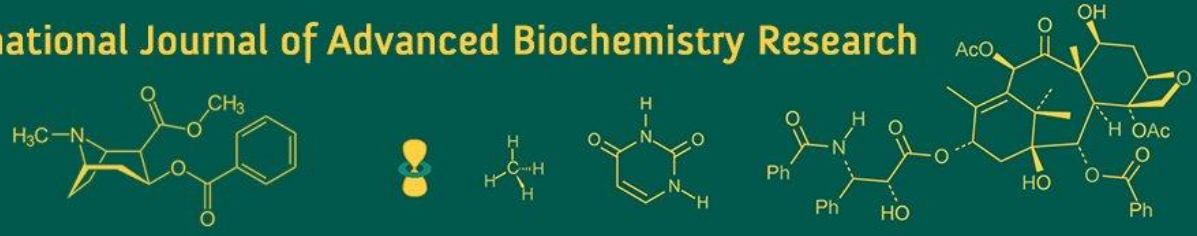
International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2025): 5.29
IJABR 2025; SP-9(12): 595-599
www.biochemjournal.com
Received: 07-09-2025
Accepted: 10-10-2025

Harish Vikram T
Department of Genetics and Plant Breeding, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, Karnataka, India

Lakshmana D
Professor, Department of Genetics and Plant Breeding, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, Karnataka, India

Santosh K Pattanashetti
Assistant Professor, Department of Genetics and Plant Breeding, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, Karnataka, India

Manjunatha B
Assistant Professor, Department of Genetics and Plant Breeding, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, Karnataka, India

Halingali BI
Professor and Head, Department of Agricultural Statistics, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, Karnataka, India

Sridhar S
Professor, Department of Agronomy, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, Karnataka, India

Corresponding Author:
Harish Vikram T
Department of Genetics and Plant Breeding, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, Karnataka, India

Cluster analysis of diverse maize germplasm panel for fodder yield and related traits

Harish Vikram T, Lakshmana D, Santosh K Pattanashetti, Manjunatha B, Halingali BI and Sridhar S

DOI: https://www.doi.org/10.33545/26174693.2025.v9.i12Sh.6570

Abstract
Maize (*Zea mays* L.) is an important fodder crop, and enhancing biomass yield, leaf area and overall forage quality requires systematic use of the genetic diversity available in global germplasm resources. In this study, the CIMMYT Asia Association Mapping (CAAM) panel was evaluated during two contrasting seasons, rabi 2023 (249 genotypes) and kharif 2024 (200 genotypes), to quantify phenotypic variability in fodder yield and related traits using k-means clustering. The panel was phenotyped under an alpha-lattice design for key phenological, morphological and biomass parameters. Across both seasons, k-means consistently grouped genotypes into six clusters, but composition of clusters during both the seasons are different, hence clustering of individual seasons is presented revealing a stable yet environmentally responsive structure of diversity, with clear shifts in cluster membership and trait expression reflecting genotype × season interactions. Distinct ideotypes were identified in each season. During rabi 2023, cluster 2 recorded the highest biomass with a green fodder yield of 26,941 kg ha^{-1} and a dry fodder yield of 9,579 kg ha^{-1}, supported by superior plant height and leaf area. In kharif 2024, cluster 3 emerged as the most productive group, achieving a green fodder yield of 38,462 kg ha^{-1}, while clusters 1 and 5 also expressed strong leafiness and high biomass potential. Inter-cluster distances were substantially greater than intra-cluster distances in both seasons, confirming meaningful divergence among clusters and highlighting clusters 1, 2 and 3 as promising parental pools for creating wide and complementary heterotic combinations. The two-season analysis clearly demonstrates the presence of well-structured and exploitable phenotypic variation in the CAAM panel. The identification of stable high-performing clusters, along with season-specific divergent groups, provides a robust foundation for cluster-based parent selection and for designing targeted breeding strategies aimed at developing high-yielding, leafier and environmentally resilient fodder maize cultivars.

