# Genome-wide association analysis of yield-related traits of soybean using haplotype-based framework

Adeboye KA[1*], Bhat JA[2], Ganie SA[3], Varshney RK[4,5] and Yu D[2]

1. Department of Agricultural Technology, Ekiti State College of Agriculture and Technology, Isan-Ekiti, Nigeria; adeboye@daad-alumni.de
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 Genet-ics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China; javid.akhter69@gmail.com (JAB); dyyu@njau.edu.cn (DY)
3. Plant Molecular Science and Centre of Systems and Synthetic Biology, Department of Biological Sci-ences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK; showkatmanzoorfor-ever@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, Australia; r.k.varshney@cgiar.org

Presented at the 2nd International Electronic Conference on Plant Sciences, Online, 1-15 December, 2021

# Introduction

- Yield characters are complex quantitative traits which posed some difficulties to breeding efforts.
- Intervention of marker-assisted breeding – marker-trait association studies
- Single marker studies through family linkage map and linkage disequilibrium analysis of yield characters in soybean.
- Recent emergence of haplotype-based breeding, which explores multimarker association to identify superior alleles from combination of many markers within a locus associated with the traits of interest.
- 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.

# Materials and methods – phenotyping

- 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)
  - Experimental Field of Jiangsu Yanjiang Institute of Agricultural Sciences in Nantong (E3 and E4) and
  - 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.
- 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).

# Materials and methods – genome-wide haplotype analysis

- Genome-wide association study (GWAS) was conducted in GAPIT v3.0 (Lipka et al. 2012) to identify significant marker associations for the studied traits across the six environments.
  - 12,617 single-nucleotide polymorphism (SNP) markers from NJAU 355K SoySNP Array.
  - GWAS models: general linear model (GLM) with PCA (Price et al. 2006), compressed mixed linear model (CMLM) (Zhang et al. 2010), multiple-locus mixed linear model (MLMM) (Segura et al. 2012), fixed and random model circulating probability unification (FarmCPU) (Liu et al. 2016) and Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) (Huang et al. 2019).
  - Population structure: principal component analysis (PCA) using Bayesian information criterion (BIC) to estimate the optimal numbers of PCA (Schwarz, 1978; Price et al. 2006).
- Multimarker association analysis within the haplotype-based framework was conducted using PLINK v1.07 (Purcell et al. 2007)
  - Stable markers from GWAS (across environments and models) were considered as reference markers for building haplotype block/loci.
  - All markers in proxy association with the stable/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 and visualized in Microsoft Excel (Microsoft corporations 2019)

