agronomy

MDPI

Article

Genome-Wide Detection of SNP Markers Associated with Four Physiological Traits in Groundnut (*Arachis hypogaea* L.) Mini Core Collection

Abdulwahab S. Shaibu [1,2,*] **, Clay Sneller** [3] **, Babu N. Motagi** [4,5] **, Jackline Chepkoech** [6] **, Mercy Chepngetich** [6] **, Zainab L. Miko** [1] **, Adamu M. Isa** [7] **, Hakeem A. Ajeigbe** [4] **and Sanusi G. Mohammed** [1,8]

[1] Department of Agronomy, Bayero University, Kano 700001, Nigeria; zainablawanmiko@gmail.com (Z.L.M.); sanusi.gaya@buk.edu.ng (S.G.M.)

[2] The National Engineering Laboratory for Crop Molecular Breeding, MOA Key Laboratory of Soybean Biology (Beijing), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[3] Department of Horticulture and Crop Science, The Ohio State University, Wooster, OH 43210, USA; snellar.5@osu.ng

[4] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Kano 700001, Nigeria; b.n.motagi@cgiar.org or motagibn@uasd.in (B.N.M.); h.ajeigbe@cgiar.org (H.A.A.)

[5] Department of Genetics and Plant Breeding, Agriculture College, Vijayapur 586101, University of Agricultural Sciences, Dharwad 580005, India

[6] Integrated Genotyping Service and Support platform, Bioscience for East and Central Africa, International Livestock Institute, Nairobi 00100, Kenya; j.chepkoech@cgiar.org (J.C.); m.chepngetich@cgiar.org (M.C.)

[7] Department of Animal Science, Usmanu Danfodiyo University, Sokoto 840231, Nigeria; isa.adamu@udusok.edu.ng

[8] Centre for Dryland Agriculture, Bayero University, Kano 700001, Nigeria

* Correspondence: asshuaibu.agr@buk.edu.ng; Tel.: +23-4806-6019-498

check for updates

Abstract: In order to integrate genomics in breeding and development of drought-tolerant groundnut genotypes, identification of genomic regions/genetic markers for drought surrogate traits is essential. We used 3249 diversity array technology sequencing (DArTSeq) markers for a genetic analysis of 125 ICRISAT groundnut mini core collection evaluated in 2015 and 2017 for genome-wide marker-trait association for some physiological traits and to determine the magnitude of linkage disequilibrium (LD). Marker-trait association (MTA) analysis, probability values, and percent variation modelled by the markers were calculated using the GAPIT package via the KDCompute interface. The LD analysis showed that about 36% of loci pairs were in significant LD ($p < 0.05$ and $r^2 > 0.2$) and 3.14% of the pairs were in complete LD. The MTAs studies revealed 20 significant MTAs ($p < 0.001$) with 11 markers. Four MTAs were identified for leaf area index, 13 for canopy temperature, one for chlorophyll content and two for normalized difference vegetation index. The markers explained 20.8% to 6.6% of the phenotypic variation observed. Most of the MTAs identified on the A subgenome were also identified on the respective homeologous chromosome on the B subgenome. This could be due to a common ancestor of the A and B genome which explains the linkage detected between markers lying on different chromosomes. The markers identified in this study can serve as useful genomic resources to initiate marker-assisted selection and trait introgression of groundnut for drought tolerance after further validation.

Keywords: DArTseq; groundnut; linkage disequilibrium; marker-assisted selection; marker-trait association; physiological traits

1. Introduction

Groundnut or peanut (*Arachis hypogaea* L.) is an important food legume grown worldwide and is a rich source of protein for both humans and animals. Groundnut seed contains high-quality edible oil (50%), easily digestible protein (25%), and carbohydrate (20%) [1]. The crop was grown on 27.9 million hectares worldwide with a total production of 47.1 million metric tons [1].

Developing countries account for 96% (26.8 million hectares) of groundnut areas and 92% of the global production with the semi-arid tropics (SAT) region cultivating about 90%. Despite the developing countries being the largest producers of groundnut, the average yield per hectare (China, 2490 kg ha^{-1} and Nigeria 840 kg ha^{-1}) is low when compared to the United State of America (3673 kg ha^{-1}) [1]. Climate change is a major threat to groundnut yield and quality in the SAT regions. Among the factors contributing to low yield, drought adversely affects the crop's performance [2]. Shortage in the amount and distribution of rainfall in SAT regions has increased in the recent past thereby exacerbating climate risks including crop failures [3]. Drought stress has an adverse influence on water relations, photosynthesis, mineral nutrition, metabolism, growth and yield of groundnut [4]. Plants, being sessile, have evolved specific acclimation and adaptation mechanisms to respond to and survive short-long-term drought stresses [4]. Some physiological responses that allow adaptation to water deficit include root traits, stomatal conductance, SPAD (Soil Plant Analysis Development) chlorophyll meter reading, leaf area, and canopy temperature and are important measures for the agronomic response of yield under moisture stress [5]. Chlorophyll content and fluorescence parameters determine the integrity of the internal apparatus for photosynthesis and provide a precise platform for detection and quantification of plant tolerance to drought stress [6] and were speculated as a useful indicator for photosynthetic capacity in groundnut [7]. Canopy temperature (CT) was reported to be a marker for drought tolerance through its negative correlation with transpiration cooling and carbon dioxide exchange rate [8]. Genotypes that maintain cooler canopies under stress conditions possess a high potential for water stress tolerance and high yield [8]. Healthy vegetation measured by normalized difference vegetative index (NDVI) under drought conditions correlates to high photosynthetic potential and yield of groundnut [8]. Drought reduces leaf area by constraining mitosis, cell proliferation, leaf expansion and carbohydrate supply [9] and genotypes with wider leaf areas under water stress have the capacity for high photosynthesis. These drought surrogate traits are adaptation mechanisms used by groundnut (and other crops) to survive drought conditions [8].

