

## Article

# Genome-Wide Association Study Revealed SNP Alleles Associated with Seed Size Traits in African Yam Bean (*Sphenostylis stenocarpa* (Hochst ex. A. Rich.) Harms)

Oluwaseyi E. Olomitutu <sup>1,2</sup> , Rajneesh Paliwal <sup>2</sup>, Ayodeji Abe <sup>1</sup> , Olubusayo O. Oluwole <sup>2</sup>,  
Olaniyi A. Oyatomi <sup>2</sup>  and Michael T. Abberton <sup>2,\*</sup>

<sup>1</sup> Department of Crop and Horticultural Sciences, University of Ibadan, Oduduwa Road, Ibadan 200132, Nigeria

<sup>2</sup> Genetic Resources Center, International Institute of Tropical Agriculture, Oyo Road, Ibadan 200001, Nigeria

\* Correspondence: m.abberton@cgiar.org

**Abstract:** Seed size is an important yield and quality-determining trait in higher plants and is also crucial to their evolutionary fitness. In African yam bean (AYB), seed size varies widely among different accessions. However, the genetic basis of such variation has not been adequately documented. A genome-wide marker-trait association study was conducted to identify genomic regions associated with four seed size traits (seed length, seed width, seed thickness, and 100-seed weight) in a panel of 195 AYB accessions. A total of 5416 SNP markers were generated from the diversity array technology sequence (DARtseq) genotype-by-sequencing (GBS)- approach, in which 2491 SNPs were retained after SNP quality control and used for marker-trait association analysis. Significant phenotypic variation was observed for the traits. Broad-sense heritability ranged from 50.0% (seed width) to 66.4% (seed length). The relationships among the traits were positive and significant. Genome-wide association study (GWAS) using the general linear model (GLM) and the mixed linear model (MLM) approaches identified 12 SNP markers significantly associated with seed size traits across the six test environments. The 12 makers explained 6.5–10.8% of the phenotypic variation. Two markers (29420334 | F | 0-52:C>G-52:C>G and 29420736 | F | 0-57:G>T-57:G>T) with pleiotropic effects associated with seed width and seed thickness were found. A candidate gene search identified five significant markers (100026424 | F | 0-37:C>T-37:C>T, 100041049 | F | 0-42:G>C-42:G>C, 100034480 | F | 0-31:C>A-31:C>A, 29420365 | F | 0-55:C>G-55:C>G, and 29420736 | F | 0-57:G>T-57:G>T) located close to 43 putative genes whose encoding protein products are known to regulate seed size traits. This study revealed significant makers not previously reported for seed size in AYB and could provide useful information for genomic-assisted breeding in AYB.

**Keywords:** African yam bean; phenotypic variation; genomic-assisted breeding; marker-trait association; seed size traits; SNP alleles



**Citation:** Olomitutu, O.E.; Paliwal, R.; Abe, A.; Oluwole, O.O.; Oyatomi, O.A.; Abberton, M.T. Genome-Wide Association Study Revealed SNP Alleles Associated with Seed Size Traits in African Yam Bean (*Sphenostylis stenocarpa* (Hochst ex. A. Rich.) Harms). *Genes* **2022**, *13*, 2350. <https://doi.org/10.3390/genes13122350>

Academic Editor: Anca Macovei

Received: 5 October 2022

Accepted: 9 December 2022

Published: 13 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Meeting global food and nutritional demands under limited growing space and changing climatic conditions is now the major target in crop breeding programs. Recently, the focus has shifted to relatively unknown indigenous legumes such as African yam bean (AYB) due to its dietary protein and mineral content and adaptive nature to wide climatic and soil conditions [1,2]. African yam bean is economically the most important species in the genus *Sphenostylis* and the most important tuberous legume of tropical Africa [3,4]. The utilization of its food substances (tubers and pulse) is a feature of cultural diversity in Africa [4,5]. The tuber contains, on average, 15.5% crude protein, 1.3% crude fat, and 68.3% carbohydrate [6]. The seed contains 22.5% protein, 53.7% carbohydrate, and 3.6% crude fat content [7]. Despite its numerous other benefits, low seed yield is one of the major constraints to its production [8,9] due to the absence of improved varieties and a lack of significant research attention.

Yield is a complex trait, difficult to improve directly. It is an expression of several component traits that are highly subject to environmental influences [10]. In crop breeding, the greatest yield improvements are associated with the selection and optimization of its component traits. Since the beginning of domestication in agriculture, increased seed size has been a major target as an important yield component trait [11]. Seed size is crucial to plant fitness in crops whose main mode of propagation is by seed and a key factor affecting eating quality and tolerance to abiotic stresses [12–15]. Compared with small seeds, large seeds accumulate sufficient nourishing substances for faster germination and stronger seedlings that can better compete for light and nutrition and have stronger tolerance to abiotic stresses [12,15,16]. Small seeds, on the other hand, are efficient at dispersal and colonization [17,18]. Seed size is the most commercially valued trait in dry grain legumes, and it varies widely among AYB accessions [4,19,20]. However, the genetic basis for variation in seed size in AYB is not yet known.

Recent advances in high-throughput genomic platforms have created the opportunity for the genome-wide-level understanding of the genetic basis of variation in complex traits at a finer resolution. Association mapping (AM), originally developed for use in mapping human disease genes [21], is now a popular method of marker-trait association studies in plants. Association mapping detects linkage disequilibrium between genetic markers and genes controlling the trait of interest by exploiting recombination events accumulating over many generations in natural populations [22]. It evaluates whether certain alleles within a population are found with specific phenotypes more frequently than expected [23]. Once genes and/or loci are identified and validated, they could be fixed to develop improved genotypes. Association mapping has several advantages over traditional linkage mapping. These include an increased resolution, a reduced research time (using existing populations rather than generating population via biparental crosses), and a higher allele number detection per locus as opposed to only two [24–26]. Association mapping also suffers some shortcomings, such as the detection of false positives in the population structure, which is a result of the linkage between causal and noncausal sites, more than one causal site, and epistasis. However, advancement in statistical methods has helped to reduce the rate of false positives [27]. Several GWAS reports have identified putative QTLs/genes in many leguminous crops [28–33] that played an important role in understanding the inheritance of quantitative traits [34,35] and trait deployment using a marker-assisted selection approach [36,37]. In legumes, several QTLs have been identified for seed size traits in cowpea [11], soybean [38], and common bean [39,40]. Unfortunately, very limited molecular research has been conducted in indigenous African legumes, including AYB. In AYB, the only available report of association mapping is the preliminary assessment for nutritional qualities by Oluwole et al. [41]. Here, a marker-trait association study was conducted to investigate the genetic basis of variation in seed size traits (seed length, seed width, seed thickness, and 100-seed weight) in AYB using diversity array technology sequence (DArTseq) genotype-by-sequencing (GBS)-based single nucleotide polymorphism (SNP) markers.

## 2. Materials and Methods

### 2.1. Germplasm

The germplasm consisted of 196 AYB accessions obtained from the existing collection of landraces at the Genetic Resource Center, International Institute of Tropical Agriculture, Ibadan, Nigeria. The passport data of the accessions can be found in Olomitutu et al. [9].

### 2.2. Phenotyping

The accessions were phenotyped under optimal field conditions across three IITA research farms in Nigeria—Ibadan, Kano, and Ubiaja—during the 2018 and 2019 cropping seasons. The experimental design was a  $14 \times 14$   $\alpha$  lattice with three replications. Each experimental unit consisted of 4 m single-row plots, with an inter-row spacing of 0.75 m and intrarow spacing of 0.5 m. Phosphorus fertilizer application in the form of triple super-

phosphate at a rate of 50 kg P/ha and staking were performed three weeks after planting. Manual weeding was carried out when necessary to keep the field clean. Details of the phenotyping methodology are described in Olomitutu et al. [9]. At harvest, data were recorded on 100-seed weight (g), seed length (mm), seed width (mm), and seed thickness (mm) on a plot basis using IITA descriptors for AYB [42].

### 2.3. Genotyping and Quality Control

Leaf samples were collected from three-week-old seedlings of each 196 AYB accession and stored at  $-80^{\circ}\text{C}$ . Genomic DNA (gDNA) was extracted using the diversity array technology (DArT) DNA extraction protocol ([https://ordering.seqart.net/files/DArT\\_DNA\\_isolation.pdf](https://ordering.seqart.net/files/DArT_DNA_isolation.pdf) accessed on 20 January 2019). The isolated gDNA were qualified on 1% agarose gel electrophoresis and quantified using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) following the manufacturer's protocol. The high-quality DNA (100 ng/ $\mu\text{L}$ ) samples were shipped to DArT Pty Ltd., Canberra, Australia, for genotyping using the whole genome profiling service of DArTseq technology [43]. Detailed methodology on complexity reduction, cloning, library construction, and cleaning was described by Egea et al. [44].

A raw dataset of 5416 DArTseq SNPs was generated. The DArTseq SNPs were filtered using call rate  $\geq 70\%$ , average reproducibility  $\geq 95\%$ , minor allele frequency (MAF)  $\geq 0.01$ , and 20% missing SNP data to remove poor-quality SNPs. After SNP quality control, a total of 2491 SNPs from 195 accessions were retained and used for the genome-wide association study (GWAS). Accession TSs-442 was filtered out due to low-quality SNPs. In the absence of the AYB genome, trimmed sequences of filtered SNPs were aligned on the common bean reference genome v1.1 (available at <https://phytozome-next.jgi.doe.gov>, (accessed on 20 January 2019)).

### 2.4. Phenotypic Data Analysis

Based on plot means across test locations, a combined analysis of variance (ANOVA) was performed using PROC GLM in SAS [45], with the RANDOM statement and TEST option. Location and year were considered fixed, while all other factors were regarded as random effects. Broad-sense heritability ( $H^2$ ) was estimated from the phenotypic ( $\sigma^2_p$ ) and genotypic ( $\sigma^2_g$ ) variances [46] and categorized as low (0–30%), moderate (30–60%), and high (>60%) according to [47]. Correlation analysis among the traits was performed using the corPlot function in R [48]. Distribution plots were also constructed for traits using the hist function in R. The plot means of the remaining 195 accessions (in each and combined locations) were used to calculate the best linear unbiased estimate (BLUE) using META-R version 6.04 [49]. The calculated BLUE value of each genotype was further used in the GWAS analysis.

### 2.5. Genome-Wide Association Study and Candidate Gene Identification

The 2491 filtered SNPs from 195 accessions were used for the assessment of population structure and GWAS using TASSEL 5 [50]. The population structure and relatedness analysis among genotypes in the AYB population were conducted using the genomic principal component analysis (PCA) matrix (P) and kinship matrix (K) [51,52]. Marker-trait associations were determined using two different models based on the estimated BLUEs for phenotypic traits (in each and combined locations) and filtered SNPs: the general linear model (GLM) with PCA as the fixed effect (GLM + PCA) and the mixed linear model (MLM) mixed linear model (MLM + PCA + K) [39]. Based on the distribution of  $p$ -values for traits, marker-trait associations were declared significant at  $p$ -values of  $\geq 10^{-4}$  [53,54]. Manhattan and quantile–quantile (Q–Q) plots were constructed using CMplot package in R.

