

RESEARCH ARTICLE

Unraveling the Hidden Patterns of Genetic Diversity in Oat (*Avena sativa* L.) Accessions: Insights from Novel Hybrid Hierarchical K-means Clustering

Rukoo Chawla^{1,2*}, Minakshi Jattan², DS Phogat², Amit Sharma² and Mandeep Redhu³

Abstract

Avena sativa is primarily grown as a food crop but also serves as a valuable feed for livestock. Oat is a versatile and nutritious grain with a range of health benefits, making it a popular choice for consumers and a valuable crop for agricultural production. In pursuit of crop improvement, cluster analysis is an analytical approach that allows researchers to uncover underlying patterns and structures within a dataset. Eventually, this aids in identifying exemplary genotypes for improved yield attributes. The present study involved a meticulous examination of phenotypic data of 62 oat genotypes taken during *Rabi* 2019-20 at CCS Haryana Agricultural University, Hisar, Haryana. A novel and superior technique, hybrid hierarchical K-means, was employed for enhanced outcomes in comparison to traditional hierarchical and K-means clustering methods. Among the four distinct clusters formed, cluster I exhibited the highest count of genotypes. Additionally, cluster I had the lowest intra-cluster distance and the highest group mean, implying the presence of additive gene action and favorable combinations of alleles for yield-related traits, respectively. Clustering offers breeders the advantage of focusing their efforts on representative genotypes within each cluster, thereby accentuating the efficiency of resource allocation and the management of breeding materials. Promising genotypes from the most diverse clusters can act as parents for hybridization programs. The research will enhance our understanding of oat genetics and facilitate the development of superior cultivars. Ultimately, this will help optimize yields and meet the growing demand for this nutritious crop.

Keywords: Clustering, Diversity, Forage, Heterosis, H-K-means, Oat.

¹Department of Genetics and Plant Breeding, MPUAT, Udaipur, Rajasthan, India.

²Department of Genetics and Plant Breeding, CCS HAU, Hisar, Haryana, India

³Department of Plant, soil and Agricultural system, college of Agriculture, Life and Physical Sciences, Southern Illinois University, Carbondale, IL, U.S.A.

***Author for correspondence:**

rukoochawla21@gmail.com

Received: 22/12/2023 **Revised:** 04/04/2024

Accepted: 01/07/2024

How to cite this article: Chawla R, M Jattan, DS Phogat, A Sharma and M Redhu (2024) Unraveling the Hidden Patterns of Genetic Diversity in Oat (*Avena sativa* L.) Accessions: Insights from Novel Hybrid Hierarchical K-means Clustering. *Indian J. Plant Genet. Resour.* 37(3): 454-459. DOI: 10.61949/0976-1926.2024.v37i03.07

Introduction

Oat (*Avena sativa*) belonging to family Poaceae is regarded as a prominent multifunctional cereal owing to its inherent versatility (Prates and Yu 2017, Chawla *et al.*, 2021). It sets itself apart from other cereals due to a plethora of unparalleled benefits, rendering it highly coveted and favored by a substantial multitude of individuals (Ahmad *et al.*, 2014). Oat covers a worldwide production of 23.3 million metric tons (USDA, 2023). The worldwide oats market experienced growth from \$6.67 billion in 2022 to \$7.13 billion in 2023, reflecting a compound annual growth rate (CAGR) of 7.0%. Oats are known for their adaptability to diverse climatic conditions and can be cultivated in a wide range of regions around the world (Isidro-Sanchez *et al.*, 2020). The crop possesses attributes that render it highly valuable for both human alimentation and animal fodder (Chawla *et al.*, 2022). As a livestock feed, it offers a combination of nutritional composition, digestibility and palatability that makes it an excellent choice for a variety of animals (Ross *et al.*, 2004; Nikolaudakis, 2016). The non-commercial character of forage crops and the generation of fodder with minimal inputs from deteriorated & marginal land results in a substantial disparity between the availability & demand of fodder (Singh *et*

et al., 2022a). The aforementioned circumstances underscore the criticality of conducting research on forage crops in order to address this disparity effectively (Surje and De, 2014; Krishna *et al.*, 2014). Furthermore, achieving high yield despite its health benefits, remains a challenge majorly due to low genetic potential, weed competition, inadequate agronomic practices and limited investment in breeding programs (Gorash *et al.*, 2017; Yang *et al.*, 2023). Addressing these factors through improved genetics, increased research & investment can help optimize oat yields and meet the growing demand for this nutritious cereal crop.

