



International Journal of Health care and Biological Sciences

Content available at www.saap.org.in

Online ISSN: 2582-7499



Open Access

Phenotypic Diversity and Multivariate Characterization of Elite Foxtail Millet (*Setaria italica* L.) Genotypes: Implications for Genomics- and AI-Assisted Climate-Resilient Breeding

B. Vigneshwara Reddy^{*1}, Bonala Kalyana Chakravarthy², Rajasekhar Chikati³, SN Meeravali⁴^{1&2}Department of Biotechnology, Yogi Vemana University, Kadapa, Andhra Pradesh – 516005,³Department of Biochemistry, Yogi Vemana University, Kadapa, Andhra Pradesh–516005⁴Department of Botany, Yogi Vemana University, Kadapa, Andhra Pradesh – 516005

Article History	Abstract
Received on: 11-02-2023 Revised on: 03-03-2023 Accepted on: 28-03-2023	
Keywords: Foxtail millet, <i>Setaria italica</i> , genetic diversity, PCA, cluster analysis, genomic selection, pan-genome, machine learning, climate-resilient breeding.	Foxtail millet (<i>Setaria italica</i> L.) is a climate-resilient C4 cereal with strong relevance to dryland agriculture, nutritional security and climate adaptation. The present paper updates the interpretation of an earlier field evaluation of 24 elite foxtail millet genotypes, including three checks, characterized for seven quantitative traits. The original experiment identified substantial phenotypic variability, with the highest coefficients of variation for fodder yield (32.73%), number of tillers per plant (20.85%) and grain yield (17.74%). Grain yield showed positive associations with panicle length, days to maturity and number of tillers per plant, whereas fodder yield was negatively associated with grain yield. Principal component analysis (PCA) showed that the first four components explained 94.59% of the observed variation, while hierarchical clustering separated the 24 genotypes into three groups. These findings remain useful for parental selection, but contemporary foxtail millet improvement can now extend such phenotypic analysis through high-throughput phenotyping, genome-wide association studies (GWAS), genomic selection, graph-based pan-genomes, multi-omics, machine learning and targeted genome editing. Importantly, the present dataset contains phenotypic measurements only; therefore, genomic or AI-based breeding values are discussed as a forward-looking framework rather than reported as new experimental results. The divergent genotypes identified in the original study provide a rational starting set for future multi-environment validation and genomics-assisted breeding.

This article is licensed under a Creative Commons Attribution-Non-commercial 4.0 International License.

Copyright © 2023 Author(s) retains the copyright of this article.



*Corresponding Author

B. Vigneshwara Reddy

1. Introduction

Foxtail millet is one of the oldest domesticated cereals and is increasingly important as agriculture confronts water limitation, rising temperatures, soil constraints and unstable production environments[5,8]. Its short duration, C4 photosynthetic pathway, relatively small genome and adaptation to marginal environments make

it both a useful crop and an important model for understanding stress-resilient cereal biology [3].

The original study focused on phenotypic diversity among elite foxtail millet lines from the All India Coordinated Small Millets Improvement Project. Its central breeding question remains relevant: which genotypes and traits provide the greatest useful diversity for recombination and selection? Classical multivariate tools such as correlation analysis, PCA and hierarchical clustering remain valuable because they

summarize complex phenotypic relationships and identify non-redundant traits and divergent groups[1,6,9].

However, crop diversity analysis has advanced substantially. A major milestone was the construction of a *Setaria* graph-based pan-genome from 110 representative genomes. That work catalogued extensive gene-content and structural variation and evaluated 68 traits across 13 environments, demonstrating how pan-genomic variation can support marker-assisted breeding, genomic selection and genome editing. More recent germplasm studies have also combined large-scale phenotyping with modern diversity analysis, and current climate-resilient breeding frameworks increasingly emphasize genomic prediction, high-throughput phenotyping and machine learning.

Accordingly, this paper retains the experimentally observed phenotypic results of the original 24-genotype study while updating their scientific interpretation. The objectives are: (i) to describe variability and trait associations; (ii) to identify major axes of phenotypic variation; (iii) to classify elite lines into divergent clusters; and (iv) to place these findings within a 2026 breeding framework integrating phenomics, pan-genomics, genomic selection, AI-assisted prediction and precision breeding.

