Genome Wide Association Studies for Four Physiological Traits in Groundnut (*Arachis hypogaea* L.) Minicore 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, Nigeria
[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, China
[3] Department of Horticulture and Crop Science, The Ohio State University, Wooster Ohio, USA
[4] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Kano, Nigeria
[5] Department of Genetics and Plant Breeding, Agriculture College, University of Agricultural Science, Dharwad, India.
[6] Integrated Genotyping Service and Support platform, Bioscience for East and Central Africa, International Livestock Institute, Nairobi, Kenya
[7] Department of Animal Science, Usmanu Danfodio University, Sokoto, Nigeria
[8] Centre for Dryland Agriculture, Bayero University, Kano, Nigeria
* Corresponding author: asshuaibu.agr@buk.edu.ng

**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 SNP markers for a genetic analysis of the ICRISAT groundnut minicore collection for genome wide marker-trait association for some physiological traits and to determine the magnitude of linkage disequilibrium (LD) present in the genetic resources. 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. There was rapid decline in LD with distance and the LD was <0.2 at a distance of 41635 bp. The marker trait association (MTAs) studies revealed 20 significant MTAs (p <0.001) with 11 markers for leaf area index (4), canopy temperature (13), chlorophyll content (1) and NDVI (2). The markers explained 2 to 21% 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. The duplications of effect observed could be due to common ancestor of the A and B genome which explains the linkage detected between markers lying on different chromosomes seen in the current study. The present study identified a total of 20 highly significant marker trait associations with 11 markers for four physiological traits of importance in groundnut; LAI, CT, SCMR and NDVI. 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 markers in this study may be useful for marker assisted selection after further validation.

**Keywords:** DArTseq; Groundnut; Linkage disequilibrium; Marker assisted selection; Marker trait association; Physiological traits

**Introduction**

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

Developing countries account for 96% groundnut areas and 92% of the global production. Despite the developing countries been the largest producers of groundnut, the average yield per hectare (China, 2490 kgha$^{-1}$ and Nigeria 840 kgha$^{-1}$) is low when compared to America (3673 kgha$^{-1}$) [1]. Among the factors contributing to low yield, drought adversely affects the crop's performance [2]. Drought stress has adverse influence on water relations, photosynthesis, mineral nutrition, metabolism, growth and yield of groundnut [3]. Plants, being sessile, have evolved specific acclimation and adaptation mechanisms to respond to and survive short-long-term drought stresses [4]. Traits that allow adaptation to water deficit include root traits, stomatal conductance, SPAD chlorophyll meter reading, leaf area, and canopy temperature [5].

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 efficiency of selection of target traits in groundnut [6–10]. Very few informative and good quality single nucleotide polymorphisms (SNPs) are available in peanut in contrast to availability of thousands of simple sequence repeats (SSRs) [9], though SNPs can be more easily generated than SSRs and SNPs are usually preferred due to low cost. In recent times, restriction site associated DNA sequencing (RADseq) [11] and genotyping by sequencing (GBS) [12] methods has allowed researchers to identify and genotype thousands of SNPs in plants.

Diversity Arrays Technology methodology (DArT) which is based on genome complexity reduction and SNP detection through hybridization of PCR fragments [13], has been used in genome-wide association studies (GWAS), construction of dense linkage maps and mapping quantitative trait loci (QTL) [7,14–16]. The DArTseq technology from DArT produces data on biallelic SNP markers as well as the older dominant DArT markers. GWAS discovers marker-trait association (MTA) involving representative markers and genetically diverse populations [17,18]. Furthermore, the magnitude of linkage disequilibrium (LD) present in genetic resources are important pre-requisites to deduce the genetic makeup, composition and genomic predictions of traits of interest during selection. 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 [19].

In this study, SNP markers were used for the genetic analysis of a groundnut minicore collection from International crop research institute for the semi-arid tropics (ICRISAT) for 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.

## Results

### Phenotypic evaluation

Weather data was not available for the test sites. However, drought is a common occurrence and each environment was exposed to drought though the extent can't not be quantified.

The groundnut mini core collection represents the global diversity of about 14,000 accessions conserved in the ICRISAT genebank. The result of the phenotypic evaluation showed highly significant differences (p<0.01) between lines for canopy temperature (CT), SPAD chlorophyll meter reading (SCMR) and normalized difference vegetative index (NDVI) but no significant genetic variation for leaf area index (LAI) (Table 1). The interaction between

lines and environment was significant (p<0.01) for only 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

correlation of SCMR with LAI and CT was negative and non-significant but was positive with

NDVI. 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).


**Marker Data**

The DArTseq genotyping produced 3,591 biallelic SNP markers of which 3,396 had a call rate

that exceeded 0.6999. The average polymorphism information content of the 3,396 markers

was 0.077.  Of the 3,396 markers, just 396 had a minor allele frequency that exceeded 0.05. A

total of 3,124 markers were given chromosome assignments:  368 (11.8%) were aligned to only

the A genome, 449 (14.4%) only to the B genome, and 2,308 (73.8%) to both genomes. Over

73% of the markers that aligned with both the A and B genomes were assigned to homeologous

chromosomes and the correlation of their position on those two set of homologues was 0.87. A

principal component (PC) analyses of the data from the 3,124 markers assigned to a

chromosome (s) did not reveal a strong discernible population structure in the first PC that

accounted for 61% of the variation (Figure 1). Cluster analysis of the marker data suggested

two groups of lines with one group having two subgroups (Figure 2).