The SNP markers that were significantly associated with seed size traits through GWAS in the combined location analysis were annotated for candidate gene identification in the LIS-legume information systems (<https://www.legumeinfo.org> accessed on 4 April 2020). Since AYB currently lacks a reference genome, which is a limitation to candidate gene mapping, the genome of a related legume, the common bean (*Phaseolus vulgaris*), was used. A blast search was performed for trimmed nucleotide sequence (60–80 bps) of significant AYB SNPs on the *Phaseolus vulgaris* genome database (*Phaseolus vulgaris* G19833 genome v2.0) in the legume information system. Synteny of related legumes (*Glycine max* 2.0, *Vigna angularis* 3.0, and *Cajanus cajan* 1.0) was also included in the search. After marking the annotated position in the *Phaseolus vulgaris* genome database, the scroll was zoomed to 1 Mb (500 Kbp up and downstream from the annotated position of the AYB SNP tag in the *Phaseolus vulgaris* genome database) to check for the surrounding candidate genes and their encoding protein products and know if they regulate the traits of interest. Identified putative candidate genes in the *Phaseolus vulgaris* genome database were also subsequently further researched in the previous crop studies literature for verification.

### 3. Results

#### 3.1. Phenotypic Evaluations

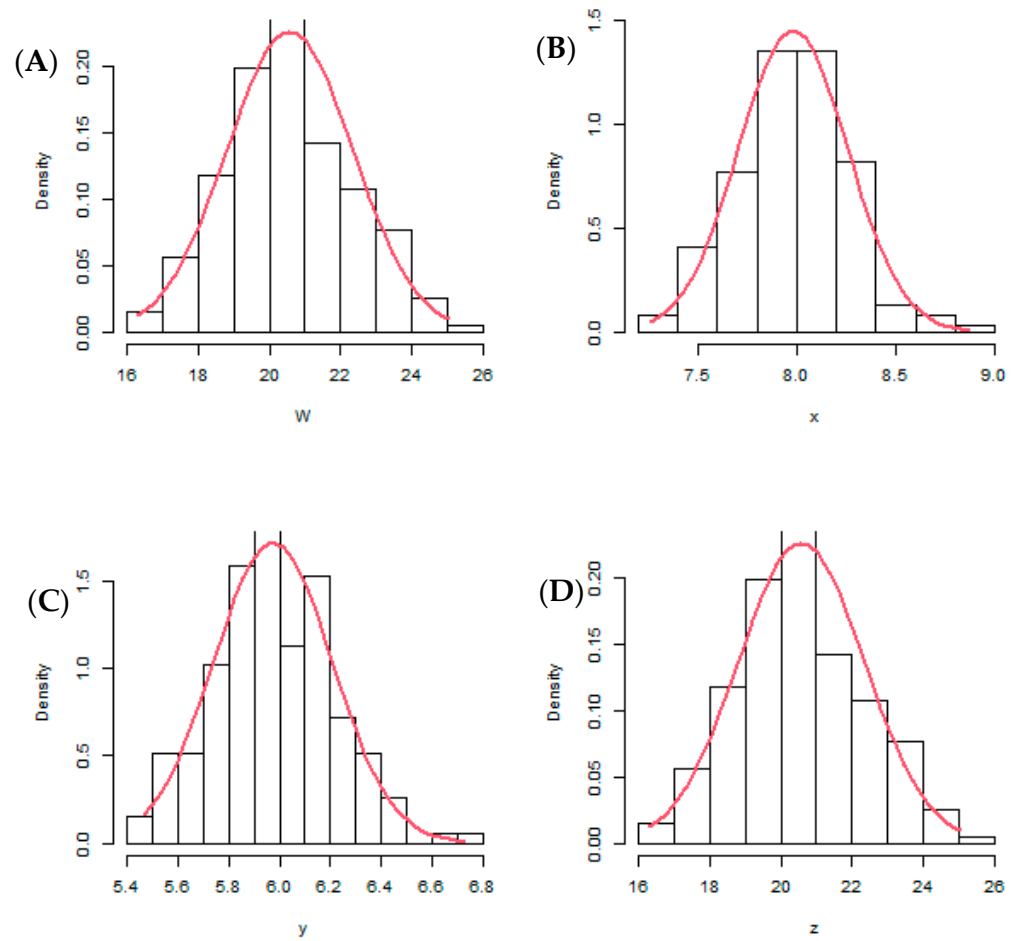
The ANOVA across the test locations revealed significant accession, location, location  $\times$  year, accession  $\times$  location, and accession  $\times$  location  $\times$  year effects for the four seed size traits. The accession  $\times$  year interactions effect was significant only for 100-seed weight and seed length, while year effect was significant for seed length and seed width. High broad-sense heritability (66.4%) was obtained in seed length, while seed thickness (57.8%), 100-seed weight (51.6%), and seed width (50.0%) had moderate heritability estimates (Table 1).

**Table 1.** Mean squares from analysis of variance for four seed size traits of 196 accessions of African yam bean evaluated at three test locations in Nigeria.

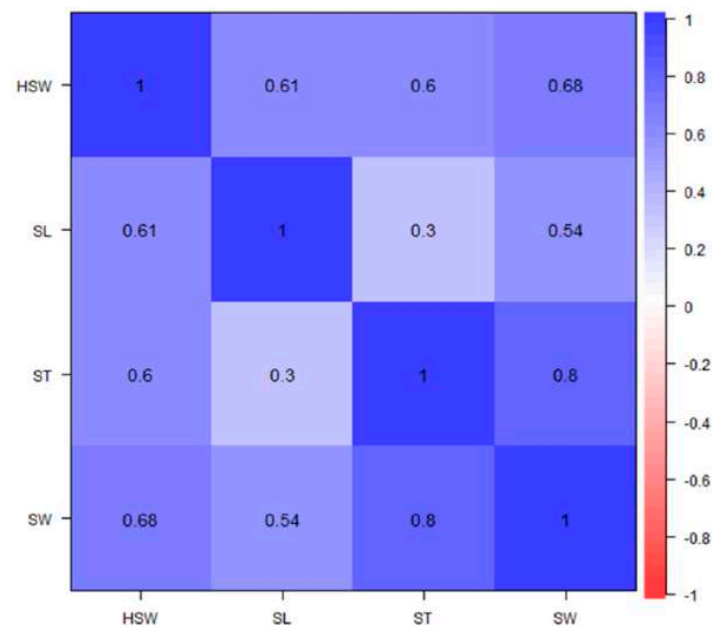
SOV	DF	HSW	SL	SW	ST
Accession	195	52.18 **	1.26 ***	0.55 **	0.88 ***
Location	2	5863.2 ***	117.51 ***	126.96 ***	148.95 ***
Year	1	78.6	23.45 ***	22.8 ***	0.36
Location $\times$ Year	2	536.66 **	19.46 ***	4.32 **	9.39 ***
Accession $\times$ Location	390	29.7 **	0.49 ***	0.32 **	0.45 **
Accession $\times$ Year	195	28.1 **	0.46 **	0.25	0.31
Accession $\times$ Location $\times$ Year	390	20.74 ***	0.33 ***	0.23 ***	0.33 ***
Replication (Location $\times$ Year)	12	66.67 ***	0.54 **	0.37 **	0.52 **
Block (Replication $\times$ Location $\times$ Year)	234	17.14 *	0.22	0.14	0.2
Error	2107	14.1	0.22	0.14	0.21
Heritability		0.52	0.66	0.5	0.58

\*, \*\*, \*\*\*, significant at 0.05, 0.01, and 0.001 levels of probability, respectively: HSW, 100-seed weight; SL, seed length; SW, seed width; ST, seed thickness.

The distribution of the seed size traits evaluated in the six environments and the correlation coefficients between the traits are presented in Figures 1 and 2. The relationships among the traits were positive and significant, and the distribution of traits was near-normal.



**Figure 1.** Distribution of seed size traits of 196 accessions of African yam bean evaluated during the 2018 and 2019 cropping seasons at three locations in Nigeria: (A) hundred seeds weight (g), (B) seed length (mm), (C) seed thickness (mm), and (D) seed width (mm).



**Figure 2.** Phenotypic correlation coefficient of seed size traits of 196 accessions of African yam bean evaluated during the 2018 and 2019 cropping seasons at three locations in Nigeria: HSW, 100-seed weight; SL, seed length; SW, seed width; ST, seed thickness.