Keywords: CAAM panel, cluster, diversity, fodder and maize

1. Introduction
Maize (*Zea mays* L.) is one of the world's most versatile cereal crops, valued not only for its grain but also as an indispensable source of green fodder and silage in livestock-based farming systems. Its extensive genetic and phenotypic diversity, shaped over thousands of years through domestication, natural adaptation and human selection, offers a rich foundation for improving complex fodder-related traits such as biomass yield, leafiness, nutritional quality and stress resilience (Matsuoka *et al.*, 2002) [6]. In recent decades, the rapid growth in global livestock demand and the intensification of mixed crop-livestock production systems have emphasized the need for high-yielding, nutritionally superior and climate-resilient fodder maize cultivars. Simultaneously, environmental challenges such as increasing temperatures, erratic rainfall and degraded soil health have reinforced the necessity of identifying diverse and adaptable maize germplasm capable of maintaining biomass production under stress (FAO, 2010) [2]. Consequently, systematic diversity analysis has become crucial not only for effective germplasm conservation but also for identifying unique accessions, selecting parents for maximizing heterosis, defining ideotypes and developing breeding strategies that match the needs of different production environments. Machine-learning approaches, particularly clustering algorithms, offer powerful solutions for handling such complex multivariate data. Among these, k-means clustering, an unsupervised learning method introduced by MacQueen (1967) [4] is especially well suited for organizing

phenotypic variability because of its computational efficiency, scalability and straightforward biological interpretation. By partitioning accessions into groups with minimized within-cluster variability, k-means provides clear summaries of trait patterns in the form of cluster centroids, enabling researchers and breeders to easily compare groups of genotypes based on biomass, plant architecture and quality-related traits. In fodder maize, this facilitates the identification of clusters representing high-biomass types, stress-resilient types or ideotypes combining multiple desirable attributes. Moreover, k-means clustering supports downstream applications such as defining breeding pools, selecting diverse parental combinations, identifying representative genotypes for multilocation testing and structuring data for association mapping or genomic prediction. Thus, integrating k-means clustering into fodder maize diversity analysis strengthens both scientific understanding and practical decision-making in breeding programs. In this context, the present study aims to analyse the phenotypic diversity of maize germplasm for fodder-related traits using k-means clustering, with the goal of identifying distinct diversity groups, characterizing their trait profiles across seasons and generating insights that can guide targeted breeding for improved fodder yield, quality and resilience.

2. Materials and Methods

The present study utilized a diverse maize germplasm panel known as the CIMMYT Asia Association Mapping Panel (CAAM panel), developed and maintained at the International Maize and Wheat Improvement Center (CIMMYT) Asia Hub, Hyderabad, India. The panel consists of genetically diverse maize accessions encompassing wide variation in tropical and subtropical genetic background. Field evaluation of the CAAM panel was carried out during two cropping seasons *i.e*., *Rabi* 2023 and *Kharif* 2024, for 249 and 200 genotypes respectively, at the College of Agriculture, Shivamogga, using alpha lattice design with two replications. Standard agronomic practices were followed for optimum crop establishment, growth and protection. Phenotyping was conducted for fodder yield and related traits including phenological, morphological and yield. The mean values across replications were obtained for subsequent analysis after rigorous data validation and removal of outliers. To characterize the phenotypic diversity within the CAAM panel, k-means clustering approach (MacQueen, 1967) [4] was employed to explore the underlying structure of variability using the 'Factoextra' package in R (Version 4.1.1). The statistical significance of differences among cluster-wise trait means was assessed using Levene's test for homogeneity of variances (Levene, 1960). Cluster-wise mean values for each trait were then computed using the standard formula,

$$J = \sum_{i=1}^{m} \sum_{k=1}^{k} \omega_{ik} \|X^2 - \mu_k\|^2$$

Where,
J = Objective (cost) function to be minimized
K = Total number of clusters
m = Total number of data points
xi-ith data point
μk = Center (mean vector) of cluster k
ωik = 1, if data point xi belongs to cluster k; otherwise, 0
$\|X^2-\mu k\|^2$ = Squared Euclidean distance between data point X_i and its cluster center μ_k