The DArTseq genotyping produced 12,693 dominant *silico* markers with a call rate that

exceeded 0.70. Just 2,349 (18.5%) of these had a minor allele frequency (MAF) > 0.05. The

average polymorphism information content of the 2,349 markers was 0.070. A total of 12,611

markers were given chromosome assignments: 1,709 (13.6%) were aligned to only the A

genome, 2,502 (19.8%) only to the B genome, and 8,400 (66.7%) to both genomes. Over 76%

of the markers aligned with both the A and B genomes were assigned to homeologous

chromosomes and the correlation of their position on those two set of homologues was 0.91. This reflects the common origin of the A and B genomes.

There was some evidence for inter-genome exchange of genes between non homeologous chromosomes. A set of 46 markers were located in a 16,714,895 bp region of chromosome A08 and a 73,947,364 bp region of B07: The correlation of positions for the 46 markers was 0.769. A set of 20 markers were located in a 3,926,875 bp region of chromosome A02 and a 11,674,301 bp region of B09 (Figure S2 and S3): The correlation of positions for the 46 markers was -0.604 showing any putative exchange involved an inversion.

**Linkage disequilibrium**

Linkage disequilibrium analysis was conducted using 305,919 loci pairs within chromosomes. The LD was significant (P < 0.05) for 36.3% of loci pairs were (Figure 3, Table S1). Furthermore, 9,592 (3.14%) of the pairs were in complete LD ($r^2 = 1$). There was 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 P-value and $R^2$ ($r = -0.5381$), revealing the existence of linkage decay (Figure 4).

**Marker-trait association**

Due to the insignificant genotype by environment interaction the GWAS was performed using phenotypic best linear unbiased predictions (BLUPs) estimated over all environments. None of the 396 biallelic SNPs were significantly associated with any trait at P<0.001. Only the marker-trait associations (MTAs) 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 leaf area index (LAI). Marker M1 was associated with chromosomes A03

and B03 while M2 was associated with chromosomes A06 and B07. The individual markers explained 6.8 to 7.3% of the total phenotypic variation observed (Table 3). For CT, seven markers (M3-M9) identified 13 possible loci: six of these markers (M3-M8) identified regions on both the A and B genomes. The CT markers explained 9.6 to 16.6% of the variation observed. One marker (M10) on chromosome B05 was found to be associated with SCMR and explained 20.8% of the observed phenotypic variation. One marker (M11) identified regions of chromosome A04 and B02 associated with NDVI. The summaries of the number of SNP markers observed are presented in Table 4 and these may be useful for marker assisted selection after further validation. From all the 20 associations found, the B genome had 11 associations and the A genome had nine. Equal number of associations was found on each genome except for canopy temperature were A genome had six and B genome seven.

**Gene identification and description**

Most of the MTAs identified were located in genes that have functions corresponding to the gene description (Table 5). Two of the MTAs (M1_A03 and M1_B03) identified for LAI were residing in genes (*Aradu.02I1H* and *Araip.E67FB*, respectively) that both have the functions of transition of mitotic cell cycle and regulation of meristem structural organization. Nine out of the 13 MTAs identified for canopy temperature were located in the coding sequence of already identified for genes which are responsible primarily for phospholipase, ATP and shoot gravitopism activities (Table 5).

The marker identified for SCMR was located in *Araip.5TB63.1* gene which is responsible for zinc ion binding. For NDVI, the two MTAs identified, M11_AA04 and M11_B02, were located in *Aradu.DA17R* and *Araip.I1WMT* which regulates structural constituent of ribosome and jasmonic acid carboxyl methyltransferase, respectively. All the

genes identified in A genome and their corresponding chromosomes on the B genome were having the same descriptions.

**Discussion**

The phenotypic evaluation showed that the groundnut panel varied significantly for the physiological traits except for LAI (Table 1). The physiological traits, LAI, CT, SCMR and NDVI are important in improving productivity and are used as indirect indices for improving drought tolerance. Traits such as NDVI have been reported to be highly associated with yield and can be effective as in-season prediction of yield [20,21]. The heritability of the traits ranged from moderate to high. The most correlations between traits were significant though the values were low suggesting that the traits are fairly independent (Table 2).

The marker data suggested that the population was not highly structured. The SNP and *silico* markers gave similar results. The two groups of minicore collections and one group having two subgroups as suggested by the marker data may be due to different origins of the collections. Both types of markers produced more markers in the B genome than the A genome. Both markers systems assigned a similar portion of the markers to both genomes. While many polymorphic markers were detected, a large portion had MAF < 0.05 and the average PIC values for both types of markers was very low, about 0.07. Groundnut is known to have a low polymorphism rate and low genetic diversity [22–24] and our results support that conclusion.

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

## Linkage disequilibrium

The 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. There was a decline in LD with distance. Several studies have reported LD decay with distance [7,25,26].

## Marker-trait association

The MTA studies revealed 20 significant MTAs (p <0.001) involving 11 markers. Eight markers (M1-M8) identified possible MTA in both the A and B genomes. Our analysis cannot distinguish if just one or if both chromosomes identified by such markers are affecting the trait. Two of these markers (M2 and M7) found MTA on non-homeologous chromosomes. It is possible that some chromosome rearrangement has caused the marker sequence to appear on different homeologous, or is possible that there is some error in the sequencing or bioinformatics alignment process. Validation studies will be needed to see if these markers are identifying one locus or perhaps a locus duplicated in the two genomes.