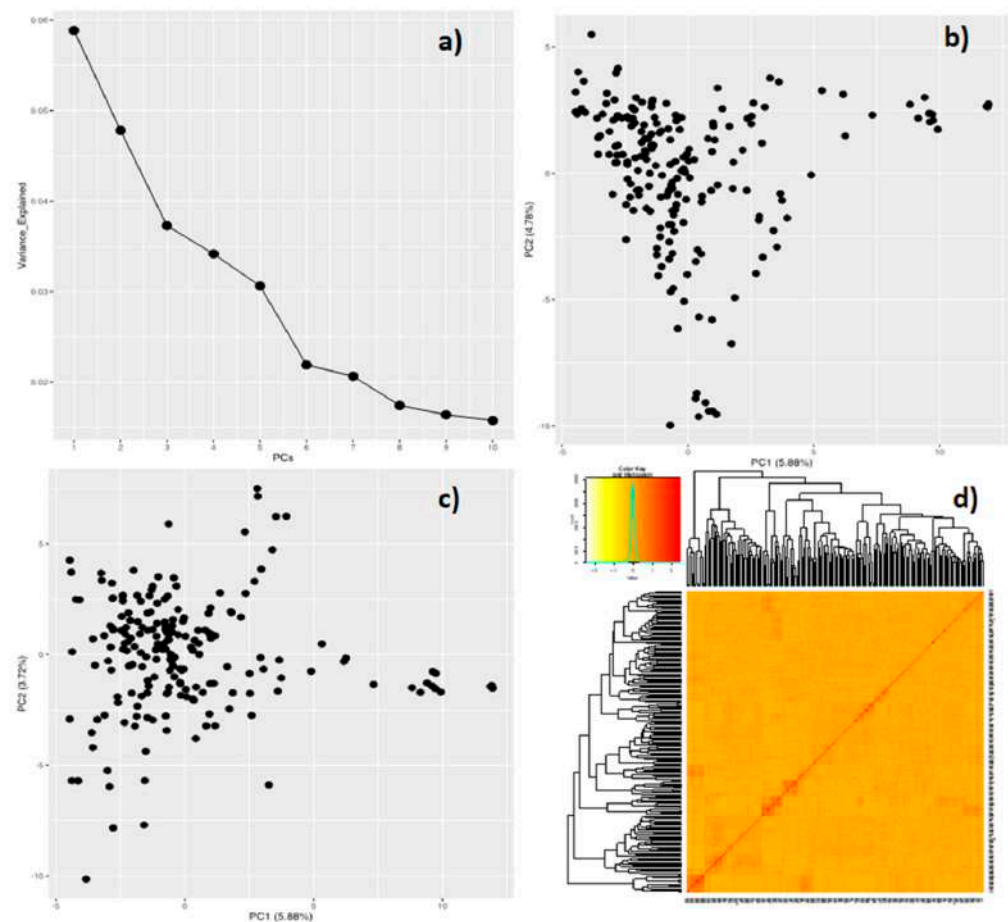


### 3.2. Genotyping and SNP Filtering

A diverse set of 196 AYB accessions were genotyped using a high-depth DArT-seq SNP approach and generated a total of 5416 SNPs. After SNP quality control (call rate  $\geq 0.70$ , marker reproducibility  $\geq 0.95$ , MAF  $\geq 0.01$ , and missing rate  $\leq 0.20$ ) (Supplementary Figures S1–S3), a total of 2491 SNPs were retained and used for GWAS analysis. The MAF ranged between 0.01 and 0.49, with an average of 0.16 (Supplementary Figure S4). The average heterozygosity of the population was 0.15, while TSs-137 accession (0.006) and TSs-10 (0.41) showed minimum and maximum heterozygosity in the AYB population (Supplementary Figures S5 and S6). The average proportion of missing data (based on genotypes) was 0.012 in the AYB population (Supplementary Figure S7). Out of 2491 AYB SNPs, only 422 showed genome-wide syntenic relationships with the common bean reference genome. In the common bean genome, most SNPs were aligned on chromosome 2 (58 SNPs), followed by chromosome 3 (51 SNPs), and the least on chromosome 10 (13 SNPs) (Supplementary Figure S8).

### 3.3. Association Analysis

Principal component analysis revealed that the first three principal components (PCs) respectively accounted for 5.9, 4.8, and 3.7% of the variation among the AYB accessions. No clear clustering (population structure) could be deduced among the accessions based on these two PCs (Figure 3). The coefficient of relatedness in the pairwise kinship matrix ranged between  $-0.33$  and  $2.52$ . The kinship heatmap plot was developed to visualize the relatedness within the population, which indicated low relatedness between accessions (Figure 3).



**Figure 3.** Principal component analysis (a–c) of AYB population and heatmap plot of kinship matrix (d) of 195 African yam bean accessions genotyped with the 2491 DArTseq SNPs.

A total of 58 marker-trait associations were detected for the four seed size traits at a threshold of  $p$ -values of  $\geq 10^{-4}$  in each and combined locations analyses (Table 2, Supplementary Table S1). The combined test environments analysis revealed that 12 significant SNP markers were associated with all four seed size traits, seven of which were codetected by the two statistical methods (GLM and MLM) used. The 12 SNPs explained 6.5–10.8% of the phenotypic variation. In the GLM, six SNPs were associated with 100-seed weight, four with seed thickness, three with seed width, and one with seed length. One SNP marker each was associated with 100-seed weight and seed length, three with seed width, and four with seed thickness in the MLM. Two markers (29420334|F|0-52:C>G-52:C>G and 29420736|F|0-57:G>T-57:G>T) with pleiotropic effects were both associated with seed width and seed thickness. Six out of the twelve markers were significant in at least one of the location and combined location analyses for the same traits (Table 2). Manhattan and Q–Q plots of the SNP-based associations mapping for the four traits based on GLM and MLM are presented in Figures 4 and 5, respectively. The observed  $p$ -values for all traits aligned with expected  $p$ -values, as shown by the Q–Q plots.

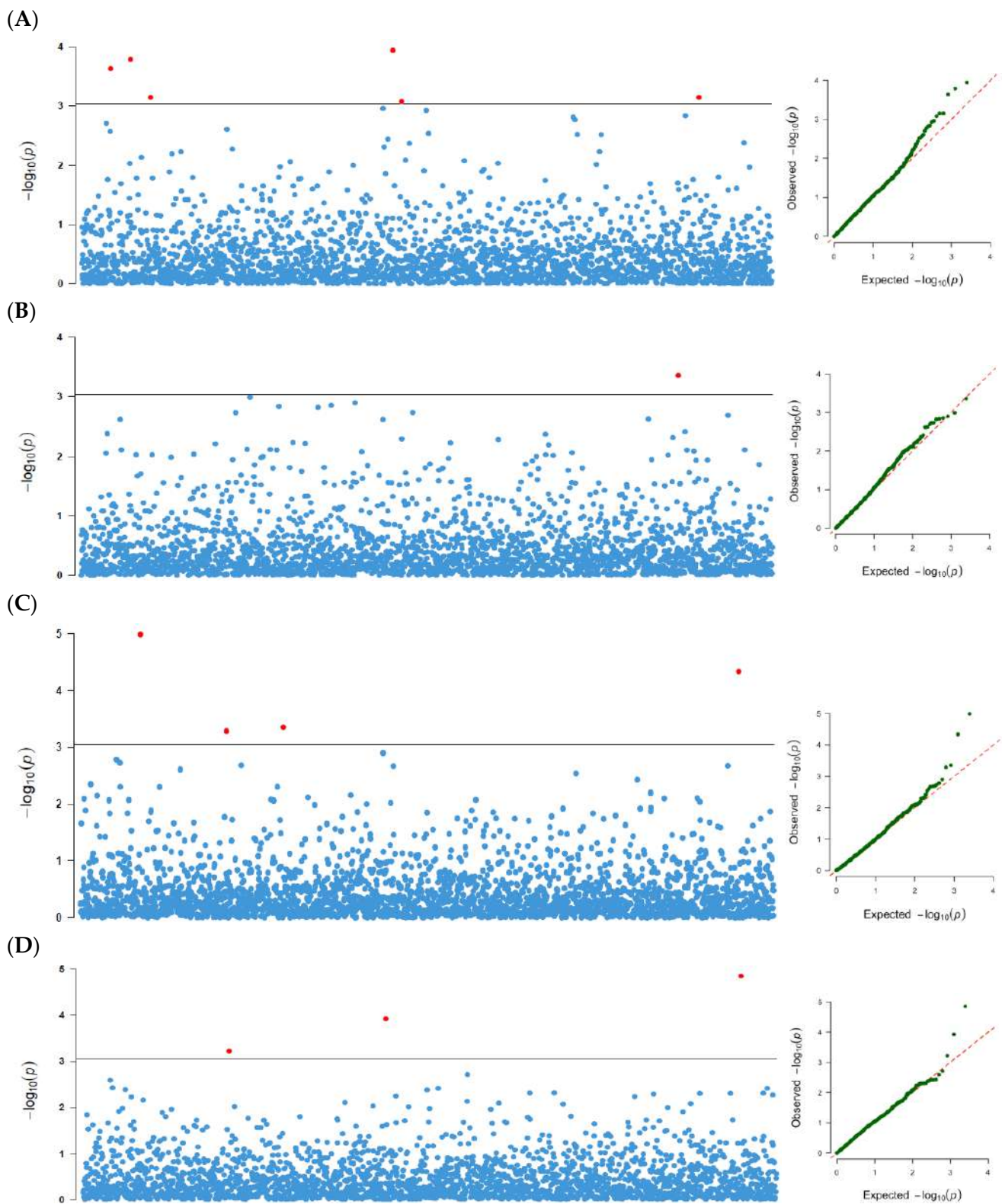
**Table 2.** DArTseq SNPs markers with significant associations with seed size traits of 195 accessions of African yam bean evaluated during the 2018 and 2019 cropping seasons at three locations in Nigeria (Ibadan, Kano, and Ubiaja).

SN	Trait	Marker	Positions	GLM		MLM		Significant in Individual Locations
				$p$ -Value	Marker R <sup>2</sup>	$p$ -Value	Marker R <sup>2</sup>	
1	HSW	29421549 F 0-25:A>C-25:A>C	1124	$1.14 \times 10^{-4}$	0.08362	$8.76 \times 10^{-4}$	0.0754	KANO, UBIAJA
2	HSW	100026424 F 0-37:C>T-37:C>T	177	$1.62 \times 10^{-4}$	0.06499			UBIAJA
3	HSW	100008851 F 0-26:T>C-26:T>C	106	$2.31 \times 10^{-4}$	0.07655			
4	HSW	100041049 F 0-42:G>C-42:G>C	249	$7.12 \times 10^{-4}$	0.06635			
5	HSW	100026423 F 0-13:A>T-13:A>T	2228	$7.12 \times 10^{-4}$	0.06635			
6	HSW	29421658 F 0-5:G>A-5:G>A	1156	$8.34 \times 10^{-4}$	0.06495			
7	SL	29420365 F 0-55:C>G-55:C>G	2152	$4.35 \times 10^{-4}$	0.07298	$8.84 \times 10^{-4}$	0.0769	
8	ST	100034480 F 0-31:C>A-31:C>A	216	$1.02 \times 10^{-5}$	0.09925	$3.42 \times 10^{-5}$	0.1036	UBIAJA
9	ST	29420736 F 0-57:G>T-57:G>T	2364	$4.59 \times 10^{-5}$	0.08693	$9.84 \times 10^{-5}$	0.0925	IBADAN, KANO
10	ST	29420680 F 0-49:T>G-49:T>G	729	$4.41 \times 10^{-4}$	0.06803	$7.61 \times 10^{-4}$	0.0712	
11	ST	29420334 F 0-52:C>G-52:C>G	525	$5.16 \times 10^{-4}$	0.06702	$7.09 \times 10^{-4}$	0.0719	
12	SW	29420736 F 0-57:G>T-57:G>T	2364	$1.40 \times 10^{-5}$	0.10405	$3.63 \times 10^{-5}$	0.1084	KANO
13	SW	29421428 F 0-9:C>A-9:C>A	1088	$1.19 \times 10^{-4}$	0.08505	$2.08 \times 10^{-4}$	0.0891	KANO
14	SW	29420334 F 0-52:C>G-52:C>G	525	$5.98 \times 10^{-4}$	0.0708	$7.94 \times 10^{-4}$	0.0746	

HSW, hundred seeds weight; SL, seed length; SW, seed width; ST, seed thickness.

