agronomy

MDPI

Article

Genetic diversity of cowpea (*Vigna unguiculata* L. Walp) landraces suggests Central Mozambique as an important hotspot of variation

Ana Maria Figueira Gomes [1,2,3], David Draper [4,5], Pedro Talhinhas [1], Paula Batista Santos [1], Fernanda Simões [6], Nascimento Nhantumbo [2], Rafael Massinga [2], José C. Ramalho [1,7,*], Isabel Marques [1,*], Ana I. Ribeiro-Barros [1,7,*]

[1] Plant Stress and Biodiversity Lab, Forest Research Center (AMFG, JCR, IM, AIRB) and Linking Landscape Environment, Agriculture and Food (PBS, PT), Instituto Superior de Agronomia (ISA), Universidade de Lisboa, Portugal. E-mail: anagomes13@gmail.com (AMFG), ptalhinhas@isa.ulisboa.pt (PT), pbbsantos@isa.ulisboa.pt (PBS), cochichor@isa.ulisboa.pt or cochichor@mail.telepac.pt (JCR), isabelmarques@isa.ulisboa.pt (IM), aribeiro@isa.ulisboa.pt (AIRB)

[2] Divisão de Agricultura, Instituto Superior Politécnico de Manica (DivAG-ISPM), Mozambique. E-mail: tonhantumbo@gmail.com (NN), rafael.massinga@gmail.com (RM)

[3] TropiKMan Doctorate Programme, Nova School of Business and Economics, Universidade NOVA de Lisboa, Portugal.

[4] National Museum of Natural History and Science and Centre for Ecology, Evolution and Environmental Change. Universidade de Lisboa, Portugal. E-mail: ddmunt@gmail.com

[5] UBC Botanical Garden & Centre for Plant Research, and Department of Botany, University of British Columbia, Canada.

[6] Unidade .de Investigação em Biotecnologia e Recursos Genéticos (UIBRG), Instituto Nacional de Investigação Agrária e Veterinária, I.P. (INIAV), Oeiras, Portugal. E-mail: fernanda.simoes@iniav.pt

[7] GeoBioSciences, GeoTechnologies and GeoEngineering (GeoBioTec), Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Monte de Caparica, Portugal.

* Correspondence: cochichor@isa.ulisboa.pt or cochichor@mail.telepac.pt (JCR), isabelmarques@isa.ulisboa.pt (IM), aribeiro@isa.ulisboa.pt (AIRB)

Received: date; Accepted: date; Published: date

Abstract: Cowpea is a multiple purpose drought-tolerant legume crop grown in several dry tropical areas. Its domestication center is thought to be East or West Africa where a high level of genetic diversity is apparently still found in many landraces. However, detailed genetic information is lacking in many African countries limiting the success of breeding programs. In this work, we have assessed the genetic variation and gene flow in 59 *Vigna unguiculata* (cowpea) landraces spanned across six agro-ecological zones from Mozambique, based on nuclear microsatellite markers. The results revealed the existence of high genetic diversity between the landraces, even in comparison to other world regions. Four genetic groups were found, with no specific geographic pattern, suggesting the presence of gene flow between landraces. In comparison, the two commercial varieties had lower values of genetic diversity, although still close from the ones found in local landraces. The high genetic diversity found in Mozambique sustains the importance of local landraces and on farm protection in order to enhance genetic diversity in modern varieties of cowpea worldwide.

Keywords: Africa; cowpea; genetic diversity; landraces; microsatellites

1. Introduction

Cowpea (*Vigna unguiculata* L. Walp), also known as black eye pea, is a major annual grain legume mostly grown in dry tropical areas of Latin America, South Asia and Africa [1]. It is cultivated mainly for its grains, which have a high content of proteins (20-32%) and carbohydrates (50-60%). Both grains and leaves, are also rich in the amino acids lysine and tryptophan, vitamin C, iron and zinc [2]. Cowpea has therefore an essential role in the human diet in many developing countries being referred as the "poor man's meat" [3]. As a legume, it is also an important component of traditional cropping systems since it fixes atmospheric nitrogen and contributes to soil fertility improvement particularly in smallholder farming systems where little or no fertilizer is used [4]. The bulk of cowpea production and consumption is sub-Saharan Africa, namely West and Central Africa [1], where its nutritional value and tolerance to drought place this crop in an unique position to the continent's efforts to establish nutrition sensitive food systems that are more likely to help curb malnutrition, particularly among the most vulnerable – pregnant or lactant women and children under five [5]. Although cowpea is known to be drought tolerant when compared to other crops, the productivity of cowpea varieties is hampered by erratic rainfall and many are sensible to heat [1]. Thus, appropriate agronomic practices could improve the performance of new varieties, under different agro-ecological zones. Indeed, physiological, and metabolic studies show a progressive acclimation of cowpea plants to stress [6] and differential drought responses of landraces with contrasting tolerance levels [7].

Despite being native to Africa [8], the domestication center of cowpea is unclear but thought to be either in East or West Africa where a high morphological and genetic diversity is found, followed by a sub-domestication region in India [8-10]. European accessions usually cluster together with those from West Africa and were likely imported from this region [10]. Breeding lines in America also show a high genetic similarity with African accessions [11] although local American landraces show a high genetic divergence [10]. In addition, regions like East Africa and Oceania show the lowest genetic diversity suggesting the presence of bottlenecks or founder effects during cowpea migration to these areas [10].

Because of this domestication history linked to a center of origin in Africa, cowpea research has been underway in several African countries for many years. Breeding activities in sub-Saharan Africa involving germplasm collection, evaluation and screening for the identification of lines with high yield potential resulted in a diverse cowpea germplasm collection constituted by more than 15000 cultivated cowpeas from 89 different countries [1]. Additionally, a core collection of more than 2000 accessions based on geographical, agronomical and botanical descriptors has been established in The International Institute for Tropical Agriculture (IITA) genebank with the aim of discovering new traits related with stress tolerance for the development of new breeding lines [12]. On the other hand, cowpea has several features of a classical model plant for genomic studies, such as a relatively small diploid (2n=2x=22 chromosomes) genome of ~613Mbp, a short annual life-cycle and a highly selfing nature [13].

The limited number of cowpea breeding programs in Mozambique has contributed to the country ineffectiveness in taking the advantage of the continent's high genetic potential. A significant pool of cowpea landraces is thought to be available, but the limited detailed information about their diversity and agronomic potential makes it difficult for breeding programs to thrive. Thus, the characterization of cowpea genetic resources available in Mozambique is of extreme importance for conservation and breeding, since it is the second most cultivated legume crop in the country, occupying an extension of ca. 380 000 ha, with an average yield of 0.275 t ha^{-1} [1]. Unlike commercial varieties, landraces maintained by farmers usually have high levels of genetic variability as they have evolved from years of uncontrolled cross-regional and infield genetic exchange, even between previously released and discontinued open pollinated varieties [14], not being subjected to selection over a long period of time. However, knowledge about their variability is usually limited [15]. Therefore, the aim of this study was to assess the genetic diversity of cowpea landraces from five agro-ecological regions across three provinces of Mozambique, using Single Sequence Repeat (SSR) markers.