Pandey et al. [7] used SSR markers to identify some significant MTAs for physiological traits in groundnut including LAI and SCMR. These two traits were also observed in the present study. Four MTA were identified for LAI among which the MTA on chromosomes A06 and B07 may be the same as those previously identified by Pandey et al. [7] for total leaf area and leaf area respectively. Marker M2 identified MTAs for LAI at similar positions on chromosome A06 and B07. Thirteen (13) MTAs were identified for CT among which six markers were located in both A and B genome and four were identified on the same chromosome in A and B genome. One MTA was found for SCMR and two for NDVI. The MTA identified on B05 is

different from the previously reported loci of A06 [7]. The marker associated with NDVI was located on both the A and B genome but on different chromosomes.

From the result of the MTAs analysis, most of the MTAs identified on the A subgenome were also identified on the respective homeologous chromosome on the B subgenome. Argawal et al. [10] reported that a significant proportion of marker loci with assigned physical locations to chromosome of one genome were mapped to respective homeologous positions on chromosomes of the other genome. Our results support this 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. [10]. Other homeologous MTAs in the present study are 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 [27]. We found some evidence for genetic exchanges occurring between the groundnuts genomes as reported earlier [24]. Also, a large number of similar markers were placed on the genetic map on different chromosomes. The possible translocation we noted do not appear to be terminal or reciprocal unlike the translocations noted by Farre et al. [28]. Translocations of markers have been previously reported in groundnut [10,27]. Some of these observed 'translocated' markers might be also due to miss-assignments because of the highly repetitive structure of groundnut genome [24].

The markers identified for LAI are needed for cell expansion and growth in plant leafs. Markers identified for canopy temperature were linked to genes associated with phospholipase, shoot gravitopism and ATP productions. In response to a cold exposure, it was shown that the phospholipase D (PLD) and the phospholipase C (PLC)/diacylglycerol kinase pathways are simultaneously activated [29]. Natalini et al. [30] showed that temperature significantly affect the production of PLD and PLC in tomato plants. Furthermore, Ryu et al. [31] and Kim et al. [32]

reported that shoot gravitopism 5 (*SGR5*) alternative splicing is accelerated at high temperatures, resulting in a high-level accumulation of SGR5β proteins. They proposed that the thermos-responsive alternative splicing of *SGR5* gene provides an adaptation strategy by which plants protect the shoots from aerial heat frequently occurring in natural habitats.

The marker identified for SCMR was located in a gene which is responsible for zinc ion binding. Zinc compounds have been shown to increase chlorophyll content by at least 50% in cilantro plant [33]. Plant growth, chlorophyll contents, crude proteins and zinc contents were noted to be higher when greater supply of zinc doses was applied [34].

The two genes identified for NDVI regulates structural constituent of ribosome and jasmonic acid carboxyl methyltransferase respectively. Defective ribosome biogenesis inhibits cell division, cell expansion, and pavement cell differentiation [35]. The depletion further downregulates auxin and up regulates Jasmonic acid (JA) [35]. It can therefore be proposed that when the nutrient condition is sufficient for the plants, the increased ribosome biogenesis will enhance cell division and expansion thereby promoting rapid growth. But if the nutrient condition is limited, then there will be an increase in JA (which causes leaf senescence) and it is associated with upregulation of the gene in the B genome.

The present study identified a total of 20 highly significant marker trait associations for four physiological traits of importance in groundnut; LAI, CT, SCMR and NDVI. 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 the significant markers identified in the present study using a larger population.

**Materials and Methods**

The research was carried out in the research field of ICRISAT located at Minjibir (2015) and Bayero University (2017) that are both in Kano State. The annual mean rainfall of both location is between 800 mm to 900 mm; and variations about the annual mean values are up to ±30%.

One hundred and twenty five groundnut minicore core collections including five check varieties were evaluated using randomized incomplete block design with three (3) replications. Each plot in a replication consisted of one row measuring 5m in length with an inter and intra row spacing of 75cm and 10cm respectively with a 1m alley between replications.

Crop specific practices such as plant protection were followed in the sowing process. Fertilizer application was done alongside planting. A basal application of Nitrogen, Phosphorous, and Potassium was done to all plots at the rate of 20 N kg/ha, 40 $P_2O_5$ kg/ha and 40 $K_2O$ kg/ha at planting. Hand weeding was done using hoes at $3^{rd}$, $8^{th}$ and $12^{th}$ week after planting to prevent weed infestation and competition between plants and weeds.

Data was collected for CT using leaf thermometer (Meco IRT550 Infrared Thermometer), SCMR using SPAD 502 PLUS Chlorophyll meter, NDVI using Hand Held optical sensor unit (Model 505 NTech Industries, Inc) and LAI was taken by the use of Leaf Area Meter (ACCUPAR LP-80). The data were collected at 8 weeks after sowing from five randomly selected plants in each plot.

**DNA extraction and genotyping**

Groundnut leaves were collected into 96 deep well samples 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. DNA extraction was done using Nucleomag Plant Genomic DNA extraction kit. The genomic DNA extracted was in the

range of 50-100ng/ul. DNA quality and quantity were checked on 0.8% agarose. Libraries were constructed according to Kilian et al. [36]. 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 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 inhouse marker scoring pipeline based on algorithms. Two types of DArTseq markers were scored, SilicoDArT markers (scored as presence or absence, 1,0) and biallelic SNP markers which were both scored for presence of the reference allele, the alternative allele, or both. In 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) to identify chromosome.

**Data Analysis**

The analysis of phenotypic data was done to obtain BLUP by fitting the following mixed linear model in R package "lme4".