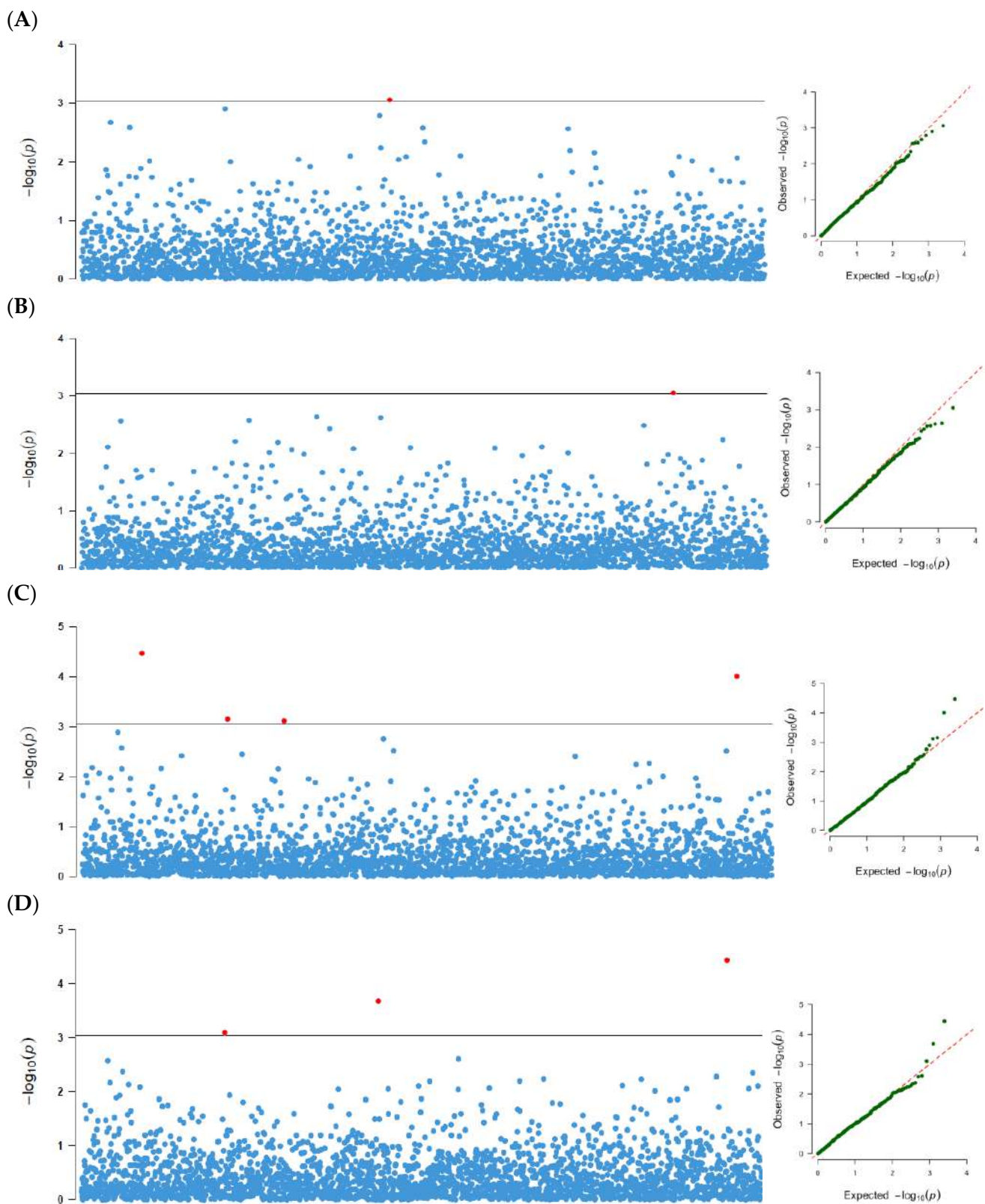
Evaluations in Ibadan over the 2 years revealed significant associations among 30 SNP markers with the four seed size traits (Supplementary Table S1). Seventeen of these SNPs were codetected by the two statistical methods. The markers explained 5.9 to 10.6% of the observed variance. Two markers with pleiotropic effects, 29421951|F|0-37:A>C-37:A>C (associated with 100-seed weight and seed width) and 100006540|F|0-20:T>G-20:T>G (associated with seed thickness and seed width), were found. In Kano, 14 markers displayed significant associations with all four seed size traits, 11 of which were codetected by both statistical methods. The variance explained by these markers ranged from 7.1 to 9.7%. Three markers, 29422320|F|0-37:T>C-37:T>C (associated with 100-seed weight and seed length), 29421428|F|0-9:C>A-9:C>A (associated with seed length and seed width) and 29420736|F|0-57:G>T-57:G>T (associated with seed thickness and seed width), were found to have pleiotropic effects. In Ubiaja, 12 markers, half of which were codetected by both statistical methods, were significantly associated with all traits. The contribution of all the markers to the phenotypic variation ranged from 5.0 and 10.5%. Evaluations in Ubiaja did not reveal any marker with a pleiotropic effect. Two marker overlaps were found: 29421549|F|0-25:A>C-25:A>C (for HSW in Kano and Ubiaja) and 29420736|F|0-57:G>T-

57:G>T (for ST in Ibadan and Kano). Chromosome positions were not given because AYB is yet to have a reference genome (Supplementary Table S1).



**Figure 4.** The Manhattan plots and Q–Q plots of the DArTSeq SNP-based associations mapping in the combined test environments analysis for: (A) hundred seeds weight (HSW), (B) seed length (SL), (C) seed thickness (ST), and (D) seed width (SW) using the GLM approach.

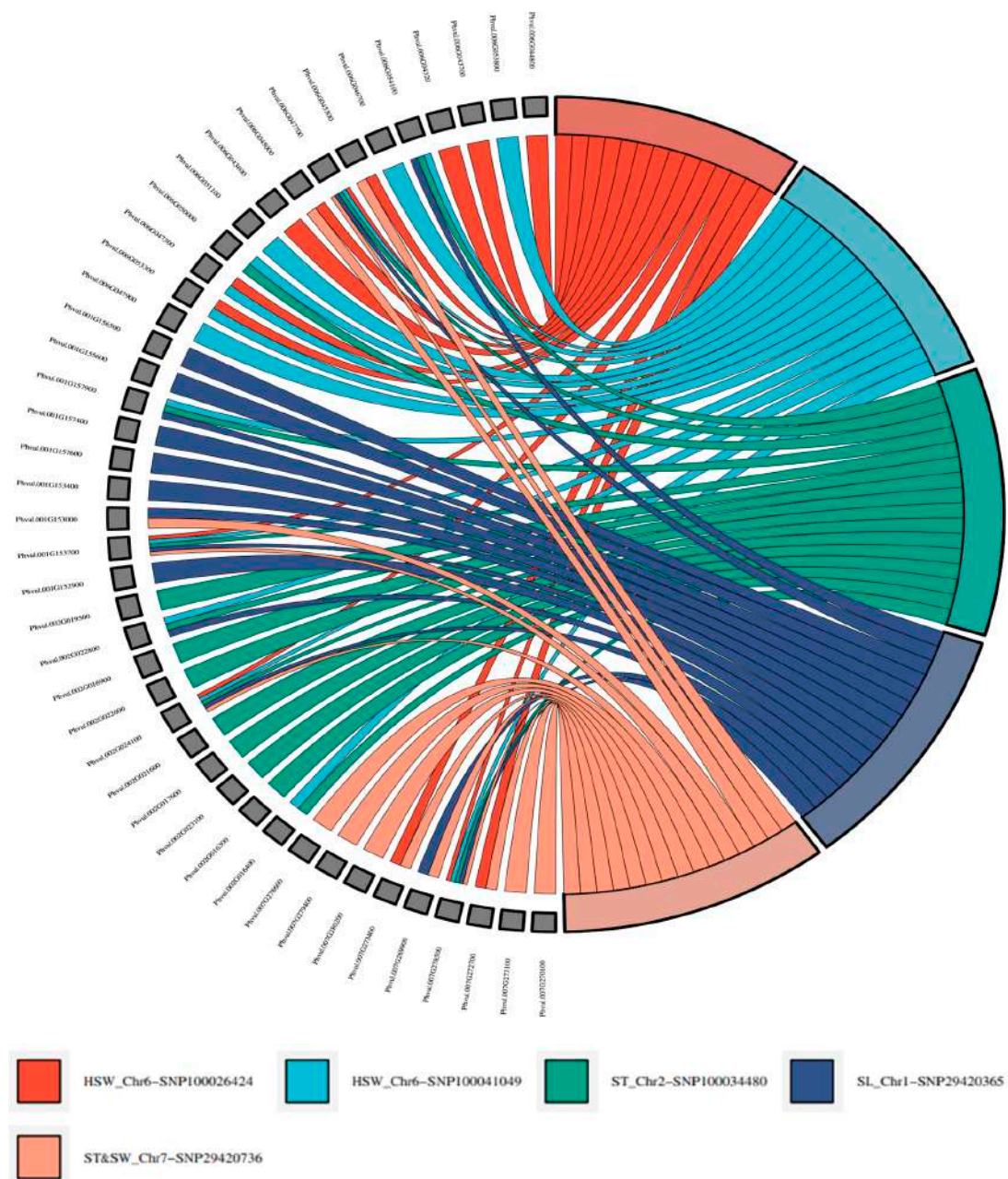




**Figure 5.** The Manhattan plots and Q-Q plots of the DArTSeq SNP-based associations mapping in the combined test environments analysis for: (A) hundred seeds weight (HSW), (B) seed length (SL), (C) seed thickness (ST), and (D) seed width (SW) using the MLM approach.

Candidate gene analysis was performed by blasting trimmed nucleotide sequences (60–80 bps) of significant AYB SNPs in the combined location analysis on the *Phaseolus*

*vulgaris* genome database (*Phaseolus vulgaris* G19833 genome v2.0) in the legume information system. Forty-three candidate genes were identified. These genes were located near (less than 500 kbp) five SNP markers (100026424|F|0-37:C>T-37:C>T, 100041049|F|0-42:G>C-42:G>C, 100034480|F|0-31:C>A-31:C>A, 29420365|F|0-55:C>G-55:C>G, and 29420736|F|0-57:G>T-57:G>T) associated with the four AYB seed traits (Figure 6, Table 3). The five SNP markers were located on chromosomes 1, 2, 6, and 7 of *Phaseolus vulgaris*. The 43 candidate genes have 34 encoding protein products, with some having similar encoding protein products (Table 3). The encoding gene products are known to regulate seed development (UDP-glycosyltransferase superfamily protein, RING-H2 finger protein 2B), seed/fruit size (cytochrome P450 superfamily protein, pentatricopeptide repeat (PPR) superfamily protein; ovate family protein 13), seed weight (cyclin-dependent kinase inhibitor family protein; ATP-binding ABC transporter), seed length ( $\beta$ -carotene isomerase D27), and grain shape (serine/threonine protein phosphatase 2A) in field crops.



**Figure 6.** Cord diagram 43 of candidate genes (based on their encoded protein products) associated with five SNP markers associated with seed size traits in the AYB population.

