biology and life sciences forum

MDPI

Proceedings Paper

Genome-Wide Association Analysis of Yield-Related Traits of Soybean Using Haplotype-Based Framework †

Kehinde Adewole Adeboye [1,*], Javaid Akhter Bhat [2], Showkat Ahmad Ganie [3], Rajeev K Varshney [4,5] and Deyue Yu [2]

[1] Department of Agricultural Technology, Ekiti State College of Agriculture and Technology, Isan-Ekiti 371106,Nigeria
[2] Soybean Research Institution, National Center for Soybean Improvement, Key Laboratory of Biology and Genetics and Breeding for Soybean, Ministry of Agriculture, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China; javid.akhter69@gmail.com (J.A.B.); dyyu@njau.edu.cn (D.Y.)
[3] Plant Molecular Science and Centre of Systems and Synthetic Biology, Department of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK; showkatmanzoorforever@gmail.com
[4] Center of Excellence in Genomics & Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India; r.k.varshney@cgiar.org
[5] State Agricultural Biotechnology Centre, Crop Research Innovation Centre, Food Futures Institute, Murdoch University, Murdoch, Western Australia 6150, Australia
* Correspondence: adeboye@daad-alumni.de; Tel.: +2348061698998

† Presented at the 2nd International Electronic Conference on Plant Sciences—10th Anniversary of Journal Plants, 1–15 December 2021; Available online: https://iecps2021.sciforum.net/.

Citation: Adeboye, K.A.; Bhat, J.A.; Ganie, S.A.; Varshney, R.K. Genome-Wide Association Analysis of Yield-Related Traits of Soybean Using Haplotype-Based Framework. *Biol. Life Sci. Forum* **2021**, *1*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Carmen Arena

Published: 2 December 2021

Abstract: Haplotype-based breeding involving multi-marker association analysis is a promising approach to developing custom-designed high-yielding crop varieties. Here, we reported multi-marker association analysis for the number of pods per plant (PNP), the number of seeds per plant (SNP), 100-seed weight (HSW), and seed yield per plant (SYP) using 211 cultivated soybean accessions. The field experiment was conducted across six environments following a randomized complete block design with three replications. Genome-wide association study (GWAS) explored 12,617 single-nucleotide polymorphism (SNP) markers from NJAU 355K SoySNP Array to identify significant marker associations for the studied traits across the six environments. Six markers that were consistently associated with the yield traits in two or more environments were considered stable and selected as the reference markers for building haplotype block/loci. The multi-marker association analysis within the haplotype-based framework revealed various allelic combinations regulating the phenotypic variations for the studied yield-related traits in soybean. These haplotype alleles may serve as genomic resources in breeding programmes aimed at improving the yield potential of soybean.

Keywords: GWAS; haplotype analysis; SNP; soybean; yield-related traits

1. Introduction

Yield characters are complex quantitative traits which posed some difficulties to breeding efforts. Analysis of family linkage map and linkage disequilibrium among unrelated individuals have been widely explored for the understanding of the genetic basis of complex quantitative traits, such as the yield characters in several plant species, including soybean [1,2]. These procedures represent the genome-wide studies of these characters for the identification of marker-trait association using single marker analysis. Recently, haplotype-based breeding has emerged as a promising approach to develop custom-designed crop varieties. It involves the identification and exploration of superior alleles from combination of many markers within a locus associated with the traits of interest.

Haplotype analysis has great potential in crop improvement programmes. It allows plant breeders to maximize the genetic variation underlying complex gene actions in a given locus. In soybean, Patil et al. [3] conducted haplotype analysis for candidate gene regulating salinity tolerance (GmCHX1). They identified various haplotypes for GmCHX1, including SV-2 which provide maximum salinity tolerance in soybean. Also, Wang et al. [4] identified superior haplotypes for grain quality, traits such as cooking and eating quality traits, in rice. Abbai et al. [5] performed haplotype analysis in rice 3K panel for 120 genes, and identified desirable haplotypes for agronomically important traits. Similarly, the five candidate genes regulating phenotypic performance of the direct-seeded rice were subjected to haplotype analysis [6] (Chen et al. 2019). Sinha et al. [7] performed haplotype analysis of five genes controlling drought tolerance in pigeonpea.