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

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 $k^{th}$ rep nested in the $j^{th}$ environment, $ge_{ij}$ is the genotype by environment interaction and $error_{ijk}$ is the error associated with the observation. Replication and environments were considered random effects. The analysis was run in R with the lme4 package [37].

Entry mean heritability was calculated as

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

Where $\sigma_g^2$ is the variance among lines, is the $\sigma_{ge}^2$ genotype by environment interaction variance,

$\sigma_{error}^2$ is the error variance, r is the number of replications, and e is the number of environments.

**Linkage disequilibrium and marker trait association**

The parameter $r^2$ was used to estimate LD between SNPs on each chromosome via the software

package TASSEL 5.0 [38]. Marker trait association analysis, probability values, and % of the

variation modelled by the markers were calculated using the GAPIT package via the

KDCompute interface (https://kdcompute.igss-africa.org/kdcompute/home). The first three

principal components and the relationship matrix were included in the model. SNPs with MAF

<5% and missing data >20% were excluded from the analyses. Missing values were imputed

using the choice of nearest neighbor algorithm using TASSEL 5.0 [38].

We used the unweighted pair-group method to cluster the lines and form a dendrogram.

We also used the marker data in a principal component analysis. Both of these analyses were

executed using KDCompute.

**Gene identification and description**

The identified markers were searched at peanut base (https://www.peanutbase.org) to identify

genes in the regions and their descriptions were retrieved from legume mine

(https://mines.legumeinfo.org).

**Supplementary materials**

Table S1: LD of markers for A and B genomes

Table S2: Significant marker trait associations

Figure S1: Histogram of BLUP of leaf area index (LAI), canopy temperature (CT), chlorophyll

content (SPAD) and NDVI from the groundnut minicore collection.

Figure S2: Distribution of markers on A genome of the groundnut accessions

Figure S3: Distribution of markers on the B genome of the groundnut accessions

## Acknowledgement

## Authors' contributions

ASS assisted in planning, data analysis and drafting of manuscript. CS analysed some data and revised the manuscript. BNM planned the field work and revised the manuscript. JC did the genotyping. MC did the bioinformatics. ZLM assisted in data processing. AMI assisted in data analysis and manuscript revision. HAA revised the manuscript. SGM revised the manuscript. All authors read and approved the final manuscript.

## Conflict of interest

The authors declare no conflict of interest.

## References

(1)     FAOSTAT. Crops. Food and Agriculture Organization of United Nations 2017.

(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).

(3)     Suther  M.S., D. M. and P. 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)     Krishnan, A.; Pereira, A. Integrative Approaches for Mining Transcriptional Regulatory Programs in Arabidopsis. *Br. Funct Genomic Proteomic* **2008**, *7* (4), 264–274. https://doi.org/10.1093/bfgp/eln035.

(5)     Price AH  Horton P, Jones HG, Griffiths H., C. 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.

(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* (3), 639–651. https://doi.org/10.1016/j.biotechadv.2011.11.001.

(7)     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* (8), e105228. https://doi.org/10.1371/journal.pone.0105228.

(8)     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* (1), 13–25.

(9)     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* (8), 1120–1134. https://doi.org/10.1016/j.biotechadv.2013.01.001.

(10)    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* (11), 1954–1967. https://doi.org/10.1111/pbi.12930.

(11)    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* (16), 2187–2193. https://doi.org/10.1093/bioinformatics/btr346.

(12)    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* (5), e19379. https://doi.org/10.1371/journal.pone.0019379.

(13)    Jaccoud, D.; Peng, K.; Feinstein, D.; Kilian, A. Diversity Arrays: A Solid State Technology for Sequence Information Independent Genotyping. *Nucleic Acids Res* **2001**, *29* (4), E25. https://doi.org/10.1093/nar/29.4.e25.

(14)    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 Genomics* **2017**, *18* (1), 423. https://doi.org/10.1186/s12864-017-3805-4.

(15)    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* (1). https://doi.org/10.1186/s12870-017-1140-1.

(16) 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* (1), 246. https://doi.org/10.1186/s12870-018-1475-2.

(17) 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* (4), 791–807. https://doi.org/10.1007/s00122-013-2257-8.

(18) 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* (2), 363–380. https://doi.org/10.2135/cropsci2008.07.0420.

(19) Baird, S. J. Exploring Linkage Disequilibrium. *Mol Ecol Resour* **2015**, *15* (5), 1017–1019. https://doi.org/10.1111/1755-0998.12424.

(20) 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. https://doi.org/10.3389/fpls.2016.01131.

(21) 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. https://doi.org/10.3389/fpls.2018.00081.

(22) 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 Genomics* **2012**, *287* (1), 21–38. https://doi.org/10.1007/s00438-011-0656-6.

(23) 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* (1), 113–126. https://doi.org/10.1093/aob/mcs237.

(24) 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* (4), 438–446. https://doi.org/10.1038/ng.3517.

(25) Neumann K  Denčić S, Varshney R, Bo¨rner A, K. B. Genome-Wide Association Mapping: A Case Study in Bread Wheat (Triticum Aestivum L.). *Mol. Breed.* **2011**, *27* (1), 37–58.

(26) 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* (2), e0171692. https://doi.org/10.1371/journal.pone.0171692.

(27) 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_ (4), 653–664. https://doi.org/10.1007/s00122-011-1737-y.

(28)   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_ (5), 1029–1037. https://doi.org/10.1007/s00122-010-1507-2.