2. Materials and Methods

2.1. Plant material

Fifty nine cowpea landraces corresponding to 10 populations were sampled in six agro-ecological zones (AEZ) in the provinces of Manica, Sofala and Zambezia, where cowpea is grown as an integral component of local cereal-legume cropping systems (Fig. 1): R3 (North and Central Gaza and Western Inhambane), R4 (Medium altitude areas of Central Mozambique), R5 (Low altitude areas of Sofala and Zambezia), R6 (Dry areas of Zambezia and Southern Tete), R7 (Mid-altitude areas of Zambezia, Nampula, Tete, Niassa and Cabo Delgado) and R10 (High altitude areas of Zambezia, Niassa, Angonia- Maravia and Manica). Additionally, two widely used commercial cultivars (IT16 and IT18) released by the Mozambican Institute of Agricultural Research (IIAM) and bred through a partnership with the International Institute of Tropical Agriculture (IITA) in Nigeria were also used in this study.

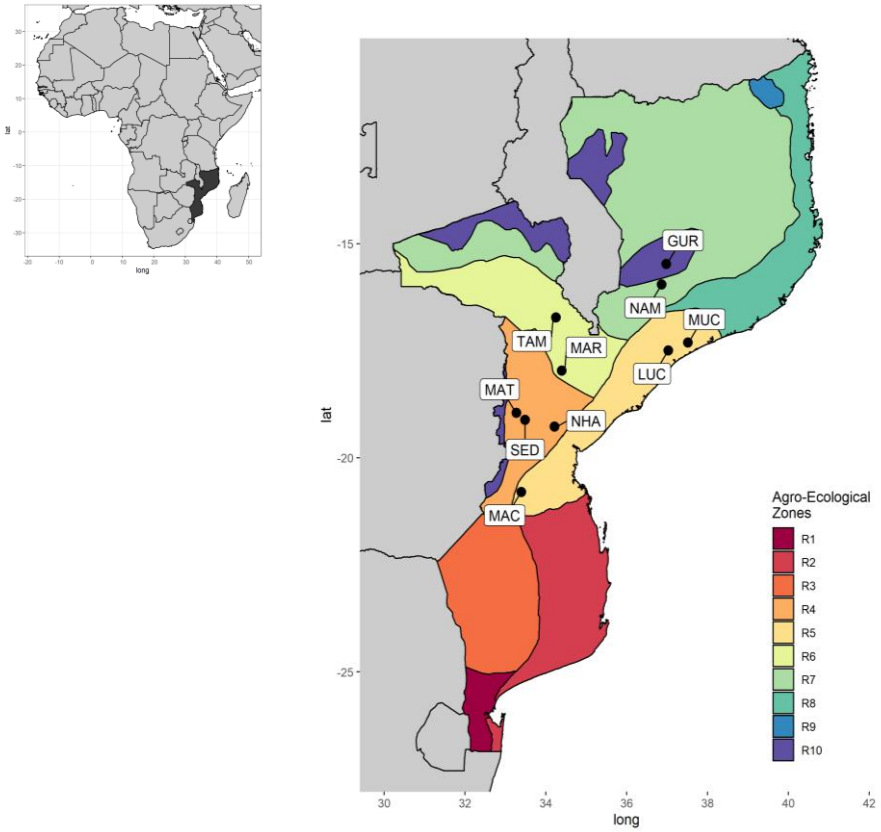

Figure 1. Left: Location of Mozambique in East Africa. Right: Studied landraces of *Vigna unguiculata*. Population codes follow Table 2. Colors indicate the different eco-geographical zones (AEZs) of Mozambique based on [16].

2.2. DNA extraction and nSSR amplification

The 61 samples used in this study were genotyped based on nine polymorphic nuclear simple sequence repeats (SSR's) previously developed by [17]: VuUGM05, VuUGM22, VuUGM31, VuUGM33, VuUGM39, VuUGM40, VuUGM68, VuUGM71 and VuUGM74. Based on an initial survey, we selected these nSSR markers since they produced robust, highly polymorphic amplified bands among the entire collection of cowpea samples. Total genomic DNA was extracted from 50 mg of ground leaves using the InnuSPEED Plant DNA Kit (Analytik Jena Innuscreen GmbH, Germany) according to the manufacturer's protocol. The average yield and purity were assessed spectrophotometrically by OD230, OD260 and OD280 readings (Nanodrop 2000, Thermo Fisher Scientific, Waltham, MA, USA) and visualized by electrophoresis in 1% agarose gels under UV light. Amplifications were performed in 15 µl reactions containing: 1.25U TaKaRa Hot startTaq polymerase, 1X Buffer I, 1mM dNTPs, 5 µM Primer F and R and 100 ng DNA under the following

PCR conditions: an initial denaturation at 95 °C for 5min followed by 35 cycles of denaturation at 65 °C (20 sec), annealing at 56 °C for 30 sec and a final extension at 60°C for 30min. Allele sizes were determined using GeneMapper 3.2 (Applied Biosystems; UK).

2.3. Genetic diversity and population structure

For each nSSR locus and landrace, genetic diversity was assessed by calculating the total number of alleles (N_a), mean expected heterozygosity (H_e), mean observed heterozygosity (H_o), allelic richness (A_R), and inbreeding coefficient (F_{IS}) using FSTAT 2.9.3.2 [18]). GenAlEx 6 software was used to estimate the mean expected heterozygosity (H_e) and mean observed heterozygosity (H_o) for each population, as well as the number of private alleles [19]. The selfing rate (s) was estimated as $s = 2F_{IS}/(1 + F_{IS})$ [20]. An analysis of variance was used to detect significant differences between sites for the measured genetic values. Grids for all significant genetic parameters were generated in R and are based on a grid with a cell size of 30 seconds (which corresponds to approximate 1 km in the study area) applying a 1.5-degree circular neighbourhood diameter. The circular neighbourhood is used to re-sample the genetic composition of a single sample to all surrounding grid cells, with a size of 30 seconds, within a diameter of 1.5 degree around its location. In this way, the genetic composition of each sample is representative for the area within the defined buffer zone.

