

Genome Sequence and Association Analysis Reveal Allelic Variants for Agronomically Important Traits in Foxtail Millet (*Setaria italica* L.) Germplasm

Sameena Shaik

Yogi Vemana University

Anand Kumar

Yogi Vemana University

Bhushan B. Dholakia

Tripura University

Konda Sravansimha Reddy

Yogi Vemana University

S. Ananda Rajakumar

Yogi Vemana University

Ramesh Palakurthi

Yogi Vemana University

Lachagari V. B. Reddy

ATGC Agri Biotech Innovation Square TSIC Kolthur Biotech Park Shamirpet Mandal, Hyderabad

Lekkala S. Prasad

ATGC Agri Biotech Innovation Square TSIC Kolthur Biotech Park Shamirpet Mandal, Hyderabad

P. Chandra Obul Reddy

Yogi Vemana University

Arjula R. Reddy

ATGC Agri Biotech Innovation Square TSIC Kolthur Biotech Park Shamirpet Mandal, Hyderabad

Venkata Chandra Mohan Reddy Chagam

Planning & Monitoring Cell, ANGRAU

A. Chandra Sekhar

chandrasedkhar9@yahoo.com

Yogi Vemana University

Research Article

Keywords: ddRAD-Seq, Foxtail millet, Genome-wide analysis, Landrace, Molecular markers, SNPs

Posted Date: September 4th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7465120/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Foxtail millet (*Setaria italica* L.), a small-grained, diploid C4 panicoid millet with a genome size of 515 Mb, is the second most widely cultivated crop globally. It serves as a model species for agronomic and nutritional research. This study employed double-digest restriction-site-associated DNA sequencing (ddRAD-seq) to characterize genomic variation within 30 foxtail millet genetic resources. The analysis identified 206,483 high-quality polymorphic SNPs through NGS-ddRAD path. Number of SNPs on individual genotypes ranged from 1184(Si-7) to 131873 (Si-3) at Read Depth 10. Results revealed three distinct clusters, effectively separating most landraces and released cultivars, thereby indicating population differentiation based on their classification and geographic origin. Among the landraces, genotypes Si-17, Si-15, Si-5, and Si-16 exhibited high yields, early flowering, and early maturation compared to release cultivars. Further examination of SNPs in landraces uncovered variations in minor allele frequency (MAF), highlighting high-frequency alleles within the *Setaria* genotype. A total 83 significant MTAs (Marker-trait associations) were identified by GWAS for Eight traits across the genome. High confidence MTAs for three important traits including total tiller number per panicle (TNPP), Plant base color (PBC), and Grain weight per panicle (GWPa) were identified. These MTAs facilitated the identification of 57 candidate genes with potential applications in molecular breeding. Additionally, a mini-core selection of 30 genotypes, representing the majority of genetic diversity indicated notable genetic traces of landraces, particularly for those three traits. These landraces show promise for use in breeding programs aimed at developing climate-resilient foxtail millet varieties.

Introduction

Foxtail millet (*Setaria italica* L.) is a nutritionally rich crop known for its resilience to harsh environmental conditions, including drought and elevated temperatures. Originating in China (Charulata, 2013), it is a staple in Asia and Africa; Because of its high nutritional content and adaptability to climate change, foxtail millet is a valuable feed and fodder source that is suited for livestock and supports sustainable farming methods (Vetriventhan et al., 2012 & Anuradha et al., 2020). This crop possesses several advantageous traits, including drought-avoidance mechanisms, low harvesting costs, nutritional value, and high tolerance to biotic stresses.

Foxtail millet's short life cycle, compact genome size (515 Mb, second only to rice), inbreeding nature, and efficient seed setting make it a valuable model species for agronomic and nutritional research (Ramesh et al., 2023). In India, it is predominantly cultivated in semi-arid regions such as Andhra Pradesh, Tamil Nadu, Karnataka, Rajasthan, Uttar Pradesh, and the North-Eastern states (Kumari et al., 2013).

Landraces are domesticated plant species that have undergone extensive local adaptation, serving as crucial genetic resources for sustainable agriculture and frequently utilized by local farmers. Compared to release cultivars, landraces generally exhibit lower yields (Pradhan et al., 2020). However, they are characterized by greater genetic heterogeneity and exhibit remarkable resilience to both abiotic and biotic stresses (Ramesh et al., 2023). Enhancing landraces for improved yields and adaptability to diverse environmental stresses holds significant potential for increasing crop productivity and improving farmers' livelihoods. Yield-focused breeding programs have been a key driver of genetic diversity loss in foxtail millet, as seen in many other crop species. DNA-based studies on millet species have demonstrated high levels of genetic diversity in foxtail millet and green millet (Le Thierry d'Ennequin et al., 2000). To mitigate genetic erosion, a strategic approach involving the collection, evaluation, and preservation of these invaluable landraces is essential for their future utilization (Ramesh et al., 2023; Palakurthi et al., 2023).

The complete genome sequence and associated bioinformatics resources of foxtail millet have been transformative for millet research, facilitating genome mapping and genetic characterization to develop elite cultivars. Advances in genomics and genetics have been significantly bolstered by high-throughput genotyping and sequencing technologies (Zhang et al., 2018). The rapid evolution of next-generation sequencing (NGS) technologies has enabled the generation of extensive single nucleotide polymorphism (SNP) datasets in both model and non-model plant species (Cloutier et al., 2019).

Early efforts to estimate genetic diversity in foxtail millet landraces utilized molecular markers (Liu et al., 2019). Genotyping-by-sequencing (GBS), an NGS-based approach for SNP discovery in reduced representations of genomes, has proven effective (Deschamps et al., 2012). Among GBS methods, double-digest restriction-site-associated DNA sequencing (ddRAD-seq) is a refined technique employing two restriction enzymes—one rare-cutting and one frequent-cutting—to generate stable, reproducible-sized genomic fragments (Peterson et al., 2012). This approach enables the generation of high-density marker datasets, crucial for advanced genetic analyses.

Concurrent advancements in statistical methods have enhanced the accuracy of genomic studies by reducing false positives arising from population structure and multiple testing corrections (Gupta et al., 2014). Genome-wide association studies (GWAS) have become a pivotal tool for dissecting the genetic basis of quantitative traits. Applications of ddRAD-seq technology in GWAS have been successfully demonstrated across various crops, including tomato (Bodanapu et al., 2019), pea (Gali et al., 2019), sesame (Ruperao et al., 2023), wheat (Arora et al., 2017), sorghum (Morris et al., 2013), maize (Xiao et al., 2017), and rice (Zhao et al., 2018). In this study, we analyzed foxtail millet germplasms, encompassing both landraces and released cultivars of *Setaria*, using an integrated approach combining ddRAD-seq with morpho-physiological, yield, and yield-related trait characterization. Germplasm separation was achieved through the single-seed descent method, enabling precise analysis of allelic variants (Ramesh et al., 2023). The resulting large-scale SNP markers and the identification of genotypes carrying novel, superior alleles provide valuable resources for foxtail millet improvement programs, particularly for enhancing climate resilience in response to future agricultural challenges.

Materials and methods

Plant material and Phenotyping

Thirty diverse genotypes of *Setaria* were collected from various locations within the Rayalaseema region of Andhra Pradesh, India. The collection comprised 20 landraces, 7 released cultivars and 3 germplasm lines. The methodology for the pure line development was previously described (Ramesh et al., 2023). These genotypes were evaluated for eight agronomic traits: plant base color, leaf base color, bristle (awn) color, grain weight per panicle, tiller number per plant, chlorophyll content, plant height, and stomatal density. The phenotyping was conducted at Yogi Vemana University, Kadapa, Andhra Pradesh, India using a randomized complete block design with three replications during the Kharif season of 2020. The average data from three replicates for all traits were used in the analysis. Descriptive statistical measures including minimum, maximum, mean, and standard error were calculated using Microsoft Excel.

DNA extraction, ddRAD sequencing and SNP calling

Genomic DNA was extracted from the leaves of five-week-old plants using a modified CTAB method (Murray and Thompson, 1980). DNA quantity was assessed based on absorbance values at 260 nm and 280 nm using a

Nanodrop spectrophotometer (Eppendorf BioSpectrometer Fluorescence, Germany). DNA quality and intact double-stranded DNA integrity were evaluated using a Qubit fluorometer (Thermo Fisher Scientific, USA) and 2% agarose gel electrophoresis. Library quality control (QC) was performed using a Bioanalyzer (Agilent Technologies, USA).

The quantified DNA was subsequently used for ddRAD-sequencing on an Illumina HiSeq2000 platform (Agri Genome Labs Pvt. Ltd., Hyderabad, India). Following sequencing, reads were processed for quality assurance. Base and adapter trimming were performed and sequences containing restriction-site-associated DNA (RAD) tags were filtered. A quality distribution plot was generated using custom scripts developed at Agri Genome Labs Pvt. Ltd. Sample reads were demultiplexed with allowance for one mismatch and low-quality bases, as well as regions with base bias at the start or end were trimmed. Illumina 5' and 3' adapter sequences were also removed. The *Setaria* reference genome was obtained from Phytozome v7.0

(ftp://ftp.jgipsf.org/pub/compugen/phytozome/v7.0/Sitalica/annotation/Sitalica_164_gene.gff3.gz). Paired-end reads were aligned to this reference genome using the Bowtie2 program (v2.1.0) (Langmead et al., 2012) with default parameters. Variant calling was performed based on these alignments to identify genomic variations.

Zygoty, diversity, phylogenetic, kinship and Principal component analyses

The diversity, zygoty, kinship and phylogenetic analyses of the 30 genotypes were conducted using genotype data filtered for a read depth (RD) > 10 and minor allele frequency (MAF) < 0.05. A radial phylogenetic tree was constructed to illustrate genetic relationships based on a similarity matrix generated using the neighbor-joining (NJ) algorithm in MEGA X (Kumari et al., 2023).