In order to integrate genomics in breeding and development of drought-tolerant groundnut genotypes, identification of genomic regions associated with drought tolerance traits is essential. With the advent of genomic tools, marker-assisted breeding (MAB) has been deployed to enhance the efficiency of the selection of target traits in groundnut [6,10–13]. Very few informative and good quality single nucleotide polymorphism (SNP) markers are available in groundnut in contrast to the availability of thousands of simple sequence repeats (SSRs) [12]. However, SNPs can be more easily generated than SSRs and are usually preferred due to low cost. In recent times, restriction site-associated DNA sequencing (RADseq) [14] and genotyping by sequencing (GBS) [15] methods have allowed researchers to identify and genotype thousands of SNPs in plants. Diversity Arrays Technology (DArT), which is based on genome complexity reduction and SNP detection through hybridization of PCR fragments [16] has been used in genome-wide association studies (GWAS), construction of dense linkage maps and mapping quantitative trait loci (QTL) [10,17–19]. DArTseq is used for SNP discovery and genotyping, which enables considerable discovery of SNPs in a wide variety of non-model organisms and provides measures of genetic divergence and diversity within the major genetic groups that comprise crop germplasm [16]. The DArTseq technology from DArT produces data on bi-allelic SNP markers as well as the older dominant DArT markers. GWAS has lent itself to extensive application in genome-environment and genome-phenotype association mapping to identify loci of local adaptation and stress conditions in crops. The GWAS approach begins with phenotyping traits of interest followed by a forward genetic analysis to identify loci and candidate genes [20] by marker-trait association (MTA); approach adopted in the present study [21,22]. GWAS has

improved the identification of MTA with genomic regions by utilizing natural populations without the need for making large numbers of combinations from bi-parental mating. Furthermore, the magnitude of linkage disequilibrium (LD) present in genetic resources is important prerequisites to deduce the genetic makeup, composition and genomic predictions of traits of interest during selection [23]. Linkage disequilibrium per se could also be used as a predictor of the resolution at which significant genomic regions with influence on traits can be detected through marker-trait-association analysis [23].

Plant genetic resources are widely used in breeding programs for imparting resistance to various stresses [24,25]. Over the years, large numbers of groundnut accessions have been evaluated for resistance to biotic and abiotic stresses [25]. However, considering the available large number of groundnut accessions available in gene banks (>15,000), many precious accessions might not get evaluated for traits of interest, as it is cumbersome to screen such a huge collection under field conditions. Hence, researchers have developed core [26] and mini core collections [27] of groundnut that represent the genetic variability of the entire collection and serve as handy germplasm sets for evaluating important biotic and abiotic stresses. The selection of resistant sources through systematic screening of mini core collection accessions is in practice for infusing genetic diversity [24]. The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) evaluated 184 accessions of the mini core collections which led to the identification of some accessions with high SPAD chlorophyll meter reading (SCMR) [28] and the accessions are currently being utilized in the breeding program for drought tolerance.

The development of climate-smart and stable groundnut varieties will assist in meeting the growing demands of the increasing population against the threats of climate change. Identification of genetic markers that are linked to important traits affected by climate change in groundnut is needed for a reliable approach to the development of new varieties. In this study, SNP markers were used for the genetic analysis of a groundnut mini core collection from ICRISAT for a genome-wide marker-trait association for some physiological traits associated with drought tolerance and to determine the magnitude of LD present in the genetic resource.

2. Materials and Methods

2.1. Plant Materials and Phenotyping

The research was carried out in the research field of ICRISAT located at Minjibir (Latitude 12°19' N and Longitude 8°63' E) in 2015 and Bayero University, Kano (BUK, Latitude 11°58' N and Longitude 8°25' E) in 2017; both locations in Sudan Savanna of Nigeria. The long term means annual rainfall of both locations is about 800 mm, and; variations about this value is up to ±30%. The recorded weather information is presented in Figure S1a,b for Minjibir and BUK, respectively. The soil at Minjibir was typic Utipsamments and Typic Kanh aplustalf at BUK.

One hundred and twenty-five groundnut mini core accessions including five check varieties were evaluated using 25×5 randomized incomplete block design with three (3) replications. The description of the mini core collections along with the checks are presented in Table S1. The mini core collection was obtained from ICRISAT Kano and used in the current study based on seed availability. Sowing was done at Minjibir on 16 July 2015 and in BUK on 21 July 2017. Each plot in a replication consisted of a single row measuring 5 m in length with a row spacing of 0.75 m, plant distance of 0.1 m and 1 m alley between replications. Recommended specific practices for growing groundnut were strictly followed. A basal application of Nitrogen, Phosphorous, and Potassium was done to all plots at the rate of 20 N kg ha^{-1}, 40 P$_2$O$_5$ kg ha^{-1} and 40 K$_2$O kg ha^{-1} at planting. Hand weeding was done using hoes at 3rd, 8th, and 12th weeks after sowing (WAS) to prevent weed infestation and competition between plants and weeds.

Data was collected for canopy temperature (CT) using leaf thermometer (Meco IRT550 Infrared Thermometer, Sunshine instruments, Tamil Nadu, India), SPAD chlorophyll meter reading (SCMR) using SPAD 502 PLUS Chlorophyll meter (Spectrum Technologies, Inc., Aurora, IL, USA), normalized difference vegetative index (NDVI) using Hand Held optical sensor unit (Model 505 NTech Industries, Inc., Ukiah, CA, USA) and leaf area index (LAI) was measured by the use of Leaf Area Meter (ACCUPAR LP-80, Meter Group, Inc., Pullman, WA, USA). The data were collected at 60 WAS (peg formation stage) from five randomly selected plants in each plot.

2.2. DNA Extraction and Genotyping

Groundnut leaves were collected at 2 WAS into 96 deep well sample collection plates and sent to Integrated Genotyping Service and Support (IGSS) platform located at Biosciences Eastern and Central Africa (BecA-ILRI) Hub in Nairobi for Genotyping. The DNA extraction was done using the Nucleomag Plant Genomic DNA extraction kit. The genomic DNA extracted was in the range of 50–100 ng/ul. DNA quality and quantity were checked on 0.8% agarose. Libraries were constructed according to Kilian et al. [29]. DArTSeq complexity reduction method through digestion of genomic DNA and ligation of barcoded adapters was done followed by PCR amplification of adapter-ligated fragments. Libraries were sequenced using Single Read sequencing runs for 77 bases. Next-generation sequencing was carried out using Hiseq2500. The DArtSeq protocol was executed according to Kilian et al. [29].