2.4. Population structure and differentiation

The Bayesian program STRUCTURE v.2.3.4 [21] was used to test whether any discrete genetic structure exists among the landraces and regions sampled. The analysis was performed assuming a number of clusters from K =1 to K = 8, with 10 repetitions per K. Models were run assuming ancestral admixture and correlated allele frequencies with 50,000 burn-in steps, followed by run lengths of 300,000 interactions for each K. The optimum K was determined using STRUCTURE HARVESTER [22], which identifies the optimal K based both on the posterior probability of the data for a given K and the ΔK [23]. To correctly assess the membership proportions (q values) for clusters identified in STRUCTURE, the results of the replicates at the best-fit K were post-processed using CLUMPP 1.1.2 [24]. POPULATION 1.2 [25] was used to calculate the Nei's genetic distance [26] among individuals and to construct an unrooted neighbour-joining tree with 1000 bootstrap replicates. A Principal Component Analysis (PCoA) was also constructed in GenAlEx6 [27] to detect the genetic relatedness among individuals based on Nei's genetic distance. We estimated genetic differentiation among locations using an analysis of molecular variance (AMOVA) with ARLEQUIN 3.11 [28]. Molecular variance was quantified among populations and within populations considering AERs and wild cowpea versus cultivars, using an AMOVA using 10,000 permutations at 0.95 significance levels in ARLEQUIN 3.11 [28].

2.5. Spatial analysis and genetic diversity rarefaction

Grids for genetic parameters were generated in DIVA-GIS (www.diva-gis.org), based on a grid with a cell size of 2.5 minutes (which corresponds to approximatly 4.5 km in the study area) and applying a circular neighborhood with a diameter buffer of one degree (corresponding to approximate 111 km). The circular neighborhood was used to illustrate the allelic composition of each sampled site representative for the area within the defined buffer zone. Genetic diversity rarefaction considered the spatial average of several population parameters such number of alleles (N_A), observed heterozygosity (H_o), inbreeding coefficient (F_{IS}) and % selfing rate (s).

3. Results

3.1. Genetic diversity

The total number of alleles varied between 49 in VuUGM74 and 145 in VuUGM40 (Table 1). For each locus, observed heterozygosity values (H_o) ranged from 0.014 in VuUGM74 to 1 in VuUGM40 and expected heterozygosity (H_e) ranged from 0.016 in VuUGM74 to 0.806 to VuUGM33. F_{IS} values

varied between -0.008 and 0.857 (respectively for loci VuUGM68 and VuUGM31; Table 1) across the loci studied.

Table 1. Characteristics and genetic diversity statistics of the nuclear microsatellite (nSSR) primers used in the genetic study of *Vigna unguiculata*. For each locus, the total number of alleles (N_a), mean expected heterozygosity (H_e), mean observed heterozygosity (H_o), and the fixation index (F_{IS}) obtained from the 61 studied samples are shown.

Primer name	Primer sequence 5′-3′	Gene Bank ID	N_a	H_o	H_e	F_{IS}
VuUGM33	F: AAAGGTGGGGGATTATGAGG R: TGTCCAATCCTGATGGATGA	FG853417	83	0.907	0.806	-0.091
VuUGM71	F: TTCACAACCTGTCCACCTCA R: GGCGTCCCAACAGATAAGAA	FG819327	125	0.143	0.548	0.783
VuUGM05	F: GCGGGATTCTATTCCAGTGA R: TCCATTGGGTTTCTCAACCT	FC459955	82	0.174	0.617	0.767
VuUGM39	F: CGAAAAAGCATGATCAACCA R: CCCCTTTCGCTAAAATTTCC	FG863845	97	0.149	0.749	0.851
VuUGM22	F: CAATCACCATTCACCAAACA R: TATTGGGACTCAGGTCTTGG	FG908248	112	0.181	0.629	0.749
VuUGM31	F: TGGTTCACTTCCCATATTGTC R: AGGCAGAGACGAAGGAGTGA	FG932695	122	0.136	0.711	0.857
VuUGM40	F: TTCTACATGGTTTTGGGGTCA R: GAGCTTGCCCTCAAGAATTG	FG864565	145	1.003	0.671	-0.426
VuUGM68	F: TGATTGATGGTGGTGTAGCC R: GCACTTCACTCATCGTTGCT	FG807949	59	0.415	0.397	-0.008
VuUGM74	F: GCCTCCTCTCACAAACTTGC	FF547768	49	0.014	0.016	0.018

A total of 327 alleles were found among the set of *V. unguiculata* landraces, varying significantly between sites (P<0.001; Table 2). The number of alleles varied geographically from 14 in the coastal area of Muchela to 71 in the dry western area of Tambara (Fig. 2). Allelic richness varied between 1.250 in Muchela and 1.751 in Gurué with no statistical differences being found between areas (P=0.452; Table 1). However, the number of private alleles varied significantly across areas (P<0.001; Table 2) with the highest number being found in Gurué, Tambara and Machaze (Fig. 3).

Table 2. Genetic diversity within the cowpea genotypes studied. The number of samples analysed (N), total number of alleles (N_A), mean allelic richness (A_R), mean observed heterozygosity (H_o) and expected heterozygosity (H_e), inbreeding coefficient (F_{IS}) and % selfing rate (s) are shown for each population.

Populations	Province	AEZ	N	N_A	A_R	H_o	H_e	F_{IS}	s
Gurué (GUR)	North Zambezia	R10	6	46	1.751	0.389	0.688	0.506	60%
Namarroi (NAM)	North Zambezia	R7	4	23	1.534	0.379	0.454	0.250	25%
Muchela (MUC)	Central Zambezia	R7	4	14	1.250	0.222	0.535	-0.412	74%
Lucas Branco (LUC)	South Zambezia	R7	4	22	1.432	0.426	0.577	-0.032	41%
Nhamatanda (NHA)	Central Sofala	R4	4	31	1.682	0.278	0.479	0.707	59%
Maringué (MAR)	Central Sofala	R5	3	22	1.503	0.407	0.494	0.310	30%
Tambara (TAM)	North Manica	R6	23	71	1.612	0.320	0.654	0.592	68%
Sede nova (SED)	North Manica	R6	3	23	1.562	0.222	0.451	0.323	69%
Matsinho (MAT)	Central Manica	R4	3	23	1.577	0.221	0.451	0.156	67%
Machaze (MAC)	South Manica	R3	5	29	1.555	0.267	0.500	0.549	64%
IT-16	Commercial cultivar	R4	1	12	1.333	0.333	0.167	-	