Kinship analysis among individuals was calculated using VanRaden's method based on their genotypic data (VanRaden, 2008). This analysis employed a centered identical-by-state (IBS) matrix and was further explored using the Trait Analysis by Association and Evolution (TASSEL v5.2.28) software (Bradbury et al., 2007). Principal component analysis (PCA) was performed on the filtered genotype data (MAF < 0.05) using PLINK v1.90 (Purcell et al., 2007) and analyzed with TASSEL to investigate genetic diversity. The PCA results highlighted the genetic differentiation among landraces, released cultivars, and germplasm lines providing insights into their evolutionary and genetic relationships.

Marker-trait association

For GWAS, SNPs with less than 30% missing data and a MAF greater than 5% were included. The Fixed and Random Model Circulating Probability Unification (FarmCPU) method (Liu et al., 2016) was employed for the association test. This recently developed technique is computationally efficient and effectively addresses challenges related to multiple testing corrections and kinship effects. Farm CPU integrates both fixed effect models (FEM) and random effect models (REM) in its framework. The FEM analyzes markers using pseudo-quantitative trait nucleotides (QTNs), which are calculated by the REM and subsequently used as covariates. The association test model included the first three principal components (PCA) as covariates to account for population structure. SNPs with a p-value < 0.005 were considered significant marker-trait associations (MTAs), and a Bonferroni-corrected p-value threshold of 0.05 was applied (Liu et al., 2016).

Quantile-quantile (Q-Q) plots were analyzed to evaluate model fitting and assess population structure correction. The Q-Q plots illustrated the distribution of observed versus expected p-values from the association tests. Proper

model fitting was indicated by a close alignment of the observed p-values with the diagonal line, suggesting minimal bias. Sharp deviations at the curve ends signified a small number of true associations among the numerous SNPs tested. The degree of deviation from the diagonal at the curve's tail served as a measure of the power of the test statistics (Turner, 2014).

Candidate gene identification

To identify potential candidate genes located near high-confidence SNPs, the associated SNPs were mapped to the *Setaria italica* reference genome (v2.2) (available at Phytozome). Transcripts within a 25 kb region flanking each associated SNP were extracted along with their functional annotations (Vandana et al., 2019).

Results

The phenotypic distribution of traits revealed a wide range of variability

Descriptive statistics, including the minimum, maximum, mean, and standard deviation, revealed substantial phenotypic variation across the eight traits evaluated in 30 genotypes (Suppl. Table S1). For instance, grain weight per panicle ranged from 4.7 g to 15.1 g, with a mean of 9.28 g and a standard deviation of 2.8 g. Similar variability was observed in other traits, such as plant height (112.55 ± 15.2 cm), tiller number per plant (6.9 ± 4.3), chlorophyll content (45.34 ± 4.08), number of stomata (72.6 ± 10.6), plant base color (1.3 ± 0.47), leaf base color (1.2 ± 0.44) and awn color (1.3 ± 0.4).

Detection of polymorphic variants indicated the presence of SNPs on all chromosomes

Detailed information about the 30 *Setaria* genotypes including variety names and collection origins is provided in Supplementary Table S2. Genomic DNA was isolated from each genotype and digested using the restriction enzymes *MluCI* and *SphI*, which have distinct recognition sites and cutting frequencies. Quality assessment of the fragmented DNA confirmed that all samples met the required screening criteria. Further processing, including quality checks, screening, and filtering of raw sequencing data, produced various read statistics (Fig. 1, Table 1). The average GC content across samples was 46%. Among the genotypes, the Si-27 landrace exhibited the highest proportion of reads containing RAD tags (47%), while Si-21 showed the lowest (43%). The percentage of uniquely aligned reads to the *Setaria italica* reference genome for all germplasms is summarized in Table 1.

Table 1
Sequence read measurements of 30 genotypes with annotated SNPs and INDELs at read depth (RD) > 10

Sample	Total reads	Reads after processing	Total reads aligned	Reads aligned (%)	Total uniquely aligned reads	Uniquely aligned reads (%)
Si-1	16,89,274	16,40,278	1,467,092	89.44%	1,409,655	96.09
Si-2	44,16,956	44,16,956	2,318,216	52.48%	1,804,356	77.83
Si-3	65,55,018	65,55,018	3,664,652	55.91%	2,897,245	79.06
Si-4	41,59,494	41,59,494	1,845,754	44.37%	1,434,943	77.74
Si-5	46,78,988	46,78,988	3,839,849	82.07%	2,906,577	75.7
Si-6	12,21,540	11,77,974	1,077,474	91.47%	1,044,431	96.93
Si-7	39,08,672	39,08,672	9,28,634	23.76%	3,91,733	42.18
Si-8	36,90,080	36,90,080	2,640,236	71.55%	2,061,422	78.08
Si-9	35,06,900	35,06,900	2,977,316	84.90%	2,377,308	79.85
Si-10	10,47,420	10,20,720	9,23,580	90.48%	8,87,660	96.11
Si-11	16,42,510	16,42,510	1,403,285	85.44%	1,091,035	77.75
Si-12	22,17,334	22,17,334	1,832,703	82.65%	1,441,068	78.63
Si-13	20,06,868	20,06,868	1,767,216	88.06%	1,416,355	80.15
Si-14	57,09,834	57,09,834	4,891,724	85.67%	3,740,042	76.46
Si-15	19,72,192	19,15,244	1,723,369	89.98%	1,663,770	96.54
Si-16	15,30,772	14,89,590	1,333,941	89.55%	1,286,660	96.46
Si-17	19,04,782	18,62,970	1,651, 673	88.66%	1,580,563	95.69
Si-18	10,39,734	10,07,330	8,98,181	89.16%	8,61,901	95.96
Si-19	20,61,392	20,35,660	1,884,005	92.55%	1,826,673	96.96
Si-20	17,68,238	17,46,624	1,659, 997	95.04%	1,620,263	97.61
Si-21	7,88,114	7,88,114	6,62,773	84.10%	4,97,729	75.1
Si-22	22,84,416	21,95,924	1,993,092	90.76%	1,921,886	96.43
Si-23	17,74,064	17,53,854	1,506,919	85.92%	1,458,523	96.79
Si-24	37,76,914	37,76,914	3,084,336	81.66%	2,326,285	75.42
Si-25	15,63,192	15,43,232	1,433,299	92.88%	1,389,854	96.97
Si-26	24,94,850	24,53,852	2,290,061	93.33%	2,231,018	97.42
Si-27	54,03,826	54,03,826	4,623,005	85.55%	3,570,870	77.24
Si-28	14,60,844	14,39,730	1,352,689	93.95%	1,320,030	97.59
Si-29	21,37,930	21,16,944	1,977,249	93.40%	1,929,895	97.61

Sample	Total reads	Reads after processing	Total reads aligned	Reads aligned (%)	Total uniquely aligned reads	Uniquely aligned reads (%)
Si-30	30,27,958	30,27,958	2,566,819	84.77%	2,014,461	78.48

In this study, landraces displayed an average number of reads aligning to the *S. italica* reference genome, with coverage spanning all nine chromosomes of *Setaria*. A total of 206,483 SNPs were identified across all germplasms when compared to the reference genome. Of these, the Si-3 landrace contributed the largest number of SNPs (131,873), while Si-7 showed the lowest contribution (1,184 SNPs) at a read depth (RD) of 10 (Fig. 2 & Table 1). Following the identification of polymorphic variants, the genetic relationships among the 30 genotypes were analyzed.

Structure and genetic diversity among genotypes

In this study, 203,422 genomic sites were predicted across 30 *Setaria* genotypes by comparison with the reference genome. Based on genotyping data with a RD > 10 and a MAF < 0.05, a total of 12,612,000 gametes were aligned with no data points excluded.

The percentage of heterozygosity was lowest in Si-17 and Si-25 (0.23%) followed by Si-16 (0.26%), Si-15 (0.28%), and both Si-10 and Si-20 (0.29%) when compared to the reference genome (Fig. 3a & Suppl. Table S3). This low heterozygosity suggested a higher degree of homology to the reference genome. The expected heterozygosity ranged from 0–0.70%, with an average of 0.25%, whereas the observed heterozygosity varied from 0–19.85%, with an average of 6.66%. (Fig. 3b).

These findings emphasized the importance of selecting appropriate parental lines with homozygous polymorphic markers, which ranged from 569 to 61,739 SNPs distributed across the entire *Setaria* genome (Table 2). This approach can significantly contribute to the development of breeding programs targeting genetic improvement in *Setaria*.

Table 2
Overall SNP Summary among the Genotypes at RD10

Sample	Total markers	No. of heterozygous loci	Heterozygous loci (%)	No. of homozygous loci	Homozygous loci (%)
<i>S. italica</i> 164	203422	0	0	203422	100
Si-1	203422	607	0.3	202815	99.7
Si-2	203422	26092	12.83	177330	87.17
Si-3	203422	38693	19.02	164729	80.98
Si-4	203422	26792	13.17	176630	86.83
Si-5	203422	25653	12.61	177769	87.39
Si-6	203422	1692	0.83	201730	99.17
Si-7	203422	693	0.34	202729	99.66
Si-8	203422	32329	15.89	171093	84.11
Si-9	203422	36274	17.83	167148	82.17
Si-10	203422	591	0.29	202831	99.71
Si-11	203422	22447	11.03	180975	88.97
Si-12	203422	19227	9.45	184195	90.55
Si-13	203422	27959	13.74	175463	86.26
Si-14	203422	31296	15.38	172126	84.62
Si-15	203422	574	0.28	202848	99.72
Si-16	203422	529	0.26	202893	99.74
Si-17	203422	476	0.23	202946	99.77
Si-18	203422	649	0.32	202773	99.68
Si-19	203422	744	0.37	202678	99.63
Si-20	203422	584	0.29	202838	99.71
Si-21	203422	14762	7.26	188660	92.74
Si-22	203422	664	0.33	202758	99.67
Si-23	203422	1620	0.8	201802	99.2
Si-24	203422	30142	14.82	173280	85.18
Si-25	203422	477	0.23	202945	99.77
Si-26	203422	705	0.35	202717	99.65
Si-27	203422	40375	19.85	163047	80.15

Sample	Total markers	No. of heterozygous loci	Heterozygous loci (%)	No. of homozygous loci	Homozygous loci (%)
Si-28	203422	5165	2.54	198257	97.46
Si-29	203422	1410	0.69	202012	99.31
Si-30	203422	31309	15.39	172113	84.61

Principal component analysis and phylogenetic relationship showed significant variations among *Setaria* genotypes

We analyzed the genetic variation among *Setaria* landraces, released cultivars, and germplasm lines using PCA based on 206,483 SNPs. The top five principal components (PCs) were extracted using default parameters (Suppl. Table S4), and the first three PCs (PC1, PC2 and PC3) were visualized in R (v. 3.3.1; R Core Team, 2016) to assess genetic diversity among the 30 genotypes (Fig. 4). The PCA results demonstrated a well-dispersed genetic structure reflecting clear differentiation among the genotypes which corresponded to their regional origins. The genotypes were primarily grouped into landraces, released cultivars and germplasm lines.