3. Results and Discussion

3.1 Clustering analysis

The 249 genotypes in season 1 (Table 1) and 200 genotypes in season 2 (Table 2) were consistently grouped into six clusters in each season (Fig. 1 and Fig. 2), demonstrating a clear and repeated pattern of multivariate divergence and confirming the presence of structured genetic variability within the panel. Although the number of clusters remained the same, the cluster sizes differed between seasons, with season 1 clusters ranging from 24 to 64 genotypes and season 2 clusters from about 22 to 43 genotypes, indicating that some groups of genotypes were more sensitive to seasonal conditions than others. The existence of the same six cluster labels (clusters 1-6) across both seasons, coupled with frequent shifts of individual genotypes from one cluster to another, shows that while the overall grouping framework is stable, the multivariate phenotypic expression of many genotypes is influenced by environmental variation, revealing genotype × season interaction for the traits used in clustering. In Rabi 2023, the largest grouping of inbreds was observed in Cluster 5 with 64 entries, while Clusters 4, 1, 6, 2 and 3 followed with 49, 42, 39, 31 and 24 lines, respectively. In Kharif 2024, Cluster 3 accommodated the highest number of inbreds (43), followed by Clusters 4, 6, 2, 5 and 1, which consisted of 38, 37, 35, 25 and 22 lines, respectively. This pattern suggests that part of the diversity is environmentally stable, represented by clusters that repeatedly harbor large groups of genotypes and part is environment-specific, represented by smaller or fluctuating clusters that may contain genotypes with more extreme or unique trait combinations.

3.2 Cluster means

Across seasons, the cluster means (Table 3 and Table 4) reveal well-differentiated ideotypes for both phenology and fodder productivity, which is a prerequisite for meaningful exploitation of divergence in a high-impact breeding programme. In season one, cluster 2 combined taller plants, greater leaf dimensions and the largest leaf area with superior green and dry fodder yields, marking it as the primary high-yielding group, while clusters 3 and 5 showed intermediate to high plant stature and leafiness with moderately high yields, suggesting useful secondary sources for yield and biomass traits. Clusters 1, 4 and 6 were generally inferior for leaf area and fodder yields, despite acceptable plant height in some cases, indicating that they largely represent low-to-moderate yielding backgrounds that may still contribute specific traits (flowering time, stem traits) rather than overall productivity. In season two, the pattern of divergence among clusters was maintained but with a clear re-ranking, reflecting genotype × season interaction for the underlying traits. Cluster 3 expressed the highest green and dry fodder yields, coupled with tall plants and adequate leaf area, thereby representing the most productive group under this environment, whereas clusters 1 and 5 also recorded high biomass with very large leaves and leaf areas, distinguishing them as complementary high-yielding ideotypes with enhanced leafiness. By contrast, cluster 2 in season two showed shorter plants, reduced leaf area and relatively lower fodder yield and

clusters 4 and 6 exhibited moderate to low yields, which together delineates a spectrum from highly productive to relatively unproductive groups that can be strategically targeted either for direct selection or as donors of earliness, reduced stature, or specific morphological traits. Considering both seasons jointly, clusters that consistently exhibit early to medium flowering with tall stature, large leaves and superior leaf area, along with high green and dry fodder yields (notably cluster 2 in season one and clusters 1, 3 and 5 in season two), can be regarded as robust donors for biomass improvement in diverse environments. The presence of clusters with contrasting phenology (DFT and DFS), plant height and yield profiles across seasons offers opportunities to design crosses between highly divergent yet complementary clusters, maximizing heterosis for fodder yield while simultaneously pyramiding favourable leaf and stem traits. Such a structured divergence in cluster means, coupled with season-dependent shifts in the identity of the top-performing clusters, underlines the need for multi-environment selection and provides a strong empirical basis to propose cluster-based parent selection schemes aimed at delivering high-yielding and stable fodder genotypes for target production environments.

