sequencing (GBS) data in common bean (*Phaseolus vulgaris* L.). *Molecular Breeding* **2016**, 36, (1), 6.
34. Campa, A.; Murube, E.; Ferreira, J. J., Genetic diversity, population structure, and linkage disequilibrium in a spanish common bean diversity panel revealed through genotyping-by-sequencing. *Genes* **2018**, 9, (11), 518.
35. Cortés, A. J.; Blair, M. W., Genotyping by Sequencing and Genome–Environment Associations in Wild Common Bean Predict Widespread Divergent Adaptation to Drought. *Frontiers in Plant Science* **2018**, 9.
36. Kilian, A.; Wenzl, P.; Huttner, E.; Carling, J.; Xia, L.; Blois, H.; Caig, V.; Heller-Uszynska, K.; Jaccoud, D.; Hopper, C., Diversity arrays technology: a generic genome profiling technology on open platforms. In *Data production and analysis in population genomics*, Springer: 2012; pp 67-89.
37. Yang, X.; Ren, R.; Ray, R.; Xu, J.; Li, P.; Zhang, M.; Liu, G.; Yao, X.; Kilian, A., Genetic diversity and population structure of core watermelon (*Citrullus lanatus*) genotypes using DArTseq-based SNPs. *Plant Genetic Resources* **2016**, 14, (3), 226-233.
38. Egea, L. A.; Mérida-García, R.; Kilian, A.; Hernandez, P.; Dorado, G., Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology “genotyping-by-sequencing” platform (DArTseq). *Frontiers in genetics* **2017**, 8, 98.
39. Ndjiondjop, M.-N.; Semagn, K.; Gouda, A. C.; Kpeki, S. B.; Dro Tia, D.; Sow, M.; Goungoulou, A.; Sie, M.; Perrier, X.; Ghesquiere, A., Genetic variation and population structure of *Oryza glaberrima* and development of a mini-core collection using DArTseq. *Frontiers in plant science* **2017**, 8, 1748.
40. Robbana, C.; Kehel, Z.; Naceur, B.; Sansaloni, C.; Bassi, F.; Amri, A., Genome-wide genetic diversity and population structure of tunisian durum wheat landraces based on DArTseq technology. *International journal of molecular sciences* **2019**, 20, (6), 1352.
41. Redjeki, E. S.; Ho, W. K.; Shah, N.; Molosiwa, O. O.; Ardiarini, N.; Kuswanto; Mayes, S., Understanding the genetic relationships between Indonesian bambara groundnut landraces and investigating their origins. *Genome* **2020**, (999), 1-9.
42. Akohoue, F.; Achigan-Dako, E. G.; Sneller, C.; Van Deynze, A.; Sibiya, J., Genetic diversity, SNP-trait associations and genomic selection accuracy in a west African collection of Kersting’s groundnut [*Macrotyloma geocarpum* (Harms) Maréchal & Baudet]. *Plos one* **2020**, 15, (6), e0234769.
43. Shaibu, A. S.; Sneller, C.; Motagi, B. N.; Chepkoech, J.; Chepngetich, M.; Miko, Z. L.; Isa, A. M.; Ajeigbe, H. A.; Mohammed, S. G., Genome-Wide Detection of SNP Markers Associated with Four Physiological Traits in Groundnut (*Arachis hypogaea* L.) Mini Core Collection. *Agronomy* **2020**, 10, (2), 192.
44. Nemli, S.; AŞÇIOĞUL, T. K.; ATEŞ, D.; EŞİYOK, D.; Tanyolac, M. B., Diversity and genetic analysis through DArTseq in common bean (*Phaseolus vulgaris* L.) germplasm from Turkey. *Turkish Journal of Agriculture and Forestry* **2017**, 41, (5), 389-404.
45. Mogga, M.; Sibiya, J.; Shimelis, H.; Lamo, J.; Yao, N., Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (*Oryza sativa* L.) using DArT markers. *PloS one* **2018**, 13, (6).
46. Ndjiondjop, M. N.; Semagn, K.; Sow, M.; Manneh, B.; Gouda, A. C.; Kpeki, S. B.; Pegalepo, E.; Wambugu, P.; Sié, M.; Warburton, M. L., Assessment of genetic variation and population structure of

- diverse rice genotypes adapted to lowland and upland ecologies in Africa using SNPs. *Frontiers in plant science* **2018**, 9, 446.