The IGSS platform uses a GBS DArTseq^TM technology, which provides rapid, high quality, and affordable genome profiling, even from the most complex polyploid genomes. DArTseq markers scoring was achieved using DArTsoft version 14, which is an in-house marker scoring pipeline based on algorithms. Two types of DArTseq markers were scored, SilicoDArT markers (scored as present or absent; 1, 0) and biallelic SNP markers which were both scored for the presence of the reference allele, the alternative allele, or both. In the genomic representation of the sample, both SilicoDArT markers and SNP markers were aligned to the reference genomes of *Arachis duranensis* (V14167, A-genome ancestor) and *A. ipaensis* (K30076, B-genome ancestor, https://www.peanutbase.org/) to identify chromosome.

2.3. Data Analysis

The analysis of phenotypic data was done to obtain the best linear unbiased prediction (BLUP) values for each accession by fitting the following mixed linear model in the *R* "lme4" package.

$$y_{ijk} = u + g_i + e_j + r(e)_{jk} + ge_{ij} + error_{ijk.} \tag{1}$$

where g_i is the effect of the *i*th line, e_j is the effect of the *j*th environment, $r(e)_{jk}$ is the effect of the *k*th replication nested in the *j*th environment, ge_{ij} is the genotype by environment interaction and $error_{ijk}$ is the error associated with each observation. The analysis was run in *R* with the lme4 package [30].

Entry means broad-sense heritability was calculated as:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_{error}^2}{re}} \tag{2}$$

where σ_g^2 is the variance among lines, σ_{ge}^2 is the genotype by environment interaction variance, σ_{error}^2 is the error variance, *r* is the number of replications, and *e* is the number of environments. The phenotypic correlation between traits was also determined.

Linkage Disequilibrium and Marker-Trait Association

Polymorphism information content (PIC) and principal component analysis (PCA) were carried out on the genotypic data using KDCompute. The unweighted pair-group method was used to cluster the accessions into a dendrogram. The parameter r^2 was used to estimate LD between SNPs on each chromosome via the software package TASSEL 5.0 [31]. Marker-trait association analysis, probability values, and percent variation modelled by both SilicoDArT and biallelic SNP markers were calculated using the GAPIT package via the KDCompute interface (https://kdcompute.igss-africa.org/kdcompute/home). The GWAS threshold for the significant marker-trait association was $p < 0.001$ without multiple testing correction due to the small population size used in the present study. The first three principal components and the relationship matrix were included in the model to account for population structure. SNPs with minor allele frequency (MAF) of <5% and missing data >20% were excluded from the analyses. Missing values were imputed using the choice of the nearest neighbor algorithm using TASSEL 5.0 [31].

3. Results

3.1. Phenotypic Evaluation

The results of the phenotypic evaluations showed highly significant differences ($p < 0.01$) between lines for CT, SCMR, and NDVI but no significant genetic variation for LAI (Table 1). The interaction between lines and the environment was significant ($p < 0.01$) only for SCMR (Table 1). The heritability of the traits was moderate to high except for LAI that had a low heritability (0.03) (Table 1). The four traits were normally distributed (Figure S2). The phenotypic correlation of SCMR with LAI and CT was negative and non-significant but was positive with NDVI. Among the mini core collections, ICG 9926 had the highest CT (38.03 °C) while some collections like ICG 1569, ICG 7190, ICG 12672, ICG 81, ICG 14106, and ICG 442 had CT that ranged from 37.67 °C–37 °C (Table S1). The lowest CT was recorded by ICG 1668 (24.1 °C) followed by ICG 76 (25 °C), ICG 1274 (25.1 °C), ICG 7437 (25.5 °C) and ICG 10092 (25.7 °C). For LAI, ICG 8567 had the highest LAI of 4.6 followed by ICG 9809 with LAI of 4.05 while ICG 11426 had the lowest LAI of 0.61. ICG 7483 had the highest SCMR of 55.43 followed by ICG 7243 with 55.03. The highest NDVI reading was recorded by ICG 3055 (0.82) followed by ICGs 7897 and 3126 both having readings of 0.81. The lowest NDVI reading of 0.45 was recorded by ICG 9449 (Table S1). A negative and significant correlation was observed between CT and LAI as well as CT with NDVI. The correlation between LAI and NDVI was negative and non-significant (Table 2). A cluster analysis on the accessions based on the geographical origin of the accessions revealed two groups with group two having sub-group 2A and 2B (Figure 1).

Table 1. Mean squares of four physiological traits of groundnut mini core collections.

Effect	df	Canopy Temperature	Leaf Area Index	SPAD Chlorophyll Meter Reading	NDVI
Rep (Environment)	2	679.62	7.28	18.73	0.47
Environment	1	4820.38 **	54.69 **	7149.59 **	13.81 **
Lines	124	7.82 **	0.66	49.14 **	0.43 **
Environment × Lines	124	7.69	0.49	26.09 **	0.001
Error	248	7.25	0.84	15.42	0.02
Heritability		0.22	0.03	0.55	0.25

** = significant at 0.01 level of probability. df = degree of freedom.

Table 2. Correlation between physiological traits of groundnut mini core collections (above value = correlation and below value are *p*-values).

	SPAD Chlorophyll Meter Reading	Leaf Area Index	Canopy Temperature	Normalized Difference Vegetative Index (NDVI)
SPAD (Soil Plant Analysis Development) chlorophyll reading				
Leaf area index	−0.03 0.72			
Canopy temperature	−0.16 0.08	−0.19 0.03		
NDVI	0.08 0.37	0.12 0.19	−0.25 0.01	
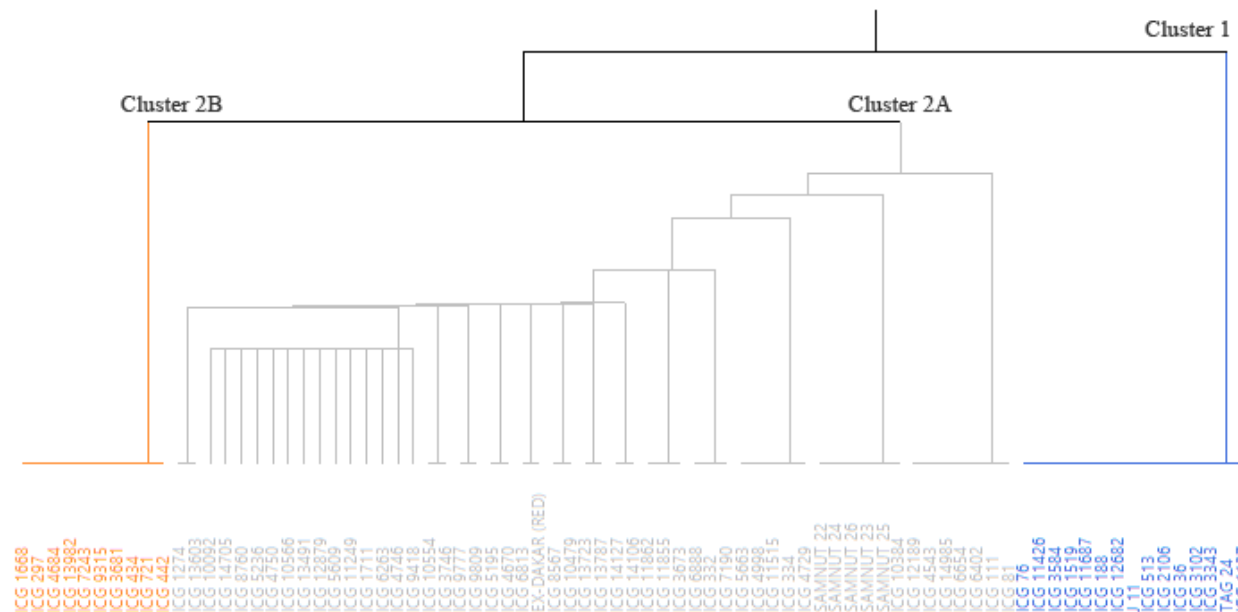
Figure 1. Dendrogram from unweighted pair-group clustering of accession based on the geographical distribution of the mini core collections.