2. Materials and Methods

2.1 Plant material and field evaluation

The experimental dataset comprised 24 elite foxtail millet genotypes, including 21 elite lines and three check varieties (SiA 326, SiA 3085 and SiA 3156), evaluated at the Regional Agricultural Research Station, Nandyal, Andhra Pradesh, India. The material originated from the All India Coordinated Small Millets Improvement Project. The original field evaluation was conducted during kharif 2014 in a randomized complete block design. Each genotype was grown in ten rows of 3 m length at 30 cm × 10 cm spacing.

2.2 Traits recorded

Observations were recorded from five randomly selected plants per accession for seven quantitative characters: days to 50% flowering (DFF), plant height (PH), number of tillers per plant (NTPP), panicle length (PL), days to maturity (DM), grain yield (GY) and fodder yield (FY)[7,8,11].

2.3 Statistical analysis

The original dataset was summarized using mean, range, standard deviation and coefficient of variation. Phenotypic correlation coefficients were used to assess

trait associations[10,12]. PCA was used to identify the principal axes explaining multivariate variation, and mean trait values of the 24 genotypes were subjected to hierarchical cluster analysis[4,5]. The numerical results reported below are retained from the original analysis and have not been replaced by simulated or newly generated values.

2.4 Modern extension recommended for future validation

For a contemporary follow-up experiment, the same germplasm should be evaluated across multiple locations and seasons using mixed-model analysis to estimate genotype, environment and genotype × environment effects. Field phenotyping can be expanded with UAV or proximal-sensor measurements of canopy temperature, vegetation indices, plant architecture and temporal growth[8]. Genotyping by sequencing or whole-genome resequencing can then support GWAS and genomic prediction. Because structural variants can be missed by a single linear reference genome, graph-based pan-genome resources are particularly relevant for *Setaria*. Machine-learning models may be compared with standard genomic best linear unbiased prediction using nested cross-validation, while explainable-AI methods can be used to identify influential genomic and phenomic features. These procedures are proposed future methods; they were not performed on the present historical dataset.

3. Results

3.1 Phenotypic variability

The 24 genotypes exhibited broad variation for the seven quantitative traits. Days to 50% flowering ranged from 33 to 52 days; plant height from 100 to 168 cm; panicle length from 12.30 to 20.70 cm; days to maturity from 58 to 87 days; grain yield from 2,204 to 5,064 kg ha⁻¹; and fodder yield from 2,809 to 10,000 kg ha⁻¹. The highest coefficients of variation occurred for fodder yield (32.73%), number of tillers per plant (20.85%) and grain yield (17.74%), indicating substantial exploitable phenotypic dispersion for productivity-related traits.

Table 01: Descriptive variability of seven quantitative traits in 24 elite foxtail millet genotypes.

Trait	Mean	Minimum	Maximum	SD	CV (%)
DFF	43.92	33	52	5.55	12.63
Plant height	141.33	100	168	16.76	11.86

(cm)					
Tillers/plant	3.46	2	4	0.72	20.85
Panicle length (cm)	18.10	12.30	20.70	2.31	12.76
Days to maturity	77.25	58	87	8.98	11.63
Grain yield (kg/ha)	3660.46	2204	5064	649.31	17.74
Fodder yield (kg/ha)	6283.46	2809	10000	2056.32	32.73

3.2 Trait associations

Grain yield was positively associated with days to 50% flowering ($r = 0.31$), plant height ($r = 0.38$), number of tillers per plant ($r = 0.62$), panicle length ($r = 0.74$) and days to maturity ($r = 0.68$). Fodder yield showed a negative association with grain yield ($r = -0.29$). Among the recorded traits, panicle length, maturity duration and tiller number therefore showed particularly useful positive relationships with grain yield in this dataset. These correlations should be interpreted as phenotypic associations rather than proof of causal genetic relationships.

3.3 Principal component analysis

The first four principal components together accounted for 94.59% of total phenotypic variation. PC1 explained 61.90%, PC2 16.15%, PC3 12.36% and PC4 4.35%. DFF (0.96), plant height (0.92), tillers per plant (0.92), panicle length (0.84), days to maturity (0.84) and fodder yield (0.56) loaded strongly on PC1. Grain yield had a dominant loading on PC2 (0.98), while fodder yield contributed strongly to PC3 (0.80). This structure indicates that vegetative and developmental traits formed a major common axis of variation, whereas grain and fodder productivity contributed additional, partially distinct dimensions.