# Results and discussion

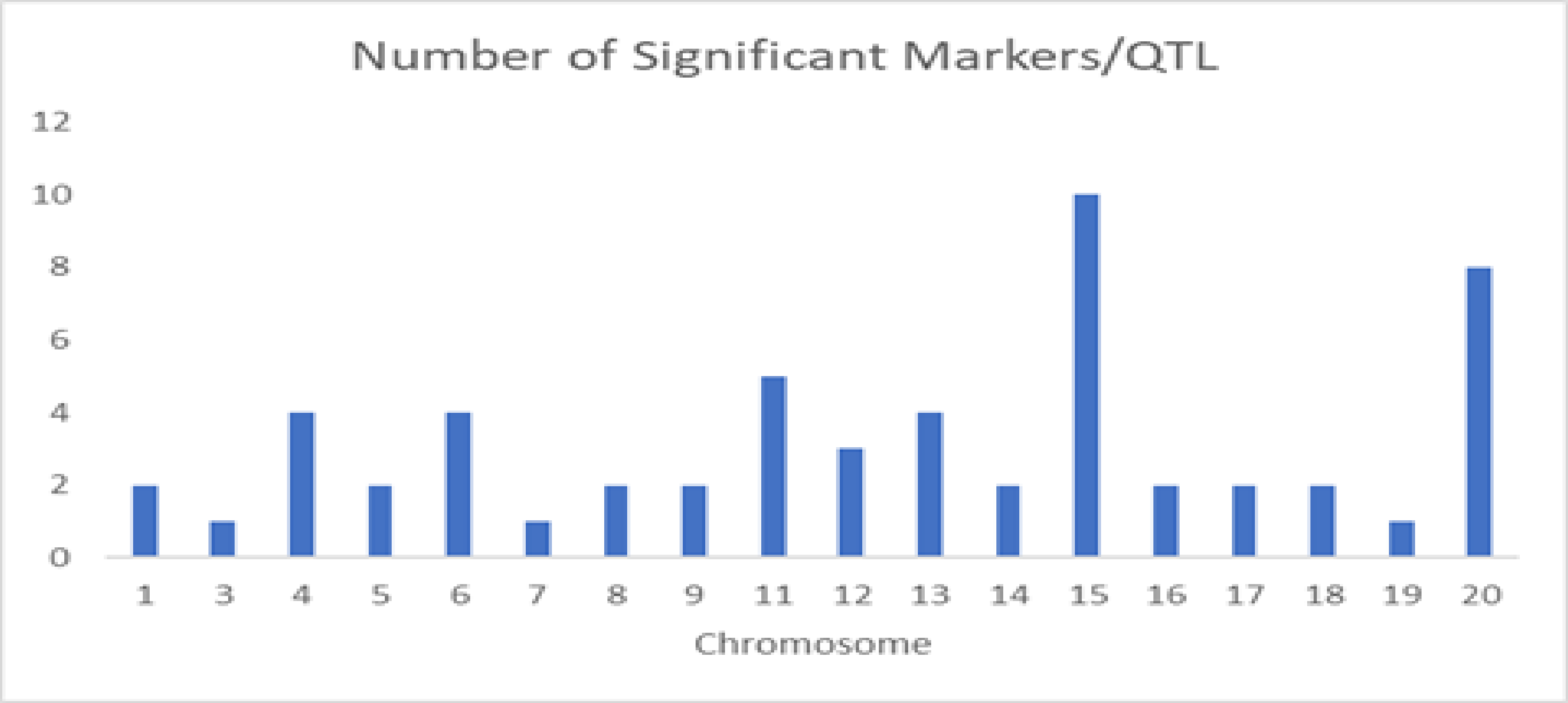

Figure 1. Distribution of significant markers/QTL across the soybean chromosomes.