Furthermore, using haplotypes for QTL mapping could compensate for several limitations of single SNPs, including their biallelic nature, and substantially improve the efficiency of QTL mapping [8]. Moreover, haplotype-traits association analyses are helpful for precise mapping of important genomic regions and location of favor alleles or haplotypes for breeding [9].

The present work is aimed at identifying superior combinations of alleles within the haplotype-based framework for yield-related traits of soybean in different environments.

2. Materials and Methods

2.1. Plant Materials and Field Experiment

A panel 211 diverse genotypes were selected from widely cultivated soybean germplasm across wide geographic areas, including the Peoples' Republic of China and the United States of America [10]. The selected genotypes were phenotyped for two years at three locations (six environments), including the Experimental Field of Nanjing Agricultural University in Nanjing (E1 and E2), the Experimental Field of Jiangsu Yanjiang Institute of Agricultural Sciences in Nantong (E3 and E4) and the Experimental Farm of the Agricultural College of Yangzhou University in Yangzhou (E5 and E6). In each of the environments, the genotypes were planted in a randomized complete block design (RBD) with three replications. Each genotype was planted in three rows per plot, each row 200 cm-long and with 50 cm row spacing. Normal agronomic cultural practices were followed for the cultivation of the soybean germplasm at each location as previously described by Zhang et al. [11]. Phenotypic data were recorded for yield-related traits including number of pods per plant (PNP), number of seeds per plant (SNP), 100-seed weight in grams (HSW), and seed yield per plant in grams (SYP).

2.2. Genome-Wide Haplotype Association Analysis

Genome-wide association study (GWAS) explored 12,617 single-nucleotide polymorphism (SNP) markers from NJAU 355K SoySNP Array to identify significant marker associations for the studied traits across the six environments. GWAS was conducted using five different statistical models, including general linear model (GLM) with PCA [12] (Price et al. 2006), compressed mixed linear model (CMLM) [13] (Zhang et al. 2010), multiple-locus mixed linear model (MLMM) [14] (Segura et al. 2012), fixed and random model circulating probability unification (FarmCPU) [15] (Liu et al. 2016) and Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) [16] (Huang et al. 2019). Population structure was corrected with principal component analysis (PCA) using Bayesian information criterion (BIC) to estimate the optimal numbers of PCA [12,17] (Schwarz, 1978; Price et al. 2006).

Haplotype analysis was conducted using PLINK v1.07 [18]. The stable markers were considered as reference markers for building haplotype block/loci. All markers that are in proxy association with the reference markers within the LD decay distance ±670 Kbp made up a haplotype block/locus. The contribution of each haplotype to the observed phenotypic variance across the environment was estimated using the "--hap-assoc" command.

3. Results and Discussion

In practical breeding, the understanding of the genetics underlying traits of interest is the ultimate objective. In this study, genome-wide association study identified a total of 57 significant markers underlying the studied traits across six individual environments plus the combined environment (Figures 1 and 2). These were distributed across 18 of the 20 soybean chromosomes, indicating a complex genetic control of these traits as similarly reported by Li et al. [19] and Hu et al. [20]. The highest number of significant markers/QTLs were detected on the Chr.15 (10) followed by Chr.20 (8) and Chr.11 (5), respectively. Four are found each on Chr.04, Chr.06 and Chr.13 while three each are located on Chr.08 and Chr.12.