3.3 Average intra and inter cluster distance

In season one (Table 5), the intra-cluster distances ranged from 2.50 (cluster 4) to 4.84 (cluster 3), indicating that cluster 3 was the most heterogeneous, followed by clusters 2 (3.63) and 1 (3.44), whereas clusters 4 (2.83), 5 (2.82) and 6 (2.67) were comparatively more homogeneous. Inter-cluster distances varied more widely, with the highest divergence recorded between clusters 1 and 2 (6.03), followed by clusters 2 and 6 (4.90) and clusters 1 and 3 (5.13), while the closest clusters included 1-4 (2.50), 4-6 (2.26) and 4-5 (2.84), reflecting a clear gradient from very divergent to relatively similar cluster pairs. In season two (Table 6), intra-cluster distances were generally of similar magnitude across clusters, ranging from 2.87 (cluster 5) to 4.04 (cluster 1), with clusters 1 (4.04), 3 (3.63) and 2 (3.50) showing relatively higher within-cluster variability and clusters 5 (2.87), 4 (3.19) and 6 (3.29) being more compact. Inter-cluster distances in this season were more balanced, with notable high values between clusters 1 and 6 (3.84), 4 and 5 (3.73) and 2 and 5 (3.42), while the smallest distances were observed between clusters 2 and 6 (2.30), 2 and 4 (2.55) and 1 and 4 (2.76), again indicating a spectrum of divergence levels among cluster pairs. Taken together, the numerical pattern across seasons shows a convergent picture: intra-cluster distances are consistently lower than inter-cluster distances, confirming that clusters represent compact groups embedded within a broader landscape of genetic divergence. Clusters with higher intra-cluster distances and large inter-cluster values (clusters 1, 2 and 3 in both seasons) represent rich sources of variability that can be exploited both for within-cluster selection and for wide crosses, whereas more homogeneous but still inter-cluster-divergent groups (such as clusters 4, 5 and 6 in several pairings) can serve as relatively uniform donors to be crossed with distant, variable clusters to maximize heterosis while maintaining manageable segregation.

Table 1: Clustering pattern of 249 genotypes of maize CAAM panel during Rabi 2023

Cluster No.	No. of genotypes	Entry No.
1	42	4, 9, 13, 29, 63, 65,75, 76, 78, 90, 92, 93, 97, 100, 101, 102, 107, 114, 115, 126, 131, 134, 135, 137, 138, 141, 145, 147, 151, 153, 163, 198, 210, 211, 213, 215, 216, 230, 233, 235, 238, 241
2	31	10, 18, 27, 50, 55, 67, 69, 70, 71, 105, 111, 127, 128, 154, 155, 164, 167, 169, 172, 177, 179, 186, 187, 188, 189, 190, 207, 208, 224, 247, 249
3	24	11, 12, 19, 21, 28, 31, 32, 36, 38, 46, 48, 49, 58, 68, 89, 94, 104, 122, 123, 129, 130, 143, 231, 239
4	49	2, 3, 6, 8, 14, 26, 35, 42, 43, 44, 47, 52, 57, 64, 66, 82, 83, 84, 85, 86, 87, 88, 96, 98, 99, 103, 112, 113, 116, 117, 121, 124, 125, 132, 133, 136, 140, 144, 146, 150, 157, 158, 160, 184, 185, 194, 203, 214, 234
5	64	15, 20, 24, 33, 51, 53, 54, 56, 77, 81, 91, 108, 109, 110, 119, 120, 142, 148, 149, 156, 161, 162, 165, 166, 168, 170, 171, 173, 174, 175, 180, 181, 182, 183, 191, 192, 196, 197, 199, 200, 201, 202, 204, 206, 212, 218, 219, 220, 221, 222, 223, 226, 227, 228, 229, 232, 236, 237, 240, 242, 243, 245, 246, 248
6	39	1, 5, 7, 16, 17, 22, 23, 25, 30, 34, 37, 39, 40, 41, 45, 59, 60, 61, 62, 72, 73, 74, 79, 80, 95, 106, 118, 139, 152, 159, 176, 178, 193,195, 205, 209, 217, 225, 244