To enhance the progress of any breeding program, it is imperative to systematically assess the available materials to identify diverse lines suitable for hybridization programs (Kiran *et al.*, 2023; Krishna *et al.*, 2014). Cultivars possessing enhanced yield potential, along with the ability to thrive at elevated altitudes beyond the existing climatic thresholds for arable farming, have the potential to alleviate the strain caused by overgrazing and deforestation (Arora *et al.*, 2021). Crop germplasm serves as a reservoir of genetic diversity that can be utilized for crop improvement programs. Different studies have highlighted the importance of genetic diversity assessment in crop plants and the recent analytical perspectives & advancements in crop breeding (Govindaraj *et al.*, 2015).

In this research, multivariate analysis was employed to expedite the identification of promising heterotic combinations suitable for harnessing diverse improved cultivars. Clustering analysis serves as a valuable tool in crop breeding, providing breeders with a systematic and data-driven approach to understand & exploiting the genetic diversity within crop populations (Mohammadi and Prasanna, 2003). By leveraging the insights gained from clustering analysis, breeders can make effective decisions, optimize breeding strategies and ultimately develop improved crop varieties with enhanced traits & productivity. This paper employs a novel and unconventional methodology for clustering analysis, i.e., hybrid hierarchical K means clustering, presenting an improved and distinctive approach. Following a comprehensive assessment of diverse oat genotypes, this investigation substantially facilitated the discernment of the most promising genotypes for yield attributes (Kaur *et al.*, 2018). As a result, this study will make a substantial contribution toward the significant attainment of food and fodder security.

Materials and Methods

Plant Material and Experimental Details

The research consists of 62 oat genotypes, including 3 checks (OS 6, JHO 851 and UPO 212) that were procured from diverse regions across the country (Supplementary Table 1). The study was undertaken at the Research Farm Area, Forage Section, Department of Genetics and Plant Breeding, CCS

Haryana Agricultural University, Hisar (Haryana) (latitude 29° 10' North and longitude 75° 46' East with an altitude of 215.2 m above the mean sea level) during *Rabi* 2019-20. The trial was conducted in an augmented design with seven blocks, each having three checks. Each genotype was sown in three rows of 3 m length each.

Data Recording

The data on phenotypic traits were collected from five randomly chosen vigorous plants for each genotype except for DFF and DM, which was on a plot basis. The data was recorded at 50% flowering stage and their mean value was taken. A total of 13 phenotypic traits were evaluated including PH- Plant height (cm), TPL- Number of tillers/plant, NOL- Number of leaves/plant, LL- Leaf length (cm), LW- Leaf width (cm), PL- Peduncle length (cm), NOS- Number of spikelets/panicle, GFY- Green fodder yield/plant (g), DMY- Dry matter yield/plant (g), DFF- Days to 50% flowering, DM- Days to maturity, SY- seed yield/plant (g), TW- Test weight (1000 seed weight in g).

Statistical Analysis

Initial hierarchical cluster analysis was performed using the methodology proposed by Ward, 1963. The k-means algorithm initiates by selecting a random set of observations as the initial cluster centers. However, the final clustering outcome is highly dependent on this initial random selection, potentially yielding different results with each computation. To mitigate this issue, a solution called hybrid hierarchical k-means clustering (k-means) has been devised, combining the strengths of hierarchical clustering and traditional k-means methods. By incorporating both approaches, k-means offers enhanced scalability, flexibility, interpretability, cluster compactness and outlier handling. These advantages render hkmeans a valuable and applicable technique for clustering analysis across diverse domains. R studio software, version 26.0 (R Core Team, 2020) was used for analysis. All the graphs were drawn using the same software. The inter and intra-cluster distances were calculated on the basis of Euclidean distance while considering the mean values of various traits as cluster mean values.

