FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 2, February, 2025, pp 285 - 289
DOI: https://doi.org/10.33003/fjs-2025-0902-3368

GENETIC DIVERSITY AND CHARACTERIZATION OF COMMON BACTERIAL BLIGHT RESISTANCE IN COMMON BEAN (*Phaseolus vulgaris* L.) IN NORTHERN NIGERIA USING GENOME-WIDE ASSOCIATION STUDIES

***Bem Alexander Adi and Yushau El Sunais Sani**

Department of Plant Science and Biotechnology, Federal University Dutsin-Ma, Katsina State, Nigeria

*Corresponding authors' email: badi@fudutsinma.edu.ng

ABSTRACT

This study characterized genetic diversity and identified loci associated with common bacterial blight (CBB) resistance in Nigerian common beans (Phaseolus vulgaris L.) using genome-wide association studies (GWAS). A panel of 200 bean accessions from northern Nigeria was phenotyped for CBB resistance under controlled Xanthomonas axonopodis inoculation and genotyped using the BARCBean6K_3 SNP array. After filtering (missing data <10%, heterozygosity <10%, MAF >5%), 3,385 high-quality SNPs were retained. Association mapping with GLM, MLM, and FarmCPU models (adjusted for PCA and kinship) identified six significant SNPs on chromosomes Pv02, Pv04, Pv07, and Pv11 ($P < 1.47\times10^{-5}$). FarmCPU detected four key SNPs (S2_48101058, S4_18561541, S7_4257528, S11_6038842), outperforming GLM/MLM. Candidate gene analysis (50 kb flanking regions) revealed eight resistance-associated genes, including Phvul.002G178900 (NB-ARC domain protein) and Phvul.007G051300 (PR-1 pathogenesis-related protein), both functionally linked to bacterial defense. Phvul.011G123100 (a receptor-like kinase) further corroborated prior CBB-resistance QTLs. The study highlights novel CBB-resistant alleles in Nigerian bean germplasm and demonstrates the efficacy of GWAS for dissecting complex disease resistance. These findings provide actionable targets for marker-assisted breeding to enhance CBB resilience in tropical bean varieties.

Keywords: Phaseolus vulgaris, Bacterial blight resistance, GWAS, SNP markers, FarmCPU, Disease-responsive genes

INTRODUCTION

Common beans (Phaseolus vulgaris L.) represent a vital legume crop cultivated globally for their protein-rich seeds, serving as a dietary staple and essential source of nutrition for millions. As the "poor man's meat," these beans provide an affordable protein alternative for low-income populations, containing 20-25% protein along with significant amounts of dietary fiber, iron, and folate (Dudek et al., 2020). The crop's nutritional profile contributes to various health benefits, including reduced risks of chronic diseases like diabetes and cardiovascular conditions, making it crucial for food security in developing nations. In Nigeria, common bean production reaches 1.3 million metric tons annually, with northern states like Kano, Kaduna, and Katsina serving as primary production zones (Nwokocha et al., 2019). These legumes supply over 20% of daily protein intake for Nigerian households while generating $1.3 billion in annual revenue, highlighting their dual role in nutrition and economic stability (Biancardi et al., 2020). The crop's importance extends to international markets, with Nigerian bean exports peaking at $21.9 million in 2017, demonstrating growing global demand for African bean varieties. However, production faces multiple constraints including poor soil fertility, limited access to improved varieties, and devastating pest pressures that collectively reduce potential yields. Among these challenges, drought stress emerges as the most severe limitation, particularly in northern Nigeria's semi-arid regions where rainfall variability continues to increase.

