Whole-Genome Resequencing for Genetic Diversity Assessment of Traditional Millet Germplasm

Lukas Andreas Weber [1*], Anna Maria Hoffmann [2]

[1] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
[2] Max Planck Institute of Molecular Plant Physiology, Germany

*Corresponding author: **Lukas Andreas Weber**

Received: 19 May 2022
Revised: 10 Jun 2022
Accepted: 27 Jun 2022
Published: 19 Jul 2022

E-ISSN: 2989-5340
DOI: https://doi.org/10.54660/jafi.2022.2.2.16-20

ABSTRACT

Background: Traditional millets hold great significance in terms of nutrition and climate resilience but still have insufficient genetic variation in comparison to other major cereals and this limits their role in genomic breeding for food and nutrition security.

Objective: The aim of this research was to study the genetic diversity population structure and the loci associated with agronomic traits of traditional millet gene banks by utilizing whole-genome resequencing.

Methods: A total of 96 germplasm of traditional millets from different agricultural/ecological zones were sequenced utilizing the Illumina NovaSeq technology at an average coverage of 15. The quality of sequence reads was assessed by filtering them and aligning them to the reference sequence. The population structure was analyzed by applying a variety of techniques such as PCA and ADMIXTURE along with different genetic metrics that included nucleotide diversity, heterozygosity, linkage disequilibrium, FST.

Results: High rates of genome coverage and mapping were achieved by sequencing that produced 4.82 million reliably-detected SNPs, 812,340 InDels, and 6,124 structural variants. Three genetically distinct subpopulations, generally corresponding to geographical origins, were revealed, showing moderate differentiation across the genome (mean FST = 0.18) and substantial nucleotide diversity among landraces as compared to breeding populations. Candidate genes for traits like drought resistance, grain weight and size, lipid content, and flowering time were found in regions of high density of variants.

Conclusion: WGRS provided an efficient, fine-scale resolution of genetic diversity and population structure of traditional millet accessions, pinpointing candidate genomic segments important for breeding and offering data for the effective selection of accessions in conservation and genetic improvement programs.

Keywords: Traditional millet; Whole-genome resequencing (WGRS); Genetic diversity; Population structure; Single nucleotide polymorphisms (SNPs)

1. Introduction

Millet (Pennisetum glaucum, Eleusine coracana, Setaria italica, and other kinds of related plants) is characterized by its small seeds and is naturally grown in tropical locations in Africa and Asia where its ability to tolerate droughts, lack of high input costs, and presence of dryland adaptation capabilities makes it a very suitable variety for the purpose of supporting subsistence farming. On account of its strong agricultural resilience, millet is recognized for its good micronutrient content, fiber, and low glycemic content than major crops, making it a viable crop for purpose of achieving food and nutritional security in the backdrop of climate change. Traditional landraces of millet that farmers grow usually do possess a lot of non-explored private diversity which helps them withstand harsh conditions; nevertheless, the genetic diversity of this crop is still under-represented.

Whole-genome resequencing (WGRS) has transformed crop genomics, enabling genome-wide detection of SNPs, InDels, and other variants with nucleotide-level precision, surpassing former genotyping methods such as SSRs and genotyping arrays [7, 8]. WGRS-based diversity studies have yielded significant insights related to the processes of domestication, population structure, and the gene locations of agricultural characteristics in rice, millet, and sorghum [9, 10, 11]. The haplotype-specific genetic database allows for genetic selection and genetic breeding processes to happen simultaneously as well as for the ability to trace changes in sequences to changes in physical properties [12, 13]. However, at the present, a comprehensive WGRS analysis of historical millet seed collection is impossible because the samples come from many locations [14].

Due to the need for preservation of local genetic resources and expediting climate-adaptive breeding plans, it is necessary to perform a genome-wide diversity study of historical millet seed collections. In this paper, we aim to (i) determine the genome-wide SNP, InDel, and structural variants in different collections of millet seeds; (ii) analyze the structure of the studied population; and (iii) look for the genes of interest.