Results and Discussion

Despite the considerable genetic variability indicated by the mean sum of squares for all traits, the analysis of variance alone does not provide a comprehensive explanation for the extent of genetic diversity observed. Understanding and harnessing genetic divergence greatly contribute to the invaluable enhancement of crops (Kumari and Jindal, 2019). The presence of a substantial multitude of clusters serves as evidence of ample diversity among the oat genotypes. Upon examination of the phenotypic data acquired from 62 oat genotypes, a comprehensive analysis led to the formation

of a total of four distinct clusters. The clustering pattern unveiled that cluster I had the highest count of genotypes, comprising 28 genotypes, while cluster IV emerged as the second largest cluster, encompassing 14 genotypes. Additionally, cluster II comprised 11 genotypes, whereas cluster III constituted the smallest count, comprising nine genotypes (Table 1). The maximum inter-cluster distance was exhibited between clusters II and III (5.93) followed by clusters II and IV (5.89). On the contrary, the least inter-cluster was demonstrated between clusters I & II. The highest intra-cluster distance was manifested by cluster II, while the minimum intra-cluster distance was showcased by cluster I (Table 2). Cluster I exhibited the most elevated mean values for economic traits, including seed yield (15.8), green fodder yield (177.5) and dry matter yield (35.8). Additionally, this cluster demonstrated the highest mean values for other attributes, namely PH (147.8), TPL (11.0), NOL (59.3), LW (2.6), DFF (115.0) and DM (145.3) (Table 3). Moreover, it was observed that the cluster mean values for these traits within cluster I exhibited a statistically significant superiority over the general mean value derived from the population of 62 oat genotypes. Cluster III had the highest mean values of NOS and DFF, whereas cluster IV had the highest mean value for LL. Poonia and Phogat, 2017 found similar findings where cluster 1 had the highest mean values for GFY, DMY and PH.

In hk-means clustering, the algorithm initially applies hierarchical clustering to the dataset, which involves creating a hierarchical structure of clusters based on similarity measures. This hierarchical structure is represented in visual presentation as a dendrogram. Four distinct clusters emerge; visually differentiated by different colors as illustrated in Figure 1. Furthermore, after generating the initial clustering

Table 1: Clustering details of 62 oat genotypes based on D² statistics

Cluster number	Genotypes
I	OS 6, UPO 212, HFO 1116, HFO 1117, HFO 1118, HFO 1121, HFO 1122, GP 65, GP 298, HFO 424, HFO 902, HFO 903, HFO 1013, HFO 1016, HFO 1109, HFO 1111, HFO 1112, HFO 1113, HFO 1114, HFO 1115, KENT, RO-11-2-2, JHO 2006-1, OL 1766-2, HFO 1107, HFO 1105, OL 125, OL 1861
II	JHO 851, GP 192, GP 492, HFO 1123, OL 1874-2, NDO-1, PLP-1, RO-11-2-6, JHO 99-1, JO-1, HFO 1106
III	HFO 529, GP 580, HFO 607, GP 68, GP 158, HFO 1101, HFO 1104, HFO 1108, OL 1869-1
IV	HFO 611, HFO 707, GP 781, HFO 806, HFO 818, GP 875, HFO 901, HFO 915, HFO 917, HFO 1003, HFO 1005, HJ 8, OS 403, JHO 822

Table 2: Intra and inter-cluster distance based on average Euclidean D² values

	I	II	III	IV
I	3.53	4.87	5.48	5.00
II	4.87	4.62	5.93	5.89
III	5.48	5.93	3.84	5.30
IV	5.00	5.89	5.30	4.49

Diagonal values indicate intra-cluster distance

hierarchy, the algorithm applies the k-means clustering approach to refine the clusters. The k-means algorithm assigns each data point to the nearest centroid and updates the centroids iteratively until convergence. This step helps to optimize the cluster assignments and improves the compactness & separation of the resulting clusters. This is presented in the form of a cluster plot having 4 distinct groups (Figure 2). The cluster membership in this plot has been adjusted and refined compared to the dendrogram, resulting in distinct and improved groupings.

Cluster III exhibited a minimum number of genotypes, signifying a constricted genetic foundation. It is characterized by their shared ancestral origin and diminished relatedness to the genotypes belonging to other clusters. Inter-cluster distance helps in identifying distinct clusters harboring divergent characteristics (Paw *et al.*, 2020). This in turn, aids in the meticulous selection of the diverse genotypes. Breeding programs frequently aspire to forge hybridizations between genetically distant clusters to instill novel traits, augment genetic diversity and foster the development of superior cultivars, as corroborated by the study conducted by Bhandari *et al.*, (2017). Clusters II and III exhibited the greatest

Table 3: Cluster mean and general mean values of 13 phenotypic traits in 62 oat genotypes

