Plant Communications
Resource article

CellPress
Partner Journal

Integrating pan-genome analysis, GWAS, and interpretable machine learning to prioritize trait-associated structural variations in *Setaria italica*

Wenying Wang[1,4], Tianhao Wu[1,4], Guangyu Fan[2,4], Shuai Zhang[2], Songyu Liu[1], Shuqin Jiang[3], Qian Cheng[1], Meiqi Shang[1], Yanfen Xu[3], Wenlin Zhang[3], Jianan Zhang[3], Xiangfeng Wang[1,*], Zhihai Zhao[2,*] and Jun Yan[1,*]

[1]State Key Laboratory of Maize Bio-breeding, National Maize Improvement Center, Frontiers Science Center for Molecular Design Breeding, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100094, China

[2]Institute of Millet, Zhangjiakou Academy of Agricultural Science, Zhangjiakou, Hebei Province 075000, China

[3]Molbreeding Biotechnology Co., Ltd, Shijiazhuang, Hebei Province 051430, China

[4]These authors contributed equally to this article.

***Correspondence: Xiangfeng Wang (xwang@cau.edu.cn), Zhihai Zhao (zhaozhihai58@163.com), Jun Yan (yanjun@cau.edu.cn)**

https://doi.org/10.1016/j.xplc.2025.101626

ABSTRACT

Structural variations (SVs), especially presence–absence variations (PAVs), are crucial in crop domestication and trait improvement. Although pan-genome analysis provides an exhaustive view of PAVs, it is often limited by high costs and restricted sample sizes. In contrast, genome-wide association studies (GWASs) can effectively identify trait–marker associations in large populations but typically overlook PAVs and face challenges in distinguishing causal variants due to linkage disequilibrium. In this study, we performed *de novo* assembly of eight reference-quality foxtail millet (*Setaria italica*) genomes and constructed a graph-based pan-genome to systematically explore PAVs. We subsequently performed a GWAS with 344 millet accessions, targeting genomic regions associated with the color of the leaf, leaf sheath, and leaf pulvinus. Using interpretable machine-learning models, we identified large-effect variants in the 26.84–26.94 Mb interval on chromosome 7, including a 5002-bp *Copia* element insertion and other key variants associated with phenotypic variations in leaf color traits. This integrative approach combines the detailed variant-detection capabilities of pan-genome analysis with the large-scale mapping potential of GWASs and enhances variant prioritization using interpretable machine learning, providing a cost-efficient yet effective framework for studying agronomic traits in crops.

Key words: foxtail millet, pan-genome, presence–absence variation, PAV, genome-wide association study, GWAS, machine learning

Wang W., Wu T., Fan G., Zhang S., Liu S., Jiang S., Cheng Q., Shang M., Xu Y., Zhang W., Zhang J., Wang X., Zhao Z., and Yan J. (2026). Integrating pan-genome analysis, GWAS, and interpretable machine learning to prioritize trait-associated structural variations in *Setaria italica*. Plant Comm. **7**, 101626.

INTRODUCTION

Setaria italica (*S. italica,* foxtail millet) is one of the oldest domesticated cereal crops, with a cultivation history spanning over 10 millennia (Zhang et al., 2012). It is valued for its high nutritional content, being rich in carbohydrates, proteins, lipids, and vitamins, and serves as both a staple food and an important source of animal feed (Hassan et al., 2021). Owing to its exceptional drought tolerance, low water requirement, and high adaptability to nutrient-poor soils, foxtail millet thrives in arid and semi-arid regions where other major cereals often fail. Its short growing cycle and resilience to extreme environmental conditions allow stable yields even under limited management inputs (Shrestha et al., 2023). These attributes make foxtail millet a strategic crop for ensuring food security and agricultural sustainability, particularly in semi-arid regions of Africa and Asia experiencing intensified climate change (Srivastava et al., 2022).

Recent advances in *Setaria* genome research have greatly enhanced our understanding of this model crop species. The first chromosomal-scale genome assembly of *S. italica* was published by Zhang et al. (2012), providing a foundation for subsequent genomic research. He et al. (2024) further advanced *Setaria* genomics by assembling an improved version of the reference

genome, Yugu1-T2T, providing a more comprehensive genomic resource. He et al. (2023) assembled reference-grade genomes from 110 *Setaria* varieties and integrated them with earlier reference genomes to create a comprehensive pan-genome. This pan-genome study identified structural variations (SVs), including presence–absence variations (PAVs), which are linked to critical agronomic traits, such as those affected by the yield-related gene *SiGW3* (He et al., 2023). Research in other crops has similarly highlighted the importance of SVs and PAVs in shaping essential traits (Qin et al., 2021; Liu et al., 2024; Jiao et al., 2025). PAVs can result in the complete loss or duplication of genes, affecting key agronomic traits such as metabolic phenotypes, stress resistance, and yield. Moreover, PAVs often involve genes crucial to domestication and adaptive evolution, making their identification essential for understanding the mechanisms that regulate important traits in crops (Yuan et al., 2021).

Although pan-genome analysis can facilitate the accurate identification of PAVs involved in species evolution, the high cost of *de novo* genome assembly often limits the number of samples analyzed, hindering the broader use of pan-genomes for the discovery of agronomically important genes (Li et al., 2022). The most efficient method for such discovery remains a genome-wide association study (GWAS) performed on a genotyped and phenotyped population, typically from a collection of core germplasm (Yang et al., 2022). However, to balance research costs and population size, GWASs often use custom genotyping arrays or medium- to low-depth sequencing. This approach may overlook key variations such as PAVs, limiting its utility for genetic improvement. Moreover, due to linkage disequilibrium (LD), a GWAS typically only identifies trait-associated genomic regions, making it difficult to assess the functions of specific variants within these regions. Machine-learning algorithms can identify complex non-linear relationships between variants and traits, overcoming the limitations of traditional statistical methods in detecting associations in high-dimensional data (Yan et al., 2023). In particular, interpretable machine-learning techniques not only establish predictive relationships between variants and traits but also enhance the transparency of models' decision-making process, revealing the specific contributions of individual variants (Azodi et al., 2020). Therefore, the integration of pan-genome analysis, GWASs, and interpretable machine learning offers new avenues for deciphering the contributions of PAVs to trait variation.

In the present study, we constructed a pan-genome atlas from eight reference-level millet genomes generated by *de novo* assembly. The eight sequenced accessions were representative elite lines widely used in millet breeding in China, and we used their genomes to identify PAVs. We then performed a GWAS using a population of 344 *Setaria* accessions genotyped with a custom array and identified chromosomal regions associated with three agronomic traits: leaf color, leaf sheath color, and leaf pulvinus color. We developed an analytical framework that uses interpretable machine-learning models to prioritize large-effect variants such as PAVs within these regions. Ultimately, we identified a 5002-bp insertion of a *Copia* element that appears to be a key factor associated with the change in leaf sheath and pulvinus color from purple to green. Beyond pan-genome construction, our study introduces a cost-effective, machine-learning–driven strategy for functional prioritization of trait-associated structural variants, demonstrating that even a limited set of representative assemblies can effectively reveal biologically meaningful variation.

RESULTS

De novo genome assembly of eight millet accessions

To construct a pan-genome atlas of foxtail millet, we selected eight genetically diverse elite lines representing distinct geographical origins and ecological adaptations (Supplemental Table 1). This panel includes six stress-resistant landraces (DXH02, DXH06, DXH07, and DXH08 from China; DXH04 and DXH05 from India), which exhibit exceptional drought tolerance, disease resistance, and environmental adaptability developed through long-term natural selection, along with two high-yield cultivars (DXH01/Yugu18 and DXH03/JP-60D) with superior agronomic performance. These lines collectively capture key evolutionary adaptations found across major millet-growing regions in northern and southern China and India (Supplemental Table 1). Reference-quality genomes for the eight accessions were generated by *de novo* assembly using a combination of PacBio circular consensus sequencing (CCS) and high-throughput chromosome conformation capture (Hi-C) sequencing (methods and Supplemental Figure 1). The resulting assemblies ranged from 426.3 to 438.8 Mb in length, with average scaffold and contig N50 values of 43.7 Mb and 15.6 Mb, respectively (Supplemental Table 2). Between 93.8% and 97.2% of the sequences were anchored to chromosomes, with the fewest gaps (38) in DXH04 and the most (171) in DXH07. Assembly completeness was assessed using benchmarking universal single-copy orthologs (BUSCO) analysis (Simão et al., 2015) with the Embryophyta gene set (1614 genes), resulting in an average complete BUSCO score of 98.5%. Compared with the foxtail millet pan-genome constructed by He et al. (2023), our assemblies exhibit higher quality, as indicated by improved BUSCO completeness scores and contig N50 values (Supplemental Figure 2). The average contig N50 across our eight assemblies was 15.6 Mb, with the lowest value still above 11.1 Mb; in contrast, among the 110 samples analyzed by He et al. (2023), only three had a contig N50 greater than 11 Mb, and the average was just 1.7 Mb.

Genome consistency was evaluated by aligning PacBio long reads to the assembled genomes, which showed that over 99.0% of all genomic regions were covered. In addition, a *k*-mer–based approach was used to assess the assembly consensus (quality value [QV]) and genome completeness. The average QV across the eight genome assemblies was 61.07, with values ranging from 56.48 to 68.42, which indicates high consensus accuracy. Long terminal repeat (LTR) retrotransposon analysis further supported the assembly quality; all eight genomes met the reference standard (Long Terminal Repeat Assembly Index [LAI] > 10), with LAI values ranging from 15.63 to 18.33. These summary statistics demonstrate the high quality of the genome assemblies and the completeness of their predicted gene sets, which are comparable to those reported in He et al. (2023). Furthermore, the eight assemblies showed high collinearity with the *S. italica* Yugu1 reference genome (Supplemental Figure 3), confirming their quality and structural consistency.

Annotation of protein-coding genes, non-coding RNAs, and transposable elements

Our gene-annotation pipeline integrated homology-based searches, *de novo* prediction, and transcriptome-based methods (see methods). All eight genomes contained a comparable number of protein-coding genes (34 222–34 841), with an average gene length of 2622 bp and an average mRNA length of 1151 bp (Supplemental Table 2). On average, 99.1% of the tested BUSCO genes were complete in each genome, further underscoring the high completeness of the eight assemblies. Transposable elements (TEs) in the eight genomes were identified using a combination of Repbase searches and *de novo* predictions. Repetitive sequences accounted for 46.6%–49.2% of each genome. On average, 61.7% and 25.1% of these repetitive regions consisted of class I LTR retrotransposons and class II DNA transposons, respectively, across the eight accessions (Supplemental Table 3). As in other crop species, *Gypsy* and *Copia* elements were the two most abundant superfamilies of class I repeats, accounting for 64.8% and 35.3% of all LTR retrotransposons, respectively. Compared with protein-coding genes and TEs, genomic regions containing rRNA sequences exhibited much greater variation, with copy numbers ranging from 3729 in DXH03 to 7509 in DXH01 (Supplemental Table 4). By contrast, the copy numbers of microRNAs (miRNAs) and small nuclear RNAs (snRNAs) showed less variation among accessions, ranging from 159 to 165 and from 436 to 463, respectively.