Table 2: Clustering pattern of 200 genotypes of maize CAAM panel during Kharif 2024

Cluster No.	No. of genotypes	Entry No.
1	22	2, 9, 10, 16, 17, 25, 27, 32, 33, 40, 46, 47, 57, 63, 81, 85, 87, 104, 117, 150, 158, 193
2	35	1, 4, 7, 11, 18, 20, 22, 29, 49, 51, 52, 60, 83, 84, 90, 92, 94, 96, 112, 114, 120, 124, 139, 142, 147, 151, 155, 163, 167, 174, 176, 179, 188, 194, 196
3	43	8, 13, 14, 23, 24, 31, 42, 50, 55, 56, 72, 91, 95, 97, 99, 102, 105, 106, 108, 111, 113, 116, 119, 121, 128, 131, 132, 136, 144, 146, 152, 159, 161, 162, 164, 169, 180, 181, 183, 185, 191, 195, 197
4	38	3, 5, 6, 12, 15, 21, 26, 30, 34, 36, 37, 39, 43, 48, 53, 54, 59, 64, 71, 73, 77, 78, 80, 82, 88, 100, 101, 103, 110, 115, 127, 135, 137, 170, 173, 186, 187, 190
5	25	44, 58, 67, 69, 70, 74, 75, 89, 109, 118, 122, 123, 126, 129, 133, 134, 141, 145, 148, 153, 157, 160, 165, 199, 200
6	37	19, 28, 35, 38, 41, 45, 61, 62, 65, 66, 68, 76, 79, 86, 93, 98, 107, 125, 130, 138, 140, 143, 149, 154, 156, 166,168, 171 172, 175, 177, 178, 182, 184, 189, 192, 198

Table 3: Cluster-wise mean performance for fodder yield and related traits in maize during Rabi 2023

Cluster	DFT	DFS	PH	LP	LL	LW	LA	SD	LSR	GFY	DFY
1	65.42	70.38	100.82	9.63	50.62	4.79	1770	16.59	0.55	16157	5292
2	64.58	68.60	129.23	11.56	59.88	6.25	3285	22.51	0.55	26941	9579
3	58.50	61.19	114.81	10.59	63.22	5.85	2941	18.69	0.47	25145	8183
4	61.39	65.14	103.62	9.93	53.54	5.11	2041	19.13	0.60	20713	7109
5	68.09	72.56	115.02	10.52	55.91	5.79	2551	18.33	0.54	21361	7335
6	61.71	65.31	117.35	9.92	52.99	5.23	2089	16.66	0.41	19256	6208

Table 4: Cluster-wise mean performance for fodder yield and related traits in maize during Kharif 2024

Cluster	DFT	DFS	PH	LP	LL	LW	LA	SD	LSR	GFY	DFY
1	57.55	60.39	121.63	11.26	73.89	5.43	3368.85	18.73	0.52	31478.96	7656.78
2	62.93	65.67	97.98	9.55	60.08	5.72	2431.51	16.01	0.46	23238.40	6925.26
3	63.19	66.10	118.07	10.56	59.66	5.17	2426.35	21.56	0.43	38462.90	8798.26
4	57.84	61.05	107.72	10.07	54.76	5.50	2259.39	18.21	0.60	30056.16	7770.07
5	64.40	67.92	127.95	11.98	65.11	5.85	3414.17	19.76	0.58	33211.40	7645.99
6	63.73	66.50	120.69	10.81	55.63	4.84	2161.56	15.96	0.59	24442.41	6583.86

