International Journal of Advance and Applied Research (IJAAR)
Peer Reviewed Bi-Monthly

ISSN – 2347-7075
Impact Factor – 7.328
Vol.8 No.4 Mar–Apr 2021

GENETIC DIVERSITY, POPULATION STRUCTURE, AND LINKAGE DISEQUILIBRIUM OF PEARL MILLET

Dr. Kumar Amit

Ph.D. VKSU. Ara. Bihar.

INTRODUCTION:

Over the last two decades, several seed-related studies have been conducted in semi-arid Africa to improve farmers' access to quality seeds of dry land cereals and legumes. These have indicated that genetic diversity which is at stake is a major resource. However, there is an undeniable evidence of the erosion of crop genetic diversity. The aim of the study is to evaluate genetic variability of pearl millet cultivars obtained from four semi-arid villages of northern-eastern Nigeria namely Dagaceri and Kaska. It should be noted that all the 42 sampled respondents in all the study areas are males and heads of households. They are most active in agricultural practices and also have the final say in the activities of their household. A total of 25 pearl millet genotypes were collected based on diverse morphological data recorded on the field using Participatory Rural Appraisal. The main approach to the present study is to link the advanced biological technique (laboratory study) on genetic characteristics with social science field methodologies. The techniques used in the laboratory analysis are the Amplified Fragment Length Polymorphism (AFLP) and Multiplexed Single Oligonucleotide Amplification. Laboratory studies revealed that genetic compositions of all inventoried pearl millet are not the same. The difference within and between the landraces was estimated using molecular marker (AFLP) and from the data it was noted that farmers' husbandry practice resulted to the isolation of group ideotypes, making landrace names quid pro quo of genetic diversity. It was recommended that because farmers' methods of selection play an important role in genetic management and conservation, it should be linked with the formal seed system to enhance genetic management and control genetic erosion.

Pearl millet is an important cereal crop extensively cultivated in arid and semiarid regions. It ranks sixth in area of production in the world after wheat, maize, rice, barley, and sorghum It is cultivated on >30 million hectares, with a

majority of the area in Africa and the Indian subcontinent (Gupta et al., 2015). It is the main component of traditional farming systems in west-African and the Indian subcontinent. More than 500 million people depend on it as their staple food (National Research Council, 1996). Its high photosynthetic efficiency and dry-matter production capacity (Yadav and Rai, 2013) make pearl millet a highly desirable crop for farmers in adverse agroclimatic regions where other cereals are likely to fail to produce economic yields. It is also grown as temporary summer pasture or cover crop in the Americas and other continents.

Pearl millet is a naturally cross-pollinating species with protogynous flowering, and traditional cultivars are random-mating populations with considerable heterozygosity and heterogeneity. Hybrid breeding has become a major approach for pearl millet improvement and it has brought a progressive yield improvement especially in India (Yadav and Rai, 2013; Kumara et al., 2014). The development of a cytoplasmic male-sterility system (Burton, 1958) has facilitated hybrid seed production. Greater productivity is possible through genetic diversification of hybrid parents if hybrids are developed based on heterosis prediction using parental genomic information (Gupta et al., 2018). There are several semidwarf inbred parental lines that were developed for hybrid breeding in the United States. Assessment of genetic variability permits the identification of genetically diverse parental materials, which can enhance hybrid vigor and yield stability in variable climates (Haussmann et al., 2012; Bashir et al., 2015). Analysis of molecular diversity, population structure, and LD in different sets of materials enables the identification of heterotic parental lines for enhanced hybrid vigor. Genetic diversity analysis in a pearl millet inbred germplasm association panel, which represents cultivated germplasm in different areas and possesses a high gene diversity, was structured into six subpopulations (Sehgal et al., 2015). Those subpopulations supported pedigree differences and different characteristics of specific lines rather than their geographic origin. Also, new germplasms introduced from various sources, mainly the Germplasm Resource Information Network (GRIN) and the Plant Gene Resources of Canada (PGRC), have been collected from different geographic areas in Africa and elsewhere by multiple scientists for the purpose of preservation and utilization. However, there is limited information as to the genetic variability and heterotic potential of these resources. To fill this void, inbred lines developed as seed and pollen parents and germplasm lines need to be assessed for molecular diversity using next-generation markers.