Phylogenetic analysis based on SNP clustering further supported these findings by dividing the 30 genotypes and the reference *Setaria* genome into three distinct clades (Fig. 5). Clade 1 included the reference genome along with 14 genotypes: Si-8, Si-21, Si-13, Si-2, Si-5, Si-3, Si-13, Si-4, Si-24, Si-11, Si-9, Si-27, Si-28 and Si-12. Clade 2 comprised 8 genotypes: Si-18, Si-22, Si-1, Si-6, Si-10, Si-15, Si-16 and Si-17. Clade 3 consisted of another 8 genotypes: Si-25, Si-23, Si-26, Si-29, Si-28, Si-20, Si-19 and Si-7. These groupings highlighted significant divergence among the selected genotypes.

The largest clade displayed two prominent subclades, potentially indicating the presence of closely related orthologs. Overall, the PCA results were consistent with the clustering patterns observed in the radial dendrogram, further corroborating the genetic variation and diversity among the analyzed *Setaria* genotypes.

GWAS identified high-confidence marker-trait association for selected agronomic traits

A total of 82 significant MTAs ($-\log_{10} p > 4.0$) were identified for eight traits using the Farm CPU model with a p-value threshold of < 0.05 (Table 3). These MTAs were distributed across all nine chromosomes of the foxtail millet genome (Fig. 6). Among these, the trait TNPP had the highest number of MTAs, with 36 associations spanning all chromosomes. The most significant MTAs for TNPP were predominantly located on chromosomes 3 and 9. For GWpa, 12 MTAs were detected across six chromosomes (3, and 5–9), with the most significant association on chromosome 3.

Table 3
Details of significant MTAs along with their chromosome, position, SNP effect and p -value cut-off

Trait	SNP	Chromosome	$-\log(p)$	SNP effect	MAF
TNPP	U1.P.7810928	1	4.51065441	-4.14611	0.137931
	U1.P.1686015	1	4.871089734	11.34653	0.327586
	U2.P.29140605	2	4.190382344	4.655844	0.293103
	U2.P.20227992	2	4.195025703	9.376994	0.344828
	U2.P.32723821	2	4.77850309	4.915417	0.155172
	U2.P.32775439	2	4.909348074	4.549813	0.155172
	U2.P.32765288	2	5.213821995	-4.21476	0.189655
	U2.P.26263756	2	5.653138206	6.287075	0.068966
	U2.P.32673002	2	5.773905487	5.186474	0.137931
	U3.P.36902456	3	4.063555849	-8.59198	0.344828
	U3.P.37623561	3	4.338051378	-5.58157	0.086207
	U3.P.88759	3	4.571251608	-4.31637	0.172414
	U3.P.17048505	3	4.871089734	11.34653	0.327586
	U4.P.39627594	4	4.044480769	3.062524	0.293103
	U4.P.8734529	4	4.403062891	-4.99981	0.448276
	U4.P.1287751	4	5.09674574	7.627616	0.051724
	U5.P.9351113	5	4.131248453	4.710379	0.103448
	U5.P.8755075	5	4.243048711	-4.00264	0.206897
	U5.P.8755078	5	4.243048711	-4.00264	0.206897
	U5.P.8755090	5	4.243048711	4.002636	0.206897
	U5.P.36939765	5	4.376501903	-7.11579	0.482759
	U6.P.21553011	6	4.08053225	3.418203	0.258621
	U7.P.21435470	7	3.952038362	2.975882	0.362069
	U7.P.22190549	7	4.528113937	8.965163	0.344828
	U7.P.30066461	7	4.958372051	-3.8794	0.206897
	U8.P.32134455	8	4.51065441	-4.14611	0.137931
	U8.P.28192415	8	5.757902993	-6.90113	0.344828
	U8.P.28192393	8	5.757902993	6.901126	0.344828
	U9.P.45558178	9	4.638588133	-5.03268	0.310345

Trait	SNP	Chromosome	-log(p)	SNP effect	MAF
	U9.P.45558191	9	4.638588133	-5.03268	0.310345
	U9.P.32110399	9	4.668270053	-5.10951	0.12069
	U9.P.21138360	9	4.871089734	11.34653	0.327586
	U9.P.48367936	9	4.871089734	11.34653	0.327586
	U9.P.424584	9	4.877187635	3.782103	0.293103
	U9.P.32110401	9	4.892237021	8.643724	0.362069
	U9.P.55319287	9	5.299583744	-9.70576	0.344828
GWPa	U3.P.26134706	3	4.342843907	-4.58263	0.396552
	U3.P.2130440	3	4.438322103	-4.34619	0.068966
	U3.P.7003009	3	4.508147784	6.732641	0.344828
	U3.P.48594755	3	4.508147784	-6.73264	0.344828
	U5.P.6504655	5	4.096600693	3.439549	0.344828
	U6.P.24872538	6	4.512850913	3.577091	0.086207
	U6.P.20678085	6	5.204750316	3.765414	0.344828
	U7.P.22568466	7	4.508147784	-6.73264	0.344828
	U8.P.32393157	8	4.048654437	3.255125	0.103448
	U9.P.43245595	9	4.438322103	4.346191	0.068966
	U9.P.48339773	9	4.508147784	6.732641	0.344828
	U9.P.8894136	9	4.508147784	6.732641	0.344828
AC	U1.P.34713158	1	3.964368665	-0.63546	0.310345
	U4.P.5383889	4	4.042255266	0.932299	0.362069
	U5.P.34217851	5	4.051430958	0.741664	0.413793
	U5.P.23902170	5	4.466401385	0.937002	0.37931
	U7.P.111108796	7	3.909406031	-0.55644	0.137931
	U9.P.22952108	9	4.466401385	0.468501	0.241379
LBC	U1.P.2465258	1	3.930104256	-0.97443	0.362069
	U1.P.2465260	1	3.930104256	-0.97443	0.362069
	U5.P.19096852	5	4.082105944	0.546267	0.37931
	U5.P.5619936	5	4.103632084	0.527403	0.310345
	U6.P.2387149	6	3.970549286	-0.42045	0.206897
NS	U1.P.10586306	1	4.083269876	17.32689	0.396552

Trait	SNP	Chromosome	-log(p)	SNP effect	MAF
	U5.P.8171593	5	4.013930874	-14.3424	0.086207
	U7.P.13558003	7	4.051098726	11.90777	0.155172
	U7.P.13558005	7	4.051098726	11.90777	0.155172
	U7.P.13558007	7	4.051098726	11.90777	0.155172
	U7.P.13558008	7	4.051098726	11.90777	0.155172
	U7.P.13558009	7	4.051098726	11.90777	0.155172
PBC	U1.P.10254412	1	4.740718101	0.815994	0.413793
	U2.P.48541889	2	3.955817718	0.532029	0.206897
	U5.P.19096852	5	3.984821803	0.54043	0.37931
	U5.P.5619936	5	4.06441339	0.525075	0.310345
	U6.P.2387149	6	5.00700628	-0.46155	0.206897
	U7.P.10981753	7	4.040905414	0.931684	0.362069
PH	U1.P.10886040	1	4.625174362	14.42422	0.275862
	U5.P.18384997	5	4.199329944	-20.9148	0.448276
	U6.P.28254812	6	4.424033326	10.92063	0.362069
	U7.P.31998098	7	3.930925109	14.94568	0.155172
	U7.P.31998099	7	3.930925109	14.94568	0.155172
	U7.P.31998096	7	3.930925109	-14.9457	0.155172
	U7.P.31998100	7	3.930925109	-14.9457	0.155172
	U7.P.31998101	7	3.930925109	-14.9457	0.155172
SPAD	U1.P.20351799	1	3.63583685	-3.42676	0.241379
	U5.P.28095651	5	4.368148363	7.807428	0.482759

Other traits showed the following distribution of MTAs: eight for PH, seven for NS, six each for PBC and AC, five for LBC, and two for SPAD. All 82 SNPs exhibited single-trait associations. Detailed information on these MTAs, including their chromosomal positions, p-values, minor allele frequencies and SNP effects, is provided in Suppl. Table S4.

To minimize false positives caused by multiple testing, MTAs were filtered using the Bonferroni correction. Of the 82 MTAs, 16 met the stringent Bonferroni threshold and were designated as high-confidence MTAs. These high-confidence MTAs were associated with three traits: TNPP (14 MTAs), GWPa (1 MTA), and PBC (1 MTA) (Table 4).