3.2. Marker Data

The DArTseq genotyping produced 3591 biallelic SNP markers of which 3396 had a call rate that exceeded 70%. The average PIC of the 3396 markers was 0.077. Of the 3396 markers, just 396 had a MAF that exceeded 0.05. A total of 3124 markers were successfully assigned to their chromosome by mapping them to the A and B genomes: 368 (11.8%) were aligned to only the A genome, 449 (14.4%) to only the B genome, and 2308 (73.8%) to both genomes. Over 73% of the markers that aligned with both the A and B genomes were assigned to homoeologous chromosomes and the correlation of their position on those two sets of homologues was 0.87. In the principal component (PC) analyses of the data from the 3,124 markers assigned to a chromosome(s), the first PC accounted for 61% of the variation and the first two PCs accounted for 78% of the variation (Figure 2). Cluster analysis of the marker data suggested two groups of lines with one group having two subgroups named 2A and 2B (Figure S3).
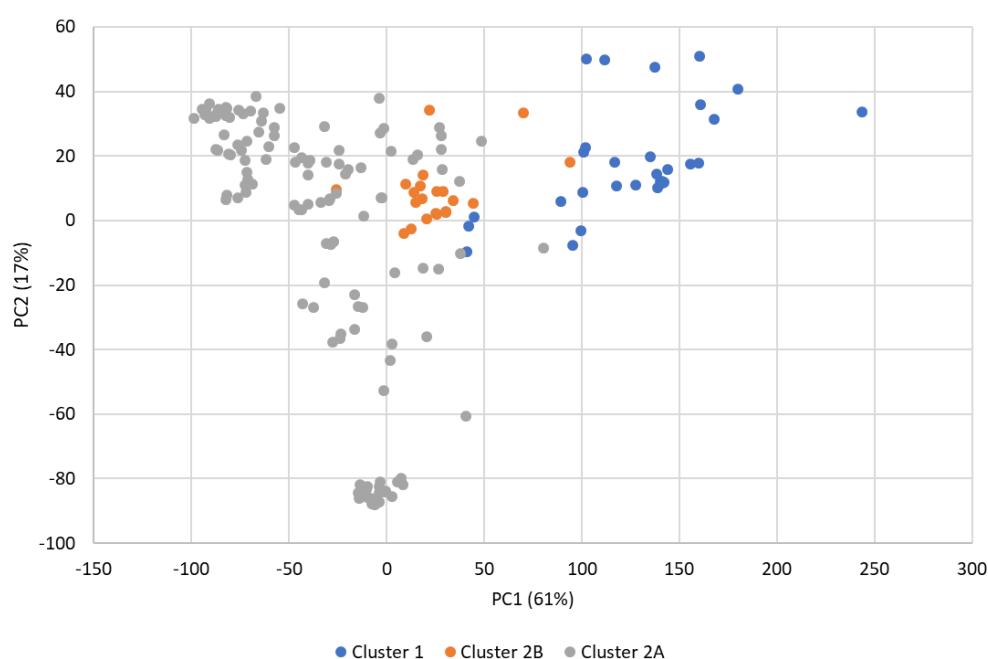

Figure 2. Graph of the scores of the first two principal components of the analysis of the marker data from the groundnut mini core collection.

The DArTseq genotyping produced 12,693 dominant silico markers with a call rate that also exceeded 70%. Only 2349 (18.5%) of these had a minor allele frequency (MAF) > 0.05. The average PIC of the 2349 markers was 0.070. A total of 12,611 markers were given chromosome assignments: 1709 (13.6%) were aligned to only the A genome, 2502 (19.8%) to only the B genome, and 8400 (66.7%) to both genomes. Over 76% of the markers aligned with both the A and B genomes were assigned to homoeologous chromosomes and the correlation of their position on those two sets of homologues was 0.91. This reflects the common origin of the A and B genomes.

There was some evidence for the inter-genome exchange of genes between non-homoeologous chromosomes. A set of 46 markers located in a 16.72 Mbp region of chromosome A08 was also found in a 73.95 Mbp region of B07. The correlation of positions for the 46 markers was 0.769. Another set of 20 markers spanning a 3.93 Mbp region of chromosome A02 and an 11.67 Mbp region of B09 were also found (Figures S4 and S5). The correlation of positions for the 46 markers was −0.604 showing any putative exchange involved an inversion.

3.3. Linkage Disequilibrium

Linkage disequilibrium analysis conducted using 305,919 loci pairs within chromosomes showed that 36.3% of loci pairs had significant LD (Table S1). Furthermore, 9592 (3.14%) of the pairs were in complete LD ($r^2 = 1$). There was a rapid decline in LD with distance and the correlation analysis revealed negative correlation ($r = -0.0795$) between the LD (R^2) and the physical distance; as well as between the R^2 and p-value ($r = -0.5381$), revealing the existence of linkage decay (Figure S6).