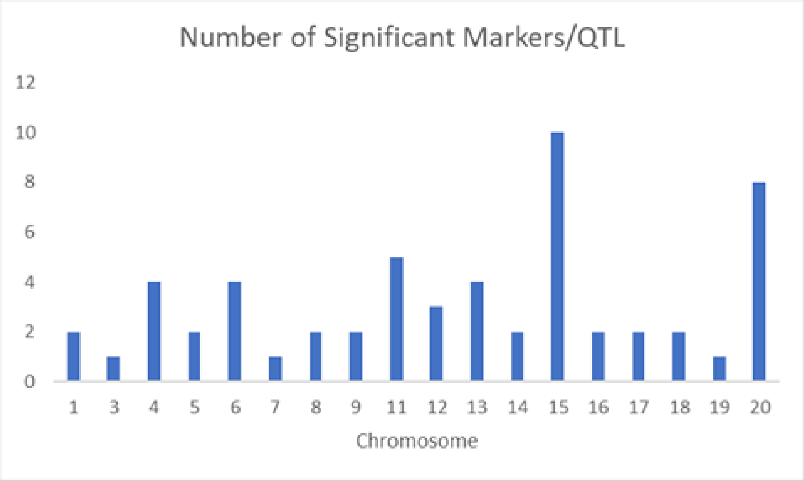

Figure 1. Distribution of significant markers/QTL across the soybean chromosomes.

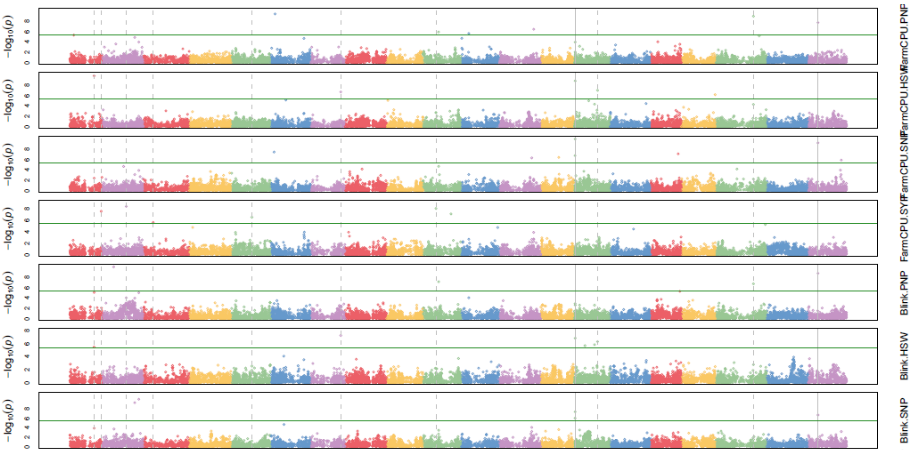

Figure 2. Mahanatan plot showing the significant association of markers with yield-related traits in the combined environment based on the five GWAS models: BLINK, CMLMM, FarmCPU, GLM and MLMM.

Furthermore, in many studies, stable genomic regions or quantitative trait loci (QTL) are defined by markers consistently associated with a given trait across multiple environments or genetic backgrounds [21,22]. In the present study, stable genomic regions were found for three of the studied traits, including HSW, SNP and PNP on chromosomes 4, 5, 11, 13, 18 and 20 (Table 1). The stable QTL on chromosomes 11 and 13 were associated with both HSW and SNP while those on chromosome 4 and 20 were associated with PNP and SNP. The stable QTL on chromosome 5 is associated with HSW and the one on chromosome 18 is associated with HSW and PNP. The stable QTLs for 100-seed weight on chromosome 5 and 11 have been respectively reported by Han et al. [23] and Du et al. [24], and Han et al. [23].

Table 1. Stable QTLs/genomic regions identified for the yield-related traits consistently across the environments.

QTL/Marker	Chromosome	Physical Position (bp)	Trait (Environment)	Related QTL
AX-93703924	4	4291705	SNP (COM and E6); PNP (E3)	No related QTL
AX-93922099	5	36599702	HSW (COM, E1 and E5)	*Seed weight 34-9* [17]; *Seed-yield 22-10* [18]
AX-93793210	11	29587057	HSW (COM, E1, E3 and E4); SNP (E2, E3 and E5)	*Seed weight 35-9* [17]
AX-93807406	13	1843185	HSW (COM, E1, E2, E4 and E5); SNP (COM, E1 and E6)	No related QTL
AX-94176727	18	46137043	PNP (COM and E1); HSW (E2)	No related QTL
AX-94199992	20	12095298	PNP (COM and E3); SNP (COM and E1)	No related QTL