IT-18	Commercial cultivar	R6	1	11	1.222	0.222	0.111	-

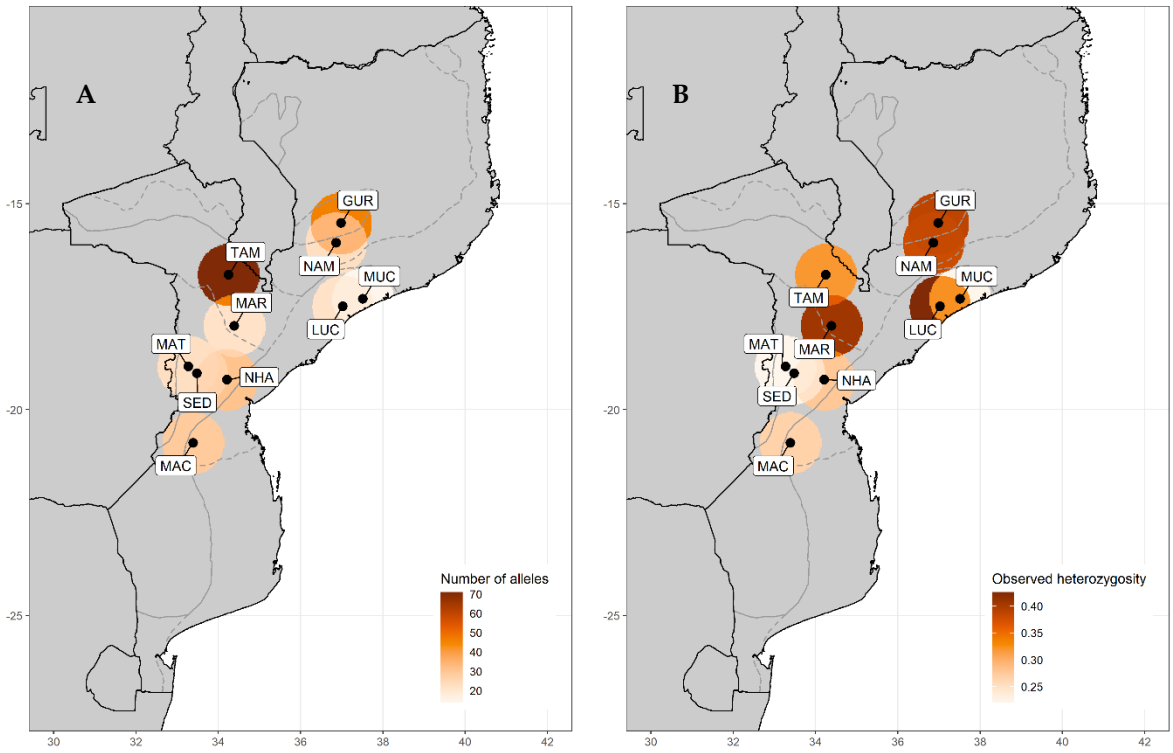

Figure 2. Map of the number of alleles (A) and observed heterozygosity (B) in 30 seconds (1km) grid cells applying a 1-degree circular neighborhood. Dashed lines indicate the agro-ecological zones [16].

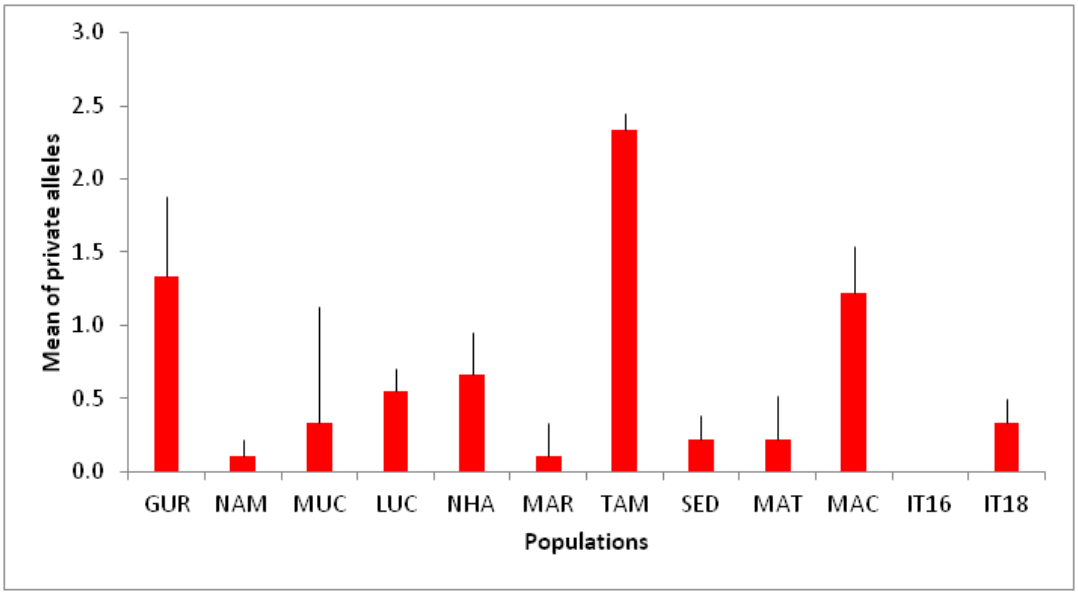

Figure 3. Population structure of *Vigna unguiculata* based on 9 SSRs and using the best assignment result retrieved by STRUCTURE ($K = 4$). Each individual sample is represented by a thin vertical line divided into K coloured segments that represent the individual's estimated membership fractions in K clusters. Landraces and province are indicated below. AEZs are indicated in individual labels with different colours for better visualization. The two cultivars are also indicated.

The mean observed heterozygosity varied significantly between 0.222 (Muchela, Sede Nova and Matsinho) and 0.426 (Lucas Blanco) ($P<0.001$; Fig. 2), and the mean expected heterozygosity varied between 0.451 (Matsinho) and 0.654 (Tambara) without statistical differences ($P=0.481$; Table 2). F_{IS}

values varied significantly between sites (P<0.001; Table 2), ranging from negative values of -0.412 in the coastal area of Muchela to positive values of 0.707 in the central area of Nhamatanda (Fig. 4). The rate of self-fertilization in *V. unguiculata* also varied significantly between sites (P<0.001; Table 2) with the lowest values found in the northern region of Namarroi (25%) and the highest in the coastal area of Muchela (74%) (Fig. 4).

The two cultivars had a low number of alleles (IT-16: 11 and IT-18: 2) and allelic richness (IT-16: 1.333 and IT-18: 1.222) constrained by the small sampling size. However, although the observed heterozygosity (IT-16: 0.333 and IT-18: 0.222) was higher than the expected one in both cultivars (IT-16: 0.167 and IT-18: 0.111; P<0.001 in both cases), it was also lower than the ones found in most local landraces (Table 2; P<0.001).