47. O'Connor, K.; Kilian, A.; Hayes, B.; Hardner, C.; Nock, C.; Baten, A.; Alam, M.; Topp, B., Population structure, genetic diversity and linkage disequilibrium in a macadamia breeding population using SNP and silicoDART markers. *Tree Genetics & Genomes* **2019**, 15, (2), 24.
  48. Zhao, J.; Bayer, P. E.; Ruperao, P.; Saxena, R. K.; Khan, A. W.; Golicz, A. A.; Nguyen, H. T.; Batley, J.; Edwards, D.; Varshney, R. K., Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnology Journal* **2020**.
  49. Farahani, S.; Maleki, M.; Mehrabi, R.; Kanouni, H.; Scheben, A.; Batley, J.; Talebi, R., Whole Genome Diversity, Population Structure, and Linkage Disequilibrium Analysis of Chickpea (*Cicer arietinum* L.) Genotypes Using Genome-Wide DARTseq-Based SNP Markers. *Genes* **2019**, 10, (9), 676.
  50. Sodedji, F. A. K.; Agbahoungba, S.; Agoyi, E. E.; Kafoutchoni, K. M.; Kim, H.-Y.; Nguetta, S.-P. A.; Assogbadjo, A. E., DART-seq based SNP analysis of diversity, population structure and linkage disequilibrium among 274 cowpea (*Vigna unguiculata* L. Walp.) accessions. *BMC Genetics* **2020**, (In press).
  51. Kaur, S.; Kimber, R. B.; Cogan, N. O.; Materne, M.; Forster, J. W.; Paull, J. G., SNP discovery and high-density genetic mapping in faba bean (*Vicia faba* L.) permits identification of QTLs for ascochyta blight resistance. *Plant Science* **2014**, 217, 47-55.
  52. Alam, M.; Neal, J.; O'Connor, K.; Kilian, A.; Topp, B., Ultra-high-throughput DARTseq-based silicoDART and SNP markers for genomic studies in macadamia. *PloS one* **2018**, 13, (8), e0203465.
  53. Islam, A. F.; Blair, M. W., Molecular Characterization of Mung Bean Germplasm from the USDA Core Collection Using Newly Developed KASP-based SNP Markers. *Crop Science* **2018**, 58, (4), 1659-1670.
  54. Zavinon, F.; Adoukonou-Sagbadja, H.; Keilwagen, J.; Lehnert, H.; Ordon, F.; Perovic, D., Genetic diversity and population structure in Beninese pigeon pea [*Cajanus cajan* (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers. *Genetic Resources and Crop Evolution* **2020**, 67, (1), 191-208.
  55. O'Leary, S. J.; Puritz, J. B.; Willis, S. C.; Hollenbeck, C. M.; Portnoy, D. S., These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular ecology* **2018**, 27, (16), 3193-3206.
  56. Priyanka, S.; Sudhagar, R.; Vanniarajan, C.; Ganesamurthy, K.; Souframanien, J., Combined mutagenic ability of gamma ray and EMS in horsegram (*Macrotyloma uniflorum* (Lam) Verdc.). *Electronic Journal of Plant Breeding* **2019**, 10, (3), 1086-1094.
  57. Morris, J., *Macrotyloma axillare* and *M. uniflorum*: descriptor analysis, anthocyanin indexes, and potential uses. *Genetic resources and crop evolution* **2008**, 55, (1), 5-8.
  58. Kafoutchoni, K. M.; Agoyi, E. E.; Dassou, G. H.; Sossou, H. S.; Ayi, S.; Ahanhanzo Glèlè, C.; Adomou, A. C.; Yédomonhan, H.; Agbangla, C.; Assogbadjo, A. E., Reproductive biology, phenology, pollen viability and germinability in Kersting's groundnut (*Macrotyloma geocarpum* (Harms) Maréchal & Baudet, Fabaceae). *Annales of Botany* **2020**, (In press).
  59. Pusadee, T.; Jamjod, S.; Chiang, Y.-C.; Rerkasem, B.; Schaal, B. A., Genetic structure and isolation by distance in a landrace of Thai rice. *Proceedings of the National Academy of Sciences* **2009**, 106, (33), 13880-13885.
  60. Sharma, S.; Upadhyaya, H.; Varshney, R.; Gowda, C., Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Frontiers in Plant Science* **2013**, 4, (309).