Genetic divergence between crossing parents is very important either to generate variation for selection or maximize hybrid vigor. Hence, formation of heterotic groups among the breeding populations is an essential breeding task to enhance hybrid vigor. However, there is limited research in evaluation of germplasm and breeding materials for heterotic groupings in pearl millet. Morphological traits and pedigree information have been used to characterize germplasm used for development of parents and open-pollinated varieties (Gupta et al., 2011). However, morphological traits are influenced by environment and do not measure diversity accurately. Assessment of genetic diversity, population structure, and LD is necessary to facilitate identification of heterotic groups, breeding via genomics-assisted breeding, and resource conservation. Knowledge of population structure and genetic diversity of breeding populations, germplasm, and parental lines used in the breeding program is also strikingly essential for association mapping studies, genomic selection, and genomics-assisted breeding.

MATERIALS AND METHODS:

Plant Materials:

A total of 400 accessions comprising 203 inbred lines that were developed as parents for hybrid breeding and 197 germplasm lines from different sources were included in this study (Supplemental Table S1). Among them, 155 were parental inbred lines developed by Kansas State University, 27 by the University of Georgia, and seven by the University of Nebraska–Lincoln; 200 germplasm accessions include 50 from the GRIN Plant Genetic Resources Conservation Unit, Griffin, GA, and 149 from PGRC. The germplasm accessions were diverse in geographic origin mainly from Africa, the Middle East, and India (Fig. 1). Two inbred lines (16-861 and 16-911) with poor quality sequences were removed from the pool and the analysis was conducted on 398 accessions.

DNA Extraction:

The seeds were germinated in 96-cell trays and grown in a greenhouse at Kansas State University. Approximately 70 to 100 mg fresh leaf tissue was collected from two to four plants per line 15 d after emergence. Freshly collected tissue in 96-well plates was freeze-dried for 48 h to rapidly remove water. A 4.5-mm steel ball was added to each sample and capped plates were oscillated on a matrix mill (Retsch) at 30 cycles per second for 4 min to grind the tissue.

Genomic DNA was extracted from leaf tissue using a standard high-throughput 2% CTAB and chloroform/isoamyl (24:1) alcohol method in which 4

mM TCEP [tris (2-carboxyethyl) phosphine] was used in place of 2-mercaptoethanol and supplemented with 2% polyvinylpolypyrrolidone and 40 µg RNase. Sample DNA concentrations were assayed using a Quant-iT PicoGreen dsDNA HS assay kit (ThermoFisher) on a FLUOstar Omega fluorescence plate reader (BMG LABTECH) and normalized to 20 ng µL−1 with 10 mM TRIS.