Table 02: Major principal components retained from the original analysis.

PC	Eigenvalue	Variation (%)	Cumulative (%)
PC1	4.33	61.90	61.90
PC2	1.13	16.15	78.05
PC3	0.87	12.36	90.41
PC4	0.30	4.35	94.59

3.4 Cluster analysis and candidate parents

Hierarchical clustering separated the 24 genotypes into three groups. Cluster I contained 12 genotypes, Cluster II eight and Cluster III four. Cluster I showed the highest mean grain yield (4,057.08 kg ha⁻¹), whereas Cluster II had the highest mean plant height, tiller number, panicle length, maturity duration and fodder yield (8,788.50 kg ha⁻¹). Cluster III was characterized by shorter stature and earlier maturity. The original analysis highlighted SiA 3145, SiA 3146, DhFtMV333, TNSI 267, SiA 3223, SiA 3221, SiA 3222 and SiA 3159 as particularly variable/divergent material. These genotypes are therefore logical candidates for re-evaluation in modern multi-environment and genomic studies.

4. Discussion: Interpretation with Current Breeding Technology

4.1 From morphological diversity to precision phenotyping

The strong phenotypic variation observed for yield and fodder traits demonstrates the continuing value of well-designed field characterization. In modern breeding, however, manual measurements can be complemented by repeated, non-destructive phenotyping. UAV imagery, multispectral and thermal sensing, and automated image analysis can quantify canopy development and stress responses over time. This changes phenotyping from a small set of end-point traits to longitudinal phenomic profiles, potentially improving selection for drought adaptation and yield stability.

4.2 Pan-genomes and structural variation

The most important genomic advance for *Setaria* is the availability of a graph-based pan-genome. A 2023 Nature Genetics study assembled 110 representative *Setaria* genomes and identified extensive structural variation, including presence/absence variants. The study also linked a promoter presence/absence variant to the yield-related gene SiGW3 and demonstrated breeding-potential prediction across multiple traits and environments. This means that future analysis of the present elite lines need not rely only on morphology or a single reference genome; their SNPs, haplotypes and structural variants can be compared against a richer species-wide genomic framework.

4.3 Genomic selection and machine learning

PCA and clustering describe diversity but do not directly predict the breeding value of untested progeny.

Genomic selection addresses this limitation by learning genome-wide marker effects from a training population and predicting genomic estimated breeding values. Machine-learning methods can additionally model nonlinear relationships among genomic, environmental and phenomic variables. For a small dataset such as the present 24-genotype panel, complex AI models would be statistically inappropriate without a much larger training population. A rigorous next phase should therefore enlarge the population, use multi-environment records, compare linear genomic prediction with machine-learning approaches, and report prediction accuracy from properly separated training and validation sets.

4.4 Multi-omics and gene editing

Transcriptomics, metabolomics and other omics layers can help connect field phenotypes to biological mechanisms. Once robust candidate genes or regulatory variants are supported by association, expression and functional evidence, genome editing can be considered for targeted improvement. In foxtail millet, pan-genome resources now provide candidate structural variants and genes that can guide such work. Nevertheless, editing targets should not be inferred from the present seven-trait phenotypic dataset alone.

4.5 Climate-resilient breeding strategy

A practical modern pipeline for these elite lines would be: multi-environment re-phenotyping → high-throughput phenomics → genome-wide genotyping or resequencing → pan-genome-aware variant calling → GWAS and haplotype analysis → genomic prediction → environment-aware machine learning → validation of superior parents and crosses → functional validation of high-confidence genes. This pipeline preserves the strengths of classical breeding while adding genomic resolution and predictive selection.

5. Breeding Implications

The original results suggest complementary breeding roles among the clusters. High-grain-yield material from Cluster I may be crossed with high-biomass material from Cluster II or early, shorter material from Cluster III, depending on the target product profile. However, cluster distance alone does not guarantee heterosis or superior progeny. Modern parental selection should combine phenotypic divergence with genomic relationship, favorable haplotypes, stress-response data and multi-environment performance. The eight

divergent genotypes identified in the original work provide a focused candidate set for such validation.