**Table 3.** Significant markers whose nucleotide sequences were found on the *Phaseolus vulgaris* genome and the encoding protein of genes found close to them.

SN	Trait	Marker	Position	Gene ID	Crop	Chromosome	Encoding Product	Role	References
1	HSW	100026424 F 0-37:C>T-37:C>T	Pv06:15,126,920..15,126,980	Phvul.006G04320	<i>P. vulgaris</i>	6	Cinnamoyl-CoA reductase	Lignin biosynthesis in seed coat	[55]
				Phvul.006G043600	<i>P. vulgaris</i>	6	RING/U-box superfamily protein	Regulate seed size	[56–58]
				Phvul.006G043700	<i>P. vulgaris</i>	6	Calcium-dependent protein kinase 33	Seed development	[59,60]
				Phvul.006G044800	<i>P. vulgaris</i>	6	3-hydroxy-3-methylglutaryl-coenzyme A reductase-like protein	Seed development	[61]
				Phvul.006G045000	<i>P. vulgaris</i>	6	RING-H2 finger protein 2B	Seed development	[62]
				Phvul.006G045300	<i>P. vulgaris</i>	6	Myb transcription factor	Regulate seed size	[63,64]
				Phvul.006G047700	<i>P. vulgaris</i>	6	Pentatricopeptide repeat (PPR) superfamily protein	Regulate seed size	[65]
				Phvul.006G047300	<i>P. vulgaris</i>	6	WRKY family transcription factor	Regulate seed size	[65]
2	HSW	100041049 F 0-42:G>C-42:G>C	Pv06:15,580,580..15,580,630	Phvul.006G046700	<i>G. max</i>	6	GDP-L-galactose phosphorylase 1-like	Regulate seed size	[66]
				Phvul.006G047300	<i>P. vulgaris</i>	6	WRKY family transcription factor	Regulate seed size	[13,65]
				Phvul.006G047700	<i>P. vulgaris</i>	6	Pentatricopeptide repeat (PPR) superfamily protein	Regulate seed size	[65]
				Phvul.006G047900	<i>P. vulgaris</i>	6	zinc finger CCCH-type with G patch domain protein	Regulate seed size	[67]
				Phvul.006G050000	<i>P. vulgaris</i>	6	UDP-Glycosyltransferase superfamily protein	Seed development	[68]
				Phvul.006G051100	<i>P. vulgaris</i>	6	Transmembrane protein, putative	Regulate seed size	[63]
				Phvul.006G053300	<i>P. vulgaris</i>	6	WRKY family transcription factor family protein	Regulate seed size	[13,65]
				Phvul.006G053800	<i>P. vulgaris</i>	6	ATP-binding ABC transporter	Regulate seed size/weight	[69]
Phvul.006G054100	<i>P. vulgaris</i>	6	Cytochrome P450 superfamily protein	Regulate seed/fruit size	[70]				

Table 3. Cont.

SN	Trait	Marker	Position	Gene ID	Crop	Chromosome	Encoding Product	Role	References
3	ST	100034480   F   0-31:C>A-31:C>A	Pv02:2,303,744..2,303,815	Phvul.002G016300	<i>P. vulgaris</i>	2	Ubiquitin-conjugating enzyme 3	Regulate seed size	[15,56]
				Phvul.002G016400	<i>P. vulgaris</i>	2	UDP-glucosyltransferase family protein	Regulate grain size	[71]
				Phvul.002G016900	<i>G. max</i>	2	Ethylene-responsive transcription factor 3-like	Mediates seed size and seed weight	[72,73]
				Phvul.002G017600	<i>G. max</i>	2	Transcription factor SPATULA-like	Regulate grain size	[74]
				Phvul.002G019500	<i>P. vulgaris</i>	2	Cyclin-dependent kinase inhibitor family protein	Regulate seed size/weight	[75]
				Phvul.002G021600	<i>G. max</i>	2	Serine/threonine protein kinase TIO-like	Regulate seed size	[76]
				Phvul.002G024100	<i>P. vulgaris</i>	2	Pentatricopeptide repeat (PPR) superfamily protein	Regulate grain size	[65]
				Phvul.002G022600	<i>P. vulgaris</i>	2	Lipase/Acylhydrolase superfamily protein	Seed development	[77]
				Phvul.002G022800	<i>P. vulgaris</i>	2	Cytochrome P450 superfamily protein	Regulate seed/fruit size	[70]
				Phvul.002G023100	<i>P. vulgaris</i>	2	Transducin/WD40 repeat-like superfamily protein	Regulate grain size	[78]
4	SL	29420365   F   0-55:C>G-55:C>G	Pv01:40,806,541..40,806,608	Phvul.001G153000	<i>P. vulgaris</i>	1	Ovate family protein 13	Regulate seed/fruit size	[79,80]
				Phvul.001G152900	<i>P. vulgaris</i>	1	RING/FYVE/PHD zinc finger superfamily protein	Regulate seed size	[81]
				Phvul.001G153400	<i>P. vulgaris</i>	1	Kelch repeat F-box protein	Regulate seed size	[82]
				Phvul.001G153700	<i>P. vulgaris</i>	1	Pentatricopeptide repeat (PPR) superfamily protein	Regulate seed size	[65]
				Phvul.001G155600	<i>G. max</i>	1	$\beta$ -carotene isomerase D27	Regulate seed length	[81]
				Phvul.001G156500	<i>P. vulgaris</i>	1	Auxin response factor 11	Regulate seed size	[12]
				Phvul.001G157400	<i>P. vulgaris</i>	1	E3 ubiquitin-protein ligase DRIP2-like	Regulate seed size	[56,83]
				Phvul.001G157600	<i>G. max</i>	1	Ethylene-responsive transcription factor 12-like	Regulate seed size	[72]
Phvul.001G157900	<i>P. vulgaris</i>	1	Cytochrome P450 superfamily protein	Regulate seed/fruit size	[70]				



Table 3. Cont.

SN	Trait	Marker	Position	Gene ID	Crop	Chromosome	Encoding Product	Role	References
5	ST, SW	29420736 F 0-57:G>T-57:G>T	Pv07:40,040,935..40,041,1	Phvul.007G269900	<i>P. vulgaris</i>	7	Ovate family protein 13	Regulate seed/fruit size	[79,80]
				Phvul.007G270100	<i>P. vulgaris</i>	7	Ubiquitin-conjugating enzyme 20	Regulate seed size	[15,56]
				Phvul.007G272700	<i>P. vulgaris</i>	7	RING-H2 finger protein 2B	Seed development	[62]
				Phvul.007G273100	<i>P. vulgaris</i>	7	Serine/threonine protein phosphatase 2A	Regulate grain shape	[76,84]
				Phvul.007G273400	<i>P. vulgaris</i>	7	Myb transcription factor	Regulate grain size	[63,64]
				Phvul.007G278500	<i>P. vulgaris</i>	7	Pentatricopeptide repeat (PPR) superfamily protein	Regulate seed size	[65]
				Phvul.007G278600	<i>P. vulgaris</i>	7	Argonaute family protein	Regulate seed size	[85]
				Phvul.007G279400	<i>P. vulgaris</i>	7	ARM repeat superfamily protein	Regulate seed size	[86]
				Phvul.007G280200	<i>P. vulgaris</i>	7	ATP-binding/protein serine/threonine kinase	Regulate seed size	[76]

HSW, hundred seeds weight; SL, seed length; SW, seed width; ST, seed thickness.



#### 4. Discussion

The significant differences observed among the accessions for the seed size traits revealed the existence of adequate genetic variability among them. Seeds of AYB are known to harbor vast genetic variability in color, shape, and size [4,19,20,87]. Adewale et al. [4] had earlier suggested the use of six seed characters (seed length, width, and thickness and their ratios) as unique indices for discriminating among AYB accessions. The significant accession  $\times$  location  $\times$  year effects for all traits indicated the distinctiveness of the environments in discriminating among the accessions. Moderate to high heritability estimates observed for all traits implied increased power of SNP detection in the accession, hence, identification of true associations between a marker and putative gene [53]. The significant positive correlations between pairs of traits suggest the possibility of simultaneous improvements in the traits to enhance seed yield.

At the genetic level, differences in the extent of relatedness among individuals in a population used for association mapping can lead to the formation of a population structure that can cause spurious associations between genotypes and the traits of interest [88]. Principal component analysis is a widely used multivariate statistical approach proposed by Price et al. [89] that can calculate population relatedness and count groups in a population in order to reduce dimensional genotype data and control population structure (by selecting the first few PCs that present most of the total variation among individuals based on their SNP data). In this study, however, population structure analysis using the PCA approach showed barely noticeable differentiation among accessions. This result of subtle population structure was also confirmed by the multidimensional scaling (MDS) approach using Tassel software. The first three principal components of PCA had a high correlation with the first three principal components of MDS (data not shown). This might have resulted from the fact that most of the accessions, especially the large proportion whose origins are unknown (102 accessions), are possibly from Nigeria. Association analyses between specific phenotypes and genotypes within a genome are an important step toward the discovery of genes controlling the traits [53,90]. Of the 195 accessions used in this study, 137 have been previously utilized for conducting genome-wide association studies for nutritional traits, and several significant SNPs were found to be associated with the studied traits [41]. In this study, model fitness for the GWAS was confirmed by the Q-Q plots. The alignment of observed and expected  $p$ -values in the Q-Q plots for all the measured traits indicated that spurious associations as a result of population structure and familial relatedness were largely corrected. Using two alternative GWAS models, we found that the MLM model had a stricter decrease in the number of significant markers than the GLM model. This is because GLM is considered a naive model with a high rate of false positives because it does not take population structure into account [91], whereas MLM takes population structure into account and avoids spurious associations [92,93]. Though subtle population structure was found in this study, both models were reported because this was the first attempt at dissecting the genetic basis of seed size traits in AYB using GWAS. Significant QTLs associated with agronomic traits have also been reported in the absence of population structure using GLM and MLM models of GWAS in rice and faba bean crops [92,94]. The contribution of all the significant markers to the phenotypic variation, which ranged from 5.0 to 10.4%, suggested that the markers could be useful for marker-assisted selection in AYB improvement. However, due to the nonexistence of a reference genome for AYB, the exact locations of the markers on the chromosomes remain unknown. The Alliance for Accelerated Crop Improvement in Africa (ACACIA) is currently undertaking the whole genome sequencing of AYB [95]. Markers with pleiotropic effects could be useful in the simultaneous improvement of the correlated traits. The six significant markers that were consistent in the one location and the combined location analyses for the same traits in this study could be considered putative makers.