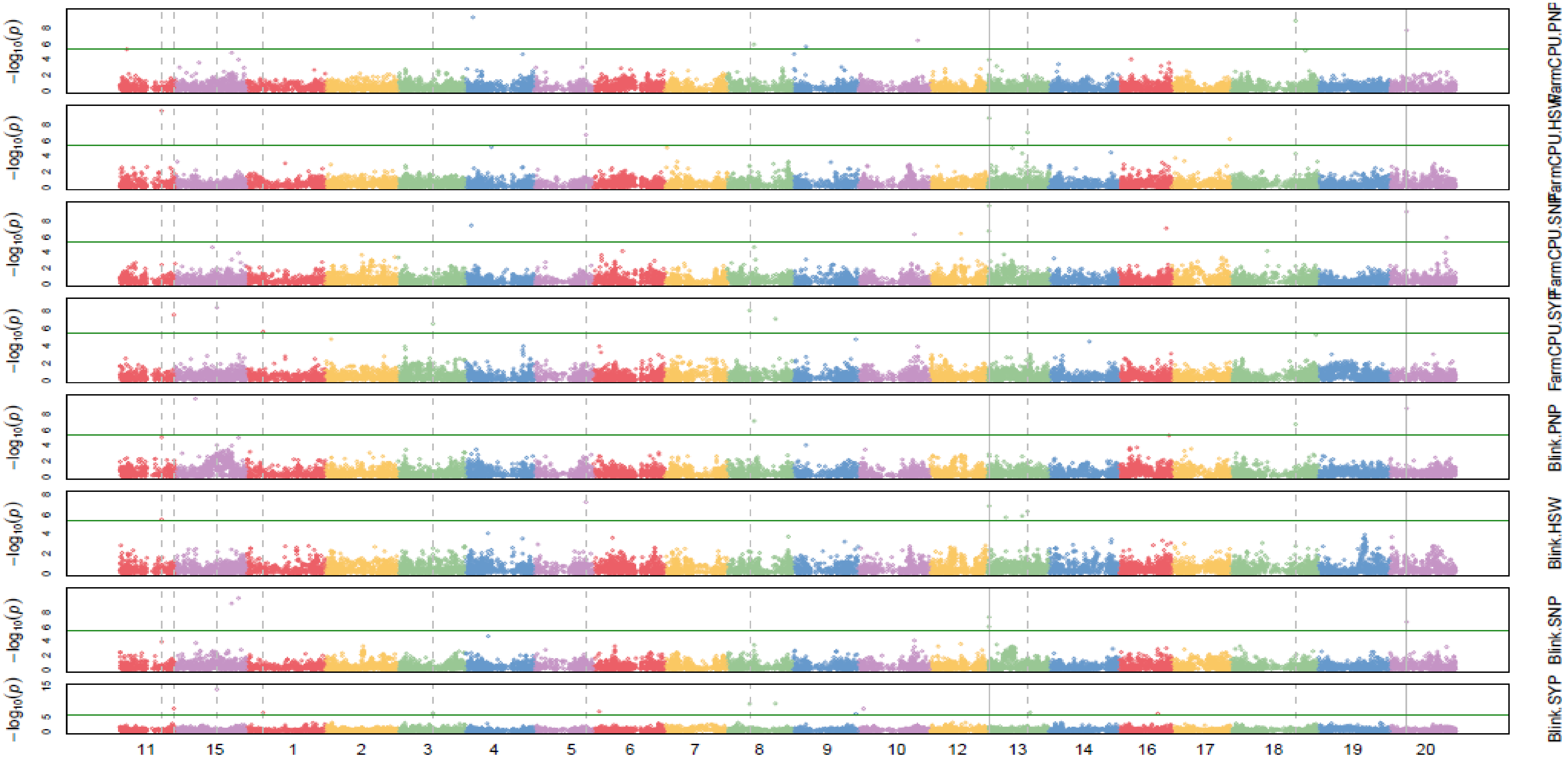

Figure 2. Mahanatan plot showing the significant association of markers with yield-related traits in the combined environment based on BLINK and FarmCPU models

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

| QTL/Marker | Chromosome | Physical Position (bp) | Trait (Environment) | Remark |
|---|---|---|---|---|
| AX-93703924 | 4 | 4291705 | SNP (COM and E6);<br>PNP (E3) | |
| AX-93922099 | 5 | 36599702 | HSW (COM, E1 and E5) | Seed weight 34-9 (Han et al. 2012);<br>Seed-yield 22-10 (Du et al. 2009) |
| AX-93793210 | 11 | 29587057 | HSW (COM, E1, E3 and E4);<br>SNP (E2, E3 and E5) | Seed weight 35-9 (Han et al. 2012) |
| AX-93807406 | 13 | 1843185 | HSW (COM, E1, E2, E4 and E5);<br>SNP (COM, E1 and E6) | |
| AX-94176727 | 18 | 46137043 | PNP (COM and E1);<br>HSW (E2) | |
| AX-94199992 | 20 | 12095298 | PNP (COM and E3);<br>SNP (COM and E1) | |