(29)   Vergnolle, C.; Vaultier, M.-N.; Taconnat, L.; Renou, J.-P.; Kader, J.-C.; Zachowski, A.; Ruelland, E. The Cold-Induced Early Activation of Phospholipase C and D Pathways Determines the Response of Two Distinct Clusters of Genes in Arabidopsis Cell Suspensions. _Plant Physiol._ **2005**, _139_ (3), 1217. https://doi.org/10.1104/pp.105.068171.

(30)   Natalini, A.; Vanesa, M.-D.; Ferrante, A.; Pardossi, A. Effect of Temperature and Ripening Stages on Membrane Integrity of Fresh-Cut Tomatoes. _Acta Physiol. Plant._ **2013**, _36_ (1), 191–198. https://doi.org/10.1007/s11738-013-1399-2.

(31)   Ryu, J. Y.; Kim, J.-Y.; Park, C.-M. Adaptive Thermal Control of Stem Gravitropism through Alternative RNA Splicing in Arabidopsis. _Plant Signal. Behav._ **2015**, _10_ (11), e1093715–e1093715. https://doi.org/10.1080/15592324.2015.1093715.

(32)   Kim, J. Y.; Ryu, J. Y.; Baek, K.; Park, C. M. High Temperature Attenuates the Gravitropism of Inflorescence Stems by Inducing SHOOT GRAVITROPISM 5 Alternative Splicing in Arabidopsis. _New Phytol_ **2016**, _209_ (1), 265–279. https://doi.org/10.1111/nph.13602.

(33)   Reddy Pullagurala, V. L.; Adisa, I. O.; Rawat, S.; Kalagara, S.; Hernandez-Viezcas, J. A.; Peralta-Videa, J. R.; Gardea-Torresdey, J. L. ZnO Nanoparticles Increase Photosynthetic Pigments and Decrease Lipid Peroxidation in Soil Grown Cilantro (Coriandrum Sativum). _Plant Physiol Biochem_ **2018**, _132_, 120–127. https://doi.org/10.1016/j.plaphy.2018.08.037.

(34)   Samreen, T.; Humaira; Shah, H. U.; Ullah, S.; Javid, M. Zinc Effect on Growth Rate, Chlorophyll, Protein and Mineral Contents of Hydroponically Grown Mungbeans Plant (Vigna Radiata). _Arab. J. Chem._ **2017**, _10_, S1802–S1807. https://doi.org/https://doi.org/10.1016/j.arabjc.2013.07.005.

(35)   Ahn, C. S.; Cho, H. K.; Lee, D.-H.; Sim, H.-J.; Kim, S.-G.; Pai, H.-S. Functional Characterization of the Ribosome Biogenesis Factors PES, BOP1, and WDR12 (PeBoW), and Mechanisms of Defective Cell Growth and Proliferation Caused by PeBoW Deficiency in Arabidopsis. _J. Exp. Bot._ **2016**, _67_ (17), 5217–5232. https://doi.org/10.1093/jxb/erw288.

(36)   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. https://doi.org/10.1007/978-1-61779-870-2_5.

(37)   Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Usinglme4. _J. Stat. Softw._ **2015**, _67_ (1). https://doi.org/10.18637/jss.v067.i01.

(38)   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_ (19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308.

Table 1: Mean squares of four physiological traits of groundnut minicore collections

| Effect | df | Canopy temperature | Leaf area index | SPAD chlorophyll reading | NDVI |
|---|---|---|---|---|---|
| Rep(Year) | 2 | 679.62 | 7.28 | 18.73 | 0.47 |
| Year | 1 | 4820.38** | 54.69** | 7149.59** | 13.81** |
| Lines*Year | 124 | 7.69 | 0.49 | 26.09** | 0.001 |
| Error | 214 | 7.67 | 0.84 | 15.42 | 0.02 |
| Heritability | | 0.22 | 0.03 | 0.55 | 0.25 |

** = significant at 0.01 level of probability

Table 2: Correlation between physiological traits of groundnut minicore collections (above value = correlation and below value are p levels)