Functional annotations were assigned to protein-coding genes by searching their protein sequences against the SwissProt, TrEMBL, InterProScan, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Annotation results from the different databases were combined; 93.0% of the 34 513 protein-coding genes received annotations from at least one source, whereas 19 876 (57.6%) were annotated by all five sources, with similar functional descriptions (Supplemental Figure 4).

Construction of a graph-based pan-genome for millet

We constructed a millet pan-genome using a graph-based approach that integrated eight genome assemblies, including two modern cultivars (DXH01 and DXH03) and six landraces (DXH02 and DXH04–DXH08). Yugu18 (DXH01), a modern cultivar with 34 841 genes and a genome size of 438.8 Mb, was selected as the reference due to its superior genome completeness and agronomic performance. As a systematically bred variety, DXH01 better represents mainstream cultivars, facilitating the identification of core genomic features. Its high yield stability, broad adaptability, and strong lodging resistance further support its suitability. Compared with DXH03, a parental line of earlier cultivars, DXH01 exhibits improved performance across multiple traits. After graph-based integration of the eight assemblies, the final pan-genome was 565.6 Mb in size, 133 Mb longer than the average genome size (432.6 Mb) of the individual assemblies. It contained 37 862 protein-coding genes, 3349 more genes than the average (34 512) of the individual assemblies (Supplemental Table 5). We clustered orthologous gene families and identified 35 240 clusters containing 276 103 genes. We then classified the 35 240 clusters as "core" (shared by all eight genomes), "soft-core" (shared by seven genomes), "dispensable" (shared by two to six genomes), or "private" (present in only one genome). The 21 645 core clusters contained 233 352 genes, representing 61.4% of all clusters and 84.5% of all genes. There were 1332 soft-core clusters (10 543 genes) and 5359 private clusters (5655 genes), representing 3.8% and 15.2% of all clusters and 3.8% and 2.0% of all genes, respectively (Figure 1A). DXH01 had the largest number of private gene clusters (1401), whereas DXH02 had the fewest (398). The remaining 6904 clusters (26 533 genes) were classified as dispensable. As additional genomes were sequentially incorporated into the pan-genome, the number of core gene clusters decreased from 27 384 to 21 645, whereas the total number of gene clusters increased from 27 384 to 35 240 (Figure 1B).

Evolution of the millet genome

To clarify the phylogenetic relationships and genetic population structure of millet, we constructed a phylogenetic tree using orthologous gene sets from the eight assembled millet accessions (DXH01–DXH08) and the *S. italica* reference genome, with five other plant species (*Sorghum bicolor*, *Zea mays*, *Oryza sativa*, *Brachypodium distachyon*, and *Arabidopsis thaliana*) as outgroups. We identified 3979 single-copy gene families among the 14 genomes, and the number of unclustered genes ranged from 838 to 1694, significantly lower than the 11 856 unclustered genes in *O. sativa* (Supplemental Table 6). With the inclusion of the six additional genomes (the reference genome and five outgroup species), the number of core gene families shared by all 14 genomes was about 9200 (Supplemental Figure 5). The eight newly assembled accessions and the *S. italica* reference genome showed similar proportions of single-copy and multi-copy orthologs, as well as unique paralogs, whereas the other five genomes showed distinct patterns (Figure 1C).

To further examine the evolutionary relationships among the millet accessions and the five outgroup species, we generated an unrooted phylogenetic tree using the single-copy orthologs identified above. The resulting tree indicated that millet and maize diverged about 27.35 (25.0–29.7) million years ago (MYA), likely falling within the Miocene epoch of the Neogene period (Figure 1D). We next analyzed gene family expansion and contraction in the millet pan-genome. Among the eight newly sequenced millet accessions, contracted gene families were generally more numerous than expanded gene families. In contrast, in rice and maize, expanded families significantly outnumbered contracted families. To investigate the biological significance of the expanded and contracted gene families in the millet genomes, we performed GO enrichment analysis (Supplemental Table 7). We specifically focused on the nine *S. italica* accessions, which exhibited 79 expansion events and 65 contraction events relative to their most recent common ancestor. The expanded gene families were significantly enriched in the GO terms "oxidoreductase activity" (GO:0016705) and "oxidation-reduction process" (GO:0055114), whereas the contracted families showed a greater diversity of enriched terms, including "protein kinase activity" (GO:0004672), "protein phosphorylation" (GO:0006468), and "ATP binding" (GO:0005524).

Because the expanded and contracted gene families may have undergone positive selection, thereby contributing to millet evolution, we next identified positively selected genes (PSGs) in millet, using the nine millets as the foreground branch and *S. bicolor*,

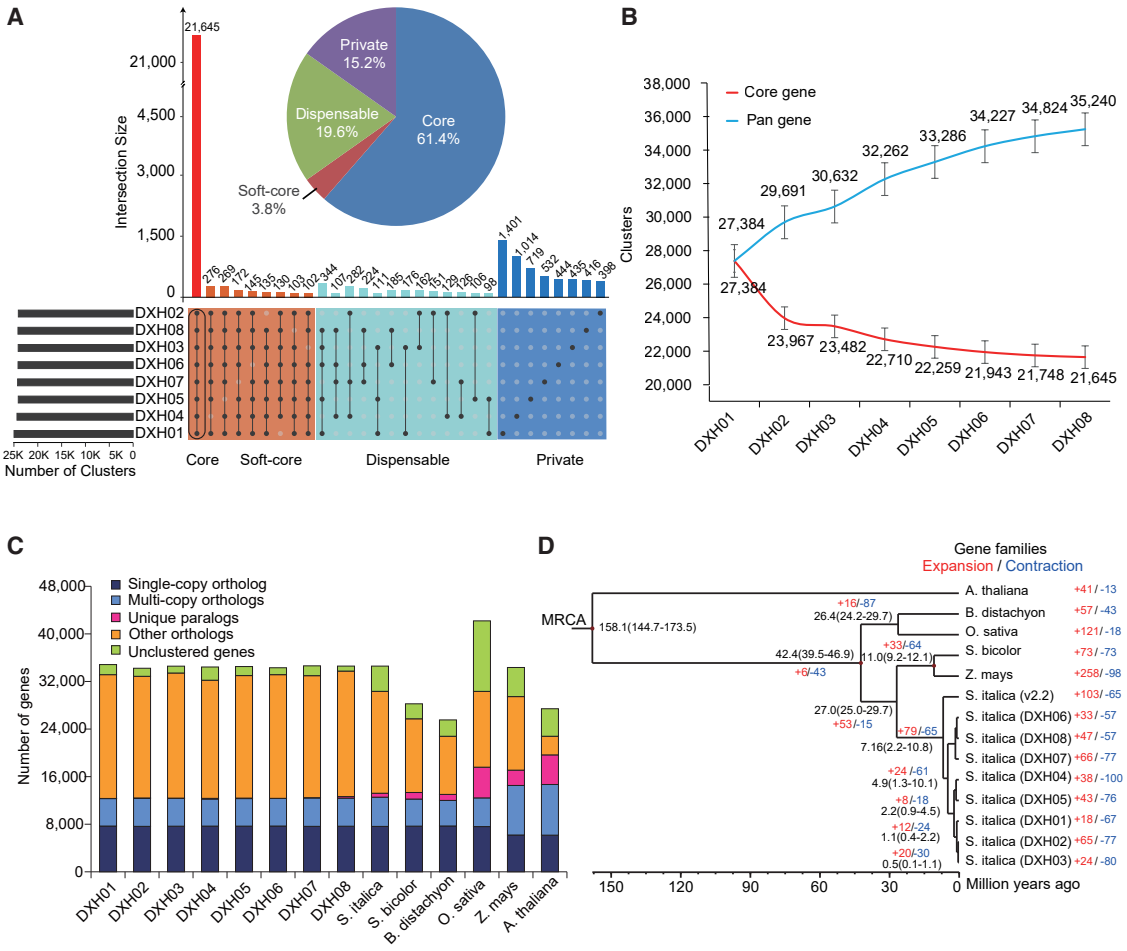


Figure 1. Pan-genome and core-genome analysis of eight millet accessions.

(A) Numbers and proportions of core, soft-core, dispensable, and private gene clusters. The UpSet plot illustrates the distribution of gene families (clusters) shared among the eight genomes. The *x*-axis represents different combinations of genomes with shared gene families, indicated by filled circles and connecting lines. The *y*-axis denotes the number of gene families present in each corresponding genome combination.

(B) Numbers of gene clusters in the pan-genome and the core genome.

(C) Statistics for orthologous and paralogous gene clusters in the eight pan-genome accessions, the *S. italica* reference genome, and the genomes of five other plant species.

(D) Phylogenetic tree showing divergence times and the evolutionary history of orthologous gene families.

Z. mays, *O. sativa*, *B. distachyon*, and *A. thaliana* as background branches. We estimated the direction and magnitude of positive selection by calculating K_{a/K_s} using the branch-site model implemented in the Phylogenetic Analysis of Maximum Likelihood (PAML) package. In total, 22 PSGs covering a wide range of biological functions (Supplemental Table 8) were detected in the nine millet genomes ($p < 0.05$, likelihood ratio test). Interestingly, seven of the 22 PSGs encoded transcription factors, which suggests that transcriptional regulatory networks may have undergone positive selection during the evolutionary adaptation of millet (Supplemental Table 9).

Identification of SVs in eight millet genomes

We directly compared the genome sequences of the eight accessions to identify five types of SVs: insertions, deletions, inversions, translocations, and duplications. Between 6585 and 11 130 SVs were identified in each assembly, with presence (insertion) and absence (deletion) variations (i.e., PAVs) accounting for most of the SVs (96.4%; Figure 2A). By contrast, translocations, inversions, and duplications accounted for only 1.2%, 0.3%, and 2.1% of SVs, respectively (Figure 2B and Supplemental Figure 6). On average, the lengths of the PAVs were relatively similar across the eight genomes and ranged from 51 bp to 30 926 bp (Figure 2C). We assessed the distribution of PAVs across genomic regions and found that they were evenly distributed among intergenic regions (27.5%) and the 2-kb flanking regions upstream (33.1%) and downstream (28.3%) of protein-coding genes (Figure 2D).