3.4. Marker-Trait Association

Due to the insignificant genotype by environment interaction for all traits except for SCMR, the GWAS was performed using phenotypic BLUPs estimated over all environments. The marker-trait association (MTA) analysis was done for both the dominant silico and biallelic SNP markers. However, significant associations were only detected from the dominant silico markers and none was detected from the biallelic SNP markers. Only the MTAs from the dominant silico markers that had p-values < 0.001 (Table 3) were considered as significant for all traits (details of significant MTAs with p-values between 0.05 and 0.001 are presented in Table S2). We found 20 MTAs with 11 markers (Table 3). Two markers (M1, M2) identified four possible loci for LAI. Marker M1 was associated with chromosomes A03 and B03 with allelic effects of −1.33 and −1.31, respectively while M2 was associated with chromosomes A06 and B07 with allelic effects of 1.96 and 1.97, respectively. The individual markers explained 6.8% to 7.3% of the total phenotypic variation observed (Table 3). Thirteen loci within seven markers (M3–M9) showed MTAs with CT and six of these markers (M3–M8) were located in chromosomal regions on both the A and B genomes. The CT markers explained between 9.6% to 16.6% of the variation observed and all had negative allelic effects. One marker (M10) on chromosome B05 was found to be associated with SCMR and explained 20.8% of the observed phenotypic variation with an allelic effect of −11.65. Another marker (M11) located on chromosomes A04 and B02 was associated with NDVI and had an allelic effect of 0.16 for both chromosomes. From all the 20 MTAs detected, the B genome had 11 and the A genome had nine. An equal number of associations were found on each genome except for CT and SCMR. The A genome had six MTAs and the B genome had seven MTAs for CT. The MTA with SCMR was exclusively detected on the B genome.

Table 3. Marker-trait associations of four physiological traits of groundnut in the groundnut mini core collection.

Trait		Marker	Chromosome	Position (Mbp)	*p*-Value	Marker R^2	Allelic Effect
Leaf area index	M1	100028731	A03	112.30	0.0007	0.073	−1.33
	M1	100028731	B03	114.70	0.00084	0.068	−1.31
	M2	100055743	A06	29.59	0.00089	0.069	1.96
	M2	100055743	B07	39.39	0.00083	0.068	1.97
Canopy temperature	M3	100002202	A02	14.18	0.00099	0.098	−1.74
	M3	100002202	B02	17.44	0.00089	0.097	−1.95
	M4	100006533	A03	12.02	2.06×10^{-5}	0.166	−4.27
	M4	100006533	B03	14.68	2.19×10^{-5}	0.163	−4.22
	M5	100057474	A05	4.01	0.00017	0.128	−2.86
	M5	100057474	B05	3.92	0.00018	0.125	−2.87
	M6	100007799	A05	4.75	0.00078	0.102	−2.33
	M6	100007799	B05	4.84	0.00046	0.109	−2.56
	M7	100008002	A08	7.71	0.00097	0.098	−2.29
	M7	100008002	B07	113.51	0.00057	0.105	−2.52
	M8	100007218	A10	78.51	0.00069	0.104	−1.51
	M8	100007218	B10	103.61	0.00062	0.104	−1.49
	M9	100002387	B01	135.56	0.00097	0.096	−4.75
SPAD chlorophyll meter reading	M10	100000269	B05	17.00	0.00099	0.208	−11.65
NDVI	M11	100118763	A04	67.54	0.00069	0.070	0.16
	M11	100118763	B02	106.74	0.001	0.066	0.16

4. Discussion

The core, mini core, and reference collections developed by several germplasm resource centers are significant sources of genetic variation. By screening these groundnut sources, sources of tolerance or resistance to traits needed for the development of climate-smart varieties can be identified. Physiological traits such as LAI, SCMR, canopy conductance and canopy temperature are important measures of agronomic response to yield under water stress [32]. The phenotypic evaluation showed that the groundnut collections varied significantly for the physiological traits except for LAI (Table 1). Nageswara Rao et al. [7] also reported similar specific leaf area between peanut genotypes. These physiological traits (LAI, CT, SCMR, and NDVI) are important in improving productivity and are used as indirect indices for improving drought tolerance in peanut [8]. Traits such as SCMR and NDVI determine the photosynthetic potential of plants and have been reported to be highly associated with yield and can be effective as in-season predictors of yield [33,34]. Some of the identified mini cores with desirable physiological traits can be integrated into breeding programs for the development of varieties that are drought tolerant. The heritability of CT, SCMR, and NDVI ranged from moderate to high and contribution of genetics to the phenotypic variation in LAI was low. The low correlation coefficients between most of the traits suggests that the traits are fairly independent (Table 2).

The low genetic diversity of both SNP and silico markers and their non-distinct separation on the PC plot suggests that the populations were not highly structured. The two groups of mini core collections and one group having two subgroups as suggested by the marker data may be due to different origins of the collections as evident from the cluster analysis of the accessions based on origins. Cluster 1 consisted solely of accessions from India while cluster 2B consisted of only USA accessions. The accessions from other countries were all grouped in cluster 2A. Pandey et al. [10] also observed three groups in groundnut using SSR and DArT markers. The A genome is more conserved because it has fewer markers than the B genome. Both markers systems assigned a similar portion of the markers to both genomes pointing to a common ancestry of the collections. While many polymorphic markers were detected, a large portion had MAF < 0.05 and the average PIC values for both types of markers were very low; about 0.07. Taken together, the results of the current study support earlier reports on the low polymorphism rate and low genetic diversity of groundnut [35–37].

In general, most markers assigned to the A and B genomes were assigned to homoeologous positions and their genetic positions in the A and B genome were highly correlated and suggest that their position in the genomes has been mostly conserved since evolving from their common ancestor. Of the markers mapped to non-homoeologous chromosomes in the A and B genome there was some evidence for the exchange of chromosome segments between chromosomes A02 and B09 and then A08 and B07.

Linkage disequilibrium analysis showed that about 36% of loci pairs were in significant LD ($p < 0.05$) and 3.14% of the pairs were in complete LD with an average distance of 31 kb among these pairs, indicating that LD extended quite some distance in groundnut. This is not surprising given the low polymorphism rate and PIC values in groundnut which means that detectable recombination is likely to be very low. Several studies have reported LD decay with distance [10,38,39] which agrees with the findings of the present study.