Traits	I	II	III	IV	General mean
PH	147.8	145.6	124.6	143.3	143.0
TPL	11.0	7.0	8.6	8.1	8.1
NOL	59.3	36.2	43.3	42.2	42.3
LL	49.9	49.8	45.4	55.6	50.9
LW	2.6	2.2	2.2	2.2	2.2
PL	31.3	33.7	23.4	32.6	31.8
NOS	66.6	59.2	74.1	60.4	62.4
GFY	177.5	131.5	119.7	150.9	143.1
DMY	35.8	24.4	21.3	28.8	27.1
DFF	115.0	113.8	115.0	110.6	113.2
DM	145.3	141.3	143.0	137.2	140.9
SY	15.8	9.3	12.8	12.9	11.7
TW	32.9	32.4	27.0	37.2	33.2

Note: The values shown in bold are the highest cluster mean values for the respective traits

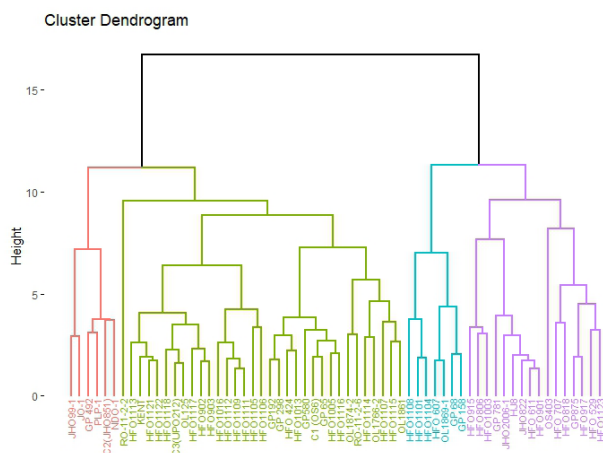


Figure 1: Dendrogram based on initial clustering (hierarchical) of 62 oat genotypes for thirteen phenotypic traits

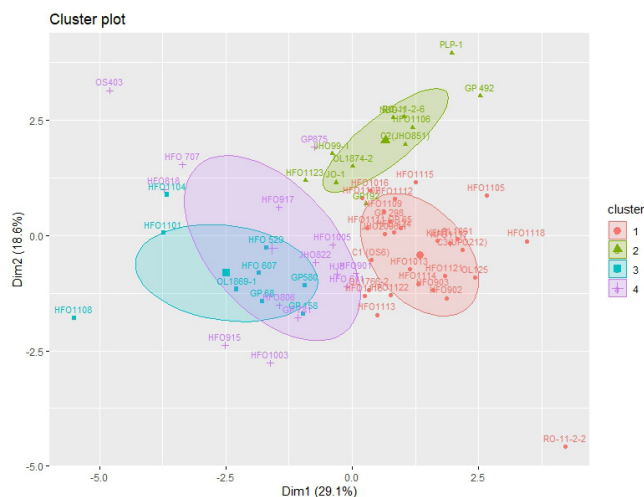


Figure 2: K means cluster plot inferred from final clustering (K means clustering). Refined genotypes which led to final cluster membership

inter-cluster distance, indicating significant dissimilarity between them. Within cluster II, the genotype JHO 851 is a multi-cut check. While genotype HFO 1108 present within cluster III showed the highest GFY and DMY among all 62 genotypes. The cross-breeding of these genotypes has the potential to yield highly heterotic hybridization, akin to the effective implementation observed in the pedigree method. Similarly, second most distant clusters were II and IV. In cluster IV, HFO 707 & HFO 818 were high yielding genotypes for GFY and OS 403 & GP 781 were high seed yielding genotypes. Henceforth, genotypes from these two clusters act as potent source of diverse yet good performing parents required for hybridization (Poonia *et al.*, 2020). Intra-cluster distances, by measuring the genetic similarity within clusters can indicate the extent to which additive genetic effects govern the observed phenotypic variation (Chakravorty *et al.*, 2012). When genotypes within a cluster exhibit small intra-cluster distance, it suggests that they share a high

degree of genetic similarity and likely they possess similar combinations of additive alleles. This indicates the presence of strong additive gene action, as the phenotypic variation within the cluster can be attributed to the additive effects of genes that are relatively fixed within the cluster (Carroll *et al.*, 2015). Understanding the extent of additive gene action is crucial in breeding programs and genetic improvement efforts. The highest intra-cluster distance was seen for cluster II (4.62); indicating greater genetic diversity within the cluster. Kumari and Jindal, 2019 observed the highest intra-cluster distance to be 6.02, depicting genetic relation among genotypes. This implies that a wider range of genetic factors, including non-additive gene effects such as dominance and epistasis, influences the observed phenotypic variation. For instance, a study assessed 25 morphological traits and found varying degrees of differentiation, indicating a broad genetic base influencing these traits. It suggested that a wider range of genetic factors influences the observed phenotypic variation within the cluster of early Polish oat cultivars (Boczkowska *et al.*, 2014).