To identify SV hotspots on the eight millet chromosomes, we calculated the densities of genes, insertions, and deletions using a 1-Mb sliding window with a 200-kb step size (Figure 2E). The densities of insertions and deletions were positively correlated with each other ($r = 0.79$, $p < 2.2 \times 10^{-16}$) and were both positively correlated with gene density ($r = 0.57$, $p < 2.$

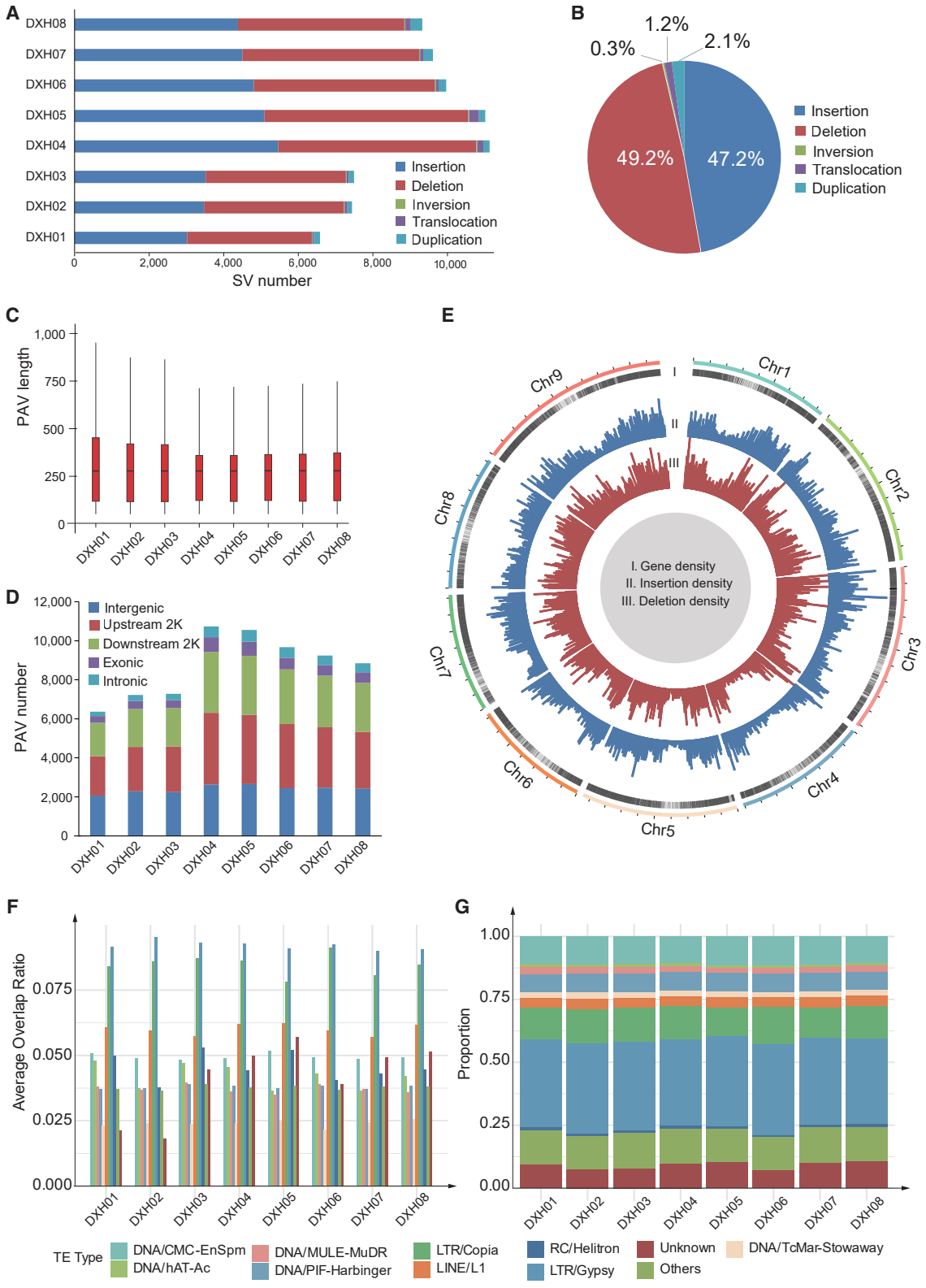


Figure 2. Structural variations in the eight pan-genome accessions.

(A) Numbers of five types of SVs in the eight genomes.

(B) Proportions of five types of SVs in the pan-genome.

(C) Boxplot showing the distribution of PAV lengths in the eight genomes.

(D) Numbers of PAVs in different regions of the eight genomes.

(legend continued on next page)

2×10^{-16}). Because large PAVs may cause the loss or gain of genes that contribute to phenotypic variation (Li et al., 2021), we paid particular attention to PAVs that resulted in the loss of entire genes in at least one genome. In total, 350 genes were lost or gained in at least one genome as a result of 290 PAVs (Supplemental Table 10).

We next analyzed the associations between structural variants and transposable elements (TEs) to infer their possible origins and functional relevance. We calculated the proportion of structural variants that overlapped with TEs. Across all accessions (DXH01–DXH08), approximately 46%–48% of structural variants overlapped with TEs, demonstrating the widespread contribution of transposon insertions to genome structural diversity (Supplemental Figure 7A). Among these, 27%–28% of structural variants overlapped with LTR-type TEs, which were the most common category of TE-associated structural variants and showed consistent proportions across samples (Supplemental Figure 7B). Overlap analysis of nine major TE types revealed that DNA, long interspersed nuclear elements (LINEs), LTRs, RCs, retrotransposons, and short interspersed nuclear elements (SINEs) were present in all pan-genome samples and were linked to structural variants. For most TE types, fewer than 10% of instances overlapped with structural variants (Supplemental Figure 7C). Among them, LTR elements showed the highest frequency of overlaps, followed by DNA elements, whereas LINEs and other elements were less involved (Supplemental Figure 7D). We further assessed overlaps with 143 annotated TE classes and found that the top 10 structural-variant-associated classes were predominantly consistent across accessions, particularly LTR/Gypsy, LTR/Copia, DNA/CMC-EnSpm, and DNA/PIF-Harbinger (Figure 2F and 2G). These findings suggest that specific TE classes, particularly LTR retrotransposons, are critical in shaping the structural variant landscape and contribute to SV and gene-content diversity in foxtail millet.

Characteristics of PAV-related genes in the eight accessions

We next assessed the frequencies of accession-specific and shared gene losses among the eight genomes. A total of 112 genes showed accession-specific loss relative to the reference, representing nearly one-third of all PAV-related losses (Figure 3A). Based on PAV patterns, 351 genes were classified into four categories: 14 core genes (present in all eight genomes but absent from the reference), 22 soft-core genes (present in seven genomes), 112 private genes (accession-specific), and 203 dispensable genes (Figure 3B and 3C). GO enrichment analysis of the lost genes (Figure 3D) revealed that the three most enriched terms were all related to ion binding, potentially reflecting divergence in ion-associated pathways during selective breeding. Notably, the GO analysis provides a functional overview of PAV loss patterns and may indicate categories under selection; however, the specific selective mechanisms and their phenotypic consequences remain to be determined.

Identification of trait-related genomic regions through GWAS

To identify candidate PAVs potentially associated with millet traits, we used genotyping data from a custom millet 96K SNP array containing 82 738 loci, together with 344 foxtail millet accessions with leaf phenotype data provided by the Zhangjiakou Academy of Agricultural Sciences (Wang et al., 2024). Phylogenetic analysis revealed that the 352 millet accessions (344 GWAS accessions and 8 newly sequenced millets) could be divided into three subgroups. Subgroup 1 included DXH01, DXH02, and DXH03, which are elite varieties used for millet breeding in northern China; subgroup 2 included DXH06, DXH07, and DXH08, which are elite varieties used for millet breeding in southern China; and subgroup 3 included the Indian varieties DXH04 and DXH05, which are also used as resources for improving millet lines in China (Figure 4A).

To increase variant density in the GWAS population, genotype imputation was performed using SNPs, InDels, and PAVs identified from the eight high-quality genome assemblies, increasing the variant set from 96 000 to 1 508 660 sites. Imputation was conducted using Beagle (v5.4) (Browning et al., 2007), which infers unobserved genotypes based on the LD structure of known markers. We used the SNPs, InDels, and PAVs from the eight reference genomes as the reference panel, and the 96K SNP array data from the 344 accessions were used as the target dataset. Beagle automatically extracted overlapping SNP loci as anchor markers and inferred missing genotypes through haplotype-based modeling using default parameters.

To assess imputation performance, we conducted benchmarking tests using both SNPs and PAVs. For SNPs, accuracy was evaluated under varying reference panel sizes by randomly sampling subsets of the 344 accessions and masking 95% of the genotype data (a 1:20 imputation ratio). A reference panel of eight samples achieved > 85% accuracy, with only marginal improvement when using larger panels (Supplemental Figure 8A). Accuracy was also assessed at different proportions of missing genotypes using 17 447 common SNPs; even when 95% of genotypes were masked, accuracy remained above 83% (Supplemental Figure 8B), demonstrating robust imputation under sparse input when using genetically diverse reference genomes. For PAV inference, we integrated 110 published accessions, the eight *de novo* assemblies, and the Yugu1 reference genome, followed by SNP and structural variant calling. Using chromosome 7 as a representative region (88 752 loci after downsampling to match SNP-based evaluations), we performed two complementary tests: (1) masking 95% of loci and first sequentially increasing the number of reference samples from one DXH line (DXH01) to all eight (DXH01–DXH08), then

(E) Density (number per Mb) of genes, insertions, and deletions in the eight newly sequenced millet genomes compared with the *S. italica* (Yugu1) reference genome.
(F) Overlap ratios of the top 10 TE classes associated with SVs.
(G) Proportions of the top 10 TE classes associated with SVs.