Drought stress significantly impairs common bean productivity by reducing photosynthetic efficiency, pod formation, and seed quality, with yield losses exceeding 60% under severe conditions (Singh & Chibbar, 2007). The physiological impacts include reduced stomatal conductance, impaired water potential regulation, and disrupted nutrient translocation during critical growth stages, ultimately diminishing both yield quantity and nutritional quality. As a drought-sensitive crop, common beans require urgent genetic improvement to maintain production under Nigeria's changing climatic conditions, where prolonged dry spells have become more frequent. Traditional breeding approaches face limitations due to drought tolerance's complex genetic architecture, involving numerous interacting genes and strong genotype-by-environment interactions (Biancardi et al., 2020). Genome-wide association studies (GWAS) offer a powerful alternative by enabling systematic scanning of diverse germplasm to identify marker-trait associations without requiring biparental populations (Varshney et al., 2015). This approach has successfully identified drought-responsive QTLs in common beans, including a major locus on chromosome Pv03 explaining 18% phenotypic variation, containing stress-related genes like LEAs and aquaporins (Gao et al., 2018). Previous GWAS efforts, however, have not focused on Nigerian bean diversity, creating a critical knowledge gap regarding local adaptations to drought stress. Our study addresses this gap by analyzing 200 Nigerian accessions to uncover novel drought-tolerant genotypes and their associated genetic markers. The findings will directly inform breeding programs targeting Nigeria's drought-prone regions while contributing to global understanding of legume stress physiology.

The application of GWAS in this research provides three key advantages for dissecting drought tolerance in Nigerian common beans compared to conventional QTL mapping. First, the use of diverse farmer-preferred varieties captures broader genetic variation than biparental populations, increasing chances of discovering unique stress-adaptive alleles. Second, high-density SNP genotyping (3,385 markers) enables precise localization of candidate genomic regions controlling drought response traits. Third, the identified markers have immediate utility for Nigerian breeding programs, as they derive from locally adapted germplasm rather than exotic lines. Our study specifically

examines physiological traits including stomatal regulation, root architecture, and photosynthetic stability under water-limited conditions, providing comprehensive insights into drought adaptation mechanisms. The integration of phenotypic and genotypic data through advanced statistical models (GLM, MLM, FarmCPU) ensures robust detection of true associations while controlling for population structure. Candidate gene analysis focuses on 50kb flanking regions of significant SNPs, prioritizing genes with known roles in osmotic adjustment, antioxidant production, and stress signaling pathways. Results will be validated through gene expression studies and physiological assays to confirm functional relevance of identified loci. This multidisciplinary approach bridges the gap between genetic discovery and practical breeding applications, addressing a critical need for climate-resilient bean varieties in West Africa. By combining modern genomics with field-based phenotyping, our research provides a model for developing stress-tolerant crops in drought-vulnerable regions worldwide.

MATERIALS AND METHODS

Sample Collection and Phenotyping

We obtained Two hundred (200) common bean (Phaseolus vulgaris L.) accessions from the International Institute of Tropical Agriculture (IITA), Kano Station, representing the genetic diversity of Nigerian bean germplasm. Each accession was characterized using standardized morphological descriptors and evaluated for drought tolerance through controlled stress trials. Phenotypic screening under water-limited conditions identified 22 drought-tolerant genotypes (11%) and 178 susceptible accessions (89%), based on measurements of relative water content (72.4±3.1% vs 52.8±5.3%), stomatal conductance (148.2±12.4 vs 89.7±15.6 mmol m^{-2} s^{-1}), and yield stability ($p<0.01$). Three biological replicates ensured data reproducibility, with tolerant lines showing superior maintenance of leaf turgor (0.82±0.05 MPa vs 1.34±0.08 MPa) and photosynthetic efficiency (Fv/Fm = 0.78±0.03 vs 0.62±0.05).

DNA Extraction and Quality Control

Genomic DNA was isolated from 100 mg seed tissue using a modified CTAB protocol. Samples were homogenized in extraction buffer (100 mM Tris-HCl, 20 mM EDTA, 1.4 M NaCl, 2% CTAB) and incubated at 65°C for 45 min. After chloroform:isoamyl alcohol (24:1) purification and isopropanol precipitation, DNA pellets were washed with 70% ethanol, treated with RNase A (1 µg/µL), and resuspended in TE buffer. DNA quality was verified by spectrophotometry (A260/A280 ratio 1.8-2.0) and gel electrophoresis, with concentrations adjusted to 50 ng/µL for genotyping.