Table 4
List of 16 high-confidence MTAs that fulfilled the Bonferroni correction

Trait	SNP	Chromosome	Position	P-value
TNPP	U1.P.1686015	1	1686015	4.87109
	U2.P.32775439	2	32775439	4.909348
	U2.P.32765288	2	32765288	5.213822
	U2.P.26263756	2	26263756	5.653138
	U2.P.32673002	2	32673002	5.773905
	U3.P.17048505	3	17048505	4.87109
	U4.P.1287751	4	1287751	5.096746
	U8.P.28192415	8	28192415	5.757903
	U8.P.28192393	8	28192393	5.757903
	U9.P.21138360	9	21138360	4.87109
	U9.P.48367936	9	48367936	4.87109
	U9.P.424584	9	424584	4.877188
	U9.P.32110401	9	32110401	4.892237
	U9.P.55319287	9	55319287	5.299584
GWP _a	U6.P.20678085	6	20678085	5.20475
PBC	U6.P.2387149	6	2387149	5.007006

The 14 MTAs linked to TNPP were located on chromosomes 1, 2, 3, 4, 8, and 9, with the most significant associations being:

- C1.P.1686015 (p-value: 1.35×10^{-5})
- C2.P.32775439 (p-value: 1.23×10^{-5})
- C2.P.32765288 (p-value: 6.11×10^{-6})
- C2.P.26263756 (p-value: 2.22×10^{-6})
- C2.P.32673002 (p-value: 1.68×10^{-6})
- C3.P.17048505 (p-value: 1.35×10^{-5})
- C4.P.1287751 (p-value: 8.00×10^{-6})
- C8.P.28192415 (p-value: 1.75×10^{-6})
- C8.P.28192393 (p-value: 1.75×10^{-6})
- C9.P.21138360 (p-value: 1.35×10^{-5})
- C9.P.48367936 (p-value: 1.35×10^{-5})
- C9.P.424584 (p-value: 1.33×10^{-5})
- C9.P.32110401 (p-value: 1.28×10^{-5})

- C9.P.55319287 (p-value: 5.02×10^{-6}).

The single high-confidence MTA for GWP_a was located on chromosome 6 (C6.P.20678085, p-value: 6.24×10^{-6}), while the MTA for PBC was also on chromosome 6 (C6.P.2387149, p-value: 9.84×10^{-6}).

Identification of candidate genes suggested the potential involvement of different pathways

A total of 57 candidate genes were identified within the 25 Kb flanking regions (upstream and downstream) of the 16 high-confidence MTAs associated with three traits: TNPP, GWP_a, and PBC (Table 4). Notably, none of the associated SNPs were located within gene-coding regions.

For the 14 MTAs linked to TNPP, 52 candidate genes were identified across various chromosomes. These genes encoded diverse proteins, including EF-hand domain-containing proteins, 1,3-beta-glucan synthase, 1,3-beta-glucan synthase components, FKS1-like domain-containing proteins, Myb-like domain-containing proteins, and cysteine dioxygenase, among others.

In the genomic regions associated with the single MTA for PBC, four candidate genes were identified. These included *Obg-like ATPase-1*, *Pentatricopeptide-repeat region of PRORP domain-containing protein*, *Photosystem II 10kDa chloroplastic polypeptide*, and *K Homology domain-containing protein*. For GWP_a, a single candidate gene with a known function, *MATH domain-containing protein*, was identified within the 25 Kb upstream region of the associated SNP (MTA). The detailed list of candidate genes, along with their chromosomal positions and putative functions, is provided in Table 5.

Table 5

Details of the identified candidate genes residing in close vicinity (25kb either side) of the associated SNPs

Trait	Associated SNP	Transcript	Start	End	Strand	Description
TNPP	U1.P.1686015	SETIT_016772mg	1686015	1662015	-	EF-hand domain-containing protein
		SETIT_016150mg	1686015	1710015	+	1,3-beta-glucan synthase
		SETIT_016487mg	1686015	1710015	+	1,3-beta-glucan synthase component FKS1-like domain-containing protein
	U1.P.7810928	SETIT_019086mg	7810928	7786928	-	Calmodulin-binding domain-containing protein
		SETIT_019554mg	7810928	7786928	-	Homeobox domain-containing protein
	U2.P.26263756	SETIT_029444mg	26239756	26263756	-	Myb-like domain-containing protein
		SETIT_030953mg	26239756	26263756	-	cysteine dioxygenase
	U2.P.32673002	SETIT_029077mg	32649002	32673002	-	Potassium transporter
		SETIT_032857mg	32649002	32673002	-	F-box domain-contain
		SETIT_030244mg	32649002	32673002	-	Protein XRI1
	U2.P.32723821	SETIT_032446mg	32699821	32723821	-	Zinc finger PHD-type
	U2.P.32765288	SETIT_029726mg	32741288	32765288	-	C2H2-type domain-co
	U2.P.32775439	SETIT_033245mg	32751439	32775439	-	Preprotein translocase subunit-SecE
	U3.P.88759	SETIT_022424mg	64759	88759	-	RING-CH-type domain-containing protein
		SETIT_025676mg	64759	88759	-	CAP-Gly domain-containing protein
		SETIT_022584mg	64759	88759	-	Sugar phosphate transporter domain-containing protein
		SETIT_021320mg	64759	88759	-	Methyltransferase
		SETIT_020982mg	64759	88759	-	1-phosphatidylinositol-

Trait	Associated SNP	Transcript	Start	End	Strand	Description
						3-phosphate 5-kinase
	U3.P.17048505	SETIT_021362mg	III:17048505–17072505		+	Protein kinase domain-containing protein
		SETIT_023473mg	III:17048505–17072505		+	Yippee domain-containing protein
		SETIT_022793mg	III:17048505–17072505		+	FCP1 homology domain-containing protein
		SETIT_023745mg	III:17048505–17072505		+	Histone H4
	U4.P.1287751	SETIT_005759mg	IV:39627594–39651594		+	Protein kinase domain-containing protein
	U7.P.22190549	SETIT_010421mg	22166549	22190549	-	SAM domain-containing protein
		SETIT_012723mg	22166549	22190549	-	DUF6598 domain-containing protein
		SETIT_009246mg	22166549	22190549	-	CRM domain-containing protein
	U7.P.30066461	SETIT_009982mg	30042461	30066461	-	F-box domain-containing protein
		SETIT_010903mg	30042461	30066461	-	Multiple organellar RNA editing factor 3, mitochondrial
		SETIT_011506mg	30042461	30066461	-	Wall-associated receptor kinase galacturonan-binding domain-containing protein
	U9.P.21138360	SETIT_038219mg	21114360	21138360	-	BHLH domain-containing protein
	U9.P.32110399	SETIT_037461mg	32086399	32110399	-	Proline-rich protein
	U9.P.424584	SETIT_034495mg	400584	424584	-	Metallo-beta-lactamase domain-containing protein
		SETIT_037834mg	400584	424584	-	H15 domain-containing protein
		SETIT_037173mg	400584	424584	-	Derlin
		SETIT_038671mg	400584	424584	-	Pentacotriptide-repeat region of PRORP domain-containing protein

Trait	Associated SNP	Transcript	Start	End	Strand	Description
		SETIT_039302mg	400584	424584	-	Transcription factor CBF/NF-Y/archaeal histone domain-containing protein
		SETIT_037626mg	400584	424584	-	Ycf20-like protein
		SETIT_035086mg	55295287	55319287	-	Pentacotriptide-repeat region of PRORP domain-containing protein
U9.P.45558178		SETIT_039321mg	45558178	45582178		Peroxidase
		SETIT_033939mg	45534178	45558178	-	NB-ARC domain-containing protein
U9.P.48367936		SETIT_036189mg	48343936	48367936	-	VAN3-binding protein-like auxin canalisation domain-containing protein
		SETIT_034343mg	48343936	48367936	-	DEAD-box ATP-dependent RNA helicase 24
		SETIT_036926mg	48343936	48367936	+	ATP-dependent Clp protease proteolytic subunit
		SETIT_034160mg	48367936	48391936	+	[histone H3]-lysine(27) N-trimethyltransferase
		SETIT_036751mg	48367936	48391936	+	UBC core domain-containing protein
		SETIT_034564mg	48367936	48391936	+	Trichome birefringence-like N-terminal domain-containing protein
		SETIT_035086mg	55295287	55319287	-	Pentacotriptide-repeat region of PRORP domain-containing protein
U9.P.55319287		SETIT_038756mg	55295287	55319287	-	Thioredoxin domain-containing protein
		SETIT_036985mg	55295287	55319287	-	DUF1997 domain-containing protein
		SETIT_036078mg	55319287	55343287	+	Serine/threonine-protein phosphatase
		SETIT_035602mg	55319287	55343287	+	Major facilitator superfamily (MFS) profile domain-containing protein

Trait	Associated SNP	Transcript	Start	End	Strand	Description
		SETIT_039016mg	55319287	55343287	+	F-box domain-containing protein
PBC	U6.P.2387149	SETIT_013409mg	2363149	2387149	-	Obg-like ATPase 1
		SETIT_013265mg	2387149	2411149	+	Pentacotriptide-repeat region of PRORP domain-containing protein
		SETIT_014601mg	2387149	2411149	+	Photosystem II 10 kDa polypeptide, chloroplastic
		SETIT_013409mg	2387149	2411149	+	K Homology domain-containing protein
GWPa	U6.P.20678085	SETIT_015259mg	20653315	20653962	-	MATH Domain containing protein

Discussion

Enhancing grain yield remains one of the most challenging objectives in crop improvement. In this study, the selection of 30 genetically diverse *Setaria* genotypes demonstrated their suitability for GWAS. To facilitate effective genetic improvement, conservation, and resource management, it is essential to harness crop diversity. The *Setaria* genotypes used in this study were sourced from various districts of the Rayalaseema region in Andhra Pradesh, including Nandyala, Kadapa, Chittoor, Kurnool and Anantapur (Ramesh et al., 2023). Descriptive statistics revealed significant phenotypic diversity across the panel for all eight agronomic traits (Suppl. Table S1).

Understanding the genetic structure of a population is crucial for genotype conservation and plant breeding efforts. A comprehensive understanding of the genetic structure of a species is critical for implementing breeding programs that prioritize the efficient preservation and utilization of plant genetic resources (Izzatullayeva et al., 2014; Tripathi et al., 2012). ddRAD-seq has emerged as a highly efficient and cost-effective method for SNP genotyping, offering a simplified protocol with high accuracy. Despite its potential, relatively few studies have utilized this advanced technology to investigate the genetic relationships and structure within the *Setaria* genus (Ramesh et al., 2023; Jaiswal, 2019). The higher-than-expected heterozygosity highlights the substantial genetic variation present in landraces compared to released cultivars.