GWAS has created a considerable need for downstream studies including genetics, physiology, and biochemistry to ascertain genotype–phenotype associations that can be used to decipher the underlying mechanisms for intricate traits such as yield and stress responses [31]. The current study revealed 20 significant MTAs ($p < 0.001$) involving 11 markers. The *p*-values used for identifying significant MTAs were not adjusted using multiple testing corrections because of the small sample size used in the study and we want to be able to identify QTL of moderate effect. Our analysis cannot distinguish if a single marker on more than one chromosome exhibits MTA independently or is a result of additive contribution from the two homoeologous chromosomes. Markers associated with physiological traits show uneven distribution among chromosomes and between the genomes. Chromosome B05 of the B genome houses three markers that showed MTA with CT and SCMR. In addition, two markers each

were found on chromosomes A03, B03 and A05. Genome-wise comparison showed that all the eleven markers detected are found on B-genome while only 9 were found on the A genome. Three of these markers (M2, M7, and M11) are found on non-homoeologous chromosomes. It is possible that some chromosome rearrangement has caused the marker sequence to appear on different homeologs in the course of evolution of groundnut, although an error in sequencing or bioinformatics alignment can result in a similar outcome. Validation studies will, therefore, be needed to see if these markers are identifying one locus or perhaps a locus duplicated in the two genomes. The allelic effects of the markers identified for CT and SCMR were negative which shows that the markers are identified in genomic regions that have decreasing effects for these traits. Plants with lower CT are preferred in drought-prone areas because the lower CT enables the plant to reduce its transpiration rate and therefore conserved moisture [8]. However, for SCMR, groundnut genotypes with higher SCMR are preferred to maintain healthy vegetation which will promote kernel yield. The marker identified for NDVI had an increasing effect on the phenotype while for LAI, M1 had a decreasing effect and M2 had an increasing effect. Though the difference between the accessions was not significant phenotypically for LAI, the two markers were identified to be associated with LAI. A previous study reported the additive and additive × additive gene actions for specific leaf area in groundnut [40]. We would assume that the contrasting effects of M1 and M2 might act additively and lead to insignificance of the overall difference in LAI between the accessions. Nevertheless, the identified markers could be used for selection of LAI genomic region in groundnut. Furthermore, after validation of all the identified markers, they can be deployed in marker-assisted breeding for the selection of groundnut genotypes with desirable physiological traits.

Pandey et al. [10] used SSR markers to identify some significant MTAs for physiological traits including LAI and SCMR in groundnut which were also observed in the present study. All the Four MTAs identified for LAI were associated with two markers (M1 and M2) found among which the MTA on chromosomes A06 and B07 may be the same as those previously identified by Pandey et al. [10] for total leaf area and leaf area respectively. Eight of the thirteen (13) MTAs identified for CT are associated with four markers located on homoeologous chromosomes in both A and B genomes. The one MTA associated with M1 was located on chromosome B05 but absent on its homeolog. Interestingly, M10 accounts for >20% of the phenotypic variation in SCMR and is different from the previously reported loci on A06 [11]. It is therefore suggested to be novel. Markers M5 and M6 were in high LD and may possibly suggest the same MTA. The two markers associated with NDVI were located on both the A and B genome but on different chromosomes.

From the analysis, most of the MTAs identified on the A subgenome were also identified on the respective homoeologous chromosome on the B subgenome. Argawal et al. [13] reported that a significant proportion of marker loci with assigned physical locations to the chromosome of one genome were mapped to respective homeologous positions on chromosomes of the other genome. Our results support this hypothesis as the correlation of markers positions between the A and B genome exceeded 0.87 for both marker types. Most of the homeologous MTAs were seen between chromosomes A03 and B03 which is similar to the study of Argawal et al. [13]. Other homeologous MTAs detected in the present study are located between chromosomes A02 and B02, and A05 and B05. Homeologous mapping of QTLs in groundnut has also been reported between chromosomes A07 and B07, and A08 and B08 [41]. We found some evidence for genetic exchanges occurring between the groundnuts genomes as reported earlier [38]. Also, many similar markers were placed on the genetic map on different chromosomes. The possible translocation we noted does not appear to be terminal or reciprocal unlike the translocations noted by Farre et al. [42]. Translocations of markers have been previously reported in groundnut [13,41]. Some of these observed 'translocated' markers might be also due to miss-assignments because of the highly repetitive structure of the groundnut genome [37].

From the present study, the DArTs markers showed the highest reproducibility and consistency than the SNP markers. The DArTseq approach generated a large set of useful SNPs with broad genome coverage which represented both coding and non-coding regions thereby allowing for the

accurate assessment of structure and quantity of genetic diversity in the mini core collections. This study identified a total of 20 highly significant marker-trait associations for four physiological traits of importance in groundnut; LAI, CT, SCMR, and NDVI. Chromosome B05 of the B genome contained more markers associated with drought surrogate physiological traits in groundnut. The markers identified in this study can serve as useful genomic resources to initiate marker-assisted selection and trait introgression of groundnut for drought tolerance. The identified MTAs can also be used for fine mapping and cloning of the underlying genes. Further studies are required to validate significant markers identified in the present study using a larger population size.

Supplementary Materials: The following are available, Table S1: Table S1-Name, botanical grouping, origin and physiological performances of the mini core collections. Table S2: LD of markers for A and B genomes, Table S3: Significant marker-trait associations, Figure S1: Weather information for (A) Minjibir 2015 and (B) BUK 2017. Figure S2: Histogram of BLUP of leaf area index (LAI), canopy temperature (CT), chlorophyll content (SPAD) and NDVI from the groundnut mini core collection. Figure S3: Dendrogram from unweighted pair-group clustering of accession from the mini core collections. Figure S4: Distribution of markers on A genome of the groundnut accessions, Figure S5: Distribution of markers on the B genome of the groundnut accessions. Figure S6: Scatter plot showing the association between linkage disequilibrium (r^2) and distance (a) and significance of the r^2 value (b).