Based on the haplotype-based framework, we conducted multimaker association analyses using the stable markers as reference loci for the identification of superior allele combinations underlying the studied traits. Superior haplotype alleles for agronomically important traits have been reported in several crop species [5–7,25–27]. Our study revealed various allelic combinations regulating the phenotypic variations for the studied yield-related traits in soybean. Figures 3–5 highlights the haplotype alleles and the proportion of Phenotypic variance contributed by these haplotypes to the associated traits across the 6 environments.

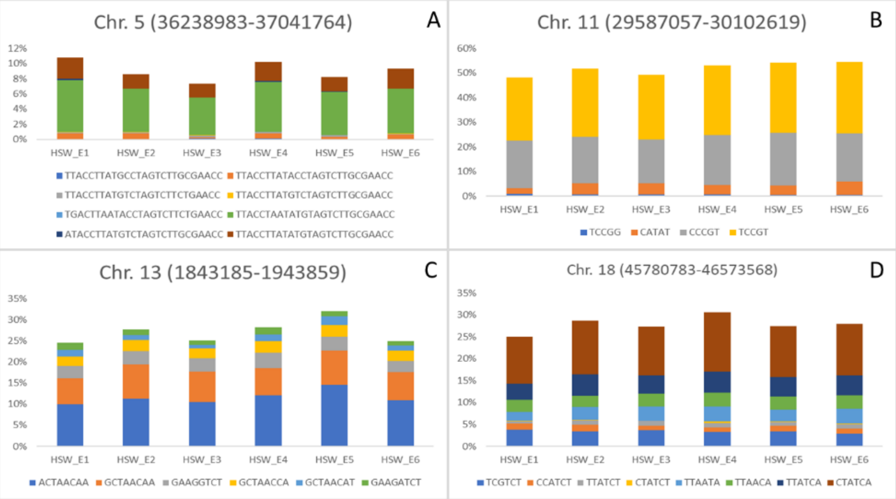

Figure 3. Haplotype alleles within the loci on chromosome 5 (**A**), 11 (**B**), 13 (**C**) and 18 (**D**), and their contribution to the phenotypic variation of hundred seed weight across the environments.

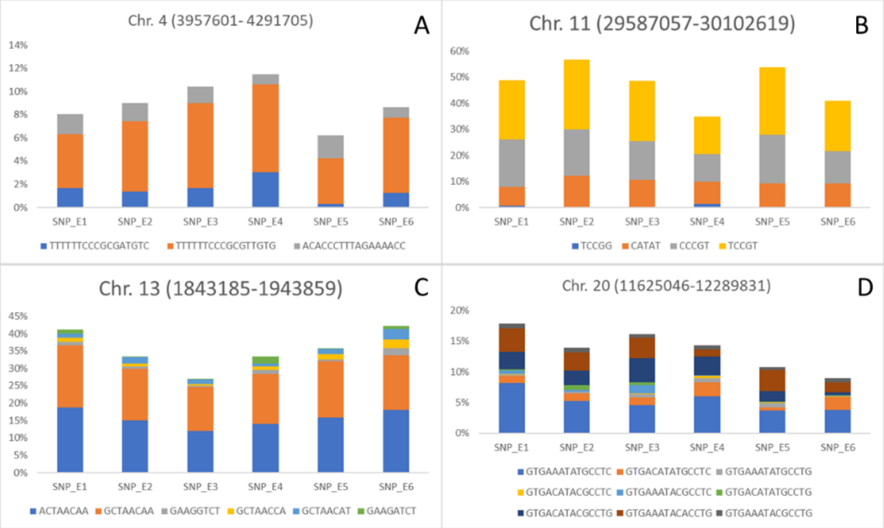

Figure 4. Haplotype alleles within the loci on chromosome 4 (**A**), 11 (**B**), 13 (**C**) and 20 (**D**), and their contribution to the phenotypic variation of seed number per plant across the environments.