In this study, ddRAD-seq was conducted using the restriction enzyme pair *Mlu*I and *Sph*I to identify genome-wide SNPs. This approach yielded high read depth suitable for reliable SNP calling across all genotypes. Approximately 86% of the genomic segments generated were uniquely aligned to the reference genome, a result comparable to findings in genome-wide SNP discovery studies in *Solanum lycopersicum* (Bodanapu et al., 2019). The experimental design and methodology employed in this research were adapted from a tomato study (Bodanapu et al., 2019), further underscoring the utility of this approach for exploring genetic diversity in crop species. Genetic diversity plays a critical role in determining a population's ability to adapt to changing environmental conditions. Insufficient genetic diversity in breeding populations poses a significant risk of severe declines under major

environmental stress events (Novaes et al., 2008). The reduced productivity, yield, and stress resilience observed in this study underscore the urgent need to manage and conserve *Setaria* resources and biodiversity effectively.

In this study, approximately 519,867 to 1,142,208 paired-end reads (100 bp each) were generated for the 30 *Setaria* genotypes, resulting in an average genome coverage of 51.99 to 114.22 MB per line, with a targeted depth of 10–20×. This extensive dataset revealed minimal contamination levels and unexpectedly low sequence redundancy. Despite certain complications, 206,483 SNPs were identified following the initial quality check. The identification of these SNPs significantly enhances the likelihood of discovering efficient molecular markers for *Setaria*. The moderate frequency of retained SNPs, particularly those located in transcribed and regulatory regions, highlights the priority of generating genome-wide SNPs for functional genomic studies in *Setaria* species (Jo et al., 2017).

Using robust tools and methodologies, this study effectively estimated heterozygosity levels across genotypes, ranging from 0 to 19.85%. While previous studies predicted heterozygosity levels between 0 and 0.70% with an average of 0.25% (Ramesh et al., 2023), the significantly higher heterozygosity observed in this study suggests that the *Setaria* population may have recently experienced a genetic bottleneck (Cornuet et al., 1996; Zhang et al., 2018). Similar findings were reported in onion inbred lines, where observed heterozygosity (H_o) exceeded expected heterozygosity (H_e) (Lee et al., 2018).

The low percentage of polymorphic loci identified across species indicates substantial genetic variation within the population, influenced by both genetic and environmental factors (Zhao et al., 2011). These findings reinforce the importance of conserving genetic diversity within *Setaria* species to support future breeding programs and environmental adaptation strategies.

Maximum likelihood phylogenetic analysis based on 206,483 SNPs identified three major clusters among the assembled *Setaria* genotypes. Cluster 1 contained the majority of the genotypes (14), while Clusters 2 and 3 each comprised eight genotypes. The results indicated that species within the same cluster exhibited the highest SNP locus similarity. These findings align with previous studies (Vetriventhan et al., 2014; Ali et al., 2016), which reported a strong correlation between geographic origin and phylogenetic relationships.

RAD-seq has proven to be an effective tool for phylogenetic reconstruction in species with sufficient orthologous restriction sites conserved across populations (Rubin et al., 2012). Similar phylogenetic analyses using RAD-seq have been conducted in onion (Lee, 2018), tomato (Bodanapu et al., 2019), grapevine (Lauco et al., 2018), and *Rhododendron meddianum* (Zhang et al., 2021), with conclusions that support the findings of the Neighbor-Joining (NJ) tree analysis in the present study.

In addition to phylogenetic analysis, cluster, and PCA analyses have proven to be highly effective for exploring genetic diversity, tracing species evolution, selecting parental lines, and identifying centers of origin (Eivazi et al., 2007). The PCA results in this study revealed three major groupings of *Setaria* genotypes, consistent with the clustering patterns observed in the phylogenetic analysis. This approach has been successfully applied in numerous population genetics studies (Gupta et al., 2012), demonstrating its utility in understanding genetic relationships and diversity in *Setaria*.

GWAS have long been a powerful tool for dissecting the genetic basis of complex traits and have been successfully applied to various crops, including cotton, rice, wheat, maize, and pearl millet (Jaiswal et al., 2016). However, there has been limited research applying GWAS to *Setaria* (foxtail millet), particularly for identifying the genetic regions controlling desirable traits (Gupta et al., 2014). The genetic diversity of the study panel was a

critical prerequisite for the success of GWAS. Additionally, kinship results posed challenges in interpreting the GWAS outcomes, and the power of GWAS is influenced by both population size and trait diversity (Flint-Garcia et al., 2005).

While the population size in this study was relatively small, it still yielded findings comparable to those from GWAS in other cereal crops (Anuradha et al., 2017). Despite the stringent Bonferroni correction, 16 significant MTAs were identified across three traits out of a total of 82 MTAs associated with eight traits. The use of thousands of markers increases the risk of false negatives, making multiple testing corrections essential. Therefore, rigorous parameters were applied to minimize false positives, and further validation of all 82 MTAs is necessary. The 16 MTAs that passed the multiple testing correction were deemed high-confidence associations. These high-confidence MTAs, identified for three key agronomic traits—TNPP, GWP_a and PBC could significantly contribute to foxtail millet breeding through marker-assisted selection, pending further validation.

Among the candidate genes identified near the associated SNP for GWP_a, the MATH domain-containing protein was of particular interest, as it plays a role in various functions, including flower development, fatty acid biosynthesis, and stress tolerance (Weber and Hellmann, 2009; Lechner et al., 2011; Chen et al., 2013; Ma et al., 2013; Chen et al., 2015). In a wheat RNAi study, MATH domain-containing proteins were found to regulate plant growth and development, especially during early embryogenesis (Bauer et al., 2019). These candidate genes, located near the associated SNPs, could be validated in future studies and deployed in breeding programs to enhance foxtail millet.

Overall, the insights gained from ddRAD sequencing in this study advance our understanding of *Setaria* species. The identified high-confidence SNPs also facilitated the discovery of 57 candidate genes associated with important traits. Identifying SNPs linked to yield-related characteristics, such as TNPP, GWP_a, and PBC, is crucial for genomics-assisted breeding in foxtail millet. Plant breeders, when supported by genotype-based relatedness analysis, can make informed decisions regarding the selection of parental lines, ultimately improving breeding programs and the long-term sustainability of millet cultivation (Lee et al., 2018). These findings offer significant insights into the genetic architecture underlying key agronomic traits in foxtail millet, elucidating the molecular basis of trait variation. They also highlight potential genetic targets for further functional validation, thereby laying the groundwork for enhancing foxtail millet improvement through precision breeding strategies.

Declarations

Credit authorship contribution statement

ACS conceived the idea; ACS, PCOR, SS, and RP collected landraces from the farmer's field; SS and AK conducted all the experiments; KSSR, VBR, and LSP performed genomic and data analysis; ACS, PCOR, BBD, SS, AK, ARK, RP, ARR and PT prepared as well as edited the entire MS; ACS and PCOR contributed consumables. All authors read and approved the final manuscript.

Ethical Statements:

This manuscript does not include any of the Human / Animal Models / Cell Lines or any other of this kind that are subjected to ethical issues. Hence this declaration is "NOT APPLICABLE"

Funding:

This work was supported by the Department of Science and Technology (DST)-INSPIRE fellowships, New Delhi, India as part of a Ph. D degree scholarship to SS (IF170752) and KSSR (IF220369). ACS and PCOR from YVU acknowledged the partial utilization of the financial support for consumables (No. CRG/2018/003280 dated May 30th, 2019) from DST-SERB, New Delhi, India. AK thanked for the JRF (No. CRG/2018/003280 dated May 30th, 2019) from DST- Science and Engineering Research Board (SERB), New Delhi.

Author Contribution

ACS conceived the idea; ACS, PCOR, SS, and RP collected landraces from the farmer's field; SS and AK conducted all the experiments; KSSR, VBR, and LSP performed genomic and data analysis; ACS, PCOR, BBD, SS, AK, ARK, RP, ARR and PT prepared as well as edited the entire MS; ACS and PCOR contributed consumables. All authors read and approved the final manuscript.

Acknowledgement

This work was supported by the Department of Science and Technology (DST)-INSPIRE fellowships, New Delhi, India as part of a Ph. D degree scholarship to SS (IF170752) and KSSR (IF220369). ACS and PCOR from YVU acknowledged the partial utilization of the financial support for consumables (No. CRG/2018/003280 dated May 30th, 2019) from DST-SERB, New Delhi, India. AK thanked for the JRF (No. CRG/2018/003280 dated May 30th, 2019) from DST- Science and Engineering Research Board (SERB), New Delhi. All the authors acknowledge Dr. Saurabh Gupta and Dr. Nisarg Vays of Department of Generative AI & Bioinformatics, Infocusp Innovations, Gala Hub, Bopal, Ahmedabad, Gujarat for their help in Bioinformatic analysis and association studies.

References

1. Ali, Asjad., et al. (2016). EST-SSR based genetic diversity and population structure among Korean landraces of foxtail millet (*Setaria italica* L.). *Korean J. Plant Res.*, 29(3),322–330. <https://doi.org/10.7732/kjpr.2016.29.3.322>
2. Anuradha, N., et al.(2017). Deciphering genomic regions for high grain iron and zinc content using association mapping in pearl millet. *Frontiers in Plant Sci.* 8, 412. <https://doi.org/10.3389/fpls.2017.00412>
3. Arora, S., et al. (2017). Genome-wide association study of grain architecture in wild wheat *Aegilops tauschii*. *Frontiers in Plant Sci.* 8, 886. <https://doi.org/10.3389/fpls.2017.00886>
4. Anuradha, N., and T. S. S. K. Patro. (2020). "Estimates of variability, heritability and genetic advance in foxtail millet." *J Pharmacogn Phytochem* 9(1),1614–1616. <https://dx.doi.org/10.22271/phyto>
5. Bauer, N., et al. (2019). The MATH-BTB protein TaMAB2 accumulates in ubiquitin-containing foci and interacts with the translation initiation machinery in Arabidopsis. *Frontiers in plant sci.* 10,1469. <https://doi.org/10.3389/fpls.2019.01469>
6. Bodanapu, R., et al.(2019). Deciphering the unique SNPs among leading Indian tomato cultivars using double digestion restriction associated DNA sequencing. *bioRxiv*,541227. <https://doi.org/10.1101/541227>
7. Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635.