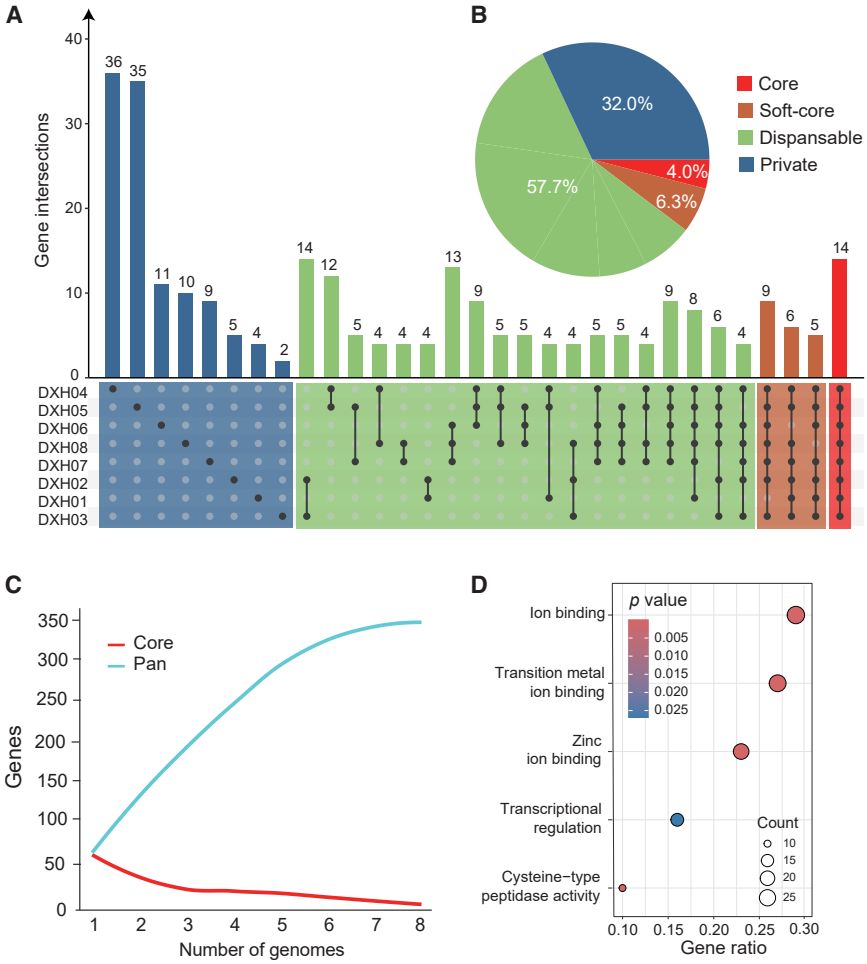


Figure 3. Analysis of genes that are completely present or absent in the eight accessions.

(A and B) (A) Numbers and **(B)** proportions of core, soft-core, dispensable, and private genes that are completely present or absent in the eight genomes. The UpSet plot illustrates the distribution of present and absent genes shared among the eight genomes. The *x*-axis represents different combinations of genomes sharing specific present or absent genes, indicated by filled circles and connecting lines. The *y*-axis denotes the number of present and absent genes in each corresponding genome combination.

(C) Variations in the numbers of completely present and absent genes across the eight genomes.

(D) GO enrichment analysis of completely present and absent genes in the eight genomes.

increasing it to 16 by adding eight more randomly selected accessions; and (2) masking different proportions of loci and imputing their genotypes using the eight DXH lines as references. For both approaches, PAV imputation accuracy closely matched the SNP-based results, confirming the reliability and robustness of our inference strategy for structural variants (Figure S8C and S8D).

We then performed GWASs for three pigmentation-related traits—leaf sheath color, leaf pulvinus color, and leaf color—using both the original 96K SNP array and the imputed 1.5M-variant dataset (Supplemental Figure 9). Although the imputed dataset identified more association signals, the strongest associations for all three traits consistently mapped to a ~100-kb region on chromosome 7 (26.84–26.94 Mb) (Figure 4B and Supplemental Figure 10). We then used haplotype data for 915 SNPs, 67 InDels, and 28 PAVs from this region to perform hierarchical clustering of all accessions, then statistically analyzed the phenotypes within each cluster. The 344 accessions were clearly separated into three major haplotype clusters (Figure 4C), each with different distributions of leaf-color phenotypes (Figure 4D). In cluster 1, 93.7% of accessions had green leaves, 68.7% had light purple leaf sheaths, and 59.3% had purple pulvini. In cluster 2, 88.3% had green leaves, 97.7% had green leaf sheaths, and 95.3% had green pulvini. In cluster 3, 74.5% had green leaves, and most leaf sheaths and pulvini were light purple or purple. These results demonstrate a close relationship between haplotypes in the 26.84–26.94 Mb interval on chromosome 7 and the colors of leaves, leaf sheaths, and leaf pulvini.

Prioritization of variants using interpretable machine-learning models

Although GWAS can effectively localize trait-associated genomic intervals, it often provides only coarse resolution, making it difficult to identify causal variants within these regions. Prioritizing variant sites adds interpretability and helps guide downstream functional validation and genome editing. In this study, we used an interpretable machine-learning strategy to prioritize candidate variants within the chromosome 7 interval (26.84–26.94 Mb) associated with leaf-related pigmentation traits.

First, we identified 75 large-effect variants in the relevant interval, including 41 non-synonymous SNPs, six frameshift InDels, and 28 PAVs (Supplemental Table 11). These features were used to construct four machine-learning models to predict the three traits across the 344 accessions. Performance evaluation using receiver-operating characteristic (ROC) curves indicated that the LightGBM model performed best, achieving area under the curve (AUC) values of 0.95, 0.89, and 0.86 for the three leaf sheath colors (green, light purple, and deep purple); 0.90, 0.86, and 0.80 for the three leaf pulvinus colors; and 0.72, 0.63, and 0.88 for the three leaf colors (Figure 5A). The random forest and XGBoost models yielded comparable predictive performance, whereas the decision-tree model performed poorly (Supplemental Figure 11). Additional evaluation metrics confirmed that LightGBM consistently outperformed the other models across all phenotypes (Supplemental Table 12); therefore, this model was selected for subsequent interpretability analyses of feature importance and genotype–phenotype associations.

To further interpret the model and prioritize individual variants, we calculated Shapley values, which quantify the contribution of each

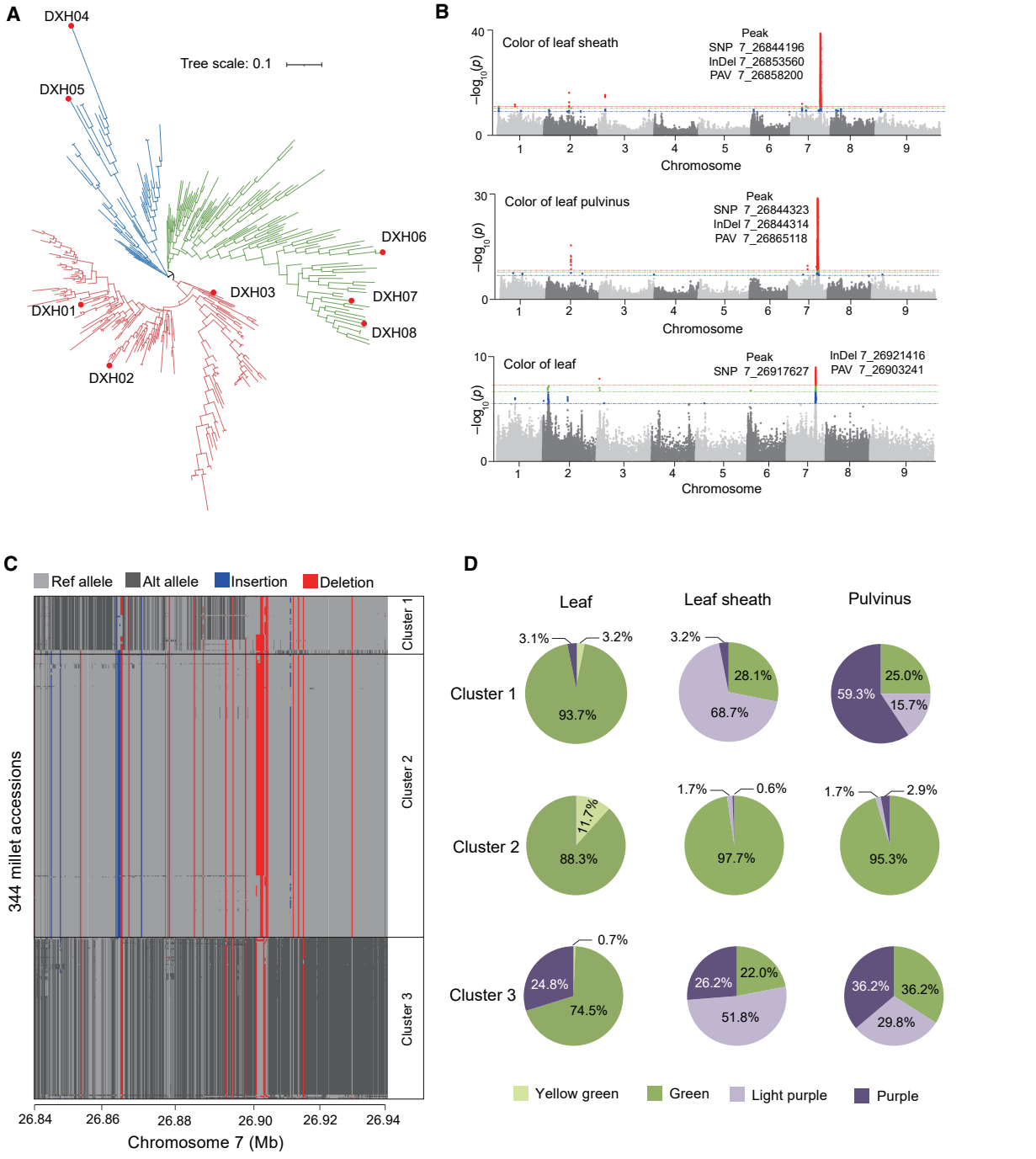


Figure 4. Mapping of trait-associated loci using the GWAS population.

(A) Phylogenetic tree of 352 millet accessions. Branches are colored red, green, and blue for subgroups 1, 2, and 3, respectively.

(B) GWAS of leaf, leaf sheath, and pulvinus color using the 1.5M genotype dataset.

(C) Clustering analysis of 344 accessions using haplotype data from the 26.84–26.94 Mb interval on chromosome 7.

(D) Proportions of accessions with green, light purple, and deep purple leaves, leaf sheaths, and leaf pulvini.

feature to the model's output (Lundberg et al., 2020). The distributions of Shapley values differed substantially across variant types: PAVs and frameshift InDels consistently had higher importance scores than non-synonymous SNPs (Supplemental Table 13), which suggests that SVs may play a more prominent role than non-synonymous SNPs in trait variation.