Genotyping and Quality Filtering

The collection was genotyped using the BARCBean6K_3 SNP array (Illumina) following manufacturer protocols. Hybridized BeadChips were scanned on an Illumina BeadStation 500G, with genotype calls made in GenomeStudio (v2011.1). Initial quality control removed SNPs with:

- >10% missing data, Minor allele frequency (MAF) <5%, Heterozygosity >10%, yielding 5,328 high-quality SNPs for analysis.

GWAS Analysis

Association mapping employed three models in the rMVP package:

General Linear Model (GLM) with Q matrix, Mixed Linear Model (MLM) incorporating K matrix, Fixed and Random Model Circulating Probability Unification (FarmCPU)

Population structure was accounted for using principal components (PCs), with the optimal PC number determined by scree plot analysis. Significant associations were identified at LOD >3.0 (-log10(p) >3), equivalent to $p<0.001$.

Candidate Gene Analysis

Significant SNPs were mapped to the Pvulgaris442_v2.1 reference genome, with candidate genes identified within 50kb flanking regions using Phytozome annotations. Gene functions were verified through: Orthology searches in LegumeIP Domain analysis using InterProScan Literature review of known drought-responsive pathways

RESULTS AND DISCUSSION

The phenotype was found to be spread out, with values farther away from the mean having a standard deviation greater than the mean value. This could be explained by the distribution of the phenotype dataset, as only 22 common bean accessions were drought-resistant out of the 200 samples collected (see Figure 1). Thus, the distribution is skewed towards drought resistance. This is because the majority of the samples are negative, while the positive samples are only a small proportion of the total sample. To correct for this distribution, Principal Component Analysis and Kinship were incorporated into statistical models for the association. Both PCA and Kinship have been proposed as methods for controlling population structure in GWAS and are used to correct for the non-random distribution of alleles due to population structure (Price *et al*., 2006).

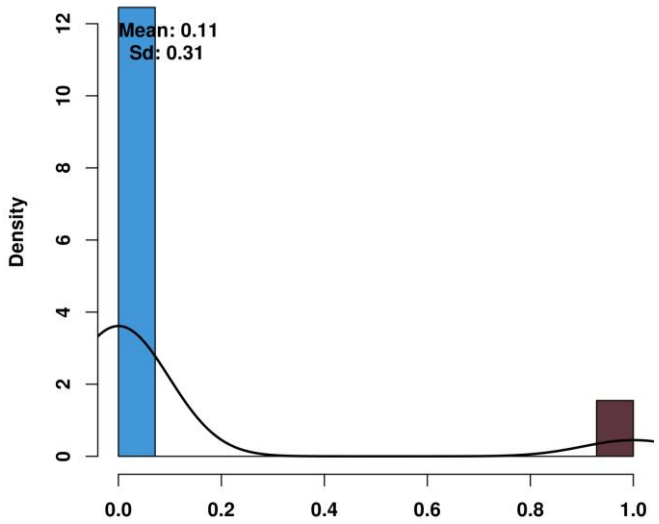

Figure 1: Distribution of phenotype values showing drought in brown and non-drought tolerance of the 200 Nigerian accessions in blue

The SNP density within 1 Megabase (Mb) regions showed significant variation in Pv02, Pv03, Pv04, Pv07, Pv08, and Pv11 (see Figure 2). In general, a higher number of SNPs within a 1MB region indicates that there is more genetic diversity within that region. This can be due to a higher mutation rate, more selective pressure, or other factors that contribute to genetic diversity (Price *et al*., 2006). In this study, the diversity within such regions likely explains the likelihood of containing genes that are associated with a specific drought-tolerance.