In this study, the common bean genome was used to align AYB SNPs, and 17% of the filtered 2491 SNPs aligned widely on common bean chr01 to chr11, including six SNPs on scaffolds. These suggest that the two crops have a syntenic relationship, which might

be due to their close evolutionary relationship. Similar results have been reported for another African indigenous legume (Bambara groundnut) having syntenic relationships with other legumes (common bean, adzuki bean, and mung bean) [96,97]. The five significant markers found in *Phaseolus vulgaris* genome at a location close to genes whose encoding proteins had been reported in other crops to regulate the same traits as those with which they are associated in AYB can also be considered candidate makers. For example, markers (29420365|F|0-55:C>G-55:C>G and 29420736|F|0-57:G>T-57:G>T) associated with seed length, seed width, and seed thickness were found located close to genes (Phvul.001G153000 and Phvul.007G269900) having ovate family protein 13 as their encoding protein. Ovate family protein 13 is known to regulate fruit shape and seed size in tomato and rice, respectively [80]. Likewise, the pentatricopeptide repeat (PPR) superfamily protein, the encoding protein of genes (Phvul.006G047700, Phvul.006G047700, Phvul.002G024100, Phvul.001G153700, and Phvul.007G278500) located close to markers (100026424|F|0-37:C>T-37:C>T, 100041049|F|0-42:G>C-42:G>C, 100034480|F|0-31:C>A-31:C>A, 29420365|F|0-55:C>G-55:C>G, and 29420736|F|0-57:G>T-57:G>T) associated with the four seed size and shape traits of AYB is known to regulate seed physical traits by participating in RNA intron splicing during seed development [66]. Though significant marker-trait associations were detected in this study, the result serves as a foundation for the genetic understanding of putative makers underlying seed size traits in AYB. The identified significant makers could be targeted by plant breeders in marker-assisted selection to accelerate the genetic improvement of AYB.

## 5. Conclusions

To the best of our knowledge, this study is the first attempt at dissecting the genetic basis of seed size traits in AYB using genome-wide association mapping. Several SNPs were significantly associated with seed size traits in AYB. The five significant SNP markers (100026424|F|0-37:C>T-37:C>T, 100041049|F|0-42:G>C-42:G>C, 100034480|F|0-31:C>A-31:C>A, 29420365|F|0-55:C>G-55:C>G, and 29420736|F|0-57:G>T-57:G>T) found on the *Phaseolus vulgaris* genome should be regarded as candidate markers. It is, therefore, recommended that efforts should be directed toward the validation of the identified significant makers using several mapping populations before they can be targeted for use in marker-assisted selection for seed size traits in AYB.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13122350/s1>, Figure S1: call rate distribution in the AYB population, Figure S2: marker reproducibility distribution in the AYB population, Figure S3: proportion of missing data (SNP markers) distribution in the AYB population, Figure S4: minor allele frequency distribution in the AYB population, Figure S5: proportion of heterozygous (SNP markers) distribution in the AYB population, Figure S6: proportion of heterozygous (based on genotypes) distribution in the AYB population, Figure S7: proportion of missing data (based on genotypes) distribution in the AYB population, Figure S8: African yam bean SNPs aligned on different chromosomes of common bean genome, Table S1: DArTseq SNP markers having significant association with seed size traits of 195 accessions of African yam bean evaluated during the 2018 and 2019 cropping seasons in Ibadan, Kano, and Ubiaja.

**Author Contributions:** Conceptualization, M.T.A., O.E.O. and R.P.; funding acquisition, M.T.A.; investigation, O.E.O.; methodology, O.E.O., A.A., R.P., O.O.O. and O.A.O.; data curation, O.E.O. and O.O.O.; formal analysis, O.E.O. and R.P.; project administration, M.T.A. and O.A.O.; supervision, M.T.A., A.A., O.A.O. and R.P.; writing—original draft, O.E.O.; writing—review and editing, A.A. and R.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is funded by the Crop Trust through the Genetic Resource Center, IITA, Ibadan, Nigeria.

**Acknowledgments:** The authors express gratitude to members of staff of the seed bank of the Genetic Resource Center, IITA, Ibadan, the cassava breeding unit, IITA, Ubiaya, the cowpea breeding unit, IITA, Kano, and the Bioscience Center, IITA, Ibadan, for their field and laboratory assistance.

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

## References

- Grela, E.R.; Samolińska, W.; Kiczorowska, B.; Klebaniuk, R.; Kiczorowski, P. Content of minerals and fatty acids and their correlation with phytochemical compounds and antioxidant activity of leguminous seeds. *Biol. Trace Elem. Res.* **2017**, *180*, 338–348. [[CrossRef](#)]
- Aremu, C.; Ige, S.A.; Ibirinde, D.; Raji, I.; Abolusoro, S.; Ajiboye, B.; Obaniyi, S.; Adekiya, A.; Asaley, A. Assessing yield stability in African yam bean (*Sphenostylis stenocarpa*) performance using year effect. *Open Agric.* **2020**, *5*, 202–212. [[CrossRef](#)]
- Potter, D. Economic botany of *Sphenostylis* (Leguminosae). *Econ. Bot.* **1992**, *46*, 262–275. [[CrossRef](#)]
- Adewale, B.D.; Kehinde, O.B.; Aremu, C.O.; Popoola, J.O.; Dumet, D.J. Seed metrics for genetic and shape determination in African yam bean. *Afr. J. Plant Sci.* **2010**, *4*, 107–115.
- Nnamani, C.V.; Ajayi, S.A.; Oselebe, H.O.; Atkinson, C.J.; Igboabuchi, A.N.; Ezigbo, E.C. *Sphenostylis stenocarpa* (ex. A. Rich.) Harms., a fading genetic resource in a changing climate: Crerequisite for conservation and sustainability. *Plants* **2017**, *6*, 30. [[CrossRef](#)]
- Ojuederie, O.B.; Ajiboye, J.A.; Babalola, O.O. Biochemical and histopathological studies of key tissues in healthy male wistar rats fed on African yam bean seed and tuber Meals. *J. Food Qual.* **2020**, *2020*, 8892618. [[CrossRef](#)]
- Baiyeri, S.O.; Uguru, M.I.; Ogbonna, P.E.; Samuel-Baiyeri, C.C.A.; Okechukwu, R.; Kumaga, F.K.; Amoatey, C. Evaluation of the nutritional composition of the seeds of some selected African yam bean (*Sphenostylis stenocarpa* Hochst Ex. A. Rich (Harms)) accessions. *Agro-Sci.* **2018**, *17*, 37–44. [[CrossRef](#)]
- Saka, J.O.; Ajibade, S.R.; Adeniyani, O.N.; Olowoyo, R.B.; Ogunbodede, B.A. Survey of underutilized grain legume production systems in the Southwest Agricultural zone of Nigeria. *J. Agric. Food Inf.* **2004**, *6*, 93–107. [[CrossRef](#)]
- Olomitutu, O.E.; Abe, A.; Oyatomi, O.A.; Paliwal, R.; Abberton, M.T. Assessing intraspecific variability and diversity in African yam bean landraces using agronomic traits. *Agronomy* **2022**, *12*, 884. [[CrossRef](#)]
- Edematie, V.E.; Fatokun, C.; Boukar, O.; Adetimirin, V.O.; Kumar, P. Inheritance of pod length and other yield components in two cowpea and tard-long bean crosses. *Agronomy* **2021**, *11*, 682. [[CrossRef](#)]
- Lo, S.; Muñoz-Amatriáin, M.; Hokin, S.A.; Cisse, N.; Roberts, P.A.; Farmer, A.D.; Xu, S.; Close, T.J. A genome-wide association and meta-analysis reveal regions associated with seed size in cowpea [*Vigna unguiculata* (L.) Walp]. *Theor. Appl. Genet.* **2019**, *132*, 3079–3087. [[CrossRef](#)]
- Li, N.; Li, Y. Maternal control of seed size in plants. *J. Exp. Bot.* **2015**, *66*, 1087–1097. [[CrossRef](#)] [[PubMed](#)]
- Gu, Y.; Li, W.; Jiang, H.; Wang, Y.; Gao, H.; Liu, M.; Chen, Q.; Lai, Y.; He, C. Differential expression of a WRKY gene between wild and cultivated soybeans correlates to seed size. *J. Exp. Bot.* **2017**, *68*, 2717–2729. [[CrossRef](#)]
- Smitchger, J.; Weeden, N.F. The ideotype for seed size: A model examining the relationship between seed size and actual yield in pea. *Int. J. Agron.* **2018**, *2018*, 9658707. [[CrossRef](#)]
- Guo, Y.; Gao, M.; Liang, X.; Xu, M.; Liu, X.; Zhang, Y.; Liu, X.; Liu, J.; Gao, Y.; Qu, S.; et al. Quantitative trait loci for seed size variation in cucurbits—A review. *Front. Plant Sci.* **2020**, *11*, 304. [[CrossRef](#)] [[PubMed](#)]
- Chen, N.L.; Gao, Y.M.; Zhang, Y.X. Effects of seed size and light intensity on seedling growth of watermelon. *China Cucurbits Vegetables.* **2006**, 1–3. [[CrossRef](#)]
- Moles, A.T.; Ackerly, D.D.; Webb, C.O.; Tweddle, J.C.; Dickie, J.B.; Westoby, M. A brief history of seed size. *Science* **2005**, *307*, 576–580. [[CrossRef](#)]
- Wu, G.; Du, G.; Shang, Z. Contribution of seed size and its fate to vegetation renewal: A review. *Ying Yong Sheng Tai Xue Bao = J. Appl. Ecol.* **2006**, *17*, 1969–1972.
- Oshodi, A.A.; Ipinmroti, K.O.; Adeyeye, E.I.; Hall, G.M. In vitro multienzyme digestibility of protein of six varieties of African yam bean flours. *J. Sci. Food Agric.* **1995**, *69*, 373–377. [[CrossRef](#)]
- Adewale, B.D.; Dumet, D.J.; Vroh-Bi, I.; Kehinde, O.B.; Ojo, D.K.; Adegbite, A.E.; Franco, J. Morphological diversity analysis of African yam bean (*Sphenostylis stenocarpa* Hochst. ex A. Rich.) Harms and prospects for utilization in germplasm conservation and breeding. *Genet. Resour. Crop Evol.* **2012**, *59*, 927–936. [[CrossRef](#)]
- Corder, E.H.; Saunders, A.M.; Risch, N.J.; Strittmatter, W.J.; Schmechel, D.E.; Gaskell, P.C.; Rimmler, J.B.; Locke, P.A.; Conneally, P.M.; Schmechel, K.E.; et al. Protective effect of apolipoprotein-E type-2 allele for late-onset Alzheimer disease. *Nat. Genet.* **1994**, *7*, 180–184. [[CrossRef](#)] [[PubMed](#)]
- Ruggieri, V.; Francese, G.; Sacco, A.; D’Alessandro, A.; Rigano, M.M.; Parisi, M.; Milone, M.; Cardi, T.; Mennella, G.; Barone, A. An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biol.* **2014**, *14*, 337. [[CrossRef](#)] [[PubMed](#)]
- Flint-Garcia, S.A.; Thuillet, A.-C.; Yu, J.; Pressoir, G.; Romero, S.M.; Mitchell, S.; Doebley, J.; Kresovich, S.; Goodman, M.; Buckler, E. Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J. Cell Mol. Biol.* **2005**, *44*, 1054–1064. [[CrossRef](#)]