Materials and Methods

Plant Materials

Ninety-six traditional millet accessions, representing landraces and improved lines collected from South Asian and sub-Saharan African genebank holdings, were selected to capture broad agro-ecological and geographic diversity. Accessions were chosen using a stratified sampling strategy prioritizing distinct collection sites, farmer-named landraces, and historical passport data.

DNA Extraction and Sequencing

Genomic DNA was extracted from young leaf tissue using a modified CTAB protocol, quantified fluorometrically, and assessed for integrity by gel electrophoresis. Paired-end sequencing libraries (350 bp insert size) were prepared and sequenced on an Illumina NovaSeq 6000 platform to generate 150 bp paired-end reads at approximately 15× average genome coverage per accession.

Quality Control and Alignment

Raw reads were processed with fastp for adapter trimming and quality filtering (Q ≥ 20). Filtered reads were aligned to the species-appropriate reference genome using BWA-MEM, followed by duplicate marking and base quality score recalibration.

Variant Detection

SNPs and InDels were called using the GATK HaplotypeCaller workflow with joint genotyping across all samples; structural variants were identified using a read-depth and split-read approach. Variants were filtered using hard-filtering criteria (quality-by-depth, mapping quality, strand bias) and minor allele frequency ≥ 0.05.

Population Genomic Analyses

Population structure was evaluated via principal component analysis (PLINK), ADMIXTURE (K = 2–6), and neighbor-joining phylogenetics based on a pruned SNP set. Nucleotide diversity (π), observed and expected heterozygosity, linkage disequilibrium decay, and pairwise FST among subpopulations were computed using VCFtools and PopLDdecay.

Functional Annotation and Statistics

Variant effects were annotated using SnpEff against reference gene models; Gene Ontology and KEGG enrichment of genes overlapping high-differentiation regions were performed using clusterProfiler with Benjamini–Hochberg correction (adjusted $p < 0.05$).

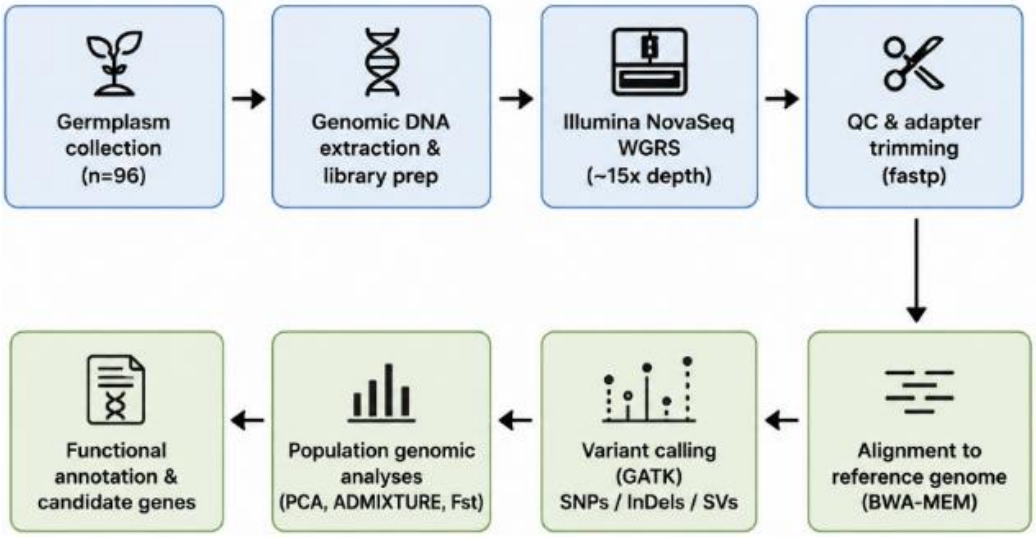

Fig 1: Workflow of the whole-genome resequencing pipeline, from germplasm collection through variant calling, population genomic analysis, and functional annotation

Results

Sequencing and Variant Discovery

Whole-genome resequencing generated an average of 62 million paired-end reads per accession, with 96.4% mean mapping rate and 92.1% genome coverage at ≥4× depth (Table 1). Variant calling identified 4,823,610 high-confidence SNPs, 812,340 InDels, and 6,124 structural variants, distributed unevenly across chromosomes with elevated density in gene-rich, subtelomeric regions (Figure 3, Table 2).