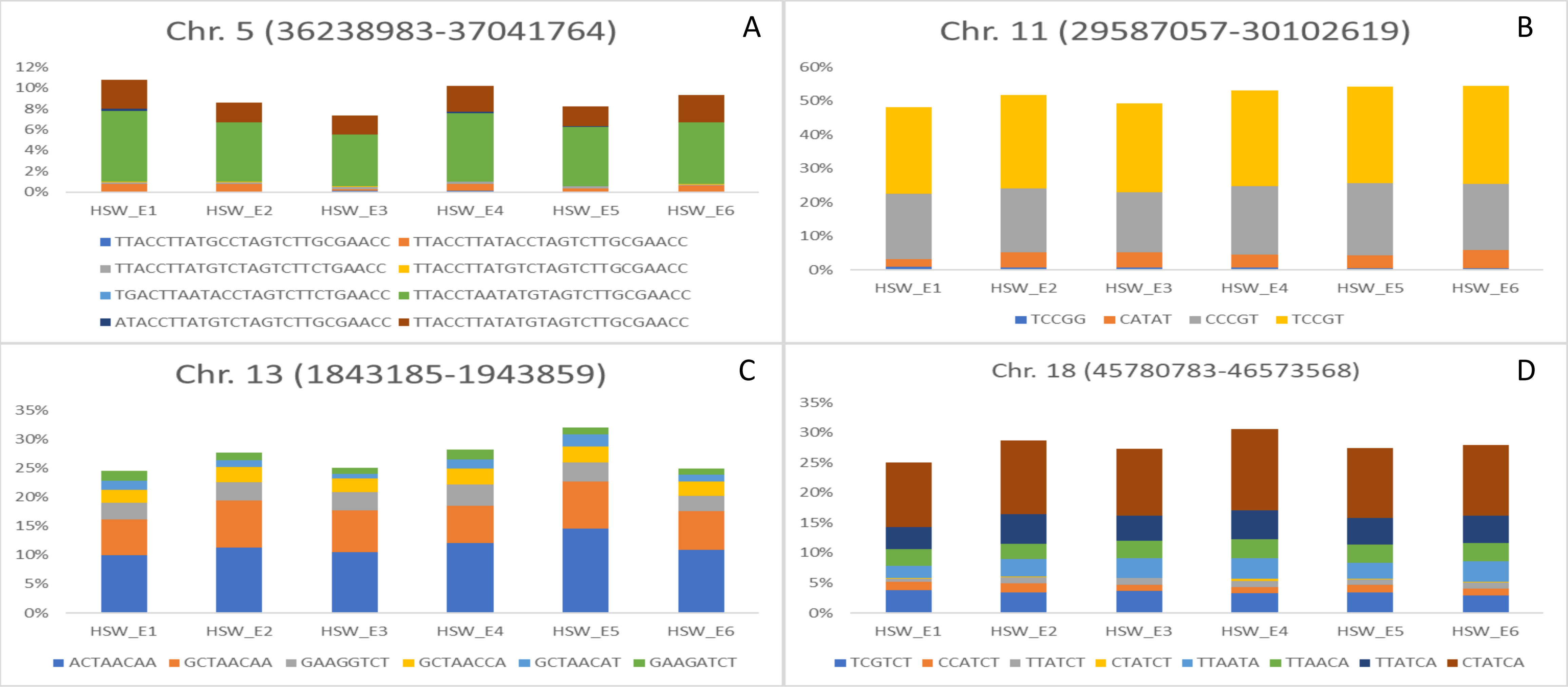

Figure 3. Haplotype alleles within the loci on chromosome 5 (A), 11 (B), 13 (C) and 18 (D), and their contribution to the phenotypic variation of hundred seed weight across the environments.