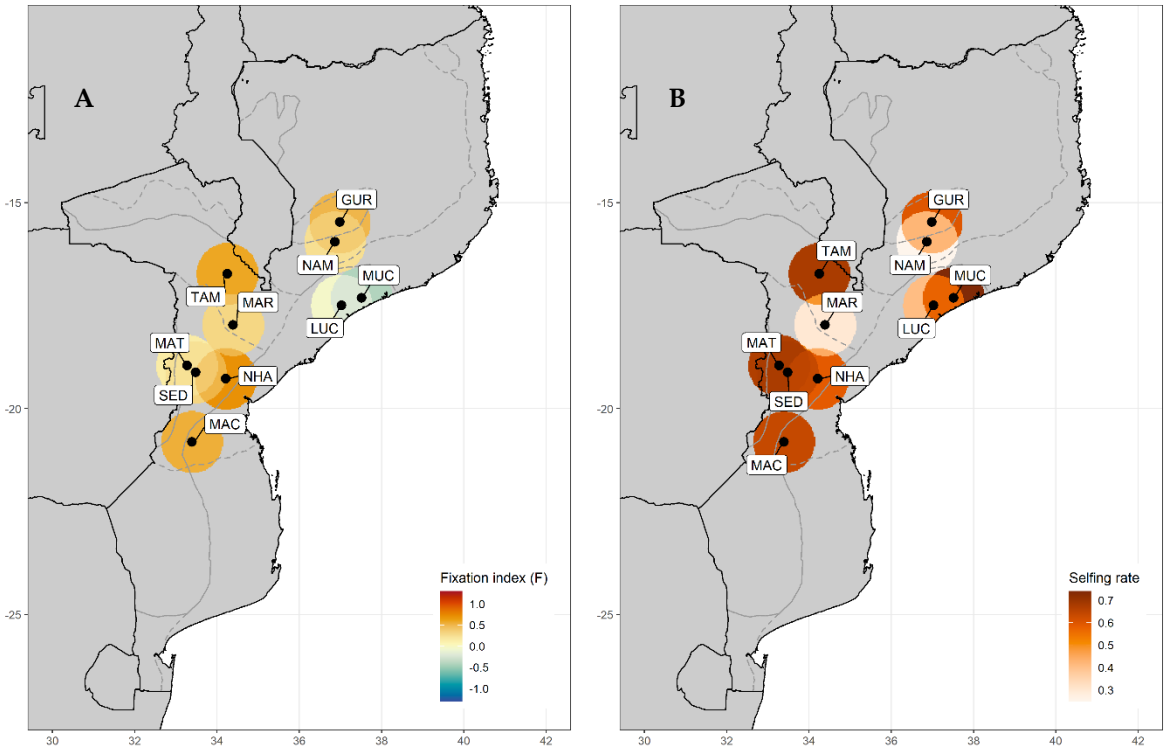

Figure 4. Map of the fixation index (A) and selfing rate (B) in 30 seconds (1km) grid cells applying a 1-degree circular neighborhood. Dashed lines indicate the agro-ecological zones [16].

3.2. Genetic structure of V. unguiculata

The Bayesian clustering program STRUCTURE found the highest LnP(D) and ΔK values for K = 4 (Fig. S1). Results showed a high degree of admixture between populations without any specific geographic pattern or clustering considering the different AEZs (Fig. 4). One cluster was predominant and grouped all landraces from North Zambezia, and most landraces from Sofala and Central Manica; the second cluster characterized Central and South Zambezia landraces; the third clustered landraces from North Manica as well as Central Sofala; the fourth cluster was exclusively composed by landraces from South Manica (Fig. 2). The two cultivars clustered with one the predominant group found in several populations, although both cultivars showed signs of admixture with the other clusters.

In accordance with these results, the NJ tree separated all groups assigned by STRUCTURE revealing again no general correlation with the geographical distribution of landraces (Fig. 5). All individuals from R3 and R7 were clustered into two different clades, one with 65% and the other with 34% bootstrap support (BS) value (Fig. 5). Most individuals from R6 clustered in the same group (57% BS) while R4, R5 and R10 were clustered into two different groups. The two cultivars were nested within the wild populations, although in two different separated groups.

The PCoA spatially separated the landraces analysed into three main groups (Fig. 6). In accordance to the NJ tree, the landraces from R3 and R7 were separated from the main group: the first being on the up-left of axis 2 that accumulated 21.14% of variance while the second on the down-left of axis 2. All remaining landraces were clustered in a heterogeneous group containing also the two cultivars.

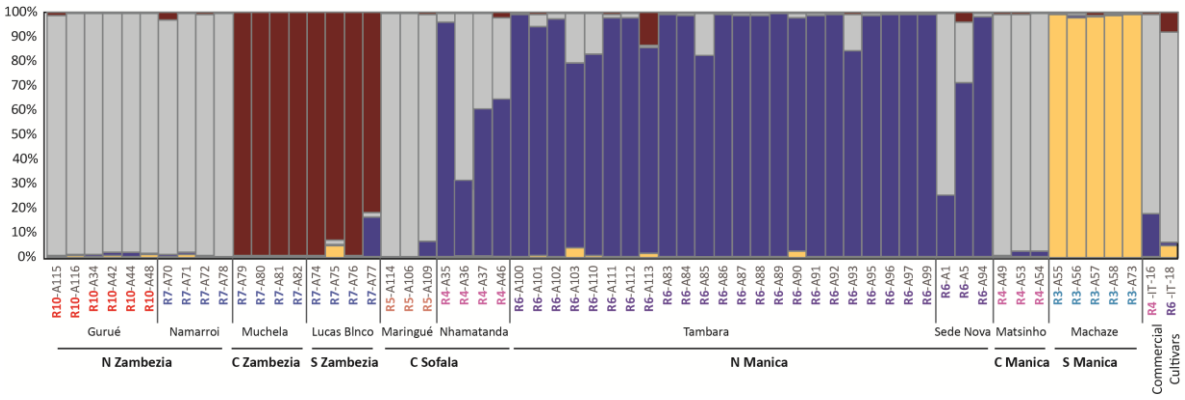

Figure 5. Population structure of *Vigna unguiculata* based on 9 SSRs and using the best assignment result retrieved by STRUCTURE ($K = 4$). Each individual sample is represented by a thin vertical line divided into K coloured segments that represent the individual's estimated membership fractions in K clusters. Landraces and province are indicated below. AEZs are indicated in individual labels with different colours for better visualization. The two cultivars are also indicated.