For the three leaf pulvinus colors, the features with the highest average Shapley values were all PAVs: Chr7_26868773, Chr7_26865118, and Chr7_26868773, which explained 37.93%, 19.79%, and 28.58% of the phenotypic variance, respectively. The top ten features collectively accounted for 81.17%, 70.06%, and 75.04% of the total phenotypic variation (Figure 5B). For the

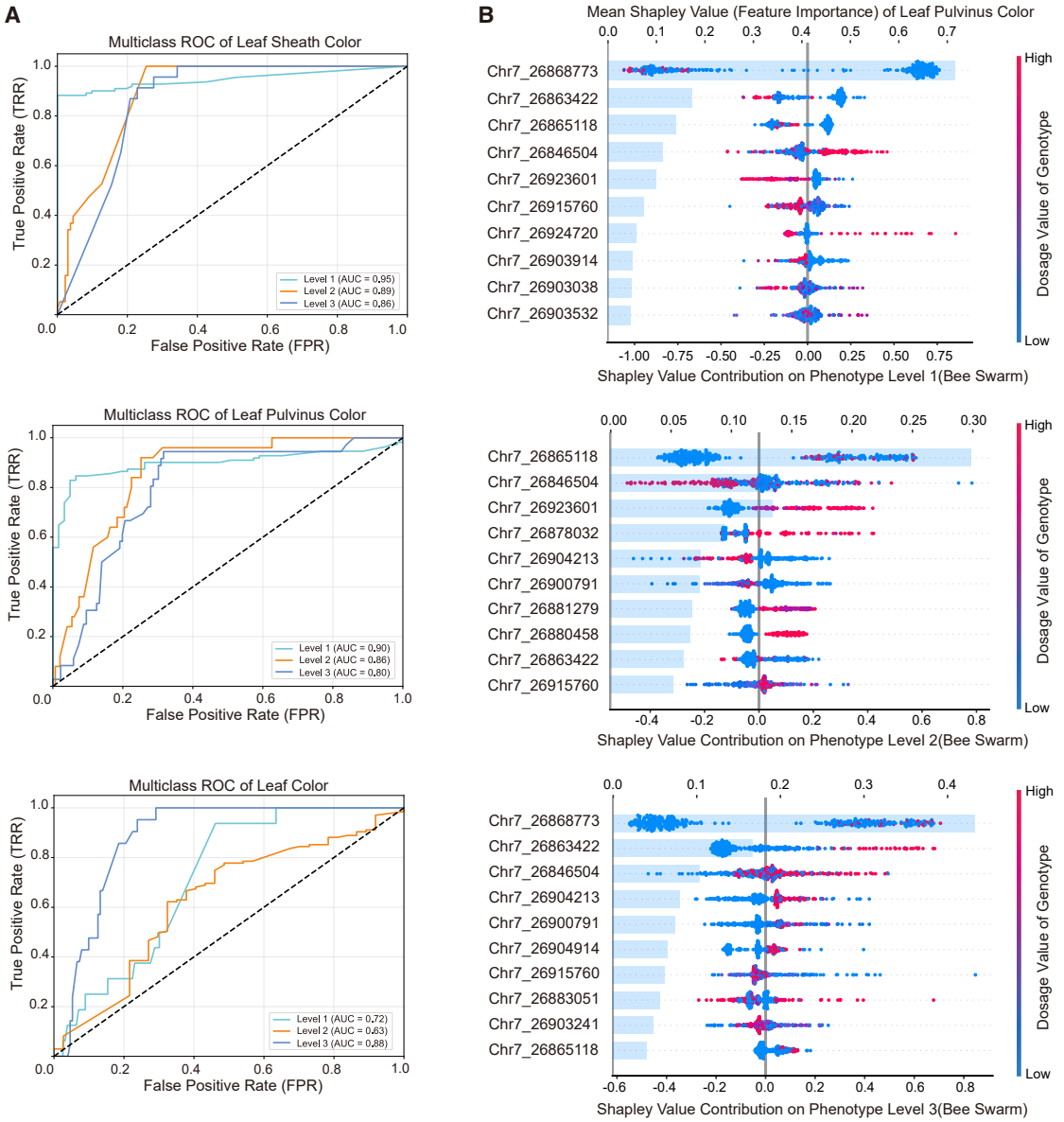


Figure 5. Prioritization of variants using the LightGBM model and Shapley values.
(A) Multiclass ROC curves for prediction of three traits using the LightGBM model.
(B) Bee-swarm plot of Shapley values with a bar plot of feature importance for the top 10 features related to the three leaf pulvinus colors.

three leaf sheath colors, two of the highest-ranking features were again Chr7_26868773 and Chr7_26865118; the third was an InDel, Chr7_26863422 (Supplemental Figure 12). For the three leaf colors, Chr7_26865118 ranked first for green leaves, Chr7_26868773 ranked second for light purple leaves, and Chr7_26863422 ranked second for purple leaves (Supplemental Figure 12).

Key PAVs that influence leaf-color traits in millet

The results of the machine-learning models indicated that the three variants Chr7_26868773, Chr7_26865118, and Chr7_26863422 had the most significant effects on the analyzed traits. We therefore performed further analyses of these variants. Examination of their genomic locations revealed that Chr7_26863422 is a 4-bp frameshift InDel located in the gene *Seita.7G194900*, whereas Chr7_26868773 and Chr7_26865118 are PAVs located within 5 kb upstream of *Seita.7G194900*. Although this gene has not been functionally characterized in millet, its rice homolog encodes a uridine diphosphate (UDP)-glucosyltransferase associated with anthocyanin and proanthocyanidin levels (Rao et al., 2019), consistent with a potential role in pigment accumulation. However, the function of this gene in millet and the effects of its variants require further experimental validation.

The two prioritized PAVs in the 26.84–26.94 Mb interval on chromosome 7, Chr7_26868773 and Chr7_26865118, were also located near the functionally validated gene *Seita.7G195400*. *Seita.7G195400* encodes the basic helix–loop–helix (bHLH) transcription factor *PPLS1*, a key regulator of purple coloration in the

leaf pulvinus and sheath (Bai et al., 2020). In a previous study, this gene was shown to be differentially expressed across materials with different leaf pulvinus and sheath colors (Bai et al., 2020); however, the reason for these differences was unclear, motivating us to further analyze this region. The *PPLS1* genomic region is rich in PAVs, which suggests that it may be a hotspot for TE activity. We annotated the TEs in this region across 10 genomes: the wild ancestor *Setaria viridis*, the *S. italica* reference line Yugu1, and the eight genomes assembled in this study (Figure 6A). We identified a *Harbinger* deletion 25 kb upstream of *PPLS1* in Yugu1, but it was absent from the other eight accessions and from the wild ancestor. DXH04 contained a unique *Gypsy* insertion in the longest intron of *PPLS1*, but this did not affect the purple phenotype of the leaf pulvinus and sheath. A 5002-bp *Copia* insertion was located 19–24 kb downstream of *PPLS1* in Yugu1; accessions DXH01, DXH02, DXH03, and DXH08, which have green leaf sheaths and pulvini, all contained this insertion. By contrast, the four accessions with purple leaf sheaths and pulvini (DXH04, DXH05, DXH06, and DXH07), as well as the wild ancestor, lacked this insertion. Importantly, the two key PAVs identified by our machine-learning models, Chr7_26868773 and Chr7_26865118, are located in the *Copia* insertion region. To further validate this TE insertion, we performed the same analysis in 12 additional high-quality genomes published by He et al. (2023), which confirmed its positional consistency (Supplemental Figure 13).

We next compared nucleotide diversity (π) in this region between accessions with green or purple leaf sheaths. Diversity was significantly lower in the green group (207 accessions) than in the purple group (137 accessions) (Figure 6B), which suggests that this region may have undergone selection during the domestication of green-sheath varieties. Because *S. viridis* and *S. italica* Yugu1 represent the wild ancestor and modern cultivated varieties, respectively, previous studies have proposed that their genomes reflect the ancestral and post-selection TE activity patterns in millet (Hu et al., 2018).

The *Copia insertion downstream of PPLS1 is* a notable shared feature of Yugu1 and the other three accessions with green leaf sheaths (DXH01, DXH03, and DXH08). DXH02, which displays a light purple sheath, carries both PAVs identified in this study. Unlike DXH01, DXH03, and DXH08, which are homozygous at the *PPLS1* locus, DXH02 shows markedly elevated heterozygosity in this region (Figure 6C and 6D). This heterozygous state may contribute to its intermediate phenotype: although it possesses the downstream *Copia* insertion associated with green-sheathed accessions, the retention of dominant purple-associated alleles in heterozygous form may partially counteract the green-sheathed phenotype, resulting in a light purple sheath. This interpretation is supported by nucleotide diversity patterns (Figure 6B): green-sheathed accessions (DXH01, DXH03, and DXH08) exhibit reduced diversity at *PPLS1*, consistent with selection-driven fixation, whereas DXH02 retains higher diversity, reflecting the maintenance of heterozygosity. Notably, the four pan-genome accessions with purple sheaths (DXH04–DXH07) lack the downstream *Copia* insertion, implicating this structural variant as the primary determinant of sheath coloration. Collectively, these findings suggest that the *Copia* insertion downstream of *PPLS1*, rather than other local variants, functions as the key regulatory element mediating pigmentation via TE-driven *cis*-regulatory effects.

DISCUSSION

A high-quality pan-genome resource for elite millet lines

In this study, we constructed a high-quality pan-genome using eight representative elite breeding lines of foxtail millet, covering the three major genetic subgroups of the species. Although limited in sample size, this panel captures the core genomic diversity present in modern breeding programs. Phylogenetic analysis integrating 110 published genomes confirmed that our selected accessions are broadly representative, making them suitable for small-scale but functionally rich pan-genome analyses (Supplemental Figure 14). To further assess the representativeness of this resource, we integrated the eight *de novo* assemblies with the 110 high-quality genomes reported by He et al. (2023) and the reference genome Yugu1. Comparative analysis revealed that the eight assemblies contributed 55.91 Mb (4.67%) of novel sequences to the combined graph pan-genome (total length ≈1.20 Gb), reflecting previously unrepresented genomic diversity (Supplemental Table 14). Notably, expanding the dataset from eight to 118 genomes only modestly increased the number of PAVs identified in the focal region on chromosome 7 (26.84–26.94 Mb), from 33 to 44, indicating that fewer than 7% of the total samples captured 75% of the region's key PAVs (Supplemental Table 15). Collectively, these findings demonstrate that a small but strategically selected panel of high-quality assemblies can recover most functionally relevant PAVs while substantially reducing sequencing cost and effort.