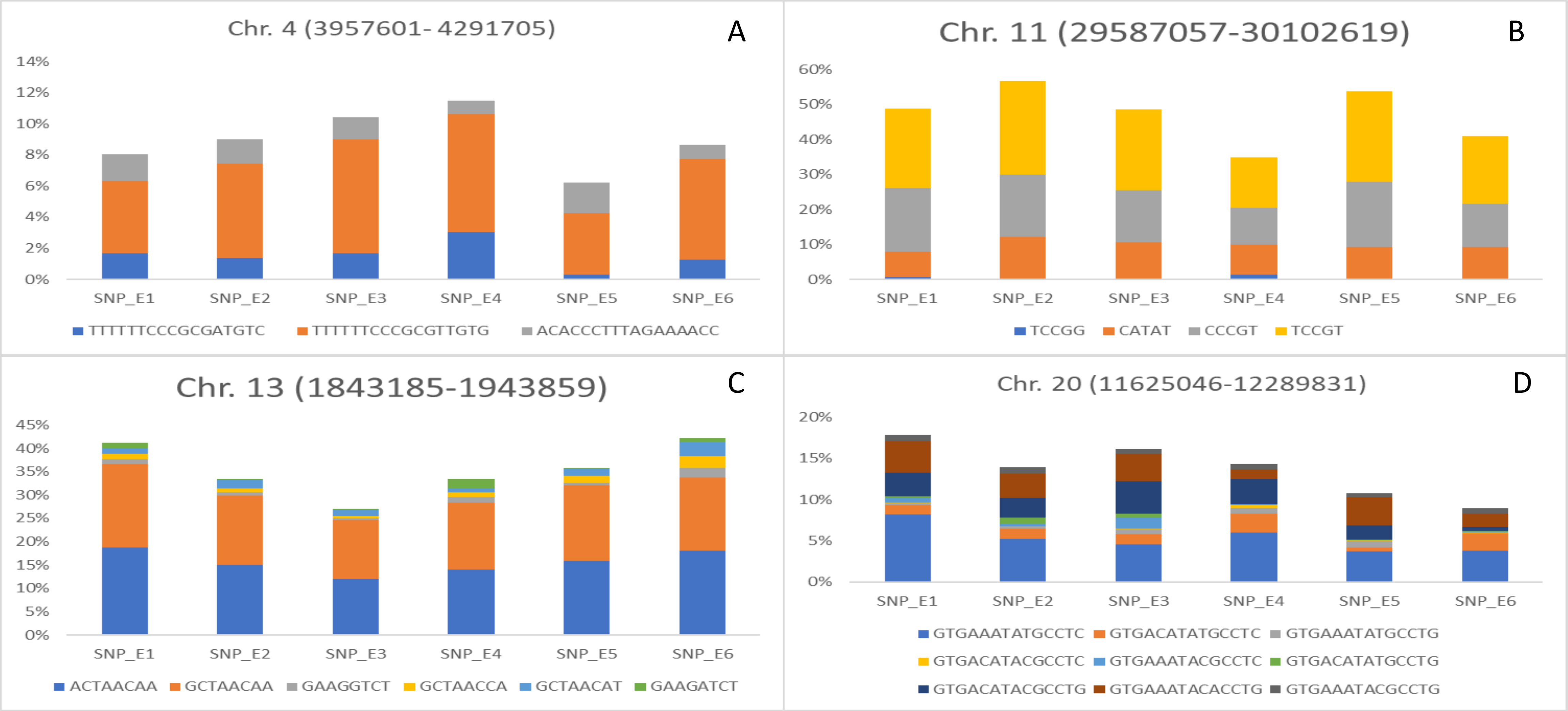

Figure 4. Haplotype alleles within the loci on chromosome 4 (A), 11 (B), 13 (C) and 20 (D), and their contribution to the phenotypic variation of seed number per plant across the environments.