The mean values of traits within a cluster represent the average performance of genotypes within that cluster (Mohammadi and Prasanna, 2003). It provides quantitative evidence of the performance and potential of genotypes for specific traits (Singh *et al.*, 2022b). This average performance can provide insights into the genetic characteristics and potential of genotypes in terms of their ability to express desirable traits or exhibit tolerance to specific environmental conditions. Cluster I demonstrated superior mean values for GFY, DMY, SY, PH, TPL, NOL, LW, DFF and DM, which suggested that the genotypes within this cluster possessed genetic factors that contributed to high-yield performance. That the genotypes in this cluster had favorable combinations of alleles for these yield-related traits. Breeders can focus on genotypes within cluster I to develop cultivars with superior yield attributes. Furthermore, the general mean (for all 62 genotypes as a population) is also mentioned for comparison in Table 3. Cluster I exhibited not only the highest mean values for the aforementioned traits but also surpassed the overall mean. This offers breeders a valuable opportunity to avoid individual genotypic assessments and concentrate their endeavors on representative genotypes within each cluster. This enables them to administer the breeding material with enhanced efficiency and allocate resources in a more judicious manner.

Conclusion

The present study employed a more advanced approach, i.e., hybrid hierarchical k-means clustering to evaluate genetic diversity and performance for various yield contributing traits. The application of hybrid hierarchical k-means clustering combined the advantages of hierarchical and k-means methods. The analysis led to the categorization

of 62 oat genotypes into four distinct clusters, as visually depicted in the cluster plot. This plot provided a clear and concise representation of the clusters and their refined memberships. These results played a crucial role in identifying promising genotypes for future breeding programs. Furthermore, the assessment of inter-cluster distances results in emphasizing the potential of hybridization between genotypes of genetically distant clusters to introduce novel traits and enhance overall genetic diversity. Cluster II and III exhibited the greatest dissimilarity, while cluster I demonstrated the highest mean values for economic traits and outperformed the general mean. These findings provide breeders with valuable insights for crop improvement and cultivar development. The identification of distinct clusters, assessment of inter-cluster and intra-cluster distances and evaluation of mean values for traits provided valuable information for breeders to select superior genotypes, enhance genetic diversity and ultimately foster the development of improved oat cultivars.

Authors Contribution

Rukoo Chawla: Investigation, data collection, writing an original draft and statistical analysis. Meenakshi Jattan: Designing of experiments, supervision, guidance. D.S. Phogat: Contribution of experimental material, supervision and editing. Amit Sharma: Reviewing and improving. Mandeep Redhu: reviewing and editing.

Acknowledgment

For carrying out the research, the authors are grateful to the Director of Research, CCS Haryana Agricultural University, Hisar (Haryana).