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

References

1. FAOSTAT. Crops. Food and Agriculture Organization of the United Nations. 2017. Available online: http://www.fao.org/faostat/en/#data/QC (accessed on 3 February 2018).
2. Padi, F.K. Genotype × Environment Interaction for Yield and Reaction to Leaf Spot Infections in Groundnut in Semiarid West Africa. *Euphytica* **2008**, *164*, 143–161. [CrossRef]
3. Suther, M.S.; Patel, D.M. Yield and Nutrient Absorption by Groundnut and Iron Availability in Soil as Influenced by Lime and Soil Water. *J. Indian Soc. Soil Sci.* **1992**, *40*, 594–596.
4. Singh, A.L.; Chaudhari, V. Macronutrient Requirement of Groundnut: Effects on Growth and Yield Components. *Indian J. Plant Physiol.* **2006**, *11*, 40–409.
5. Price, A.H.; Horton, P.; Jones, H.G.; Griffiths, H.; Cairns, J.E. Linking Drought-Resistance Mechanisms to Drought Avoidance in Upland Rice Using a QTL Approach: Progress and New Opportunities to Integrate Stomatal and Mesophyll Responses. *J. Exp. Bot.* **2002**, *53*, 989–1004. [CrossRef] [PubMed]
6. Pandey, M.K.; Monyo, E.; Ozias-Akins, P.; Liang, X.; Guimaraes, P.; Nigam, S.N.; Upadhyaya, H.D.; Janila, P.; Zhang, X.; Guo, B.; et al. Advances in Arachis Genomics for Peanut Improvement. *Biotechnol. Adv.* **2012**, *30*, 639–651. [CrossRef]
7. Nageswara Rao, R.C.; Talwar, H.S.; Wright, G.C. Rapid Assessment of Specific Leaf Area and Leaf Nitrogen in Peanut (*Arachis hypogaea* L.) Using a Chlorophyll Meter. *J. Agron. Crop Sci.* **2001**, *186*, 175–182. [CrossRef]
8. Luis, J.M.; Ozias-Akins, P.; Holbrook, C.C.; Kemerait, R.C., Jr.; Snider, J.L.; Liakos, V. Phenotyping Peanut Genotypes for Drought Tolerance. *Peanut Sci.* **2016**, *43*, 36–48. [CrossRef]

9. Hussain, M.; Malik, M.A.; Farooq, M.; Ashraf, M.Y.; Cheema, M.A. Improving drought tolerance by exogenous application of glycinebetaine and salicylic acid in sunflower. *J. Agron. Crop. Sci.* **2008**, *194*, 193–199. [CrossRef]

10. Pandey, M.K.; Upadhyaya, H.D.; Rathore, A.; Vadez, V.; Sheshshayee, M.S.; Sriswathi, M.; Govil, M.; Kumar, A.; Gowda, M.V.; Sharma, S.; et al. Genomewide Association Studies for 50 Agronomic Traits in Peanut Using the "reference Set" Comprising 300 Genotypes from 48 Countries of the Semi-Arid Tropics of the World. *PLoS ONE* **2014**, *9*, e105228. [CrossRef]

11. Janila, P.; Venuprasad, R.; Abhishek, R.U.A.; Reddy, K.R.; Waliyar, F.; Nigam, S.N. Genetic Analysis of Resistance to Late Leaf Spot in Interspecific Groundnuts. *Euphytica* **2013**, *193*, 13–25.

12. Varshney, R.K.; Mohan, S.M.; Gaur, P.M.; Gangarao, N.V.; Pandey, M.K.; Bohra, A.; Sawargaonkar, S.L.; Chitikineni, A.; Kimurto, P.K.; Janila, P.; et al. Achievements and Prospects of Genomics-Assisted Breeding in Three Legume Crops of the Semi-Arid Tropics. *Biotechnol. Adv.* **2013**, *31*, 1120–1134. [CrossRef] [PubMed]

13. Agarwal, G.; Clevenger, J.; Pandey, M.K.; Wang, H.; Shasidhar, Y.; Chu, Y.; Fountain, J.C.; Choudhary, D.; Culbreath, A.K.; Liu, X.; et al. High-Density Genetic Map Using Whole-Genome Resequencing for Fine Mapping and Candidate Gene Discovery for Disease Resistance in Peanut. *Plant Biotechnol. J.* **2018**, *16*, 1954–1967. [CrossRef] [PubMed]

14. Willing, E.M.; Hoffmann, M.; Klein, J.D.; Weigel, D.; Dreyer, C. Paired-End RAD-Seq for de Novo Assembly and Marker Design without Available Reference. *Bioinformatics* **2011**, *27*, 2187–2193. [CrossRef] [PubMed]

15. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379. [CrossRef] [PubMed]

16. Jaccoud, D.; Peng, K.; Feinstein, D.; Kilian, A. Diversity Arrays: A Solid State Technology for Sequence Information Independent Genotyping. *Nucleic Acids Res.* **2001**, *29*, E25. [CrossRef] [PubMed]

17. Valdisser, P.; Pereira, W.J.; Almeida Filho, J.E.; Muller, B.S.F.; Coelho, G.R.C.; de Menezes, I.P.P.; Vianna, J.P.G.; Zucchi, M.I.; Lanna, A.C.; Coelho, A.S.G.; et al. In-Depth Genome Characterization of a Brazilian Common Bean Core Collection Using DArTseq High-Density SNP Genotyping. *BMC Genom.* **2017**, *18*, 423. [CrossRef]

18. Al-Abdallat, A.M.; Karadsheh, A.; Hadadd, N.I.; Akash, M.W.; Ceccarelli, S.; Baum, M.; Hasan, M.; Jighly, A.; Abu Elenein, J.M. Assessment of Genetic Diversity and Yield Performance in Jordanian Barley (*Hordeum Vulgare* L.) Landraces Grown under Rainfed Conditions. *BMC Plant Biol.* **2017**, *17*, 191. [CrossRef]

19. Abu Zaitoun, S.Y.; Jamous, R.M.; Shtaya, M.J.; Mallah, O.B.; Eid, I.S.; Ali-Shtayeh, M.S. Characterizing Palestinian Snake Melon (*Cucumis Melo* Var. *Flexuosus*) Germplasm Diversity and Structure Using SNP and DArTseq Markers. *BMC Plant Biol.* **2018**, *18*, 246.