6. Limitations

The dataset is based on 24 genotypes, seven quantitative traits and an original field experiment conducted in 2014. It does not contain genome-wide markers, transcriptomic profiles, high-throughput phenomic measurements or direct AI predictions. Therefore, the genomic, pan-genomic, machine-learning and genome-editing sections of this updated paper are evidence-based extensions and recommendations, not newly generated experimental findings. Before journal submission as a new original research article, additional contemporary experiments would be required to support claims of genomic or AI-assisted selection.

7. Conclusion

Classical multivariate analysis of the 24 elite foxtail millet genotypes revealed substantial diversity in flowering, plant architecture, maturity, grain yield and fodder yield. Four principal components captured 94.59% of total variation, and hierarchical clustering separated the material into three phenotypically distinct groups. The genotypes SiA 3145, SiA 3146, DhFtMV333, TNSI 267, SiA 3223, SiA 3221, SiA 3222 and SiA 3159 remain promising candidates for parental selection and renewed evaluation. In a 2026 breeding context, the strongest path forward is to integrate these phenotypic results with multi-environment trials, high-throughput phenotyping, pan-genome-aware genotyping, GWAS, genomic selection, machine learning and functional validation. Such integration can convert descriptive diversity into more precise and predictive climate-resilient foxtail millet improvement.

8. Acknowledgements

The original study acknowledged AICSMIP and ANGRAU for supporting the research work.

9. Conflict of Interest

The authors declare that they have no conflict of interest.

10. Funding

The authors received no specific funding for this study.

11. Ethical Approval

Ethical approval was not required for this study because it was based on previously conducted field data.

12. Informed Consent

Informed consent was not applicable because this study did not involve human participants.

13. Institutional Review Board Statement

Institutional Review Board approval was not required because this study did not involve human participants or animals.

14. Data Availability Statement

The data used in this study are based on previously generated phenotypic data from foxtail millet genotypes.

15. References

1. Adams MW. An estimate of homogeneity in crop plants with special reference to genetic vulnerability in the dry bean, *Phaseolus vulgaris* L. *Euphytica*. 1995;26:665-679.
2. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, et al. Reference genome sequence of the model plant *Setaria*. *Nature Biotechnology*. 2012;30(6):555-561.
3. Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, et al. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature Genetics*. 2013;45(8):957-961.
4. Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, et al. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature Biotechnology*. 2012;30(6):549-554.
5. Vetriventhan M, Azevedo VCR, Upadhyaya HD, Nirmalakumari A, Kane-Potaka J, Anitha S, et al. Genetic and genomic resources, and breeding for accelerating improvement of small millets: current status and future interventions. *The Nucleus*. 2020;63:217-239.
6. Upadhyaya HD, Vetriventhan M, Deshpande SP, Sivasubramani S, Wallace JG, Buckler ES, et al. Population genetics and structure of a global foxtail millet germplasm collection. *The Plant Genome*. 2015;8(3):1-13.
7. Srilatha J, Haritha T, Reddy CVC, Ramesh D. Genetic variability studies for yield and nutritional traits in foxtail millet (*Setaria italica* (L.) Beauv.). *Journal of Pharmacognosy and Phytochemistry*. 2020;9(6):596-600.
8. Zhang X, et al. Trait selection for yield improvement in foxtail millet (*Setaria italica* Beauv.) under climate change in the North China Plain. *Agronomy*. 2022;12(7):1500.
9. Reddy BV, Srinivasulu K, Chandra Mohan Reddy CV, Chandra Sekhar A, Shanthi P. Assessment of genetic diversity in foxtail millet using clustering and principal component analysis. *Plant Cell Biotechnology and Molecular Biology*. 2020;21(45-46):74-85.
10. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop Science*. 2003;43(4):1235-1248.
11. Panse VG, Sukhatme PV. *Statistical Methods for Agricultural Workers*. New Delhi: Indian Council of Agricultural Research; 1954.
12. Johnson HW, Robinson HF, Comstock RE. Genotypic and phenotypic correlations in soybeans and their implications in selection. *Agronomy Journal*. 1955;47(10):477-483.