61. Duminil, J.; Fineschi, S.; Hampe, A.; Jordano, P.; Salvini, D.; Vendramin, G. G.; Petit, R. J., Can population genetic structure be predicted from life-history traits? *The American Naturalist* **2007**, 169, (5), 662-672.
62. Hamrick, J. L.; Godt, M. W., Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **1996**, 351, (1345), 1291-1298.
63. Szczecińska, M.; Sramko, G.; Wołosz, K.; Sawicki, J., Genetic diversity and population structure of the rare and endangered plant species *Pulsatilla patens* (L.) Mill in East Central Europe. *PLoS One* **2016**, 11, (3), e0151730.
64. Yang, Y.; Pan, Y.; Gong, X.; Fan, M., Genetic variation in the endangered Rutaceae species *Citrus hongheensis* based on ISSR fingerprinting. *Genetic resources and crop evolution* **2010**, 57, (8), 1239-1248.
65. Hamrick, J.; Godt, M.; Murawski, D.; Loveless, M., Correlations between species traits and allozyme diversity: implications for conservation biology. *Genetics and conservation of rare plants* **1991**, 3, 30.
66. Eltaher, S.; Sallam, A.; Belamkar, V.; Emara, H. A.; Nower, A. A.; Salem, K. F.; Poland, J.; Baenziger, P. S., Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. *Frontiers in genetics* **2018**, 9, 76.
67. Kabbaj, H.; Sall, A. T.; Al-Abdallat, A.; Geleta, M.; Amri, A.; Filali-Maltouf, A.; Belkadi, B.; Ortiz, R.; Bassi, F. M., Genetic diversity within a global panel of durum wheat (*Triticum durum*) landraces and modern germplasm reveals the history of alleles exchange. *Frontiers in plant science* **2017**, 8, 1277.
68. Hassani, S. M. R.; Talebi, R.; Pourdard, S. S.; Naji, A. M.; Fayaz, F., In-depth genome diversity, population structure and linkage disequilibrium analysis of worldwide diverse safflower (*Carthamus tinctorius* L.) accessions using NGS data generated by DArTseq technology. *Molecular Biology Reports* **2020**, 47, (3), 2123-2135.
69. Yelome, O. I.; Audenaert, K.; Landschoot, S.; Dansi, A.; Vanhove, W.; Silue, D.; Van Damme, P.; Haesaert, G., Analysis of population structure and genetic diversity reveals gene flow and geographic patterns in cultivated rice (*O. sativa* and *O. glaberrima*) in West Africa. *Euphytica* **2018**, 214, (11), 215.
70. Campoy, J. A.; Lerigoleur-Balsemin, E.; Christmann, H.; Beauvieux, R.; Girollet, N.; Quero-García, J.; Dirlewanger, E.; Barreneche, T., Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC plant biology* **2016**, 16, (1), 1-15.
71. Korte, A.; Farlow, A., The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **2013**, 9, (1), 29.
72. Barabaschi, D.; Tondelli, A.; Desiderio, F.; Volante, A.; Vaccino, P.; Vale, G.; Cattivelli, L., Next generation breeding. *Plant Sci* **2016**, 242, 3-13.
73. Isidro, J.; Jannink, J.-L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M. E., Training set optimization under population structure in genomic selection. *Theoretical and applied genetics* **2015**, 128, (1), 145-158.
74. Wojciechowski, M. F.; Lavin, M.; Sanderson, M. J., A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. *American journal of botany* **2004**, 91, (11), 1846-1862.
75. R Core Team *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2020.
76. Peakall, R.; Smouse, P. E., GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **2012**, 28, (19), 2537.



77. Adamack, A. T.; Gruber, B., PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* **2014**, 5, (4), 384-387.
78. Ritland, K., Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution* **1990**, 44, (5), 1230-1241.
79. Goudet, J.; Jombart, T., hierfstat: estimation and tests of hierarchical F-statistics. *R package version 0.04-22* **2015**, 10.
80. Wright, S., *Evolution and the genetics of populations, volume 4: variability within and among natural populations*. University of Chicago press: 1978; Vol. 4.
81. Kamvar, Z. N.; Tabima, J. F.; Grünwald, N. J., Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, 2, e281.
82. Dray, S.; Dufour, A.-B., The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* **2007**, 22, (4), 1-20.
83. Jombart, T., adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, 24, (11), 1403-1405.
84. Nei, M., Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **1973**, 70, (12), 3321-3323.