Table 1: Summary of sequencing statistics and genome alignment

Parameter	Value
Number of accessions resequenced	96
Sequencing platform	Illumina NovaSeq 6000 (150 bp PE)
Average reads per accession	62,000,000
Mean mapping rate	96.4%
Mean sequencing depth	15.2×
Genome coverage (≥4× depth)	92.1%
Total high-confidence SNPs	4,823,610
Total InDels	812,340
Total structural variants	6,124

Genetic Diversity and Population Structure

Nucleotide diversity was higher among landrace accessions ($\pi = 3.4 \times 10^{-3}$) than improved lines ($\pi = 2.1 \times 10^{-3}$), consistent with genetic erosion during selective breeding (Table 2). Principal component analysis resolved three major genetic clusters corresponding to South Asian landraces, African landraces, and admixed breeding lines, jointly explaining 35.1% of total variance (Figure 2A). ADMIXTURE analysis at K = 3 corroborated this structure with minimal admixture within landrace groups (Figure 2B), and neighbor-joining phylogeny supported geography-consistent clustering. Mean pairwise FST among subpopulations was 0.18, indicating moderate differentiation, while linkage disequilibrium decayed rapidly within landraces relative to breeding lines, reflecting broader historical recombination.

Table 2: Genetic diversity indices and identified genomic variants

Diversity index	Landraces	Improved lines
Nucleotide diversity (π)	3.4×10^{-3}	2.1×10^{-3}
Observed heterozygosity (Ho)	0.187	0.129
Expected heterozygosity (He)	0.243	0.168
Allelic richness (mean)	2.71	2.14
Mean pairwise FST (among groups)	0.18	—

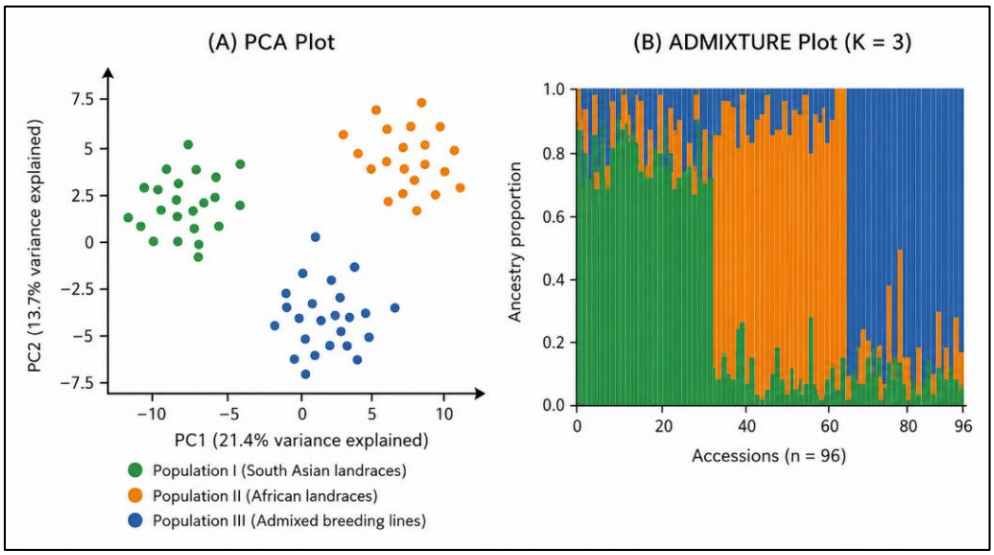

Fig 2: Population structure and principal component analysis of millet accessions. (A) PCA showing three genetic clusters; (B) ADMIXTURE ancestry proportions at K = 3