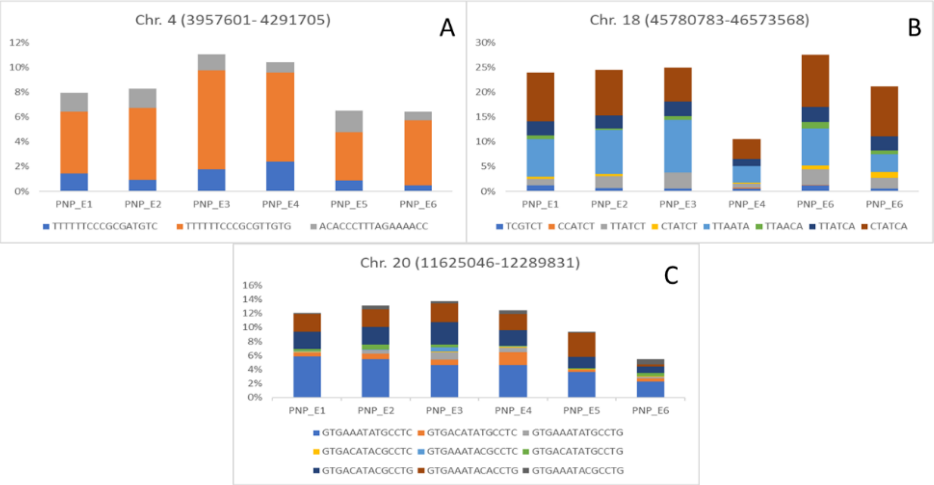

Figure 5. Haplotype alleles within the loci on chromosome 4 (**A**), 18 (**B**) and 20 (**C**), and their contribution to the phenotypic variation of Panicle number per plant across the environments.

4. Conclusions

The six stable QTL/Markers and the haplotype alleles identified in the present study may serve as genomic resources in breeding programmes aimed at improving the yield potential of soybean.

Author Contributions: Conceptualization, J.A.B., D.Y., and R.K.V.; methodology, J.A.B., D.H.; formal analysis, K.A.A.; data curation, J.A.B., K.A.A., S.A.G., D.H.; visualization, J.A.B., K.A.A., S.A.G., D.H.; writing—original draft preparation, K.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: This work was supported in part by National Natural Science Foundation of China (No. 32090065), Ministry of Science and Technology (No. 2017YFE0111000) and Horizon 2020 of European Union (No. 727312).