24. Yu, J.; Buckler, E.S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* **2006**, *17*, 155–160. [[CrossRef](#)] [[PubMed](#)]
25. Semagn, K.; Björnstad, Å.; Xu, Y. The genetic dissection of quantitative traits in crops. *Electron. J. Biotechnol.* **2010**, *13*, 16–17. [[CrossRef](#)]
26. Zhao, K.; Tung, C.W.; Eizenga, G.C.; Wright, M.H.; Ali, M.L.; Price, A.H.; Norton, G.J.; Islam, M.R.; Reynolds, A.; Mezey, J.; et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2011**, *2*, 467. [[CrossRef](#)] [[PubMed](#)]
27. Larsson, S.J.; Lipka, A.E.; Buckler, E.S. Lessons from *Dwarf8* on the strengths and weaknesses of structured association mapping. *PLoS Genet.* **2013**, *9*, e1003246. [[CrossRef](#)] [[PubMed](#)]
28. Gali, K.K.; Sackville, A.; Tafesse, E.G.; Lachagari, V.B.; McPhee, K.; Hybl, M.; Mikić, A.; Smýkal, P.; McGee, R.; Burstin, J.; et al. Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.). *Front. Plant Sci.* **2019**, *10*, 1538. [[CrossRef](#)]
29. Raggi, L.; Caproni, L.; Carboni, A.; Negri, V. Genome-wide association study reveals candidate genes for flowering time variation in common bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* **2019**, *10*, 962. [[CrossRef](#)]
30. Ahmed, S.M.; Alsamman, A.M.; Jighly, A.; Mubarak, M.H.; Al-Shamaa, K.; Istanbuli, T.; Momtaz, O.A.; El Allali, A.; Hamwiah, A. Genome-wide association analysis of chickpea germplasm differing for salinity tolerance based on DArTseq markers. *PLoS ONE* **2021**, *16*, e0260709. [[CrossRef](#)]
31. Chen, Z.; Lancon-Verdier, V.; Le Signor, C.; She, Y.M.; Kang, Y.; Verdier, J. Genome-wide association study identified candidate genes for seed size and seed composition improvement in *M. truncatula*. *Sci. Rep.* **2021**, *11*, 4224. [[CrossRef](#)] [[PubMed](#)]
32. de Almeida, C.P.; Santos, I.L.; de Carvalho Paulino, J.F.; Barbosa, C.C.F.; Pereira, C.C.A.; Carvalho, C.R.L.; Gonçalves, G.D.M.C.; Song, Q.; Carbonell, S.A.M.; Chiorato, A.F.; et al. Genome-wide association mapping reveals new loci associated with light-colored seed coat at harvest and slow darkening in carioca beans. *BMC Plant Biol.* **2021**, *21*, 343. [[CrossRef](#)] [[PubMed](#)]
33. Rajendran, K.; Coyne, C.J.; Zheng, P.; Saha, G.; Main, D.; Amin, N.; Ma, Y.; Kisha, T.; Bett, K.E.; McGee, R.J.; et al. Genetic diversity and GWAS of agronomic traits using an ICARDA lentil (*Lens culinaris* Medik.) Reference Plus collection. *Plant Genet. Resour.* **2021**, *19*, 279–288. [[CrossRef](#)]
34. Soller, M.; Beckmann, J.S. Molecular markers in varietal identification and genetic improvement. *Theor. Appl. Genet.* **1983**, *17*, 25–33. [[CrossRef](#)] [[PubMed](#)]
35. Goddard, M.E.; Kemper, K.E.; MacLeod, I.M.; Chamberlain, A.J.; Hayes, B.J. Genetics of complex traits: Prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. R. Soc. B Biol. Sci.* **2016**, *283*, 20160569. [[CrossRef](#)] [[PubMed](#)]
36. Varshney, R.K.; Mohan, S.M.; Gaur, P.M.; Chamarthi, S.K.; Singh, V.K.; Srinivasan, S.; Swapna, N.; Sharma, M.; Pande, S.; Singh, S.; et al. Marker-assisted backcrossing to introgress resistance to Fusarium wilt race 1 and Ascochyta blight in C 214, an elite cultivar of chickpea. *Plant Genome* **2014**, *7*, plantgenome2013-10. [[CrossRef](#)]
37. Uwera, A.; Rusagara, J.N.; Msolla, S.N.; Musoni, A.; Assefa, T. Molecular marker-assisted backcrossing of anthracnose resistance genes into common beans (*Phaseolus vulgaris* L.) varieties. *Am. J. Plant Sci.* **2021**, *12*, 771–781. [[CrossRef](#)]
38. Li, X.; Zhang, X.; Zhu, L.; Bu, Y.; Wang, X.; Zhang, X.; Zhou, Y.; Wang, X.; Guo, N.; Qiu, L.; et al. Genome-wide association study of four yield-related traits at the R6 stage in soybean. *BMC Genet.* **2019**, *20*, 39. [[CrossRef](#)] [[PubMed](#)]
39. Mir, R.R.; Choudhary, N.; Bawa, V.; Jan, S.; Singh, B.; Bhat, M.A.; Paliwal, R.; Kumar, A.; Chitikineni, A.; Thudi, M.; et al. Allelic Diversity, structural analysis, and Genome-Wide Association Study (GWAS) for yield and related traits using unexplored common bean (*Phaseolus vulgaris* L.) germplasm from western Himalayas. *Front. Genet.* **2021**, *11*, 609603. [[CrossRef](#)]
40. Giordani, W.; Gama, H.C.; Chiorato, A.F.; Garcia, A.A.F.; Vieira, M.L.C. Genome-wide association studies dissect the genetic architecture of seed shape and size in common bean. *G3* **2022**, *12*, jkac048. [[CrossRef](#)] [[PubMed](#)]
41. Oluwole, O.O.; Olomitutu, O.E.; Paliwal, R.; Oyatomi, O.A.; Abberton, M.T.; Obembe, O.O. Preliminary assessment of the association between DArT-SEQ SNP and some nutritional traits in African yam bean. *Trop. J. Nat. Prod. Res.* **2020**, *4*, 877–879.
42. Adewale, D.B.; Dumet, D.J. *Descriptors for African Yam Bean, Sphenostylis stenocarpa (Hochst ex. A. Rich.) Harms*; Genetic Resources Center, International Institute of Tropical Agriculture (IITA): Ibadan, Nigeria, 2011; pp. 1–13.
43. Jaccoud, D.; Peng, K.M.; Feinstein, D.; Kilian, A. Diversity Arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* **2001**, *29*, e25. [[CrossRef](#)] [[PubMed](#)]
44. 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). *Front. Genet.* **2017**, *8*, 98. [[CrossRef](#)] [[PubMed](#)]
45. SAS Institute Inc. *Base SAS® 9.4 Procedures Guide: Statistical Procedures*; SAS Institute Inc.: Cary, NC, USA, 2017.
46. Hallauer, A.R.; Carena, M.J.; Miranda Filho, J.D. *Quantitative Genetics in Maize Breeding*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; Volume 6.
47. Robinson, H.F.; Comstock, R.E.; Harvey, P.H. Estimates of heritability and the degree of dominance in corn 1. *Agron. J.* **1949**, *41*, 353–359. [[CrossRef](#)]
48. R Core Team R. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: <https://www.R-project.org> (accessed on 20 January 2019).