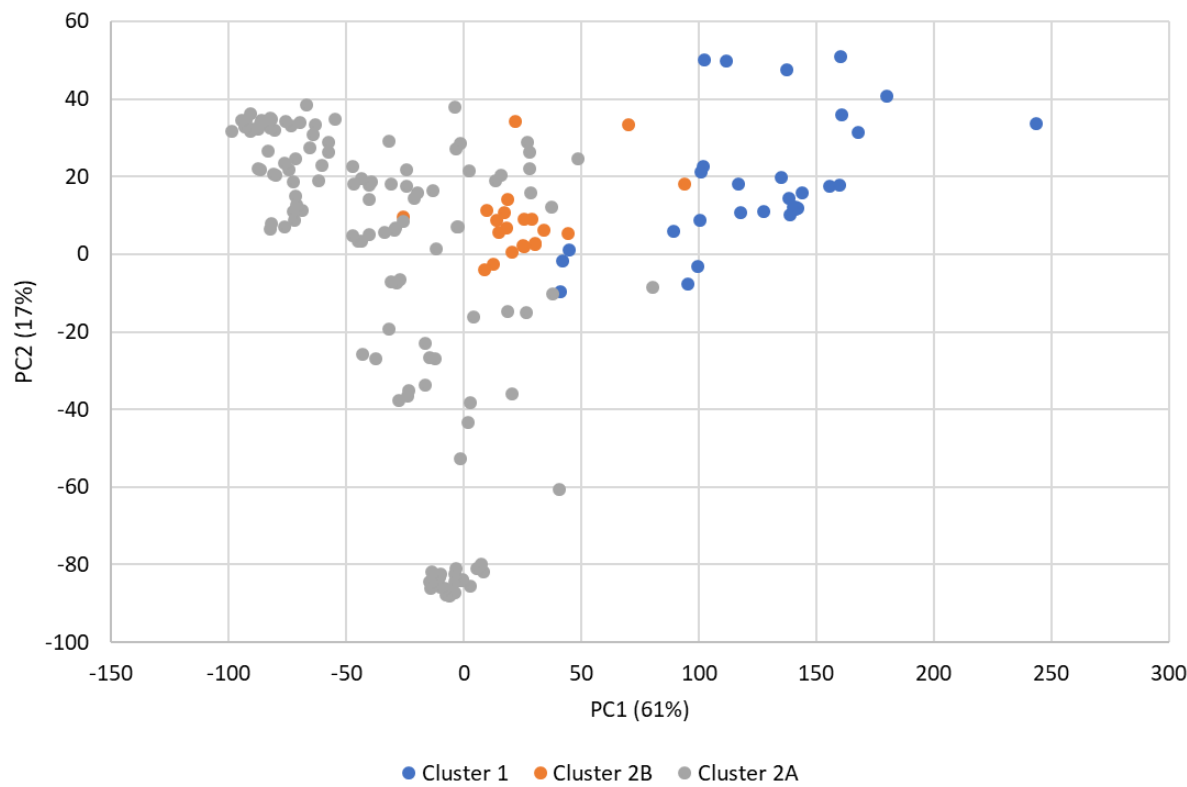| | SPAD chlorophyll reading | Leaf area index | Canopy temperature | NDVI |
|---|---|---|---|---|
| SPAD chlorophyll reading | 1 | | | |
| Leaf area index | -0.032 | 1 | | |
| | 0.72 | | | |
| Canopy temperature | -0.16 | -0.191 | 1 | |
| | 0.075 | 0.033 | | |
| NDVI | 0.081 | 0.117 | -0.245 | 1 |
| | 0.37 | 0.193 | 0.006 | |