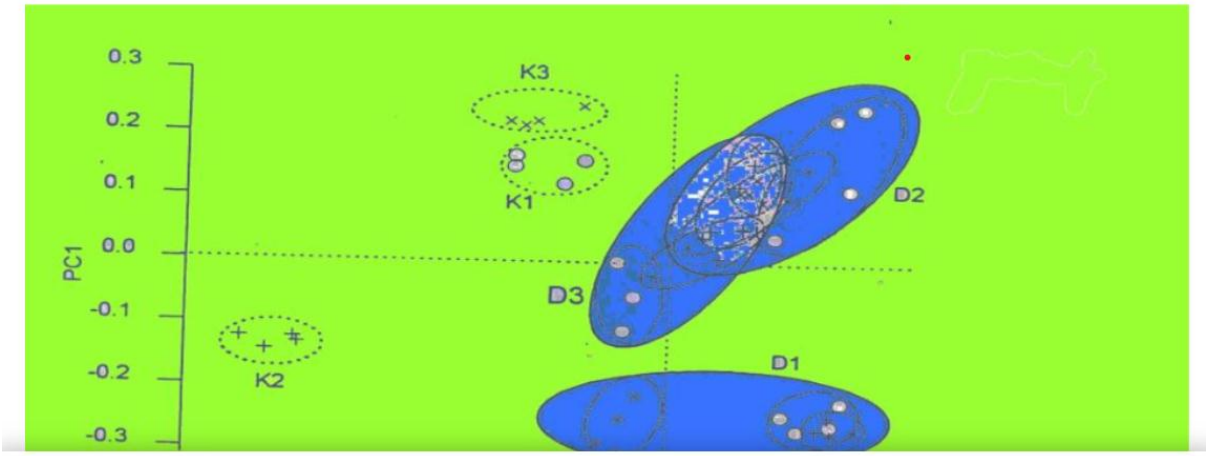


Figure: Principal Coordinates (Pc) Scores Variance

Trace = 85.435

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	Σ
Latent Roots	9.157	5.372	5.059	3.791	3.48	2.946	2.608	2.205	
Percent of Trace	10.7	17	22.9	27.4	31.4	34.9	37.9	40.5	222.7
Individual %	10.7	6.3	5.9	4.5	4	3.5	3	2.6	

Figure: Principal Coordinates (Pc) Scores Variance

Genotyping-by-Sequencing Library Construction, Sequencing, and Single-Nucleotide Polymorphism Calling:

Approximately 200 ng of genomic DNA was digested with PstI (5′-CTGCA/G-3′) and MspI (5′-C/CGG-3′) restriction enzymes (New England Biolabs). The DNA fragments from each sample were ligated to unique barcoded-adapters for identification and to allow pooling of samples for DNA sequencing and analysis.

The GBS libraries were constructed as described by Mascher et al. (2013). All adaptors and primers used for library construction and sequencing were described for Ion Torrent sequencing in Mascher et al. (2013). The concentration

of adenosine 5′-triphosphate (Millipore Sigma) used in the ligation reaction was increased to 1.25 mM, purified ligated DNA pools were quantified using the Qubit dsDNA HS assay kit (Thermo Fisher Scientific), and 7.5 ng DNA was used per 25 µL PCR reaction. After amplification, the libraries were purified using the QIAquick PCR purification kit (Qiagen), resuspended in a 30 µL elution buffer, and then quantified using the Qubit dsDNA HS assay kit (Thermo Fisher Scientific).

Libraries were size selected using a E-Gel system (Thermo Fisher Scientific) and 200 to 300 bp long fragments were recovered, quantified using the Qubit fluorometric quantitation system (Thermo Fisher Scientific), and normalized to a working concentration of 60 pM. Libraries were prepared for sequencing and loaded onto chips (PI v3) using the CHEF system (Thermo Fisher Scientific) and sequenced on an Ion Torrent Proton sequencer (Thermo Fisher Scientific) following manufacturer's instructions and using default analysis parameters. Each library was sequenced three times. Sequence reads from the Ion Torrent system were of variable length. Prior to analysis, all sequencing reads had 80 poly-A bases appended to their 3′ end so that TASSEL 5.0 would attempt to use reads shorter than 64 bases rather than discarding short reads.

The draft pearl millet genome sequence (Varshney et al., 2017) was used as a reference to map GBS reads and identify SNPs using the TASSEL 5.0 GBSv2 discovery pipeline (Bradbury et al. 2007; www.maizegenetics.net). The minimum locus coverage for SNP calls was 0.19 and the minimum minor allele frequency (MAF) was 0.002. All other TASSEL 5.0 settings were the defaults.