49. Alvarado, G.; López, M.; Vargas, M.; Pacheco, Á.; Rodríguez, F.; Burgueño, J.; Crossa, J. META-R (Multi Environment Trial Analysis with R for Windows) Version 6.04. CIMMYT Research Data & Software Repository Network, V23. 2015. Available online: <https://hdl.handle.net/11529/10201> (accessed on 20 January 2019).
50. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]
51. Li, H.; Peng, Z.; Yang, X.; Wang, W.; Fu, J.; Wang, J.; Han, Y.; Chai, Y.; Guo, T.; Yang, N.; et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* **2013**, *45*, 43–50. [[CrossRef](#)]
52. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399. [[CrossRef](#)]
53. Adewale, S.A.; Badu-Apraku, B.; Akinwale, R.O.; Paterne, A.A.; Gedil, M.; Garcia-Oliveira, A.L. Genome-wide association study of Striga resistance in early maturing white tropical maize inbred lines. *BMC Plant Biol.* **2020**, *20*, 203. [[CrossRef](#)]
54. 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*, e0198012. [[CrossRef](#)]
55. Yu, A.; Wang, Z.; Zhang, Y.; Li, F.; Liu, A. Global gene expression of seed coat tissues reveals a potential mechanism of regulating seed size formation in castor bean. *Int. J. Mol. Sci.* **2019**, *20*, 1282. [[CrossRef](#)]
56. Li, N.; Li, Y. Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* **2014**, *5*, 332. [[CrossRef](#)]
57. Hina, A.; Cao, Y.; Song, S.; Li, S.; Sharmin, R.A.; Elattar, M.A.; Bhat, J.A.; Zhao, T. High-resolution mapping in two RIL populations refines major “QTL Hotspot” regions for seed size and shape in soybean (*Glycine max* L.). *Int. J. Mol. Sci.* **2020**, *21*, 1040. [[CrossRef](#)]
58. Brinton, J.; Simmonds, J.; Uauy, C. Ubiquitin-related genes are differentially expressed in isogenic lines contrasting for pericarp cell size and grain weight in hexaploid wheat. *BMC Plant Biol.* **2018**, *18*, 22. [[CrossRef](#)]
59. Ray, S.; Agarwal, P.; Arora, R.; Kapoor, S.; Tyagi, A.K. Expression analysis of calcium-dependent protein kinase gene family during reproductive development and abiotic stress conditions in rice (*Oryza sativa* L. ssp. indica). *Mol. Genet. Genom.* **2007**, *278*, 493–505. [[CrossRef](#)]
60. Wei, S.; Hu, W.; Deng, X.; Zhang, Y.; Liu, X.; Zhao, X.; Luo, Q.; Jin, Z.; Li, Y.; Zhou, S.; et al. A rice calcium-dependent protein kinase OsCPK9 positively regulates drought stress tolerance and spikelet fertility. *BMC Plant Biol.* **2014**, *14*, 133. [[CrossRef](#)] [[PubMed](#)]
61. Moore, K.B.; Oishi, K.K. Characterization of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity during maize seed development, germination, and seedling emergence. *Plant Physiol.* **1993**, *101*, 485–491. [[CrossRef](#)]
62. Kang, Y.J.; Shim, K.C.; Lee, H.S.; Jeon, Y.A.; Kim, S.H.; Kang, J.W.; Yun, Y.T.; Park, I.K.; Ahn, S.N. Fine mapping and candidate gene analysis of the quantitative trait locus gw8.1 associated with grain length in rice. *Genes Genom.* **2018**, *40*, 389–397. [[CrossRef](#)] [[PubMed](#)]
63. Zhang, W.; Sun, P.; He, Q.; Shu, F.; Wang, J.; Deng, H. Fine mapping of GS2, a dominant gene for big grain rice. *Crop J.* **2013**, *1*, 160–165. [[CrossRef](#)]
64. Watt, C.; Zhou, G.; Li, C. Harnessing transcription factors as potential tools to enhance grain size under stressful abiotic conditions in cereal crops. *Front. Plant Sci.* **2020**, *11*, 1273. [[CrossRef](#)]
65. Li, N.; Xu, R.; Li, Y. Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* **2019**, *70*, 435–463. [[CrossRef](#)] [[PubMed](#)]
66. Ma, J.; Wang, L.; Cao, Y.; Wang, H.; Li, H. Association mapping and transcriptome analysis reveal the genetic architecture of maize kernel size. *Front. Plant Sci.* **2021**, *12*, 273. [[CrossRef](#)] [[PubMed](#)]
67. Radkova, M.; Revalska, M.; Kertikova, D.; Iantcheva, A. Zinc finger CCHC-type protein related with seed size in model legume species *Medicago truncatula*. *Biotechnol. Biotechnol. Equip.* **2019**, *33*, 278–285. [[CrossRef](#)]
68. Rehman, H.M.; Nawaz, M.A.; Bao, L.; Shah, Z.H.; Lee, J.M.; Ahmad, M.Q.; Chung, G.; Yang, S.H. Genome-wide analysis of Family-1 UDP-glycosyltransferases in soybean confirms their abundance and varied expression during seed development. *J. Plant Physiol.* **2016**, *206*, 87–97. [[CrossRef](#)] [[PubMed](#)]
69. Basu, U.; Upadhyaya, H.D.; Srivastava, R.; Daware, A.; Malik, N.; Sharma, A.; Bajaj, D.; Narnoliya, L.; Thakro, V.; Kujur, A.; et al. ABC transporter-mediated transport of glutathione conjugates enhances seed yield and quality in chickpea. *Plant Physiol.* **2019**, *180*, 253–275. [[CrossRef](#)]
70. Qi, X.; Liu, C.; Song, L.; Li, Y.; Li, M. PaCYP78A9, a cytochrome P450, regulates fruit size in sweet cherry (*Prunus avium* L.). *Front. Plant Sci.* **2017**, *8*, 2076. [[CrossRef](#)] [[PubMed](#)]
71. Dong, N.Q.; Sun, Y.; Guo, T.; Shi, C.L.; Zhang, Y.M.; Kan, Y.; Xiang, Y.H.; Zhang, H.; Yang, Y.B.; Li, Y.C.; et al. UDP-glycosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nat. Commun.* **2020**, *11*, 2629. [[CrossRef](#)] [[PubMed](#)]
72. Jiang, W.; Zhang, X.; Song, X.; Yang, J.; Pang, Y. Genome-Wide Identification and Characterization of APETALA2/Ethylene-Responsive element binding factor superfamily genes in soybean seed development. *Front. Plant Sci.* **2020**, *11*, 1348. [[CrossRef](#)]
73. Sharma, P.; Singh, R.; Sehrawat, N. A critical review on: Significance of floral homeotic APETALA2 gene in plant system. *J. Appl. Pharm. Sci.* **2020**, *10*, 124–130.
74. Liu, B.; Hua, C.; Song, G.; Wu, M.; Cui, R.; Zhang, A.; Liu, Y.; Huang, L.; Yan, A.; Ali, I.; et al. The SPATULA transcription factor regulates seed oil content by controlling seed specific genes in *Arabidopsis thaliana*. *Plant Growth Regul.* **2017**, *82*, 111–121. [[CrossRef](#)]



75. Ajadi, A.A.; Tong, X.; Wang, H.; Zhao, J.; Tang, L.; Li, Z.; Liu, X.; Shu, Y.; Li, S.; Wang, S.; et al. Cyclin-Dependent kinase inhibitors KRP1 and KRP2 are involved in grain filling and seed germination in rice (*Oryza sativa* L.). *Int. J. Mol. Sci.* **2020**, *21*, 245. [[CrossRef](#)] [[PubMed](#)]
76. Hu, Z.; He, H.; Zhang, S.; Sun, F.; Xin, X.; Wang, W.; Qian, X.; Yang, J.; Luo, X. A Kelch motif-containing serine/threonine protein phosphatase determines the large grain QTL trait in rice. *J. Integr. Plant Biol.* **2012**, *54*, 979–990. [[CrossRef](#)] [[PubMed](#)]
77. Ma, R.; Yuan, H.; An, J.; Hao, X.; Li, H. A *Gossypium hirsutum* GDSL lipase/hydrolase gene (GhGLIP) appears to be involved in promoting seed growth in Arabidopsis. *PLoS ONE* **2018**, *13*, e0195556. [[CrossRef](#)]
78. He, S.L.; Ho, S.L. Functional characterization of a WD-Repeat protein gene (OsWD1) in rice. *Am. J. Agric. For.* **2018**, *6*, 18–27.
79. Ma, Y.; Yang, C.; He, Y.; Tian, Z.; Li, J.; Sunkar, R. Rice OVATE family protein 6 regulates plant development and confers resistance to drought and cold stresses. *J. Exp. Bot.* **2017**, *68*, 4885–4898. [[CrossRef](#)]
80. van der Knaap, E.; Østergaard, L. Shaping a fruit: Developmental pathways that impact growth patterns. In *Seminars in Cell & Developmental Biology*; Academic Press: Cambridge, MA, USA, 2018; Volume 79, pp. 27–36.
81. Wang, Y.; Wang, C.; Han, H.; Luo, Y.; Wang, Z.; Yan, C.; Xu, W.; Qu, S. Construction of a high-density genetic map and analysis of seed-related traits using specific length amplified fragment sequencing for *Cucurbita maxima*. *Front. Plant Sci.* **2020**, *10*, 1782. [[CrossRef](#)] [[PubMed](#)]
82. Chen, Y.; Xu, Y.; Luo, W.; Li, W.; Chen, N.; Zhang, D.; Chong, K. The F-box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. *Plant Physiol.* **2013**, *163*, 1673–1685. [[CrossRef](#)]
83. Choi, B.S.; Kim, Y.J.; Markkandan, K.; Koo, Y.J.; Song, J.T.; Seo, H.S. GW2 functions as an E3 ubiquitin ligase for rice expansin-like 1. *Int. J. Mol. Sci.* **2018**, *19*, 1904. [[CrossRef](#)]
84. Wang, Q.; Sun, G.; Ren, X.; Du, B.; Cheng, Y.; Wang, Y.; Li, C.; Sun, D. Dissecting the genetic basis of grain size and weight in barley (*Hordeum vulgare* L.) by QTL and comparative genetic analyses. *Front. Plant Sci.* **2019**, *10*, 469. [[CrossRef](#)]
85. Zhong, J.; He, W.; Peng, Z.; Zhang, H.; Li, F.; Yao, J. A putative AGO protein, OsAGO17, positively regulates grain size and grain weight through OsmiR397b in rice. *Plant Biotechnol. J.* **2020**, *18*, 916–928. [[CrossRef](#)] [[PubMed](#)]
86. Xie, F.T.; Niu, Y.; Zhang, J.; Bu, S.H.; Zhang, H.Z.; Geng, Q.C.; Feng, J.Y.; Zhang, Y.M. Fine mapping of quantitative trait loci for seed size traits in soybean. *Mol. Breed.* **2014**, *34*, 2165–2178. [[CrossRef](#)]
87. Kay, D.E. *Crop and Product Digest No. 2—Root Crops*; Tropical Development and Research Institute: London, UK, 1987; pp. 12–13.
88. Alqudah, A.M.; Sallam, A.; Baenziger, P.S.; Börner, A. GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: Lessons from Barley—A review. *J. Adv. Res.* **2020**, *22*, 119–135. [[CrossRef](#)] [[PubMed](#)]
89. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909. [[CrossRef](#)] [[PubMed](#)]
90. Mwadzingeni, L.; Shimelis, H.; Rees, D.J.G.; Tsilo, T.J. Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. *PLoS ONE* **2017**, *12*, e0171692. [[CrossRef](#)] [[PubMed](#)]
91. Ezenwaka, L.; Del Carpio, D.P.; Jannink, J.L.; Rabbi, I.; Danquah, E.; Asante, I.; Danquah, A.; Blay, E.; Egesi, C. Genome-Wide Association study of resistance to cassava green mite pest and related traits in cassava. *Crop Sci.* **2018**, *58*, 1907–1918. [[CrossRef](#)]
92. Sallam, A.; Martsch, R. Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica* **2015**, *143*, 501–514. [[CrossRef](#)] [[PubMed](#)]
93. Zhang, Z.; Ersoz, E.; Lai, C.Q.; Todhunter, R.J.; Tiwari, H.K.; Gore, M.A.; Bradbury, P.J.; Yu, J.; Arnett, D.K.; Ordovas, J.M.; et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360. [[CrossRef](#)]
94. Bandillo, N.; Raghavan, C.; Muyco, P.A.; Sevilla, M.A.L.; Lobina, I.T.; Dilla-Ermita, C.J.; Tung, C.W.; McCouch, S.; Thomson, M.; Mauleon, R.; et al. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice* **2013**, *6*, 11. [[CrossRef](#)]
95. Paliwal, R.; Abberton, M.; Faloye, B.; Olaniyi, O. Developing the role of legumes in West Africa under climate change. *Curr. Opin. Plant Biol.* **2020**, *56*, 242–258. [[CrossRef](#)] [[PubMed](#)]
96. Ho, W.K.; Chai, H.H.; Kendabie, P.; Ahmad, N.S.; Jani, J.; Massawe, F.; Kilian, A.; Mayes, S. Integrating genetic maps in bambara groundnut [*Vigna subterranea* (L.) Verdc.] and their syntenic relationships among closely related legumes. *BMC Genom.* **2017**, *18*, 192. [[CrossRef](#)] [[PubMed](#)]
97. Osundare, O.T.; Akinyele, B.O.; Odiyi, A.C.; Paliwal, R.; Oyatomi, O.A.; Abberton, M.T. Genetic diversity and population structure of some Nigerian accessions of Bambara groundnut (*Vigna subterranea* (L.) Verdc.) using DArT SNP markers. *Genet. Resour. Crop Evol.* **2022**, 1–15. [[CrossRef](#)]