Figure 1:  Graph of the scores of the first two principal components of the analysis of the marker data from the groundnut minicore collection.
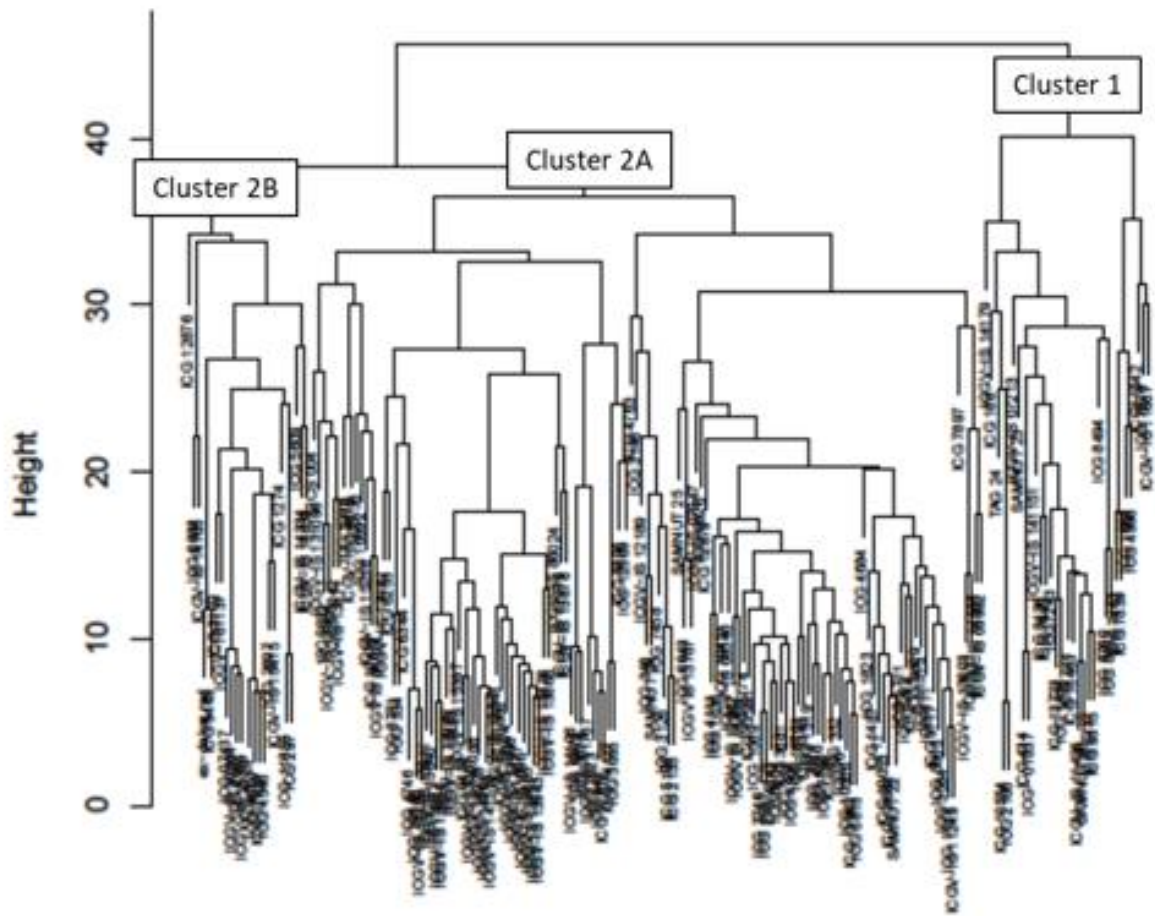
Figure 2.  Dendrogram from unweighted pair-group clustering of accession from the minicore
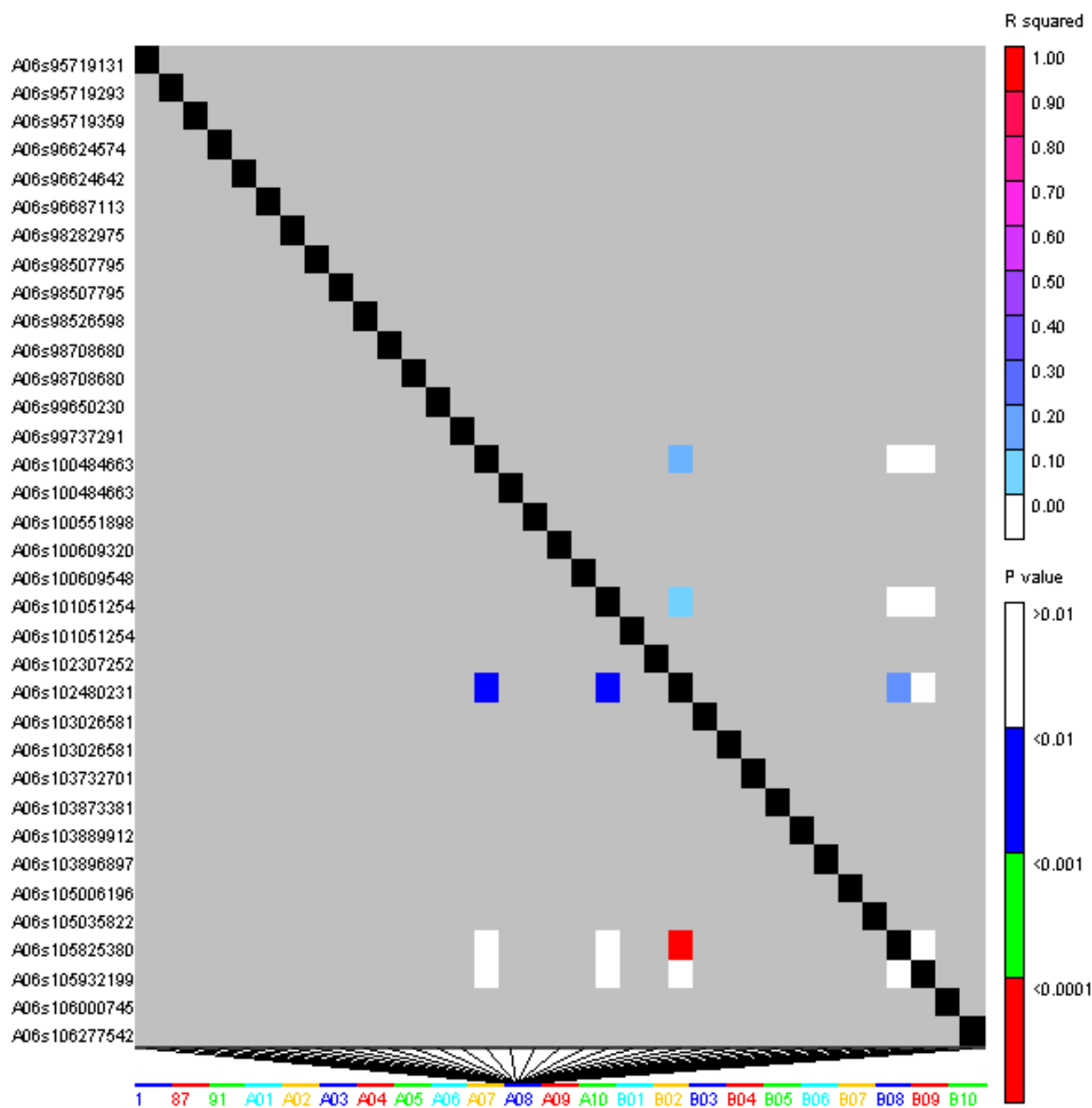
collections

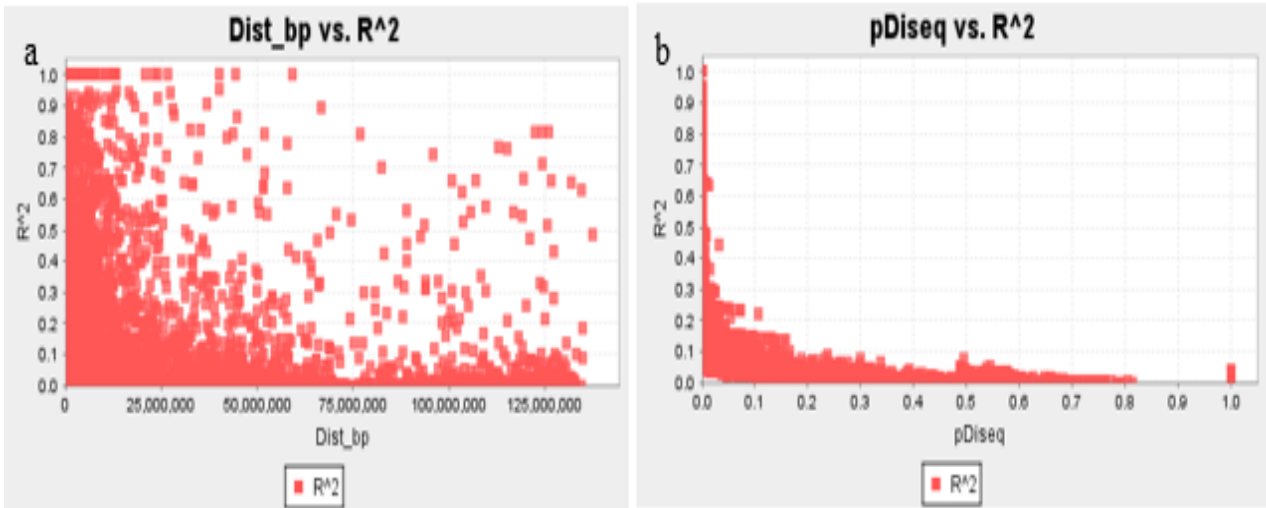Figure 3: Linkage disequilibrium among markers scored in the groundnut minicore collection