Table 5: Intra and inter cluster distance values for fodder yield and related traits in maize during Rabi 2023

	1	2	3	4	5	6
1	3.44	6.03	5.13	2.50	3.13	2.56
2	6.03	3.63	3.22	4.43	3.33	4.90
3	5.13	3.22	4.84	3.22	3.79	3.29
4	2.50	4.43	3.22	2.83	2.84	2.26
5	3.13	3.33	3.79	2.84	2.82	2.99
6	2.56	4.90	3.29	2.26	2.99	2.67

Table 6: Intra and inter cluster distance values for fodder yield and related traits in maize during Kharif 2024

	1	2	3	4	5	6
1	4.04	3.54	3.35	2.76	2.98	3.84
2	3.54	3.50	3.09	2.55	3.42	2.30
3	3.35	3.09	3.63	2.92	2.88	3.14
4	2.76	2.55	2.92	3.19	3.73	2.79
5	2.98	3.42	2.88	3.73	2.87	3.34
6	3.84	2.30	3.14	2.79	3.34	3.29

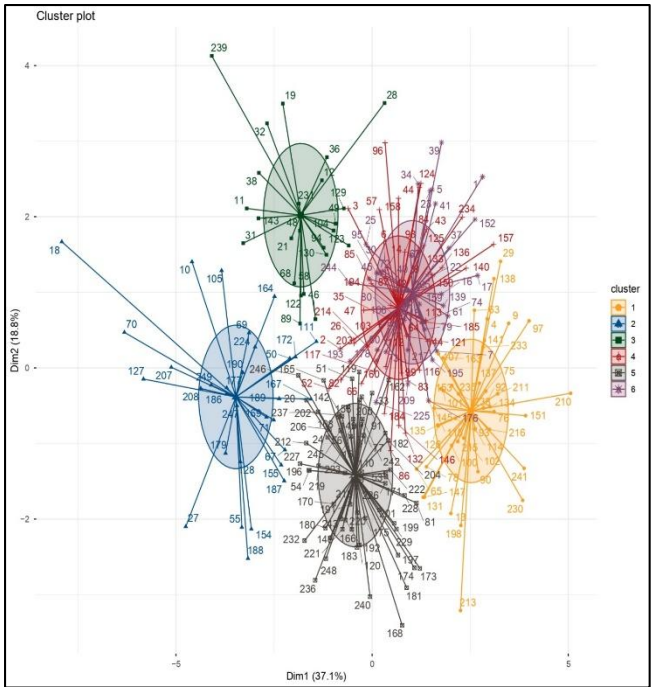


Fig 1: Cluster analysis for fodder yield and related traits in maize during Rabi 2023

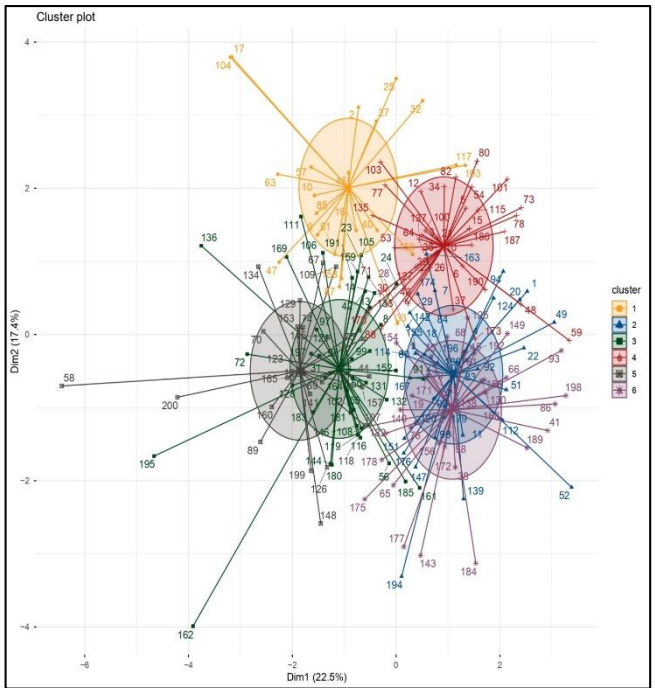


Fig 2: Cluster analysis for fodder yield and related traits in maize during Kharif 2024