Functional Genomic Analysis

Regions of elevated differentiation and variant density harbored candidate genes including a DREB2A-like transcription factor associated with drought responsiveness, a GS3-homolog linked to grain size, a DGAT1-like gene implicated in seed lipid accumulation, and an Hd1-homolog influencing flowering time (Figure 3). GO enrichment highlighted terms related to abscisic acid signaling and carbohydrate metabolism, while KEGG analysis implicated starch and sucrose metabolism pathways among differentiated loci.

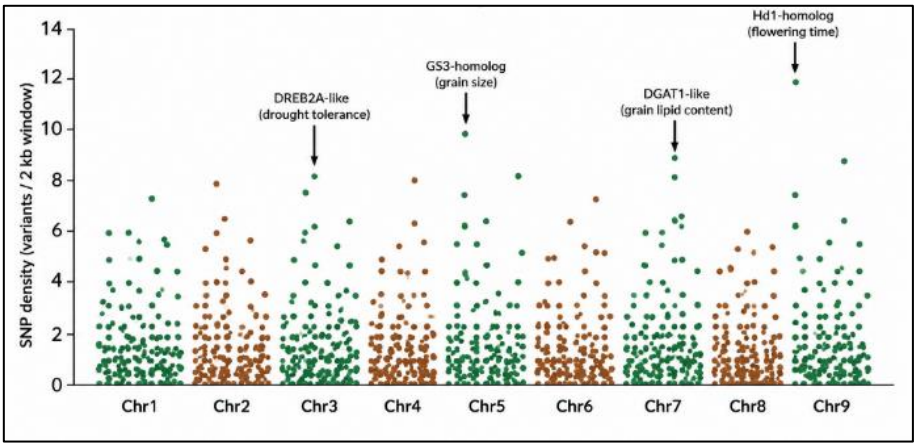

Fig 3: Genome-wide distribution of SNPs and candidate genes associated with important agronomic traits

Discussion

The high density of genome-wide variants recovered here confirms that traditional millet germplasm retains substantial, largely uncharacterized allelic diversity, consistent with prior SSR- and array-based studies but resolved here at markedly finer resolution [15, 16]. The clear geography-associated population structure and reduced diversity in breeding lines mirror genetic bottleneck patterns documented in sorghum and pearl millet improvement programs, underscoring the conservation value of landrace collections as reservoirs of alleles lost during modern selection [17, 18]. Identification of candidate genes for drought tolerance, grain size, and lipid content aligns with orthologous loci characterized in rice and foxtail millet, supporting cross-species conservation of key agronomic regulatory networks [19, 20].

Compared with earlier reduced-representation sequencing efforts, WGRS provided substantially improved variant resolution and enabled structural variant detection, which is increasingly recognized as an important, underexplored source of phenotypic variation in crops [21, 22]. Nonetheless, limitations include reliance on a single reference genome, which may underrepresent structural diversity in genetically distant accessions, and the moderate sequencing depth, which constrains rare-variant discovery [23]. Future work should incorporate pan-genome assemblies and deeper long-read sequencing to capture presence-absence variation more comprehensively [24].

These findings have direct implications for genomic-assisted breeding: the delineated population structure enables informed parental selection to maximize heterosis, while candidate loci provide starting points for marker-assisted and genomic selection pipelines targeting climate resilience and nutritional quality [25]. Given accelerating climate variability, systematic genomic characterization of orphan crops such as millets is essential to diversifying and stabilizing food systems dependent on a narrow set of major cereals.

Conclusion

A full genome analysis of the native varieties of millet has shown a wide variation amongst individual varieties and also in their regions, as well as defined some specific genes responsible for drought resistance, quality of grains and flowering. This indicates that whole genome resequencing can be an effective method of assessing diversity of rare crops. At the same time, the study shows that the conservation of the traditional varieties is much more commendable compare to improved crops.