20. Mousavi-Derazmahalleh, M.; Bayer, P.E.; Hane, J.K.; Valliyodan, B.; Nguyen, H.T.; Nelson, M.N.; Erskine, W.; Varshney, R.K.; Papa, R.; Edwards, D. Adapting Legume Crops to Climate Change Using Genomic Approaches. *Plant Cell Environ.* **2017**, *42*, 6–9. [CrossRef]

21. Edae, E.A.; Byrne, P.F.; Haley, S.D.; Lopes, M.S.; Reynolds, M.P. Genome-Wide Association Mapping of Yield and Yield Components of Spring Wheat under Contrasting Moisture Regimes. *Theor. Appl. Genet.* **2014**, *127*, 791–807. [CrossRef]

22. Sneller, C.H.; Mather, D.E.; Crepieux, S. Analytical Approaches and Population Types for Finding and Utilizing QTL in Complex Plant Populations. *Crop Sci.* **2009**, *49*, 363–380. [CrossRef]

23. Baird, S.J. Exploring Linkage Disequilibrium. *Mol. Ecol. Resour.* **2015**, *15*, 1017–1019. [CrossRef] [PubMed]

24. Upadhyaya, H.D.; Yadav, D.; Dronavalli, N.; Gowda, C.L.L.; Singh, S. Mini Core Germplasm Collections for Infusing Genetic Diversity in Plant Breeding Programs. *Electron. J. Plant Breed.* **2010**, *1*, 1294–1309.

25. Waliyar, F.; Kumar, K.V.K.; Diallo, M.; Traore, A.; Mangala, U.N.; Upadhyaya, H.D.; Sudini, H. Resistance to Pre-Harvest Aflatoxin Contamination in ICRISAT's Groundnut Mini Core Collection. *Eur. J. Plant Pathol.* **2016**, *145*, 901–913. [CrossRef]

26. Holbrook, C.C.; Anderson, W.F.; Pittman, R.N. Selection of a Core Collection from the U.S. Germplasm Collection of Peanut. *Crop Sci.* **1993**, *33*, 859–861. [CrossRef]

27. Upadhyaya, H.D.; Bramel, P.J.; Ortiz, R.; Singh, S. Developing a Mini Core of Peanut for Utilization of Genetic Resources. *Crop Sci.* **2002**, *42*, 2150–2156. [CrossRef]

28. Upadhyaya, H.D.; Dwivedi, S.L.; Nadaf, H.L.; Singh, S. Phenotypic Diversity and Identification of Wild Arachis Accessions with Useful Agronomic and Nutritional Traits. *Euphytica* **2011**, *182*, 103. [CrossRef]

29. Kilian, A.; Wenzl, P.; Huttner, E.; Carling, J.; Xia, L.; Blois, H.; Caig, V.; Heller-Uszynska, K.; Jaccoud, D.; Hopper, C.; et al. Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. *Methods Mol. Biol.* **2012**, *888*, 67–89.

30. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Usinglme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]

31. 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]

32. Gangurde, S.S.; Kumar, R.; Pandey, A.K.; Burow, M.; Laza, H.E.; Nayak, S.N.; Guo, B.; Liao, B.; Bhat, R.S.; Madhuri, N.; et al. Climate-Smart Groundnuts for Achieving High Productivity and Improved Quality: Current Status, Challenges, and Opportunities. In *Genomic Designing of Climate-Smart Oilseed Crops*; Kole, C., Ed.; Springer: Cham, Switzerland, 2019; pp. 133–172.

33. Tattaris, M.; Reynolds, M.P.; Chapman, S.C. A Direct Comparison of Remote Sensing Approaches for High-Throughput Phenotyping in Plant Breeding. *Front. Plant Sci.* **2016**, *7*, 1131. [CrossRef] [PubMed]

34. Sukumaran, S.; Reynolds, M.P.; Sansaloni, C. Genome-Wide Association Analyses Identify QTL Hotspots for Yield and Component Traits in Durum Wheat Grown under Yield Potential, Drought, and Heat Stress Environments. *Front. Plant Sci.* **2018**, *9*, 81. [CrossRef] [PubMed]

35. Nielen, S.; Vidigal, B.S.; Leal-Bertioli, S.C.; Ratnaparkhe, M.; Paterson, A.H.; Garsmeur, O.; D'Hont, A.; Guimaraes, P.M.; Bertioli, D.J. Matita, a New Retroelement from Peanut: Characterization and Evolutionary Context in the Light of the Arachis A-B Genome Divergence. *Mol. Genet. Genom.* **2012**, *287*, 21–38. [CrossRef] [PubMed]

36. Moretzsohn, M.C.; Gouvea, E.G.; Inglis, P.W.; Leal-Bertioli, S.C.; Valls, J.F.; Bertioli, D.J. A Study of the Relationships of Cultivated Peanut (*Arachis hypogaea*) and Its Most Closely Related Wild Species Using Intron Sequences and Microsatellite Markers. *Ann. Bot.* **2013**, *111*, 113–126. [CrossRef] [PubMed]

37. Bertioli, D.J.; Cannon, S.B.; Froenicke, L.; Huang, G.; Farmer, A.D.; Cannon, E.K.; Liu, X.; Gao, D.; Clevenger, J.; Dash, S.; et al. The Genome Sequences of *Arachis duranensis* and *Arachis ipaensis*, the Diploid Ancestors of Cultivated Peanut. *Nat. Genet.* **2016**, *48*, 438–446. [CrossRef] [PubMed]

38. Neumann, K.; Denčić, S.; Varshney, R.; Börner, A.; Kobiljski, B. Genome-Wide Association Mapping: A Case Study in Bread Wheat (*Triticum aestivum* L.). *Mol. Breed.* **2011**, *27*, 37–58. [CrossRef]

39. Mwadzingeni, L.; Shimelis, H.; Rees, D.J.; 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]

40. Nigam, S.N.; Upadhyaya, H.D.; Chandra, S.; Nageswara Rao, R.C.; Wright, G.C.; Reddy, A.G.S. Gene Effects for Specific Leaf Area and Harvest Index in Three Crosses of Groundnut (*Arachis hypogaea*). *Ann. Appl. Biol.* **2001**, *139*, 301–306. [CrossRef]

41. Qin, H.; Feng, S.; Chen, C.; Guo, Y.; Knapp, S.; Culbreath, A.; He, G.; Wang, M.L.; Zhang, X.; Holbrook, C.C.; et al. An Integrated Genetic Linkage Map of Cultivated Peanut (*Arachis hypogaea* L.) Constructed from Two RIL Populations. *Theor. Appl. Genet.* **2012**, *124*, 653–664. [CrossRef]

42. Farre, A.; Benito, I.L.; Cistue, L.; de Jong, J.H.; Romagosa, I.; Jansen, J. Linkage Map Construction Involving a Reciprocal Translocation. *Theor. Appl. Genet.* **2011**, *122*, 1029–1037. [CrossRef]