<https://doi.org/10.1093/bioinformatics/btm308>

8. Chen, K., et al.(2015). Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proceedings of the National Academy of Sci.* 112(27), E3535-E3544.
<https://doi.org/10.1073/pnas.1504232112>
9. Chen, L., et al. (2013). Arabidopsis BPM proteins function as substrate adaptors to a cullin3-based E3 ligase to affect fatty acid metabolism in plants. *The Plant Cell*, 25(6), 2253–2264.
<https://doi.org/10.1105/tpc.112.107292>
10. Cloutier, S., et al. (2019). Linum genetic markers, maps, and QTL discovery. *Genetics and Genomics of Linum*, 97–117. https://doi.org/10.1007/978-3-030-23964-0_7
11. Cornuet, Jean Marie, and Gordon Luikart (1996). "Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data." *Genetics* 144.4: 2001–2014.
<https://doi.org/10.1093/genetics/144.4.2001>
12. Deschamps, S., et al. (2012). Genotyping-by-sequencing in plants. *Biology*, 1(3), 460–483.
<https://doi.org/10.3390/biology1030460>
13. Eivazi, A.R., et al.(2008). Assessing wheat (*Triticum aestivum* L.) genetic diversity using quality traits, amplified fragment length polymorphisms, simple sequence repeats and proteome analysis. *Annals of Applied Biology*, 152(1), 81–91. <https://doi.org/10.1111/j.1744-7348.2007.00201.x>
14. Flint-Garcia, S.A., et al.(2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *The plant journal*, 44(6), 1054–1064. <https://doi.org/10.1111/j.1365-313X.2005.02591.x>
15. Gali, K.K., et al.(2019). Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.). *Frontiers in Plant Sci.* 10, 1538. <https://doi.org/10.3389/fpls.2019.01538>
16. Gupta, P., et al. (2012). Discovery and use of single nucleotide polymorphic (SNP) markers in *Jatropha curcas* L. *Molecular Breeding*, 30, 1325–1335. <https://doi.org/10.1007/s11032-012-9719-6>
17. Jaiswal, V., et al. (2019). Genome-wide association study of major agronomic traits in foxtail millet (*Setaria italica* L.) using ddRAD sequencing. *Scientific reports*, 9(1), 5020. <https://doi.org/10.1038/s41598-019-41602-6>
18. Izzatullayeva, V., et al.(2014). Efficiency of using RAPD and ISSR markers in evaluation of genetic diversity in sugar beet. *Turkish Journal of Biology*, 38(4), 429–438. <https://doi.org/10.3906/biy-1312-35>
19. Jaiswal, V., et al. (2016). Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*T. aestivum* L.). *PloS one*, 11(7), e0159343.
<https://doi.org/10.1371/journal.pone.0159343>
20. Jaiswal, V., et al. (2019). Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (*Setaria italica* L.). *J. Cereal Sci.* 85, 48–55. <https://doi.org/10.1016/j.jcs.2018.11.006>
21. Jo, J., et al.(2017). Development of a genetic map for onion (*Allium cepa* L.) using reference-free genotyping-by-sequencing and SNP assays. *Frontiers in Plant Sci.* 8, 1606. <https://doi.org/10.3389/fpls.2017.01606>
22. Kumari, K., et al.(2013). Development of eSSR-markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. *PloS one*, 8(6), e67742 <https://doi.org/10.1371/journal.pone.0067742>.
23. Kumari, P., et al. (2023). Genome-wide identification of GRAS transcription factors and their potential roles in growth and development of rose (*Rosa chinensis*). *J. Plant Growth Regulation*, 42(3), 1505–1521.
<https://doi.org/10.1007/s00344-022-10635-z>

24. Langmead, B. and Salzberg, S.L.,(2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4),357–359. <https://doi.org/10.1038/nmeth.1923>
25. Lata, C., et al. (2013). Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. *Critical reviews in biotechnology*, 33(3), 328–343. <https://doi.org/10.3109/07388551.2012.716809> Laucou, V., et al.(2018). Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PloS one*, 13(2), e0192540. <https://doi.org/10.1371/journal.pone.0192540>
26. Le Thierry d'Ennequin, M., et al. (2000). Assessment of genetic relationships between *Setaria italica* and its wild relative *S. viridis* using AFLP markers. *Theoretical and Applied Genetics*, 100, 1061–1066. <https://doi.org/10.1007/s001220051387>
27. Lechner, E., et al. (2011). MATH/BTB CRL3 receptors target the homeodomain-leucine zipper ATHB6 to modulate abscisic acid signaling. *Developmental Cell*, 21(6), 1116–1128. DOI: 10.1016/j.devcel.2011.10.018
28. Lee, J.H., et al.(2018). SNP discovery of Korean short day onion inbred lines using double digest restriction site-associated DNA sequencing. *PloS one*, 13(8), e0201229. <https://doi.org/10.1371/journal.pone.0201229>
29. Li, H., et al.(2009). The sequence alignment/map format and SAMtools. *bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
30. Liu, T.Y., et al. (2019). Rhizosheath formation and involvement in foxtail millet (*Setaria italica*) root growth under drought stress. *J.Integrative Plant Bio.* 61(4), 449–462. <https://doi.org/10.1111/jipb.12716>
31. Liu, X.,et al.(2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS genetics*, 12(2), e1005767. <https://doi.org/10.1371/journal.pgen.1005767>
32. Morris, G.P., et al.(2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sci.* 110(2), 453–458. <https://doi.org/10.1073/pnas.1215985110>
33. Murray, M.G. and Thompson, W., (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic acids research*, 8(19),4321–4326. <https://doi.org/10.1093/nar/8.19.4321>
34. Novaes, E., et al. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC genomics*, 9, 1–14. <https://doi.org/10.1186/1471-2164-9-312>
35. Palakurthi, R., et al.(2023). Molecular genetics and taxonomical relationship among selected *Setaria* species using inter simple sequence repeat (ISSR's) and microsatellite (SSRs) markers. *Genetic Resources and Crop Evolution*, 70(3), 903–917. <https://doi.org/10.1007/s10722-022-01474-8>
36. Peterson, B.K., et al. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
37. Pradhan, S., et al. (2020). CRISPR/Cas9-based genome editing, with focus on transcription factors, for plant improvement. In *Transcription Factors for Abiotic Stress Tolerance in Plants*. Academic Press .pp. 63–84
38. Purcell, S., et al.(2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American j. human genet.* 81(3), 559–575. <https://doi.org/10.1086/519795>
39. Ramesh, P., et al.(2023). Molecular genetics and phenotypic assessment of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces revealed remarkable variability of morpho-physiological, yield, and yield-related traits. *Frontiers in Genet.*, 14, 1052575. <https://doi.org/10.3389/fgene.2023.1052575>
40. Rubin, B.E., et al. (2012). Inferring phylogenies from RAD sequence data. *PloS one*, 7(4), e33394. <https://doi.org/10.1371/journal.pone.0033394>

41. Ruperao, P., et al.(2023). A pilot-scale comparison between single and double-digest RAD markers generated using GBS strategy in sesame (*Sesamum indicum* L.). *PloS One*, 18(6), e0286599. <https://doi.org/10.1371/journal.pone.0286599>
42. The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Res.* 49,480–489 <https://doi.org/10.1093/nar/gkaa1100>
43. Tripathi, N., et al. (2012). Assessment of genetic diversity among *Withania somnifera* collected from central India using RAPD and ISSR analysis. *Med Aromatic Plant Sci Biotech.* 6(1), 33–39.
44. Turner, S.D., (2014). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*, 005165. <https://doi.org/10.1101/005165>
45. VanRaden, P.M., (2008). Efficient methods to compute genomic predictions. *J. dairy sci.* 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
46. Vetriventhan, M., et al.(2012). Assessing genetic diversity, allelic richness and genetic relationship among races in ICRISAT foxtail millet core collection. *Plant Genetic Res.* 10(3), 214–223. [doi:10.1017/S1479262112000287](https://doi.org/10.1017/S1479262112000287)
47. Vetriventhan, M., et al. (2014). Population structure and linkage disequilibrium of ICRISAT foxtail millet (*Setaria italica* (L.) P. Beauv.) core collection. *Euphytica*, 196, 423–435. <https://doi.org/10.1007/s10681-013-1044-6>
48. Weber, H. and Hellmann, H., (2009). Arabidopsis thaliana BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. *The FEBS J.*, 276(22), 6624–6635. <https://doi.org/10.1111/j.1742-4658.2009.07373.x>
49. Xiao, Y., et al. (2017). Genome-wide association studies in maize: praise and stargaze. *Molecular plant*, 10(3), 359–374. <http://dx.doi.org/10.1016/j.molp.2016.12.008>
50. Zhang, H., et al. (2018). Developing naturally stress-resistant crops for a sustainable agriculture. *Nature plants*, 4(12), 989–996. <https://doi.org/10.1038/s41477-018-0309-4>
51. Zhang, X.J., et al. (2021). Genetic diversity and structure of *Rhododendron meddianum*, a plant species with extremely small populations. *Plant diversity*, 43(6), 472–479. <https://doi.org/10.1016/j.pld.2021.05.005>
52. Zhao, K., et al.(2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature communications*, 2(1), 467. <https://doi.org/10.1038/ncomms1467>
53. Zhao, Y., et al.(2018). Genetic architecture and candidate genes for deep-sowing tolerance in rice revealed by non-syn GWAS. *Frontiers in Plant Science*, 9, 332. <https://doi.org/10.3389/fpls.2018.00332>