Population Structure Analysis:

The millet accessions were first categorized based on their origin or source to assess the diversity within and among geographic areas and breeding programs. The Bayesian model-based quantitative assessment of population subclustering among the 398 pearl millet accessions was assessed using ADMIXTURE (Alexander et al., 2009). The analysis was performed based on a subset of genotypic data obtained by pruning adjacent SNP markers that are in strong LD according to the criterion of a 50-SNP window size and r = 0.5 using PLINK 1.9 program (Purcell et al., 2007). The percentage membership of each of the accession to a subpopulation was assessed assuming hypothetical subpopulations (K) ranging from 1 to 10. The most probable value of K corresponding to the number of subpopulations in the accessions was determined based on the cross-validation error parameters in the ADMIXTURE

program. A cross-validation fold at 10% and a block bootstrap with 2000 iterations were used in the analysis.

Minor allele frequency and allelic combination of SNPs were analyzed. The number of loci with a MAF <10% is exceedingly larger than more frequent loci (Supplemental Fig. 1A). Allele combinations indicate that the translation (A-G, C-T) rate was more than double that of the transversion rate (A-C, G-T) (Supplemental Fig. 1B).

Table. Number of single nucleotide polymorphisms (SNPs) detected and marker gaps from genotyping-by-sequencing of 398 genotypes.

Chromosome	Chromosome length bp	No. of markers on chromosome	Average marker gap on chromosome bp	Max. marker gap on chromosome bp
1	275,468,192	38,710	7116	1,251,556
2	242,893,347	36,854	6591	626,660
3	300,905,882	35,477	8482	867,928
4	191,808,916	28,053	6837	669,411
5	158,669,458	27,191	5835	865,327
6	240,561,232	29,395	8184	992,253
7	154,007,176	26,573	5796	582,530
U	515,119,487	35,714	14,423	1,849,934
Total	2,079,433,690	257,967	–	–

Population structure was further examined with principal components analysis (PCA) using the R package SNPRelate (Zheng et al., 2012). The genetic relationship between accessions was also determined based on the neighbor joining tree algorithm according to shared-allele distance between each pair of accession using the phylogenetic tree analysis in TASSEL software v5.2.35 (Bradbury et al., 2007). The neighbor-joining tree cladogram generated by TASSEL was visualized in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Genome-wide SNP variations, including MAF, observed, and expected heterozygosity for SNP markers, were examined using VCFtools (Danecek et al., 2011). To identify genomic regions shaped by natural selection in the pearl millet population, possible reductions of nucleotide diversity between population subgroups was investigated by analyzing different ratios of nucleotide diversity (π) across the entire genome. In

addition, pairwise genome-wide π, and Tajima's D-test statistics (Tajima, 1989) were calculated across the genome using VCFtools (Danecek et al., 2011).

Linkage Disequilibrium Analysis:

Genome-wide LD was estimated for the panel of 398 genotypes and for each subgroup (as determined by the population structure, which mostly overlapped with geographic origin). The LD between pairs of SNP markers was investigated as squared allele frequency correlation (r2) between pairs of intrachromosomal SNPs with known genomic positions. The LD among SNP markers across the genome was estimated using TASSEL v5.2.35 (Bradbury et al., 2007). The average pattern of genome-wide LD decay over genetic distance was constructed as a scatterplot of r2 values against the corresponding genetic distance between markers. The LD decay curve was fitted using a nonlinear regression developed by Hill and Weir (1988), as modified by Remington et al. (2001).

Genome-Wide Genetic Differentiation And Nucleotide Diversity:

Pairwise estimates of genetic differentiation (FST) between different subgroups defined by population structure and geographic origin were calculated using the method of Weir and Cockerham (1984). Using the VCFtools program (Danecek et al., 2011), specific outlying variants were filtered out from genetic variation data, and genome-wide FST estimates were compared between one subpopulation and all remaining populations. Genome-wide distribution of selection signature was visualized by plotting Weir and Cockerham's FST against chromosomes positions. The top 0.1% FST was used to set the threshold to highlight regions for signature of selection. Nucleotide diversity within each subpopulation was calculated based on a nonoverlapping sliding window of 1 Mbp using VCFtools.