Overcoming PAV imputation challenges through strategic design

Imputing PAVs from sparse genotype data and small reference panels remains a major challenge, especially in non-model crops. In this study, we used a haplotype-aware strategy that integrates high-quality assemblies using bi-allelic variant imputation. Using a ∼96K SNP array and only eight reference accessions, we achieved high imputation accuracy ($\geq$ 83%) even under 95% genotype masking. In foxtail millet, LD decays to $r^2 \approx 0.2$ at ∼200 kb, which indicates an overall long-range LD pattern but also a substantial fraction of genomic regions with low LD ($r^2 < 0.2$) (Figure S15A and S15B). The focal region on chromosome 7 showed a similarly heterogeneous LD landscape (Supplemental Figure 15C). Although low LD can reduce local imputation confidence, Beagle's haplotype-based algorithm outputs posterior genotype probabilities, which were directly incorporated into our machine-learning models. Because low-confidence genotypes exert limited influence on model training, the overall predictive performance remained stable. Collectively, these results demonstrate that our framework performs robustly across heterogeneous LD environments; future studies should investigate its applicability to populations with weaker LD structure.

Fine-scale variant prioritization: A practical step beyond gene-level analysis

Although gene-level prioritization remains a cornerstone of GWAS interpretation, our study demonstrates that variant-level prioritization offers greater precision and translational value,

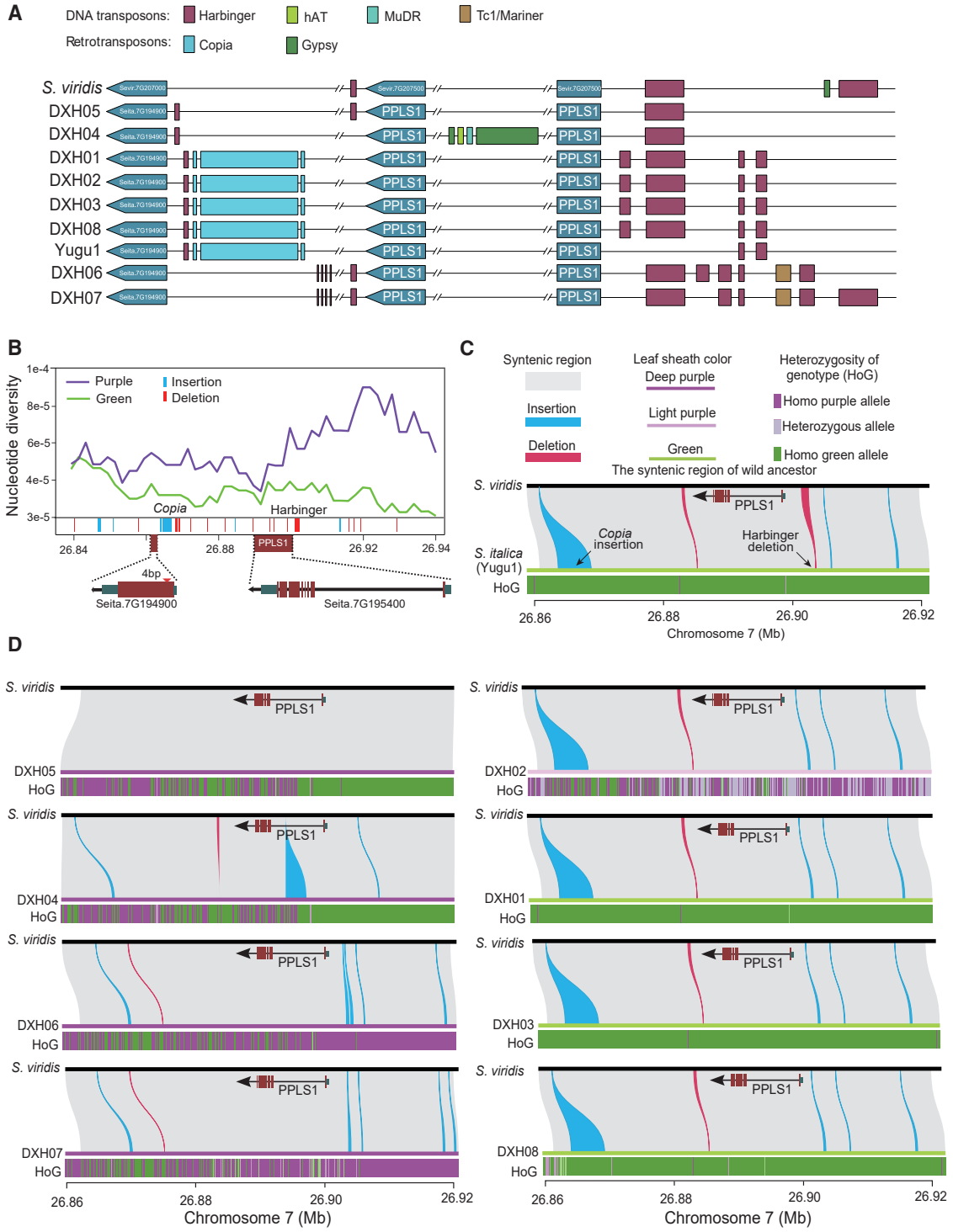


Figure 6. A *Copia* insertion is likely a key factor associated with color differences.

(A) Transposable elements identified in syntenic regions of *PPLS1* across ten genomes.

(B) Distribution of nucleotide diversity in the 26.84–26.94 Mb interval on chromosome 7 between two groups of accessions with different leaf sheath and pulvinus colors.

(C) Comparison of the syntenic region containing *PPLS1* between *S. italica* Yugu1 and the wild ancestor of foxtail millet (*S. viridis*, Me34V).

(D) Comparison of the syntenic region containing *PPLS1* among eight pan-genome accessions and the wild ancestor of foxtail millet (*S. viridis*, Me34V).

especially for genome editing and molecular breeding. Within a single gene, different variant types (e.g., missense SNPs, frameshift InDels, and regulatory PAVs) can have vastly different functional consequences. Our interpretable machine-learning framework effectively distinguishes large-effect variants from background noise, offering fine-scale insight into functional variation. Moreover, variants with high Shapley values may also serve as indirect indicators of nearby regulatory genes; this is especially valuable in incompletely annotated plant genomes.

Trait-associated PAVs revealed by integrated analysis

An integrated GWAS–machine-learning framework identified a ~100-kb locus on chromosome 7 strongly associated with pigmentation differentiation in foxtail millet. The top-ranked variants included two PAVs located within a 5002-bp *Copia* insertion ~20 kb downstream of *PPLS1* (*Seita.7G195400*) and a 4-bp InDel within *Seita.7G194900*. *PPLS1* encodes a bHLH transcription factor that regulates anthocyanin biosynthesis. Previous gene expression analyses reported differential *PPLS1* expression among foxtail millet accessions with different pigmentation patterns (Bai et al., 2020). Notably, three of those accessions—Huangdanzigu, Jinxiangyu, and SLX—were also included in our panel and displayed co-occurrence of the *Copia* insertion and elevated *PPLS1* expression. This pattern suggests a potential *cis*-regulatory role of the *Copia* element in modulating *PPLS1* activity. The differing nucleotide diversity observed between purple- and green-sheathed accessions in this genomic region further suggests that selection may have acted on the *PPLS1*-associated haplotype, likely reflecting the historical use of sheath color as a marker trait in breeding programs. *Seita.7G194900* is annotated as a putative UDP-glucosyltransferase based on sequence homology. Although its biochemical function remains unverified, its predicted enzymatic activity suggests a possible role in anthocyanin modification or related metabolic pathways. Functional validation through expression analysis and targeted mutagenesis will be essential for elucidating its contribution.

TE-mediated *cis* regulation as a driver of trait differentiation

TEs can modulate transcription through enhancer-like activity, chromatin looping, and local epigenetic remodeling, even when located tens of kilobases from their targets (Morgante et al., 2007). The *Copia* insertion located approximately 20 kb downstream of *PPLS1* likely functions as a distal *cis*-regulatory element that influences pigment biosynthesis. Although not immediately adjacent to the gene, it falls within the typical range at which long-range enhancers operate in plant genomes. These findings are consistent with previous studies showing that TEs can reshape transcriptional networks and enhance transcript stability when inserted near 3′ UTRs (Niu et al., 2019), and when inserted near functional genes, can directly influence phenotypic traits through *cis*-regulatory mechanisms (Wang et al., 2023). We propose that this insertion functions as a regulatory switch that enhances anthocyanin accumulation, highlighting the dual role of TEs as both sources of genetic diversity and evolutionary drivers of phenotypic change. Future integrative analyses combining chromatin accessibility, histone modification, and DNA methylation profiling will be essential for dissecting the underlying regulatory architecture at this locus.

Implications and future directions

The methodological pipeline developed here—integrating pan-genome construction, imputation, GWASs, and interpretable machine learning—offers a scalable and cost-effective strategy for functional variant discovery. Although demonstrated in foxtail millet, this framework can be extended to other crops and traits, particularly those for which whole-genome sequencing remains impractical. Whereas He et al. (2023) focused on high-quality genome assembly and variant cataloging, our study integrates pan-genomic data with machine learning and GWAS analyses to develop a scalable, cost-effective approach for identifying functionally important SVs in crop populations. Using only eight representative accessions, we captured most major PAV events; the resulting resource can be directly used with existing population datasets for large-scale genotyping. When compared with the standard resequencing of 344 accessions, the framework achieved an approximately 40-fold reduction in sequencing cost and data volume while maintaining high detection accuracy. This cost-efficiency greatly enhances its feasibility for population-scale functional genomics in resource-limited crop research systems. Future research aimed at improving model generalizability, broadening reference panels, and incorporating experimental validation will enable the framework to address more complex traits, such as those characterized by subtle associations or weak LD structure.

In conclusion, this study provides a comprehensive framework for identifying trait-associated PAVs, establishes a regulatory role in leaf pigmentation for a *Copia* insertion near *PPLS1*, and offers a robust, low-cost strategy for functional genomics in under-resourced crop research systems.

METHODS

Sample collection, DNA extraction, library construction, and sequencing

All eight *S. italica* accessions sequenced in this study were obtained from the Zhangjiakou Academy of Agricultural Sciences, Hebei Province, China. Genomic DNA was extracted from fresh leaves of each accession using a modified cetyl trimethylammonium bromide (CTAB) protocol (Murray and Thompson, 1980). After assessment of DNA quality, 15-kb libraries were constructed using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, CA, USA). The library construction procedure included DNA shearing, AMPure PB bead purification, removal of single-stranded DNA overhangs, damage repair, end repair, hairpin adapter ligation, and bead purification. After quality control, the resulting SMRTbell libraries were sequenced using an 8M SMRT Cell on the PacBio Sequel II platform (Pacific Biosciences).