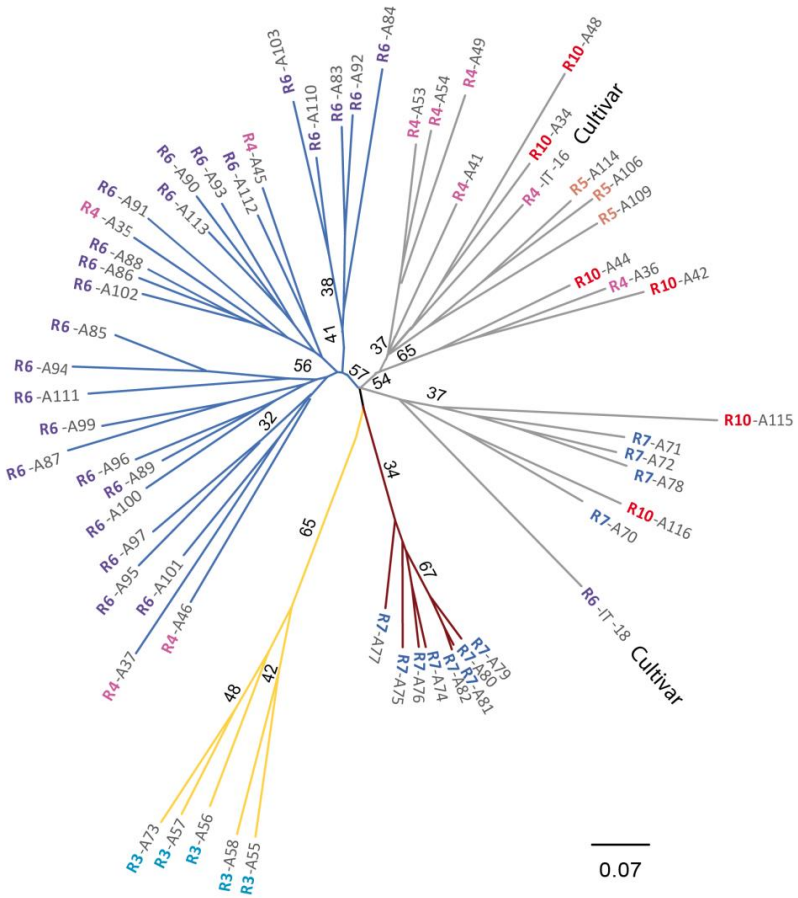

Figure 6. Unrooted neighbour-joining tree of the studied *Vigna unguiculata* landraces including the two cultivars, based on Nei's Da genetic distance. Numbers associated with branches indicate bootstrap values (BS) based on 1000 replications. Only BS above 30 are shown. Colours of branches

indicate the four genetic groups found in STRUCTURE. AEZs are indicated in branch labels with different colours following Fig. 3.

3.3. Genetic differentiation between populations

Overall, genetic differentiation was significantly low (AMOVA FST = 0.199, $P < 0.001$). The analysis performed over the landraces sampled indicated that only 19.92% of the genetic variation was attributed among AEZs (Table 3). The highest molecular variance was found among genotypes within landraces (47.39%), followed by the one found within genotypes (32.69%; P<0.001; Table 3). Remarkably, a very low molecular variance was found between wild cowpea versus the cultivars (0.12%) being most of the variance found among individuals within samples (65.58%; Table 3).

Table 3. Analysis of molecular variance (AMOVA) for the sampled populations of *Vigna unguiculata*.

Source of variance		d.f.	Sum of squares	% of variance
Among landraces				
	Among AEZs	6	77.612	19.92
	Among genotypes within landraces	54	207.109	47.39
	Within genotypes	61	60.001	32.69
Among cowpea landraces vs. cultivars				
	Among samples	1	4.772	0.12
	Among individuals within samples	58	279.949	65.58
	Within individuals	61	60.000	34.30

4. Discussion

Landraces harbor a genepool of unexplored alleles that constitute an unique set of genetic resources for breeding to improve productivity, nutritional value, adaptation and resilience to climate change [29-32]. Given their evolutionary history and adaptation to local conditions, landraces usually have higher genetic diversity and environmental resilience than modern varieties [33-36]. However, such richness tends to be lost because most of the current intensive agricultural systems is based on few high-input and high-yielding cultivars [37]. Thus, a comprehensive characterization of landraces towards the development of conservation and breeding strategies, is among the main clues to face the major agricultural challenges related to population growth and environmental risks.

Despite the ongoing agricultural changes in Africa, according to our data, the nine microsatellites employed in this study were highly polymorphic and revealed the existence of high genetic diversity between landraces of *V. unguiculata* landraces from Mozambique (Table 1). A total of 327 alleles were found among the 59 cowpea landraces, which can be attributed to high genetic heterogeneity (Table 2). Indeed, the genetic diversity values found within the studied landraces (Ho: 0.222- 0.426; He: 0.451- 0.654) were much higher than the ones reported for cultivated cowpeas. For instance, high-density single nucleotide polymorphism (SNP) genotyping using the Cowpea iSelect Consortium Array studied population structure and genetic diversity in a set of 91 worldwide cowpea accessions and found an average PIC and He of 0.25 and 0.31, respectively [8]. Similar results were obtained by Huynh et al. [10] and Xiong et al. [9] using respectively, 422 cowpea landraces and 768 cowpea genotypes, collected in 56 countries.

In comparison, the two commercial cultivars (IT-16 and IT-18) had a very low number of alleles and heterozygosity values, and cluster analyses (PcoA or NJ tree) showed no clear differentiation between these modern varieties and landraces. Pairwise genetic distances reported in other studies have also shown that African landraces were close to wild cowpea samples [10]. This suggests that genetic diversity of these two commercial varieties is still close from the ones found in landraces although more individuals are needed to accurately determine if genetic erosion is occurring.

Population structure analysis using worldwide cowpea samples usually delineate African landraces into two major gene pools separated by the Congo River basin, the East/South and the West Africa [8-10], although nothing has been reported for cowpea genetic structure within these regions. Our study, focused on Mozambican (East Africa) landraces, found four genetic groups with a high degree of admixture (Fig. 4). No specific geographic pattern or clustering was found considering the different AEZs either in the NJ tree or the PcoA (Fig. 5,6), which supports the presence of gene flow between these regions. The rate of self-fertilization in *V. unguiculata* varied across populations (25-74%; Table 2; Fig. 3) supporting the possibility of gene flow between individuals. In fact, two populations (Lucas and Muchela) exhibited negative FIS values indicating that these populations are less related than expected under random mating (Fig. 3) which could imply fewer homozygotes and consequently cross-breeding. Nonetheless, most of the remaining populations had low FIS values (0.1-0.3) which indicates that inbreeding might not be prevalent.

The analysis of genetic differentiation indicated that most of the genetic variation was explained by differences among genotypes within landraces (Table 3), which also supports the hypothesis of gene flow. This low genetic differentiation and the absence of a geographical pattern associated with AEZs might be due to crossbreeding between individuals but also to seed exchange by farmers. Seed exchange is a common practice between African farmers of neighbouring areas [38] and could explain the specific genetic cluster found in the isolated landraces of South Manica that shows no admixture with the remaining ones. It is economical unfeasible for seed companies to distribute small amounts of seeds over long rural distances in Africa, and therefore certified, commercial seeds do not reach the farmers [39] In addition, certified seeds are generally expensive and farmers are unwilling to buy them at a cost twice or more than that of the grain [39]. Nonetheless, continuous recycling of seeds decreases results in poor grain yields [38] highlighting the importance of conserving landraces and their seed stock.