RESULTS:

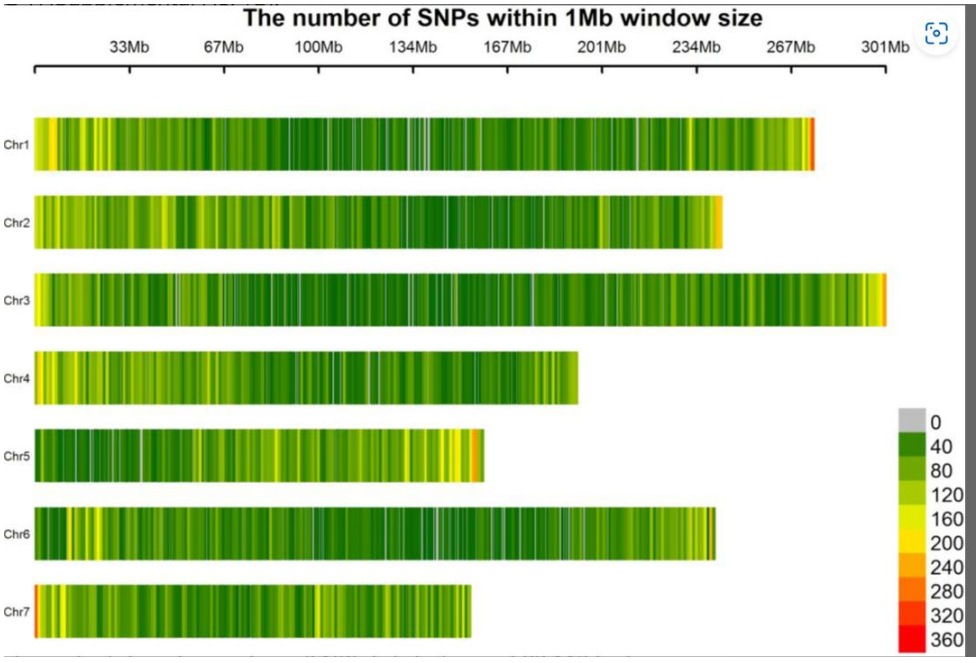
Genome-Wide Single-Nucleotide Polymorphism Discovery:

Ion proton sequencing GBS libraries of 400 samples generated >540 million unique reads and 103,186,800 SNP data points. All the raw sequencing reads for all the accessions have been submitted to the NCBI sequence read archive and deposited under the accession ID “BioProject ID”: PRJNA532596. After filtering the SNPs for >20% missing, 1% MAF, and InDels, we obtained 82,112 SNPs markers (Supplemental Table S3) that were distributed over all seven chromosomes. The largest number of SNPs was discovered on chromosome

1 (38,710) followed by chromosome 2 (36,854) (Table 1). An additional 35,714 SNPs were mapped to the scaffolds not yet assigned to specific chromosomes. Marker density ranged from 0 to 360 per Mb across the genome. Average marker density was ~48.3 SNPs per Mb of the genome. Markers were plotted to visualize the density and distribution of SNPs across all chromosomes (Fig. 2). Genome-wide marker density showed that SNPs are more abundant in the telomeric regions of the chromosome arms than the pericentromeric regions (exact centromere location unknown). In some cases, such as on chromosome 5, more SNPs were discovered on one arm of the chromosome than on the other.

Linkage Disequilibrium:

The degree to which alleles at two loci are associated was assessed to elucidate the patters of genome-wide LD decay within each subpopulation and across the whole population. Genome-wide LD decay in the west-African subpopulation was shorter than in all other subpopulations. The initial (maximum) value of average genome-wide LD (r2) in the west-African accessions declined to 0.1 at 60 kb (Fig. 5). Conversely, accessions from India showed the longest LD decay (r2 remained above 0.2 even at 200 kb). The extent of LD decay in the rest of the subpopulation ranged between the west-African and Indian subpopulations. The respective genetic distances at which initial LD decreased to r2 = 0.1 were 500, 350, and 82 kb in the accessions from southern Africa, the Middle East, and breeding population from the United States, respectively. The initial value of average genome-wide LD (r2) in the total population was reduced to 0.1 at ~18 kb.