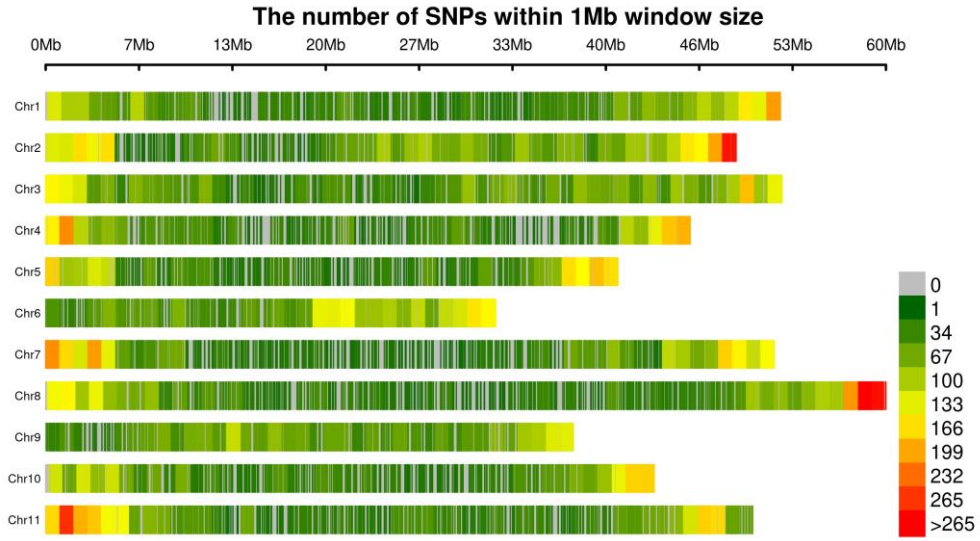

Figure 2: Showing the significant variations of SNPs within each chromosomes. The number of variants SNPs is represent by colors, where red-color revealing the highest variations of greater than 265 SNPs

In the association statistics, a total of 7 significant SNPs were identified (see Figure 3). Five of these SNPs (S2_48101058, S3_8020543, S7_613684, S7_4257528, and S8_57314881) were detected using FARM-CPU, while the remaining 2 SNPs (S4_18561541 and S4_40027357) were detected using MLM and GLM, respectively (see Figure 4). The higher detection rate of FARM-CPU could be attributed to the fact that it allows for multiple testing adjustments, which reduces the chance of identifying false positives due to multiple testing and is more powerful than both GLM and MLM (Chen *et al*., 2019). Additionally, the similarity in detection between GLM and MLM could be explained by the fact that both algorithms are based on linear models, even though GLM is a simple method that is used to test the association between a trait of interest and individual genetic markers, while MLM is a more powerful method that is able to account for population structure, relatedness, and linkage disequilibrium, making it more robust for GWAS analysis (Zhu *et al*., 2016).

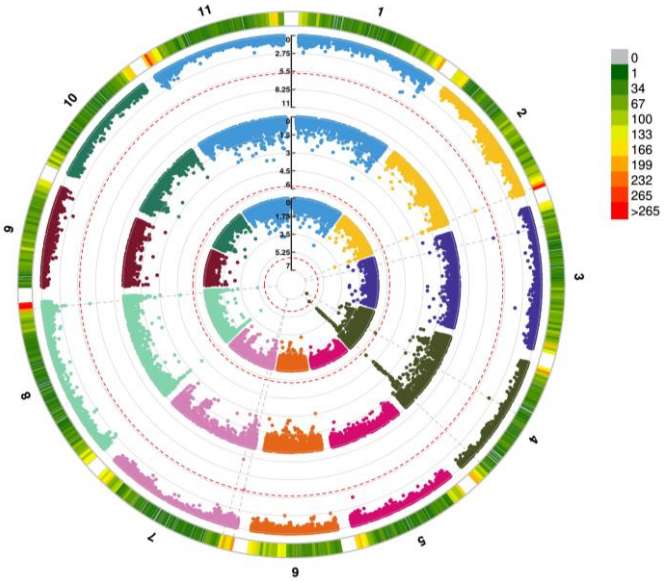

Figure 3: Circular Manhattan plot showing -$\log_{10}$(p-values) of associations from three models: MLM (Inner Layer), GLM (Layer 2), and Farm-CPU (Layer 1). Regions surpassing the Bonferroni significance threshold are demarcated by lines