References

- Ahmad M, ZA Dar and Habib M (2014) A review on oat (*Avena sativa* L.) as a dual-purpose crop. *Sci. Res. Essays* 9: 52-59.
- Arora A, VK Sood, HK Chaudhary, DK Banyal, S Kumar, R Devi, A Kumari and S Yograj (2021) Genetic diversity analysis of oat (*Avena sativa* L.) germplasm revealed by agro-morphological and SSR markers. *Range Manag. Agrofor.* 42: 38-48.
- Bhandari HR, AN Bhanu, K Srivastava, MN Singh and HA Shreya (2017) Assessment of genetic diversity in crop plants - an overview. *Adv. Plants Agric. Res.* 7: 279-286.
- Boczkowska M, J Nowosielski, D Nowosielska and W Podyma (2014) Assessing genetic diversity in 23 early Polish oat cultivars based on molecular and morphological studies. *Genet. Resour. Crop Evol.* 61: 927-941.
- Carroll SB, J Doebley, AJ Griffiths and SR Wessler (2015) Introduction to genetic analysis. *WH Freeman*.
- Chakravorty A, PD Ghosh and PK Sahu (2012) Multivariate analysis of phenotypic diversity of landraces of rice of West Bengal. *Am J. Exp. Agric.* 3: 110-123.
- Chawla R, M Jattan, DS Phogat, N Kumari, S Kumar and A Poonia (2021) Genetic correlation and path analysis for yield and quality attributes in Oat (*Avena sativa* L.). *Biol. Forum* 13: 940-945.
- Chawla R, A Poonia and S Kumar (2022) Recent advances in yield and quality of dual purpose oat. *Forage Res.* 47: 383-389.
- Gorash A, R Armoniene, J Mitchell Fetch, Z Liatukas and V Danyte (2017) Aspects in oat breeding: nutrition quality, nakedness and disease resistance, challenges and perspectives. *Ann. Biol.* 171: 281-302.
- Govindaraj M, M Vetriventhan and M Srinivasan (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res. Int.* 2015: 431487.
- Isidro-Sanchez J, E Prats, C Howarth, T Langdon and G Montilla-Bascon (2020) Genomic Approaches for Climate Resilience Breeding in Oats. In: C Kole (ed.) *Genomic Designing of Climate-Smart Cereal Crops*, Springer, Cham.
- Kaur R, R Kapoor, Y Vikal and K Kaur (2018) Assessing genetic diversity in dual purpose oat (*Avena sativa* L.) cultivars based on morphological and quality traits. *Int. J. Curr. Microbiol. Appl. Sci.* 7: 1574-1586.
- Kiran, Suvarna, BV Tembhurne, G Girish, Shivaleela and M Lakshmikanth (2023) Genetic Diversity Studies for Yield and Yield attributing Characters in Colored Sorghum Genotypes. *Indian J. Plant Genetic Resources.* 36(2): 208-215.
- Krishna A, S Ahmed, HC Pandey and V Kumar (2014) Correlation, path and diversity analysis of oat (*Avena sativa* L.) genotypes for grain and fodder yield. *J. Plant Sci.* 1: 1-9.
- Kumari T and Y Jindal (2019) Genetic diversity and variability analysis in oats (*Avena* sp) genotypes. *Electron. J. Plant Breed.* 10: 1-8.
- Mohammadi SA and BM Prasanna (2003) Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci.* 43: 1235-1248.
- Nikoloudakis N, K Bladenopoulos and A Katsiotis (2016) Structural patterns and genetic diversity among oat (*Avena sativa*) landraces assessed by microsatellite markers and morphological analysis. *Genet. Resour. Crop Evol.* 63: 801-811.
- Paw M, S Munda, A Borah, SK Pandey and M Lal (2020) Estimation of variability, genetic divergence, correlation studies of *Curcuma caesia* Roxb. *J. Appl. Res. Med. Aromat. Plants* 17: 100251.
- Poonia A and DS Phogat (2017) Genetic divergence in fodder oat (*Avena sativa* L.) for yield and quality traits. *Forage Res.* 43: 101-105.
- Poonia A, DS Phogat and A Bhuker (2020) Comparative diversity analysis of oat genotypes under multi-cut system. *Range Manag. Agrofor.* 41: 242-249.
- Prates LL and P Yu (2017) Recent research on inherent molecular structure, physiochemical properties, and bio-functions of food and feed-type *Avena sativa* oats and processing-induced changes revealed with molecular microspectroscopic techniques. *Appl. Spectrosc. Rev.* 52: 850-867.
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ross SM, JR King, JT O'Donovan and D Spaner (2004) Forage potential on intercropping berseem clover with barley, oat or triticale. *J. Agron.* 96: 1013-1021.
- Singh D, TS Dhillon, T Javed, R Singh, J Dobarra, SK Dhankhar and U Kumar (2022b) Exploring the genetic diversity of carrot genotypes through phenotypically and genetically detailed germplasm collection. *Agronomy* 12: 1921.
- Singh DN, JS Bohra, V Tyagi, T Singh, TR Banjara and G Gupta

- (2022a) A review of India's fodder production status and opportunities. *Grass Forage Sci.* 77: 1-10.
- Surje DT and DK De (2014) Correlation coefficient study in oat (*Avena sativa* L.) genotypes for fodder and grain yield characters. *J. Agric. Sci. Technol.* 1: 89-93.
- USDA, Foreign Agriculture Service. (2023) World agriculture production, Circular Series.
- Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244.
- Yang Z, C Xie, Y Bao, F Liu, H Wang and Y Wang (2023) Oat: Current state and challenges in plant-based food applications. *Trends Food Sci. Technol.* 134: 56-71.