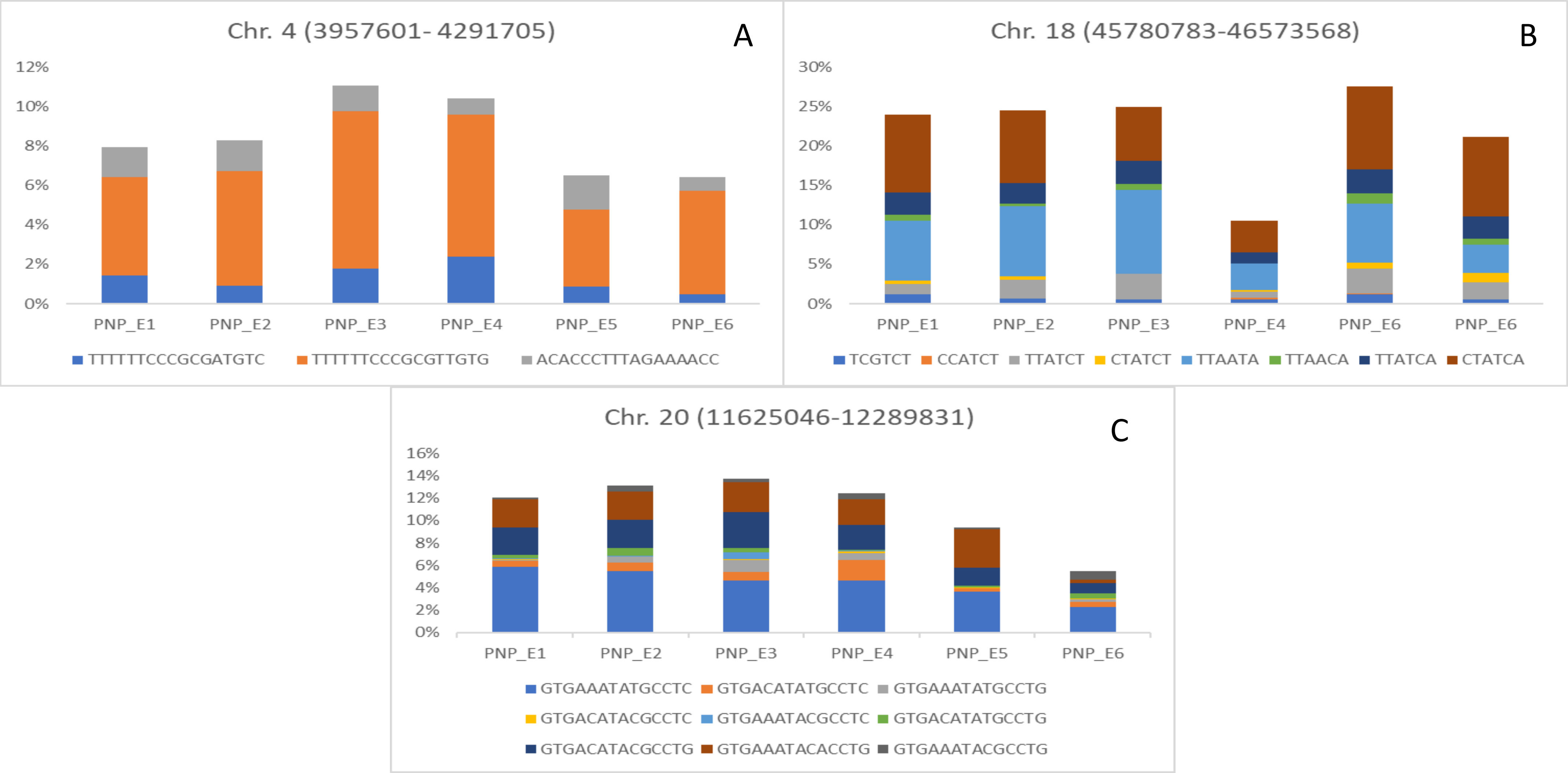

Figure 5. Haplotype alleles within the loci on chromosome 4 (A), 18 (B) and 20 (C), and their contribution to the phenotypic variation of Panicle number per plant across the environments.

# Conclusion

- 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.

# Acknowledgement

- 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).

Thank you for your kind attention