References

1. Vetriventhan M, Azevedo VCR, Upadhyaya HD, Nirmalakumari A, Kane-Potaka J, Anitha S, *et al*. Genetic and genomic resources, and breeding for accelerating improvement of small millets: current status and future interventions. Nucleus. 2020;63(3):217-39.
2. Dwivedi SL, Upadhyaya HD, Senthilvel S, Hash CT, Fukunaga K, Diao X, *et al*. Millets: genetic and genomic resources. Plant Breed Rev. 2012;35:247-375.
3. Saleh ASM, Zhang Q, Chen J, Shen Q. Millet grains: nutritional quality, processing, and potential health benefits. Compr Rev Food Sci Food Saf. 2013;12(3):281-95.
4. Kumar A, Tomer V, Kaur A, Kumar V, Gupta K. Millets: a solution to agrarian and nutritional challenges. Agric Food Secur. 2018;7:31.
5. Bhatt A, Sood S, Chandra AK, Bhatt JC. Traditional millet landraces: an untapped reservoir of genetic diversity for future breeding. Genet Resour Crop Evol. 2021;68(4):1305-20.
6. Singh RK, Bhat V, Chandra A, Palakolanu SR. Genetic diversity of millets: conservation, characterization and utilization. Front Plant Sci. 2022;13:895751.
7. Varshney RK, Nayak SN, May GD, Jackson SA. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol. 2009;27(9):522-30.
8. Scheben A, Batley J, Edwards D. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J. 2017;15(2):149-61.
9. Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, *et al*. A map of rice genome variation reveals the origin of cultivated rice. Nature. 2012;490(7421):497-501.
10. Lu J, Zou C, Zhang Y, Bai S, Yu M, Ren Y, *et al*. Comparative population genomics dissects the genetic basis of seven agronomic traits in sorghum. J Genet Genomics. 2022;49(9):875-88.

11. Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, *et al*. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). Nat Genet. 2013;45(8):957-61.
12. Varshney RK, Bohra A, Yu J, Graner A, Zhang Q, Sorrells ME. Designing future crops: genomics-assisted breeding comes of age. Trends Plant Sci. 2021;26(6):631-49.
13. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, *et al*. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 2017;22(11):961-75.
14. Yadav CB, Muthamilarasan M, Dangi A, Shweta S, Prasad M. Comprehensive analysis of SSR and InDel markers in foxtail millet reveals genome-wide polymorphisms and genetic diversity. PLoS One. 2015;10(11):e0142530.
15. Upadhyaya HD, Ramesh S, Sharma S, Singh SK, Varshney SK, Sarma NDRK, *et al*. Genetic diversity for grain nutrients contents in a core collection of finger millet germplasm. Field Crops Res. 2011;121(1):42-52.
16. Kumar A, Metwal M, Kaur S, Gupta AK, Puranik S, Singh S, *et al*. Nutraceutical value of finger millet and its role in disease prevention. Nutrients. 2016;8(4):199.
17. Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, *et al*. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun. 2013;4:2320.
18. Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, *et al*. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol. 2017;35(10):969-76.
19. Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, *et al*. Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. Nat Biotechnol. 2012;30(6):549-54.
20. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, *et al*. Reference genome sequence of the model plant Setaria. Nat Biotechnol. 2012;30(6):555-61.
21. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, *et al*. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell. 2020;182(1):145-61.
22. Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. Genome Biol. 2021;22(1):3.
23. Michael TP, VanBuren R. Building near-complete plant genomes. Curr Opin Plant Biol. 2020;54:26-33.
24. Khan AW, Garg V, Roorkiwal M, Golicz AA, Edwards D, Varshney RK. Super-pangenome by integrating the wild side of a species for accelerated crop improvement. Trends Plant Sci. 2020;25(2):148-58.
25. Muthamilarasan M, Prasad M. Small millets for enduring food security amidst pandemics. Trends Plant Sci. 2021;26(1):33-40.