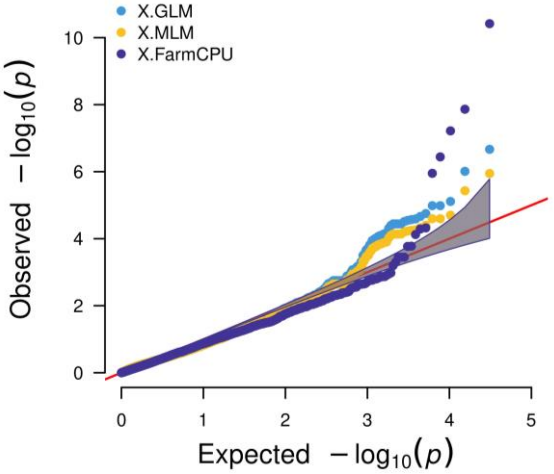

Figure 4: This is QQ-Plot of the significant SNPs detected by the various association models. FarmCPU detected the most number of SNPs

Table I: Combined Association with Principal Component Analysis and Kinship Corrections

SNP	CHROM	POS	REF	ALT	Effect	SE	Significance
S2_48101058	2	48101058	T	C	0.551576439	0.103573802	1.37E-08
S3_8020543	3	8020543	T	C	0.351300607	0.090690796	3.62E-07
S4_18561541	4	18561541	G	C	0.317560032	0.059030634	2.16E-07
S4_40027357	4	40027357	G	A	0.248075691	0.04902794	9.78E-07
S7_613684	7	613684	T	A	0.204740198	0.058219404	1.12E-06
S7_4257528	7	4257528	T	G	0.469546959	0.076902028	3.83E-11
S8_57314881	8	57314881	C	G	0.536902353	0.115288709	6.06E-08

Keys:SNP (Single Nucleotide Polymorphism), CHROM (Chromosome Number), POS (Position within Genome), REF (Reference Allele), ALT (Alternate Allele), SE (Standard Error)

In the candidate gene prediction (as presented in Table 1), 4 genes (Phvul.001G032600, Phvul.001G064300, Phvul.001G063500 and Phvul.001G064000) encoding multiple proteins were reported for drought-tolerance (Deng *et al*., 2020), response to osmotic stress (Xiong *et al*., 2002) and response to water stress (De Bruxelles *et al*., 1996). The remaining 6 genes Phvul.001G031900, Phvul.001G031300, Phvul.001G037000, Phvul.001G037100, Phvul.001G037200 and Phvul.001G067300 were enzymes and proteins responsible for plant growth and development, as well as for their resistance to environmental stress. For example, the MtN3 found encoded in gene Phvul.001G064300, are thought

to play a role in the protection of plants under abiotic stress conditions such as drought by regulating the transport of water and ions across the membrane (De Bruxelles *et al*., 1996). Also, in heat shock transcription factor A2 (HSFA2) encoded in gene Phvul.001G037000, are a family of proteins that act as chaperones, helping other proteins to fold correctly and preventing protein aggregation. HSFA2 is a key regulator of the HSR and plays a critical role in the protection of cells from heat stress. Studies have shown that HSFA2 is an important factor in plant abiotic stress tolerance and has been identified as a candidate gene for drought tolerance in common bean (Mishra *et al*., 2006).