Hi-C libraries were also prepared for the eight accessions. In brief, nuclear chromatin from plant tissues was fixed with formaldehyde and extracted. Fixed chromatin was digested with DpnII, then the resulting sticky ends were filled in with biotinylated nucleotides and ligated. Crosslinks were then reversed, and purified DNA was treated to remove any free biotin from the ligated fragments. The DNA was sheared to ~350 bp, and biotinylated fragments were enriched using streptavidin bead pull-down. Finally, PCR amplification was performed to generate the libraries. After library construction, Qubit 2.0 and Agilent 2100 instruments were used to assess concentration and insert size, and high-throughput sequencing was performed on the Illumina NovaSeq 6000 platform to obtain 150-bp paired-end reads.

De novo genome assembly and genome quality assessment

Long reads were obtained from PCR-free SMRTbell DNA libraries sequenced on the PacBio Sequel platform in CCS mode. PacBio CCS reads were first generated using the PacBio CCS software with default parameters, then draft phased contigs were constructed using Hifiasm (v0.13-r308) (Cheng et al., 2021). To obtain chromosome-level genome assemblies for each millet sample, we scaffolded the assembled draft genomes using Hi-C data. The data-processing procedure was as follows. (1) Paired-end Hi-C reads were mapped to the draft genome assembly using HiC-Pro (v2.11.1) (Servant et al., 2015); default parameters were used to filter the raw Hi-C reads, and self-ligated, non-ligated, and other invalid reads (such as PCR artifacts or pairs that failed size-filtering criteria) were discarded. (2) Juicer (v1.6.2) (Durand et al., 2016a, 2016b) and 3D-DNA (Dudchenko et al., 2017) were used to cluster the contigs into putative chromosomal groups. (3) Juicebox (v1.11.8) (Durand et al., 2016a, 2016b), combined with manual inspection, was used to confirm contig orientations and remove ambiguous fragments. (4) Purge Haplotigs (v1.0) (Roach et al., 2018) was used to remove redundant sequences.

The quality of each genome assembly was evaluated using a comprehensive set of approaches. Assembly completeness was assessed using 1614 BUSCOs from the Embryophyta_odb10 database (Simão et al., 2015). Genome consistency was evaluated by aligning PacBio sequencing reads to the assembled genomes using Minimap2 (v2.21-r1071) (Li et al., 2018), and alignment rates were calculated using SAMtools (v1.13) (Danecek et al., 2021). To further evaluate assembly consensus and completeness, a *k*-mer–based approach was applied using the Merqury program (v1.1) (Rhie et al., 2020). In addition, the abundance and integrity of long terminal repeat retrotransposons (LTRs) were analyzed using LTR_finder (v1.0) (Valencia et al., 2019) and LTR_retriever (v2.0) (Ou et al., 2018), generating LAI values as an additional measure of genome assembly quality.

Prediction and annotation of protein-coding genes, TEs, and non-coding RNAs

Protein-coding genes were identified using a combination of homology-based, *de novo*, and transcriptome-based prediction methods. Proteins from six plant genomes (*S. italica*, *S. bicolor*, *Z. mays*, *O. sativa*, *B. distachyon*, and *A. thaliana*) were downloaded from the Phytozome and NCBI databases and aligned to each new millet assembly using GenBlastA (v1.0.4) (She et al., 2009). GeneWise (v2.4.1) (Birney et al., 2004) was then used to predict the exact gene structures for the genomic regions corresponding to each GenBlastA hit. Three *ab initio* gene-prediction programs—Augustus (v3.2.1) (Hoff et al., 2019), GlimmerHMM (v3.0.4) (Majoros et al., 2004), and SNAP (v1.0) (Korf et al., 2004)—were used to predict coding regions in the repeat-masked genome. RNA sequencing data from Zhang et al. (2012) were mapped to the assembly using HISAT2 (v2.0.1) (Kim et al., 2019), then StringTie (v1.2.2) (Pertea et al., 2015) and TransDecoder (v3.0.1) (Grabherr et al., 2011) were used to assemble the transcripts and identify candidate coding regions. All gene models predicted by these three approaches were integrated using EVidenceModeler (v1.0) (Haas et al., 2008) to generate a non-redundant set of gene models. Protein-coding genes were functionally annotated by performing BLASTP searches (E value 1×10^{-5}) against two integrated protein sequence databases, SwissProt and TrEMBL. InterProScan (v5.3) (Jones et al., 2014) was used to annotate protein domains and assign GO terms to each gene, and KEGG pathways were assigned to genes using BLAST searches (E value 1×10^{-5}) against the KEGG database (release 84.0) (Kanehisa et al., 2017).

TEs were identified in the eight millet assemblies using both *de novo* prediction of repetitive sequences and database searches for homologous TEs. RepeatModeler (v1.0.8) (Flynn et al., 2020) and LTR_FINDER (v1.0.6) (Xu et al., 2007) were used for *de novo* prediction of repetitive sequences in each assembly, and RepeatMasker (v4.0.7) (Tarailo-Graovac et al., 2009) was then used to search the millet assemblies against the combined TE database. For the homology-based approach, RepeatMasker and the Repbase database (v21) (Bao et al., 2015) were used to identify TEs at both the DNA and protein levels. RepeatMasker was used for DNA-level identification, and RepeatProteinMask was used for protein-level identification (Tarailo-Graovac et al., 2009).

tRNA genes were predicted using tRNAscan-SE (v1.3.1) (Chan et al., 2019). rRNA genes were predicted by aligning the millet sequences to rRNA sequences from *Arabidopsis* and rice using BLASTN (E value 1×10^{-5}). miRNA and snRNA genes were identified by searching against the Rfam database (release 12.0) using INFERNAL (v1.1.1) (Nawrocki et al., 2013).

Pan-genome construction

The graph-based millet pan-genome was constructed using the Minigraph toolkit (v0.20-r550) (Li et al., 2020) with the DXH01 assembly designated as the reference genome. Assemblies DXH02 through DXH08 were then sequentially added to the reference as nodes. For each cycle of assembly addition, nodes that could be aligned to the pan-genome graph were colored, and unmapped sequences were designated as non-reference nodes and incorporated into the pan-genome. Nodes traversed by all eight assemblies were classified as core genome segments. The final pan-genome output in rGFA format was converted to FASTA using gfatools (https://github.com/lh3/gfatools). The nodes aligned to the pan-genome graph from each assembly were colored separately to determine the support for each node, as described in Crysnanto et al. (2021). All nodes absent from the DXH01 reference genome were classified as non-reference nodes. For core- and pan-gene analysis, we sequentially clustered genes from the eight assemblies using OrthoFinder (v2.0) (Emms et al., 2019) with the DIAMOND alignment method, using an E value of 1×10^{-5} and an inflation factor of 1.

To evaluate the representativeness of our panel and expand the pan-genome framework, we integrated the 110 high-quality foxtail millet genomes published by He et al. (2023) with the eight *de novo* assemblies generated in this study and the reference genome Yugu1. The workflow is summarized as follows. (1) SVs (including insertions, deletions, translocations, and inversions) were identified among the 118 genomes using SyRI (v1.2) after pairwise alignment to Yugu1 with Minimap2 (v2.21-r1071). (2) SVs $\geq$ 50 bp were retained, and insertions and deletions were also classified as PAVs. (3) Large PAVs and inversions were incorporated into the Yugu1 reference using Minigraph (v0.20-r550) to construct a graph-based pan-genome.

Phylogenetic analysis of gene-family expansion and contraction

We used OrthoMCL (v1.0) (Li et al., 2003) to cluster orthologous gene groups from the proteomes of the eight newly assembled millet accessions (DXH01–DXH08), the *S. italica* reference genome Yugu1, *S. bicolor*, *Z. mays*, *O. sativa*, *B. distachyon*, and *A. thaliana*. Prior to clustering, protein sequences shorter than 50 amino acids or containing internal stop codons were removed to ensure data quality. An all-versus-all BLASTP search was performed using an E-value cutoff of 1×10^{-5}, and the resulting similarity scores were processed using the OrthoMCL pipeline with the Markov cluster (MCL) algorithm (inflation parameter = 1.5). All other parameters were left at their default settings. The resulting 3979 single-copy gene families were used to construct a phylogenetic tree for the nine millet accessions and five model plant species. Four-fold degenerate sites were extracted from each gene family and concatenated to generate one supergene per species. A phylogenetic tree was then constructed using the GTR+gamma substitution model in MrBayes (v3.1.2) (Ronquist et al., 2003). Divergence times among taxa were estimated using MCMCtree (v4.4) from the PAML package (Yang et al., 2007), with an independent rates clock and the JC69 nucleotide substitution model. Calibration times for the divergences between *A. thaliana* and *O. sativa* (~160 MYA), *Z. mays* and *O. sativa* (~47 MYA), and *Z. mays* and *S. bicolor*

(~10.8 MYA) were obtained from the TimeTree database (Kumar et al., 2022). Changes in gene-family size along the phylogenetic tree were analyzed using CAFE (v2.1) (De Bie et al., 2006).

Detection of PSGs

The branch-site model in the PAML package (Yang et al., 2007) was used to detect PSGs in the millet genomes, with all nine millets designated as the foreground branch and *S. bicolor*, *Z. mays*, *O. sativa*, *B. distachyon*, and *A. thaliana* designated as background branches. The null model used for the branch-site test assumed that the K_a/K_s values for all codons across all branches were ≤ 1, whereas the alternative model allowed codons in the foreground branch to evolve with $K_{a/K_s} > 1$. A maximum-likelihood ratio test was used to compare the two models, and p values were calculated using a chi-squared distribution with 1 degree of freedom (df = 1). The resulting p values were corrected for multiple testing using the false discovery rate (FDR) method, with FDR < 0.05 considered indicative of positive selection. We also required at least one amino acid site to have a high probability of being under positive selection (Bayes probability > 95%). If none of the amino acid sites in a PSG met this cutoff, the gene was classified as a false positive and excluded. GO enrichment analysis of the PSGs was performed using GOATOOLS (v0.7.2) (Klopfenstein et al., 2018).

Detection of variations

We aligned the *de novo* assemblies of the eight pan-genome accessions to the *S. italica* v2.2 reference genome using Minimap2 (v2.21-r1071) (Li et al., 2018). We then used SyRI (v1.2) (Goel et al., 2019) for SV detection and selected deletions, insertions, inversions, translocations, and duplications larger than 50 bp. We used PanPop (Zheng et al., 2024) to segment and merge SVs located near the same positions across the eight accessions, thereby generating an integrated pan-genome SV set. During this process, deletions and insertions were classified as PAVs. SNPs and InDels were called using the GATK Best Practices pipeline (Van der Auwera et al., 2013). Variants with a minor allele frequency (MAF) < 0.05 or a missing rate > 0.1 were filtered out using PLINK (v1.9) (Purcell et al., 2007) before population genetic analyses.