Figures

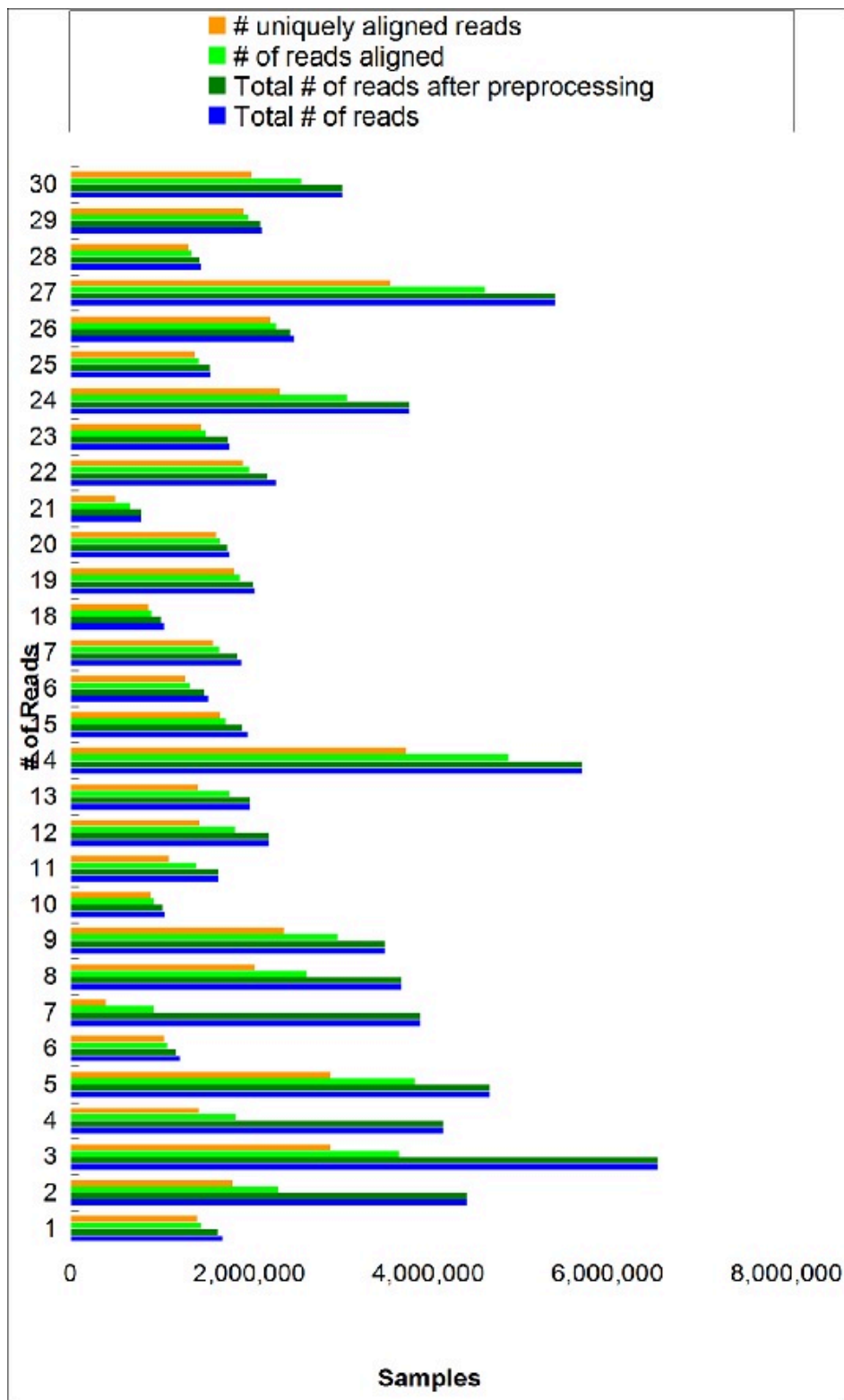


Figure 1

Sequence read measurements of 30 lines and annotated SNPs and INDELs at read depth (RD) >10

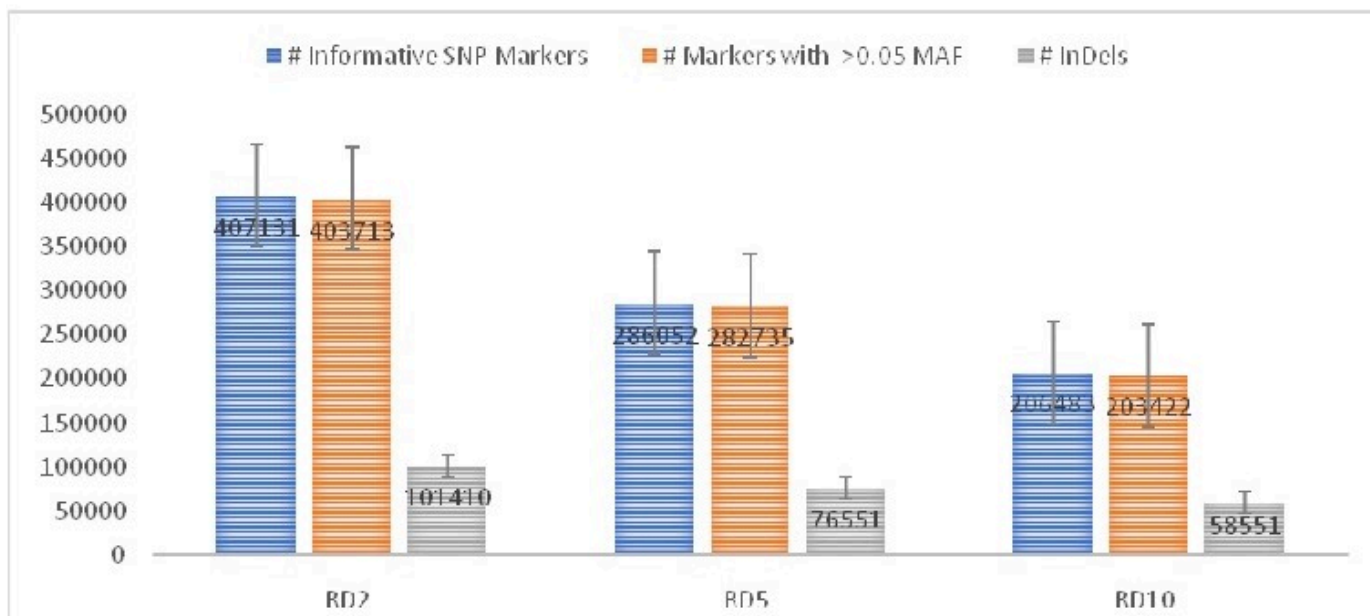


Figure 2

Number of SNP's / InDels at 2X, 5X and 10X depths

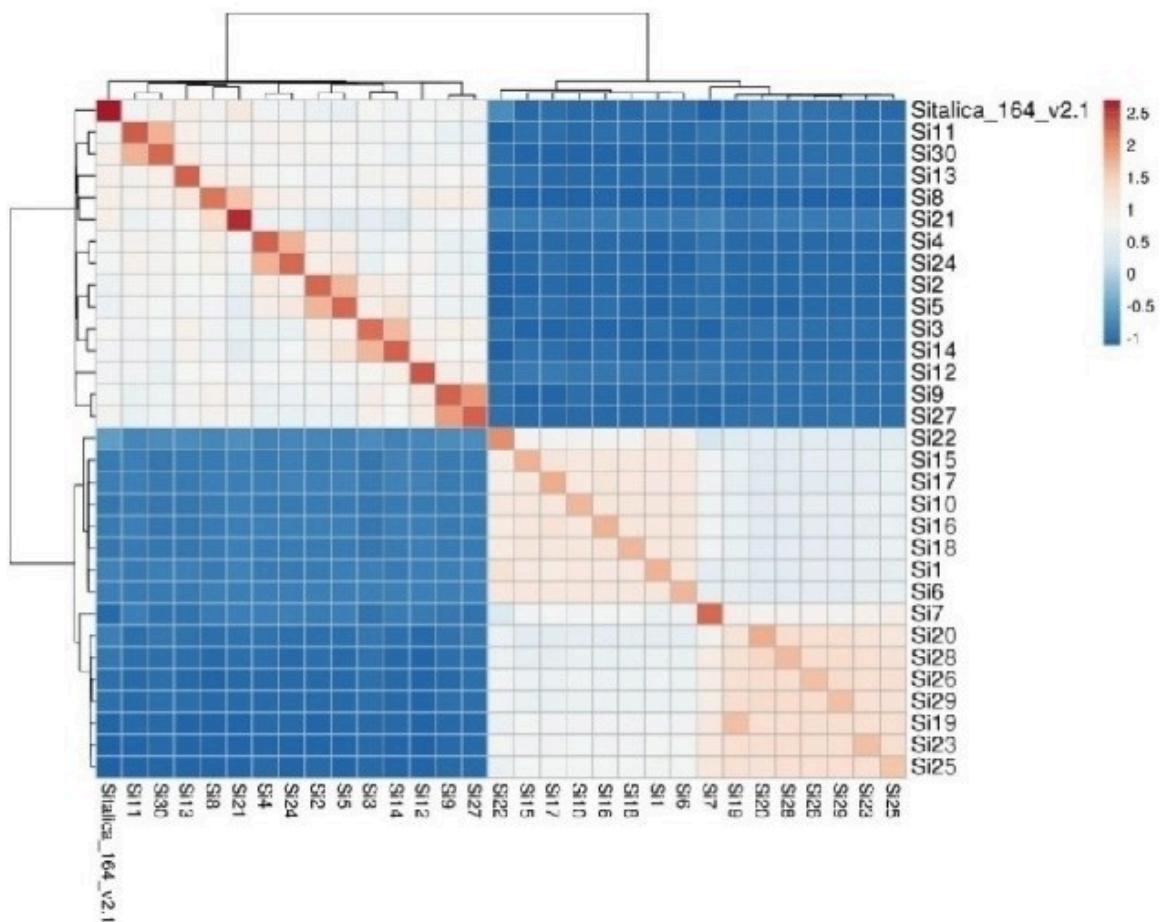


Figure 3

(A) Heatmap of kinship analysis identified SNPs across all 30 lines and (B) Percent heterozygosity across the genotype studied.