The high genetic diversity found in Mozambique, in comparison to other world regions reinforces the importance of local landraces to widen the genetic base of modern varieties of cowpea. The results of this study underline the hidden genetic diversity in local landraces, which should be conserved as sublines in genebanks to avoid the expected reduction of genetic diversity within successive regeneration of bulk samples. The high levels of genetic differentiation found within landraces (but not among AEZs) could imply the presence of different phenotypes, which should be conserved to retain the full pool of genes and morphological combinations within landraces. These suggest the existence of a valuable gene pool in Mozambican landraces, which might exhibit desired traits for exploitation in future breeding programs. In fact, according to Gomes et al [7], the comparison of landraces A55 from R3, A80 from R7, and A116 from R10, clustering in different groups (Fig. 6), revealed contrasting responses, respectively leading to high sensitivity, mild sensitivity and high tolerance to drought stress related to the regulation of photosynthesis, C/N metabolism and antioxidative status [7].

A priority for *in situ, on farm* conservation should be given to the landraces of Gurué, Tambara and Machaze, that showed a high number of private alleles (Fig. 3), and belong to different genetic groups according to STRUCTURE (Fig. 4). *On farm* conservation allows the evolution of landraces, retaining potentially useful genetic variation needed to maintain crops ability to adapt to changes [40]. However, genetic diversity conserved *on farm* is complementary to that found in the genebank, and both systems are required for efficient conservation of cowpea. Thus, further to molecular tools, farmer's knowledge should be employed to optimize sampling of sublines within landraces for *ex situ* conservation. A core germplasm collection should include most of cowpea genetic diversity, which can be used from the results outlined in this study. The results of this work encourage a broad network of *on farm* activities that should be enrolled in a socio-economic framework to complement genebank collections. This is also the best way to prevent genetic erosion in the genebank while maintaining and expanding cultivation of cowpea in a wide range of environmental conditions.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: List of the sampled 59 cowpea accessions sorted by locality and province. The agro-ecological zone (AEZ) is indicated, as

well as the number of landraces studied within each population (59 landraces). The two commercial cultivars are also indicated.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, A.M.G.F., N.N., R.M., J.C.R., I.M., A.I.R.B; methodology, A.M.F.G., D.D., P.B.S., P.T., F.S.; software, D.D.; I.M.; validation, A.M.F.G., D.D., P.T., F.S. and I.M.; formal analysis, A.M.F.G., D.D., P.T., F.S. and I.M. .; investigation, A.M.F.G., D.D., P.B.S., I.M. and A.I.R-B.; resources, N.N., P.T., F.S., R.M., J.C.R. and A.I. R-B.; data curation, D.D., P.T., J.C.R., I.M. and A.I.R-B.; writing—original draft preparation, A.M.F.G, I.M. and A.I.R-B; writing—review and editing, all co-authors; visualization, D.D. and I.M.; supervision, R.M., J.C.R., I.M. and A.I.R-B.; project administration, R.M., J.C.R. and A.I.R-B.; funding acquisition, N.N., R.M. and A.I.R-B.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by funds from the Mozambican FUNDO NACIONAL DE INVESTIGAÇÃO (Project 201-Inv-FNI), NUFFIC, the Netherlands (Project NICHE-Moz-151), and by Fundação para a Ciência e a Tecnologia, I.P., through the PhD fellowship SFRH/BD/113952/2015 (A.M.F.G.) and the post-doctoral fellowship SFRH/BPD/100384/2014 (D.D), the research units UID/04129/2020 (LEAF), UIDP/04035/2020 (GeoBioTec), and UIDB/00239/2020 (CEF), and the APC.

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflict of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

References

1. Boukar, O.; Belko, N.; Chamarthi, S.; Togola, A.; Batieno, J.; Owusu, E.; Haruna, M.; Diallo, S.; Umar, M.L.; Olufajo, O.; Fatokun, C. Cowpea (*Vigna unguiculata*): genetics, genomics and breeding. *Plant Breed*. **2019**, *138*, 415–424.
2. Goncalves, A.; Goufo, P., Barros, A.; Dominguez-Perles, R.; Trindade, H.; Rosa, E.A.; Ferreira, L.; Rodrigues, M. Cowpea (*Vigna unguiculata* L. Walp), a renewed multipurpose crop for a more sustainable agri-food system: nutritional advantages and constraints. *J. Sci. Food Agric.* **2016**, *96*, 2941-2951.
3. Jayathilake, C.; Visvanathan, R.; Deen, A.; Bangamuwage, R.; Jayawardana, B.C.; Nammi, S.; Liyanage, R. Cowpea: an overview on its nutritional facts and health benefits. *J. Sci. Food Agric*. **2018**, *98*, 4793-4806.
4. Bado, B.; Bationo, A.; Cescas, M. Assessment of cowpea and groundnut contributions to soil fertility and succeeding sorghum yields in the Guinean savannah zone of Burkina Faso (West Africa). *Biol. Fertil. Soils* **2006**, *43*, 171-176.
5. Gomes, A.M.F.; Nhantumbo, N.; Ferreira-Pinto, M.; Massinga, R.; Ramalho, J.C.; Ribeiro-Barros, A. Breeding Elite Cowpea [*Vigna unguiculata* (L.) Walp] Varieties for Improved Food Security and Income in Africa: Opportunities and Challenges. In *Legume Crops*; El-Esawi, M.A., Ed.; IntechOpen: London, U.K., 2019; 636.
6. Goufo, P.; Moutinho-Pereira, J.M.; Jorge, T.F.; Correia, C.M.; Oliveira, M.R.; Rosa, E.A.S.; António, C.; Trindade, H. Cowpea (*Vigna unguiculata* L. Walp.) metabolomics: osmoprotection as a physiological strategy for drought stress resistance and improved yield. *Front. Plant Sci*. **2017**, *8*, 586.
7. Gomes, A.M.F.; Rodrigues, A.P.; António, C.; Rodrigues, A.M.; Leitão, A.E.; Batista-Santos, P.; Nhantumbo, N.; Massinga, R.; Ribeiro-Barros, A.I.; Ramalho, J.C. Drought response of cowpea (*Vigna unguiculata* (L.) Walp.) landraces at leaf physiological and metabolite profile levels. *Environ. Exp. Bot*. **2020**, *175*, 104060.
8. Carvalho, M.; Lino-Neto, T.; Rosa, E.; Carnide, V. Cowpea: a legume crop for a challenging environment. *J. Sci. Food Agric*. **2017**, *97*, 4273-4284.
9. 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*, e0160941.
10. Huynh, B.; Close, T.J.; Roberts, P.A.; Hu, Z.; Wanamaker, S.; Lucas, M.R.; Chiulele, R.; Cissé, N.; David, A.; Hearne, S.; Fatokun, C.; Diop, N.N., Ehlers, J.D. Gene pools and the genetic architecture of domesticated cowpea. *Plant Genome* **2013**, *6*, 1–8.