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

References

1. Karikari, B.; Chen, S.; Xiao, Y.; Chang, F.; Zhou, Y.; Kong, J.; Bhat, J.A.; Zhao, T. Utilization of Interspecific High-Density Genetic Map of RIL Popolation for the QTL Detection and Candidate Gene Mining for 100-Seed Weight in Soybean. *Front. Plant Sci.* **2019**, *10*, 1001. https://doi.org/10.3389/fpls.2019.01001.
2. 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. https://doi.org/10.1186/s12863-019-0737-9.
3. Patil, G.; Do, T.; Vuong, T.D.; et al. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. *Sci. Rep*. **2016**, *6*, 1–13.
4. Wang, X.; Pang, Y.; Wang, C.; Chen, K.; Zhu, Y.; Shen, C.; Ali, J.; Xu, J.; Li, Z. New candidate genes affecting rice grain appearance and milling quality detected by genome-wide and gene-based association analyses. *Front. Plant Sci*. **2017**, *7*, 1998.
5. Abbai, R.; Singh, V.; Nachimuthu, V.V.; Sinha, P.; Selvaraj, R.; Vipparla, A.K.; Singh, A.K.; Singh, U.M.; Varshney, R.K.; Kumar, A. Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* **2019**, *17*, 1612–1622.
6. Chen, K.; Zhang, Q.; Wang, C.-C.; Liu, Z.-X.; Jiang, Y.-J.; Zhai, L.-Y.; Zheng, T.-Q.; Xu, J.-L.; Li, Z.-K. G Genetic dissection of seedling vigour in a diverse panel from the 3,000 Rice (*Oryza sativa* L.) Genome Project. *Sci. Rep.* **2019**, *9*, 1–15.
7. Sinha, P.; Singh, V.K.; Saxena, R.K.; Khan, A.W.; Abbai, R.; Chitikineni, A.; Desai, A.; Molla, J.; Upadhyaya, H.D.; Kumar, A.; et al. Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (*Cajanus cajan* L.). *Plant Biotechnol. J.* **2020**, *18*, 2482–2490.
8. Lu, X.; Xiong, Q.; Cheng, T.; Li, Q.; Liu, X.-L.; Bi, Y.-D.; Li, W.; Zhang, W.-K.; Ma, B.; Lai, Y.-C.; et al. A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* **2011**, *10*, 670–684.
9. Barrero, R.A.; Bellgard, M.; Zhang, X. Diverse approaches to achieving grain yield in wheat. *Funct. Integr. Genom.* **2011**, *11*, 37–48.
10. Wang, J.; Chu, S.; Zhang, H.; Zhu, Y.; Cheng, H.; Yu, D. Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Sci. Rep.* **2016**, *6*, 1–10.
11. Zhang, J.; Song, Q.; Cregan, P.B.; Nelson, R.L.; Wang, X.; Wu, J.; Jiang, G.-L. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. *BMC Genom.* **2015**, *16*, 1–11.
12. 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.
13. 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.
14. Segura, V.; Vilhjálmsson, B.J.; Platt, A.; Korte, A.; Seren, Ü.; Long, Q. and Nordborg, M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* **2012**, *44*, 825–830.
15. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S. and Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* **2016**, *12*, e1005767.
16. Huang, M.; Liu, X.; Zhou, Y.; Summers, R.M. and Zhang, Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **2019**, *8*, giy154.
17. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, 461–464.
18. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575.
19. Li, D.; Zhao, X.; Han, Y.; Li, W.; Xie, F. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* **2019**, *111*, 90–95.
20. Hu, D.; Zhang, H.; Du, Q.; Hu, Z.; Yang, Z.; Li, X.; Wang, J.; Huang, F.; Yu, D.; Wang, H.; et al. Genetic dissection of yield-related traits via genome-wide association analysis across multiple environments in wild soybean (*Glycine soja* Sieb. and Zucc.). *Planta* **2020**, *251*, 39.
21. Marri, P.R.; Sarla, N.; Reddy, L.V.; Siddiq, E.A. Identification and mapping of yield and yield related QTLs from an Indian accession of Oryza rufipogon. *BMC Genet.* **2005**, *6*, 33.
22. Adeboye, K.A.; Semon, M.; Oyetunde, O.A.; et al. Diversity array technology (DArT)-based mapping of phenotypic variations among recombinant inbred lines of WAB638-1/PRIMAVERA under drought stress. *Euphytica* **2021**, *217*, 130.
23. Han, Y.; Li, D.; Zhu, D.; Li, H.; Li, X.; Teng, W.; Li, W. QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theor Appl Genet* **2012**, *125*, 671–683.
24. Du, W.; Wang, M.; Fu, S.; Yu, D. Mapping QTLs for seed yield and drought susceptibility index in soybean (Glycine max L.) across different environments. *Journal of Genetics and Genomics* **2009**, *36*, 721–731.
25. Guan, Y. Detecting structure of haplotypes and local ancestry. *Genetics* **2014**, *196*, 625–642.
26. Mishra, S.; Singh, B.; Misra, P.; Rai, V.; Singh, N.K. Haplotype distribution and association of candidate genes with salt tolerance in Indian wild rice germplasm. *Plant Cell Rep.* **2016**, *35*, 2295–2308.
27. Kuroha, T.; Nagai, K.; Gamuyao, R.; Wang, D.R.; Furuta, T.; Nakamori, M.; Kitaoka, T.; Adachi, K.; Minami, A.; Mori, Y.; et al. Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. *Science* **2018**, *361*, 181–186.