Genome-wide single-nucleotide polymorphism (SNP) distributions of 82,112 high-quality SNPs detected by GBS of 398 inbred lines and accessions on the seven chromosomes of pearl millet.

DISCUSSION:

Pearl millet is one of several understudied species, referred to as orphan crops (National Research Council, 1996). Limited genetic diversity studies have been made in pearl millet especially using NGS-based high-throughput SNP genotyping methods. Polymorphisms at the single-nucleotide level are responsible for most of the diversity among individuals, and they often influence the expression of genes and genome evolution of a species (Shastry, 2009). The SNPs are also an ideal high-throughput marker for identifying genes associated with important traits. The SNPs are the simplest and the most abundant of all genetic polymorphisms and can be found in coding, noncoding, and intronic regions of genes possessing diverse biological functions (Zheng et al., 2011). The SNPs may affect transcription factor binding, gene splicing, protein folding, and many other factors at gene and transcript levels (Deng et al., 2017). The application of genomics enables the study of the genotypes and their relationship with complex phenotypic traits in plant breeding (Pérez-de-Castro et al., 2012). Integrating molecular genetics with traditional breeding significantly shortens the breeding cycle and improves selection accuracy (Lande and Thompson, 1990). Genome-wide analyses of SNPs also enable a better understanding of the selective forces that operate on a population and form a strong link between genotype and phenotype.

CONCLUSIONS:

Genotyping-by-sequencing SNPs captured much of the genome variation within the populations studied. Analysis of population structure revealed that diversity was high within the pearl millet subgroups studied. Linkage disequilibrium decay was shortest in populations from the main center of diversity, likely because of historical recombination among landraces. Signature of selection analysis revealed a number of outlying SNPs, which may be associated with important traits, local to each subgroup. The high level of admixture in the US and Indian subgroups indicted that a large number of similar germplasm lines has been shared between those subpopulations for parental inbred-line development. Incorporating more of the east-African and

Middle Eastern germplasm lines will benefit the development of diverse parental lines for increased heterosis in the US and Indian breeding programs.

REFERENCES:

1. Akhunov, E.D., Goodyear, A.W., Geng, S., Qi, L.L., Echalier, B., Gill, B.S., Miftahudin, Gustafson, J.P., Lazo, G., Chao, S., Anderson, O.D., Linkiewicz, A.M., Dubcovsky, J., La Rota, M., Sorrells, M.E., Zhang, D., Nguyen, H.T., Kalavacharla, V., Hossain, K., Kianian, S.F., Peng, J., V Lapitan, N.L., Gonzalez-Hernandez, J.L., Anderson, J.A., Choi, D.W., Close, T.J., Dilbirligi, M., Gill, K.S., Walker-Simmons, M.K., Steber, C., McGuire, P.E., Qualset, C.O., Dvorak, J.. 2003. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res. 13: 753–763.
2. Barendse, W., Harrison, B.E., Bunch, R.J., Thomas, M.B., Turner, L.B.. 2009. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10: 178.
3. Burgarella, C., Cubry, P., Kane, N.A., Varshney, R.K., Mariac, C., Liu, X., Shi, C., Thudi, M., Couderc, M., Xu, X., Chitikineni, A., Scarcelli, N., Barnaud, A., Rhoné, B., Dupuy, C., François, O., Berthouly-Salazar, C., Vigouroux, Y.. 2018. A western Sahara centre of domestication inferred from pearl millet genomes. Nat. Ecol. Evol. 2: 1377–1380.