11. Fang, J.; Chao, C.T.; Roberts, P.A.; Ehlers, J.D. Genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] in four West African and USA breeding programs as determined by AFLP analysis. *Genet. Resour. Crop Evol.* **2007**, *54*, 1197–1209.
12. Mahalakshmi, V.; Ng, Q.; Lawson, M.; Ortiz, R. Cowpea [*Vigna unguiculata* (L.) Walp.] core collection defined by geographical, agronomical and botanical descriptors. *Plant Genet. Resour*. **2007**, *5*, 113-119.
13. Ehlers, J.D.; Hall, A.E. Cowpea (*Vigna unguiculata* (L.) Walp.). *Field Crops Res.* **1997**, *53*, 187-204.
14. Ficiciyan, A.; Loos, J.; Sievers-Glotzbach, S.; Tscharntke, T. More than yield: ecosystem services of traditional versus modern crop varieties revisited. *Sustainability* **2018**, *10*, 28-34.
15. Hagenblad, J., Zie, J., Leino, M.W. Exploring the population genetics of genebank and historical landrace varieties. *Genet. Resour. Crop Evol*. **2012**, *59*,1185–1199.
16. Ministry of Agriculture and Fisheries (MAF). Agroecological zones and production systems. Working Document 2/B, Program of Investment in Agricultural Extension, Process of the Formulation of Proagri. Ministry of Agriculture and Fisheries: Maputo, Mozambique, 1996.
17. Gupta, S.K.; Gopalakrishna, T. Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other *Vigna* species. *Genome* **2010**, *53*, 508-523.
18. Goudet, J. FSTAT, a program to estimate and test gene diversities and fixation indices (Version 2.9.3.2). 2002. Available at: http://www.unil.ch/izea/softwares/fstat (accessed 03 of November of 2019).
19. Peakall, R.; Smouse, P.E. GenAlEx 6 genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295.
20. Ritland, K. Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution* **1990**, *44*, 1230-1241.
21. Pritchard, J.K.; Stephens; M.; Rosenberg, N.A.; Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet*. **2000**, *67*, 170–181.
22. Earl, D.A.; von Holdt, B.M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet. Resour*. **2012**, *4*, 359-361.
23. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol*. **2005**, *8*, 2611-2620.
24. Jakobsson, M.; Rosenberg, N.A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806.
25. Langella, O. (2000). POPULATIONS 1.2.30. Laboratoire Populations. Genetique et Evolution. Centre National de la Recherche Scientifique, CNRS UPR9034: Gif Sur Yvette, France, 2007.
26. Nei, M.; Tajima, F.; Tateno, Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol*. **1983**, *19*, 153-170.
27. Peakall, R.; Smouse, P.E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295.
28. Excoffier, L.; Laval, G.; Schneider, S. ARLEQUIN, Ver 3.0. An integrated software package for population genetic data analysis. *Evol. Bionformatics Online* **1983**, *1*, 47-50.
29. Lopes, M.S.; El-Basyoni, I.; Baenziger, P.S.; Singh, S.; Royo, C.; Ozbek, K.; Aktas, H.; Ozer, E.; Ozdemir, F.; Manickavelu, A.; Ban, T.; Vikram, P. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot*. **2015**, *66*, 3477–3486.
30. Dwivedi, S.L.; Ceccarelli, S.; Blair, M.W.; Upadhyaya, H.D.; Are, A.K.; Ortiz, R. Landrace germplasm for improving yield and abiotic stress adaptation. *Trends Plant Sci*. **2016**, *21*, 31–42.
31. Monteagudo, A.; Casas, A.M.; Cantalapiedra, C.P.; Contreras-Moreira, B.; Gracia, M.P.; Igartua, E. Harnessing novel diversity from landraces to improve an elite barley variety. *Front. Plant Sci*. **2019**, *10*, 434.
32. Hour, A.; Hsieh, W.; Chang, S.; Wu, Y.; Chin, H.; Lin, Y. |Genetic diversity of landraces and improved varieties of rice (*Oryza sativa* L.) in Taiwan. *Rice*. Preprint available at https://doi.org/10.21203/rs.3.rs-44140/v1.
33. Ganeva, G.M; Korzun, V.; Landjeva, S.; Popova, Z.; Christov, N.K. Genetic diversity assessment of Bulgarian durum wheat (*Triticum durum* Desf.) landraces and modern cultivars using microsatellite markers. *Genet. Resour. Crop Evol*. **2010**, *57*, 273–285.
34. Kumbhar, S.D.; Kulwal, P.L.; Patil, J.V.; Sarawate, C.D.; Gaikwad, A.P.; Jadhav, A.S. Genetic diversity and population structure in landraces and improved rice varieties from India. *Rice Sci*. **2015**, *22*, 99-107.

35. Soriano, J.M.; Villegas, D.; Aranzana, M.J.; García del Moral, L.F.; Royo, C. Genetic structure of modern durum wheat cultivars and Mediterranean landraces matches with their agronomic performance. *PLoS ONE* **2016**, *11*, e0160983.
36. Pascual, L.; Ruiz, M.; López-Fernández, M.; Pérez-Peña; H.; Benavente, H.; Vázquez, J.F.; Sansaloni, C.; Giraldo, P. Genomic analysis of Spanish wheat landraces reveals their variability and potential for breeding. *BMC Genomics* **2020**, *21*, 122.
37. FAO/IPGRI. Review and development of indicators for genetic diversity, genetic erosion and genetic vulnerability (GDEV): summary report of a joint FAO/IPGRI workshop. FAO/IPGRI: Rome, Italy, 2002.
38. Njonjo, M.W.; Muthomi, J.W.; Mwang'ombe, A.W. Production practices, postharvest handling, and quality of cowpea seed used by farmers in Makueni and Taita Taveta counties in Kenya. *Int. J. Agron*. **2019**, article ID1607535.
39. Kimani, P.; Kamundia, D.K.; Narla, R.D.; Mwang'ombe, A.W. An integrated seed delivery system and seed research in Kenya: case studies of pigeon pea, onion and dry bean. In *Seed Systems, Science, and Policy in East and Central Africa*; Francis, J.A., Ed.; CTA: The Hague, The Netherlands, 2014.
40. Bellon, M.R.; van Etten, J. Climate change and on-farm conservation of crop landraces in centres of diversity In *Plant Genetic Resources and Climate Change*; CABI Publishing: Egham, U.K., 2014