Figure 4: Scatter plot showing association between linkage disequilibrium ($r^2$) and distance (a)

and significance of the $r^2$ value (b)

Table 3: Marker trait associations of four physiological traits of groundnut in the groundnut minicore collection

| Trait | | Marker | Chromo-some | Position | P value | Marker R$^2$ |
|---|---|---|---|---|---|---|
| Leaf area index | M1 | 100028731 | A03 | 112302037 | 0.0007 | 0.073 |
| | M1 | 100028731 | B03 | 114699750 | 0.00084 | 0.068 |
| | M2 | 100055743 | A06 | 29592787 | 0.00089 | 0.069 |
| | M2 | 100055743 | B07 | 39388361 | 0.00083 | 0.068 |
| Canopy temperature | M3 | 100002202 | A02 | 14182722 | 0.00099 | 0.098 |
| | M3 | 100002202 | B02 | 17437688 | 0.00089 | 0.097 |
| | M4 | 100006533 | A03 | 12024382 | 2.06E-05 | 0.166 |
| | M4 | 100006533 | B03 | 14679916 | 2.19E-05 | 0.163 |
| | M5 | 100057474 | A05 | 4009729 | 0.00017 | 0.128 |
| | M5 | 100057474 | B05 | 3919472 | 0.00018 | 0.125 |
| | M6 | 100007799 | A05 | 4754635 | 0.00078 | 0.102 |
| | M6 | 100007799 | B05 | 4840242 | 0.00046 | 0.109 |
| | M7 | 100008002 | A08 | 7712880 | 0.00097 | 0.098 |
| | M7 | 100008002 | B07 | 113507761 | 0.00057 | 0.105 |
| | M8 | 100007218 | A10 | 78514031 | 0.00069 | 0.104 |
| | M8 | 100007218 | B10 | 103607600 | 0.00062 | 0.104 |
| | M9 | 100002387 | B01 | 135556895 | 0.00097 | 0.096 |
| SPAD chlorophyll meter reading | M10 | 100000269 | B05 | 16995186 | 0.00099 | 0.208 |
| NDVI | M11 | 100118763 | A04 | 67543304 | 0.00069 | 0.070 |
| | M11 | 100118763 | B02 | 106735216 | 0.001 | 0.066 |

Table 4: Summary of DArT markers with high associations with four physiological traits of groundnut

| Trait | No of MTAs | Chromosomes | Markers |
|---|---|---|---|
| Leaf area index | 4 | A03, A06, B03, B07, | 100055743 (2), 100028731 (2) |
| Canopy temperature | 13 | A02, A03, A05 (2), A08, A10, B01, B02, B03, B05 (2), B07, B10 | 100002202 (2), 100006533 (2), 100007218 (2), 100007799 (2), 100008002 (2), 100057474 (2), 100002387 |
| Chlorophyll content | 1 | B05 | 100000269 |
| NDVI | 2 | A04, B02 | 100118763 (2) |

Table 5: Gene descriptions for identified marker trait associations of four physiological traits of groundnut

| Trait | Marker | Chromo-some | Gene | Description |
|---|---|---|---|---|
| Leaf area index | M1 | A03 | Aradu.02I1H | Transition of mitotic cell cycle; regulation of meristem structural organization |
| | M1 | B03 | Araip.E67FB | Transition of mitotic cell cycle; regulation of meristem structural organization |
| | M2 | A06 | - | - |
| | M2 | B07 | - | - |
| Canopy temperature | M3 | A02 | - | - |
| | M3 | B02 | - | - |
| | M4 | A03 | Aradu.3X3DN | shoot gravitropism 2 (SGR2) |
| | M4 | B03 | - | - |
| | M5 | A05 | Aradu.UM7P3 | phospholipase D P2 |
| | M5 | B05 | Araip.5S0ML | phospholipase D P2 |
| | M6 | A05 | Aradu.QD3GI | HXXXD-type acyl-transferase family protein |
| | M6 | B05 | Araip.4G4HV | HXXXD-type acyl-transferase family protein |
| | M7 | A08 | Aradu.C4U4P | ATP-dependent clp protease ATP-binding subunit |
| | M7 | B07 | Araip.B1GNH | ATP-dependent clp protease ATP-binding subunit |
| | M8 | A10 | Aradu.VFV5Q | Putative endonuclease or glycosyl hydrolase |
| | M8 | B10 | - | - |
| | M9 | B01 | Araip.Z7V9S | GPI mannosyltransferase-like protein |
| SPAD chlorophyll meter reading | M10 | B05 | Araip.5TB63.1 | Zinc ion binding |
| NDVI | M11 | A04 | Aradu.DA17R | Structural constituent of ribosome |
| | M11 | B02 | Araip.I1WMT | Jasmonic acid carboxyl methyltransferase |