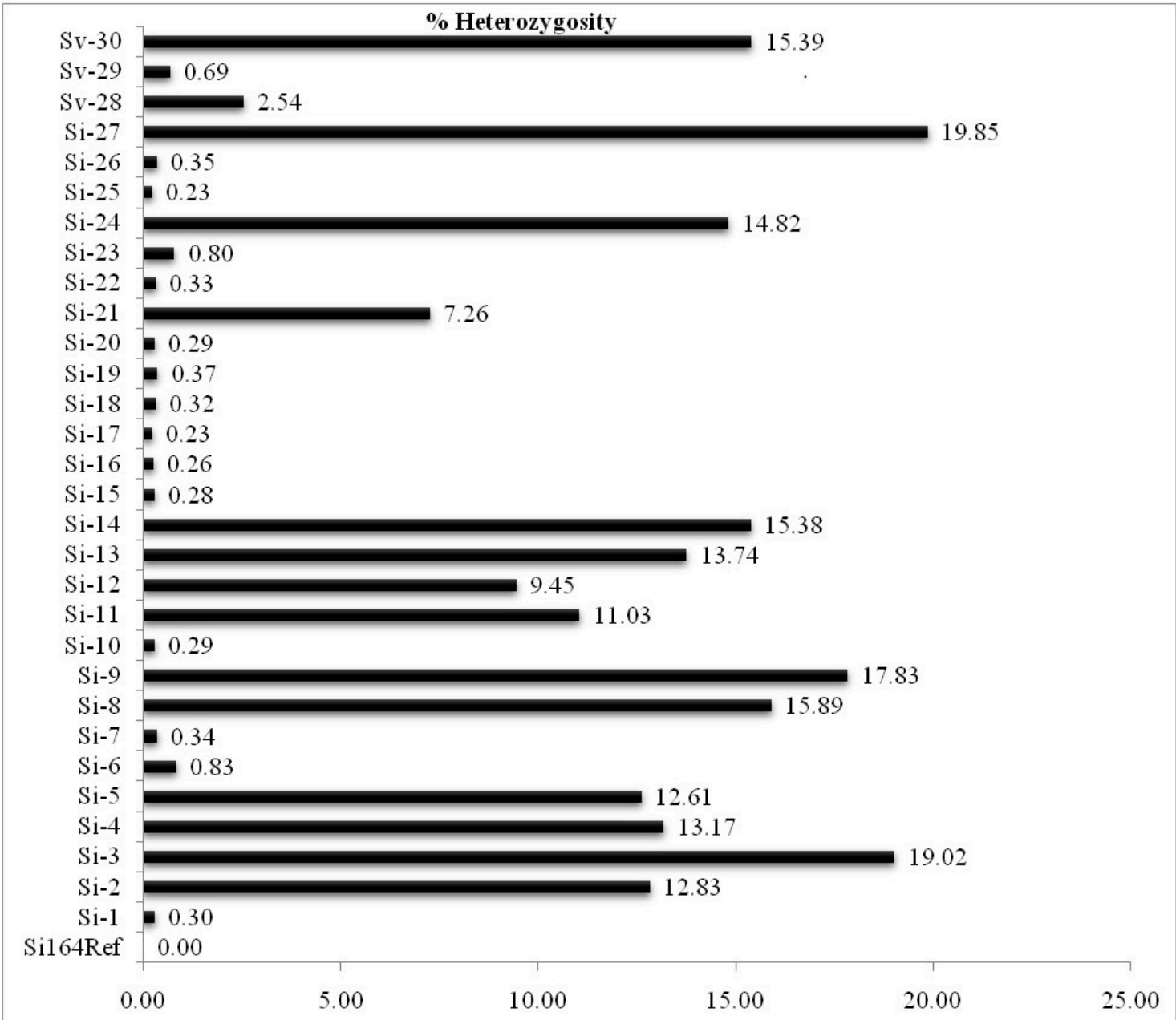


Figure 4

PCA showing the distribution of the germplasms studied.

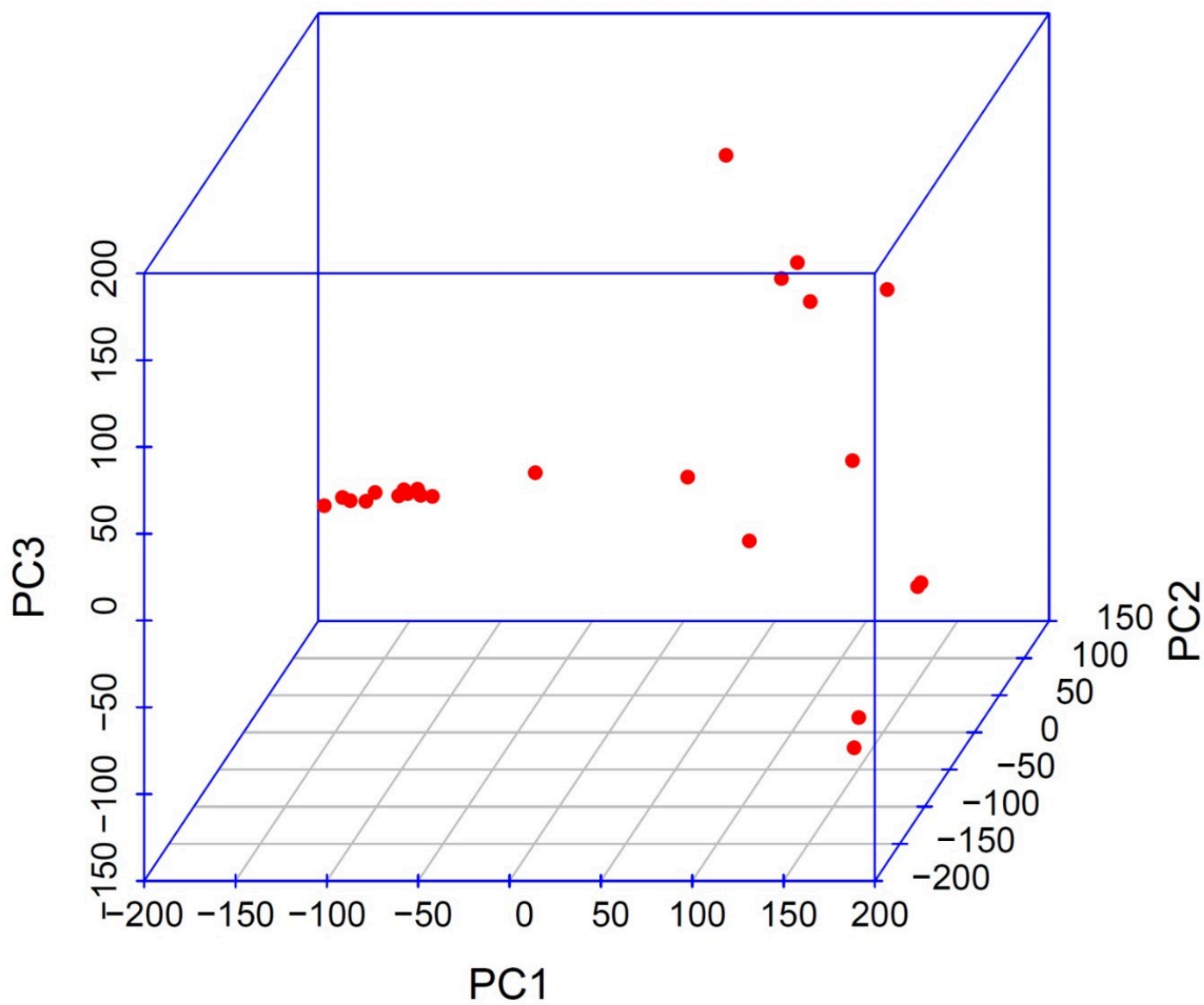


Figure 5

Radial phylogenetic plot showed evolutionary divergence among the *Setaria* germplasms

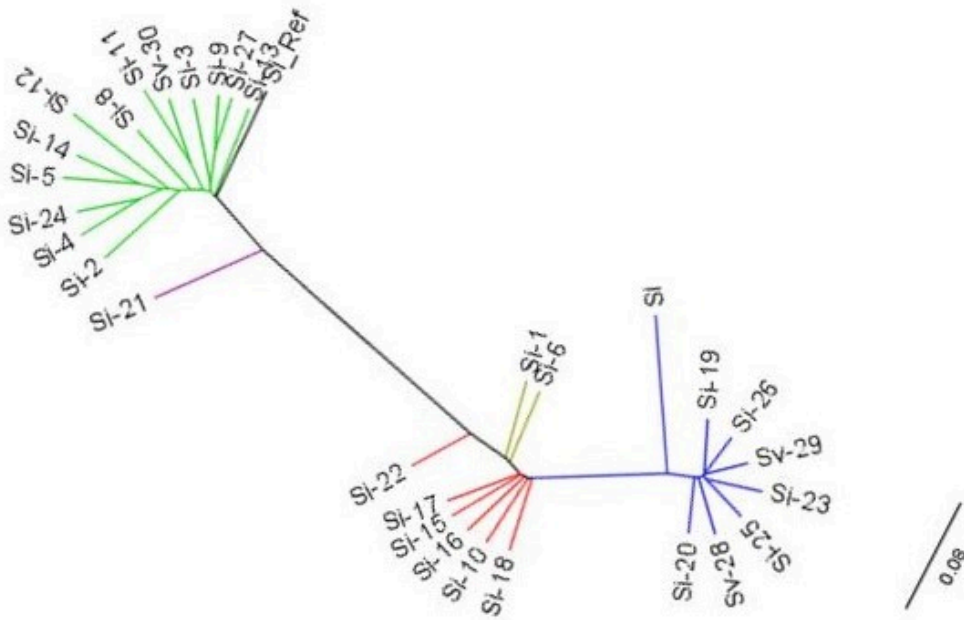


Figure 6

Manhattan plots and quantile-quantile (Q-Q) plots of the GWAS results for the eight traits of *Seteria*. Significant MTA threshold [$-\log_{10}(p) < 10^{-03}$] and Bonferroni threshold were represented by dash (red) and continuous (green) lines, respectively. x-axis represented chromosomes