4. Conclusion

The comprehensive two-season evaluation of the CAAM maize panel revealed a stable yet environmentally responsive pattern of multivariate divergence, with genotypes consistently resolving into six distinct clusters across both seasons. While the clustering structure remained constant, substantial re-ranking of cluster sizes, mean performances and divergence metrics highlighted pronounced genotype × season interactions governing key fodder-related traits. Distinct and contrasting ideotypes emerged, ranging from tall, leafy, high-biomass groups to compact, low-yielding but potentially trait-specific donors, thereby offering multiple avenues for targeted improvement. Clusters 2 (season one) and clusters 1, 3 and 5 (season two) consistently represented high-yielding groups with favourable phenology, leaf area and plant stature, marking them as robust donor pools for biomass enhancement. Inter- and intra-cluster distances further confirmed the presence of substantial and exploitable divergence, especially among clusters 1, 2 and 3, which provide opportunities for wide, heterotic cross-combinations. The structured variability

documented across environments underscores the importance of multi-season selection and validates cluster-based parental selection as an effective strategy for assembling complementary traits. Overall, the results establish a clear empirical foundation for using k-means-based diversity analysis to guide the development of high-yielding, stable and widely adaptable fodder maize genotypes in breeding programmes targeting variable production environments. Similar studies on analysing diversity on maize was done by Bai *et al.* (2025) [1], Mastrangelo *et al.* (2024) [5], Banakara *et al.* (2024) [8], James *et al.* (2024) [3], Raman *et al.* (2024) [7] and Khan *et al.* (2022) [9].

References

1. Bai L, Wang K, Zhang Q, Zhang Q, Wang X, Pan S, *et al.* A study of maize genotype-environment interaction based on deep K-means clustering neural network. Agriculture. 2025;15(4):358.
2. FAO. The second report on the state of the world's plant genetic resources for food and agriculture. Rome: Food and Agriculture Organization of the United Nations; 2010.
3. James M, Krishna TPA, Priyadarsini V, Gladina A, Senthil Kumar P. Identification of diverse maize inbred lines using molecular and phenotypic genetic divergence analysis. Int J Bio-resour Stress Manag. 2024;15(10):01-08.
4. MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol 1. Berkeley: University of California Press; 1967. p. 281-297.
5. Mastrangelo AM, Milano M, Ciani E, Paolacci AR, Porceddu A, De Pace C, *et al.* Genetic diversity within a collection of Italian maize inbred lines: a resource for maize genomics and breeding. Plants. 2024;13(3):336.
6. Matsuoka Y, Vigouroux MM, Goodman J, Sanchez G, Buckler E, Doebley J. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A. 2002;99(9):6080-6084.
7. Raman RB, Samhita C, Ashutosh S, Alok K, Jyoti K, Mala K, *et al.* Genetic diversity analysis in maize (*Zea mays* L.) germplasm based on yield and yield attributing characters. Int J Adv Biochem Res. 2024;8(12):1093-1098.
8. Banakara S, Lohithaswa HC, Sunilkumar B, Kalmeshwer G, Sinchana K, Sowmya MS, *et al.* Genetic variability for per se grain yield potential and its attributing traits among inbred lines of maize (*Zea mays* L.). Plant Arch. 2024;24(2):485-490.
9. Sumayea K, Firoz M, Tahmina A. Genetic diversity with cluster analysis of maize genotypes (*Zea mays* L.). Adv Biosci Biotechnol. 2022;13(7).