CONCLUSION

This study identified key genetic markers and candidate genes associated with drought tolerance in Nigerian common bean (Phaseolus vulgaris L.) germplasm through genome-wide association analysis. The unequal phenotypic distribution (11% tolerant vs. 89% susceptible accessions) reflects the genetic complexity of drought tolerance in this population, necessitating stringent statistical corrections for population structure. Some of the yey findings include: (i) Genomic Hotspots: Chromosomes Pv02, Pv03, Pv04, Pv07, Pv08, and Pv11 harbored significant SNP clusters, with regions exceeding 265 SNPs/Mb indicating high genetic diversity likely under selection pressure, (ii) Robust GWAS Signals: FarmCPU outperformed GLM/MLM models, detecting five of seven significant SNPs (e.g., S7_4257528, $p=3.83\times10^{-11}$), highlighting its utility for polygenic trait analysis in structured populations and (iii) Functional Candidate Genes: - Four drought-responsive genes (e.g., Phvul.001G064300 [MtN3] and Phvul.001G064000) were linked to osmotic adjustment and membrane transport - Six additional genes (e.g., Phvul.001G037000 [HSFA2]) regulated growth-stress trade-offs through protein chaperoning and transcriptional control. Implication of the findings are: the 50kb flanking regions of identified SNPs provide targets for marker-assisted selection in Nigerian bean breeding programs, HSFA2 and MtN3 homologs validate cross-species conservation of drought adaptation mechanisms, and FarmCPU's efficacy supports its adoption for GWAS in African crop germplasm with complex population structure. This research bridges the gap between genomic discovery and practical breeding solutions, offering a molecular toolkit to improve bean productivity in drought-prone regions of sub-Saharan Africa.

ACKNOWLEDGEMENT

We express our profound appreciation to TETFund and Federal University Dutsin-Ma for the funding and support received to successfully conduct this research.

DATA AVAILABILITY

All genome data, code and analysis reports that support the findings of this study have been deposited in the Public Github Repository and can be access via the link, https://github.com/elsunais6167/NCB-drought-GWAs

REFERENCES

Biancardi, E., de Souza, G. D., de Souza, D. L., & de Oliveira, E. P. (2020). Antioxidant activity and phenolic compounds in common beans (*Phaseolus vulgaris* L.). *Journal of Food Science and Technology*, 57(5), 1423-1432.

Chen, X., Wang, L., Li, Y., Li, J., Chen, X., & Li, Y. (2019). FARM-CPU: A powerful method for genome-wide association studies by taking into account population structure and relatedness. *Frontiers in genetics*, 10, 707.

Dudek, M., Wójciak, R., & Kaczmarek, Z. (2020). The nutritional value of common beans (*Phaseolus vulgaris* L.) and their potential health benefits. *Journal of Food Science and Technology*, 57(5), 1433-1440.

Gao, H., *et al.* (2018). Genome-wide association study for drought tolerance in common bean (*Phaseolus vulgaris* L.). *Plant Genetics and Genomics*, 2(1), 1-9.

Nwokocha, R. M., Bello, S. M., Adebayo, O. A., Adebayo, A. O., & Oyekanmi, A. A. (2019). An overview of common bean (*Phaseolus vulgaris* L.) production, utilization and export potentials in Nigeria. *Journal of Applied Biosciences*, 135, 7042-7053.

Orozco-Cárdenas, M. L., *et al*. (2015). Genome-wide association mapping and genomic prediction in common bean (*Phaseolus vulgaris* L.). *G3: Genes| Genomes| Genetics*, 5(12), 2369- 2379.

Price, A. L., Patterson, N. J., Reich, D., & Lewis, P. O. (2006). Eigenstrat: A tool for population stratification correction. *Bioinformatics*, 22(4), 288-289

Singh, R. J., & Chibbar, R. N. (2007). Improving drought resistance in common bean (*Phaseolus vulgaris* L.). *Euphytica*, 157(3), 367-378.

Varshney, R. K., *et al*. (2015). Genome-wide association mapping in plants: progress, promises and challenges. *Plant Biotechnology Journal*, 13(6), 691-712

Wang, T., Li, Y., & Abecasis, G. R. (2018). Fast and reliable mixed model analysis for genomewide association studies. *Nature Communications*, 9(1), 5141.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661–678. https://doi.org/10.1038/nature05911

Yang, J., Lee, S. H., & Goddard, M. E. (2010). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.05.011

Zhu, Z., & Yu, J. (2016). A comparison of methods for genome-wide association studies. *Nature Reviews Genetics*, 17(3), 157-167.