GWAS and population genetic analysis

We used 96K SNP array data and seedling-stage phenotypic data for three leaf traits from 344 foxtail millet samples provided by the Zhangjiakou Academy of Agricultural Sciences (Wang et al., 2024). Details on plant growth conditions, sample extraction, and genotyping methods are found in Wang et al. (2024). We imputed genotype data for the 344 GWAS samples using Beagle (v5.4) (Browning et al., 2021), based on SNP, InDel, and PAV data obtained from the pan-genome analysis, generating dosage values. We constructed a maximum-likelihood phylogenetic tree for all samples using MEGA7 (v7.0) (Kumar et al., 2016). LD was calculated using PLINK (v1.9) (Purcell et al., 2007), and genomic regions were classified as high LD or low LD based on three criteria: (1) LD decay patterns ($r^2 > 0.2$ for high LD; $r^2 < 0.2$ for low LD); (2) genome-wide r^2 distributions in 50-kb windows relative to the median value (0.219); and (3) LD heatmaps of a representative interval on chromosome 7 (26.84–26.94 Mb), in which red and yellow/white blocks indicate strong and weak LD, respectively. This classification enabled assessment of LD heterogeneity and its potential effects on imputation accuracy and downstream trait–SV prediction. GWAS was performed on both the 96K chip genotypes and the imputed genotypes of the 344 samples using a mixed linear model in GEMMA (v0.1) (Zhou et al., 2012), identifying significant SNP, InDel, and PAV loci. From these results, we identified haplotypes in the 26.84–26.94 Mb interval on chromosome 7, calculated a Euclidean distance matrix, and performed clustering using the "hclust" function in R. Variants in this interval were annotated using ANNOVAR (Wang et al., 2010), focusing on large-effect variants such as missense SNPs, nonsense SNPs, start- and stop-codon SNPs, splice-site SNPs, frameshift InDels, and PAVs. In addition, nucleotide diversity (π) in this interval was calculated using VCFtools (v0.1.6) (Danecek et al., 2011) with a 30-kb window size and a 2-kb step size.

Construction of interpretable machine-learning models

To construct interpretable models, we used dosage values of large-effect variant genotypes from the 344 samples as features. The phenotypic values of the three leaf-color traits served as labels to train four machine-learning models: decision tree, random forest, XGBoost, and LightGBM. To ensure consistency, the same random seed was used during data splitting, with identical training (70%) and test (30%) sets across all models. Each model underwent a rigorous grid search using the training data to identify the optimal parameters. The models were then trained with these parameters, and their performance on the test set was evaluated based on accuracy and recall. To comprehensively assess model performance, we also calculated precision, F1 score, macro average, weighted average, overall accuracy, and AUC. These metrics evaluated class-specific prediction accuracy, sensitivity to true positives, the balance between precision and recall, and overall performance under class imbalance. We also plotted ROC curves to assess the discriminative ability of each classifier under varying thresholds. Shapley values, derived from game theory, were used to interpret the machine-learning models by quantifying the contribution of each feature to the model predictions (Lundberg et al., 2020). Finally, the optimal model, LightGBM, was further analyzed using the SHAP package in Python. This analysis calculated Shapley values for each trait at every level, revealing the importance of individual features in the model and illustrating the phenotypic contribution of each variant.

DATA AND CODE AVAILABILITY

The genome assemblies and annotations of the eight foxtail millet accessions have been deposited in the NCBI GenBank database under accession number PRJNA1139385 (https://www.ncbi.nlm.nih.gov/). The integrated pan-genome in GFA format and the associated SV datasets are publicly available on Zenodo (https://zenodo.org/records/17345803). All code related to this study is publicly available on GitHub at https://github.com/Cauwth/CodeForMillet.

FUNDING

This work was supported by the National Key Research and Development Program of China (2023YFF1000100), the Hebei Provincial Science and Technology Plan Modern Breeding Industry Science and Technology Innovation Special Project (21326316D and 21326302D), the Guiding Special Fund for Central Universities to Build World-Class Universities and Promote Characteristic Development (2025AC030), and the Pinduoduo–China Agricultural University Research Fund (PC2024A01003).

ACKNOWLEDGMENTS

No conflict of interest declared.

AUTHOR CONTRIBUTIONS

X.W. and J.Y. conceived and designed the study; Z.Z., G.F., and S.Z. collected the accessions; J.Z., Y.X., and W.Z. performed the experiments; T.W., W.W., S.L., S.J., Q.C., and M.S. analyzed the data; W.W. and T.W. wrote the manuscript; and J.Y., Z.Z., and X.W. revised the manuscript. All authors reviewed and approved the final version of the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: June 20, 2025
Revised: October 17, 2025
Accepted: November 27, 2025
Published: November 29, 2025

REFERENCES

Azodi, C.B., Tang, J., and Shiu, S.H. (2020). Opening the black box: interpretable machine learning for geneticists. Trends Genet. **36**:442–455. https://doi.org/10.1016/j.tig.2020.03.005.

Bai, H., Song, Z., Zhang, Y., Li, Z., Wang, Y., Liu, X., Ma, J., Quan, J., Wu, X., Liu, M., et al. (2020). The bHLH transcription factor PPLS1 regulates the color of pulvinus and leaf sheath in foxtail millet (*Setaria italica*). Theor. Appl. Genet. **133**:1911–1926. https://doi.org/10.1007/s00122-020-03566-4.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA **6**:11. https://doi.org/10.1186/s13100-015-0041-9.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. **14**:988–995. https://doi.org/10.1101/gr.1865504.

Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. **108**:1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. **81**:1084–1097. https://doi.org/10.1086/521987.

Chan, P.P., and Lowe, T.M. (2019). tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol. Biol. **1962**:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods **18**:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Crysnanto, D., Leonard, A.S., Fang, Z.H., and Pausch, H. (2021). Novel functional sequences uncovered through a bovine multiassembly graph. Proc. Natl. Acad. Sci. USA **118**:e2101056118. https://doi.org/10.1073/pnas.2101056118.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience **10**:giab008. https://doi.org/10.1093/gigascience/giab008.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics **22**:1269–1271. https://doi.org/10.1093/bioinformatics/btl097.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science **356**:92–95. https://doi.org/10.1126/science.aal3327.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. **3**:99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. **3**:95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238. https://doi.org/10.1186/s13059-019-1832-y.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA **117**:9451–9457. https://doi.org/10.1073/pnas.1921046117.

Goel, M., Sun, H., Jiao, W.B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. **20**:277. https://doi.org/10.1186/s13059-019-1911-0.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. **29**:644–652. https://doi.org/10.1038/nbt.1883.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. **9**:R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Hassan, Z.M., Sebola, N.A., and Mabelebele, M. (2021). The nutritional use of millet grain for food and feed: a review. Agric. Food Secur. **10**:16. https://doi.org/10.1186/s40066-020-00282-6.

He, Q., Tang, S., Zhi, H., Chen, J., Zhang, J., Liang, H., Alam, O., Li, H., Zhang, H., Xing, L., et al. (2023). A graph-based genome and pan-genome variation of the model plant Setaria. Nat. Genet. **55**:1232–1242. https://doi.org/10.1038/s41588-023-01423-w.

He, Q., Wang, C., He, Q., Zhang, J., Liang, H., Lu, Z., Xie, K., Tang, S., Zhou, Y., Liu, B., et al. (2024). A complete reference genome assembly for foxtail millet and Setaria-db, a comprehensive database for Setaria. Mol Plant **17**:219–222. https://doi.org/10.1016/j.molp.2023.12.017.

Hoff, K.J., and Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. Curr. Protoc. Bioinformatics **65**:e57. https://doi.org/10.1002/cpbi.57.

Hu, H., Mauro-Herrera, M., and Doust, A.N. (2018). Domestication and improvement in the model C4 grass, Setaria. Front. Plant Sci. **9**:719. https://doi.org/10.3389/fpls.2018.00719.

Jiao, C., Xie, X., Hao, C., Chen, L., Xie, Y., Garg, V., Zhao, L., Wang, Z., Zhang, Y., Li, T., et al. (2025). Pan-genome bridges wheat structural variations with habitat and breeding. Nature **637**:384–393. https://doi.org/10.1038/s41586-024-08277-0.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics **30**:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. **45**:D353–D361. https://doi.org/10.1093/nar/gkw1092.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**:907–915. https://doi.org/10.1038/s41587-019-0201-4.

Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., et al. (2018). GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep. **8**:10872. https://doi.org/10.1038/s41598-018-28948-z.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinf. **5**:59. https://doi.org/10.1186/1471-2105-5-59.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. **33**:1870–1874. https://doi.org/10.1093/molbev/msw054.

Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., and Hedges, S.B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. Mol. Biol. Evol. **39**:msac174. https://doi.org/10.1093/molbev/msac174.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H., Feng, X., and Chu, C. (2020). The design and construction of reference pangenome graphs with minigraph. Genome Biol. **21**:265. https://doi.org/10.1186/s13059-020-02168-z.

Li, J., Yuan, D., Wang, P., Wang, Q., Sun, M., Liu, Z., Si, H., Xu, Z., Ma, Y., Zhang, B., et al. (2021). Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. Genome Biol. **22**:119. https://doi.org/10.1186/s13059-021-02351-w.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189. https://doi.org/10.1101/gr.1224503.

Li, W., Liu, J., Zhang, H., Liu, Z., Wang, Y., Xing, L., He, Q., and Du, H. (2022). Plant pan-genomics: recent advances, new challenges, and roads ahead. Journal of Genetics and Genomics **49**:833–846. https://doi.org/10.1016/j.jgg.2022.06.004.

Liu, Z., Wang, N., Su, Y., Long, Q., Peng, Y., Shangguan, L., Zhang, F., Cao, S., Wang, X., Ge, M., et al. (2024). Grapevine pangenome facilitates trait genetics and genomic breeding. Nat. Genet. **56**:2804–2814. https://doi.org/10.1038/s41588-024-01967-5.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., et al. (2020). From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. **2**:56–67. https://doi.org/10.1038/s42256-019-0138-9.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics **20**:2878–2879. https://doi.org/10.1093/bioinformatics/bth315.

Morgante, M., De Paoli, E., and Radovic, S. (2007). Transposable elements and the plant pan-genomes. Curr. Opin. Plant Biol. **10**:149–155. https://doi.org/10.1016/j.pbi.2007.02.001.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res **8**:4321–4325. https://doi.org/10.1093/nar/8.19.4321.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**:2933–2935. https://doi.org/10.1093/bioinformatics/btt509.

Niu, X.M., Xu, Y.C., Li, Z.W., Bian, Y.T., Hou, X.H., Chen, J.F., Zou, Y.P., Jiang, J., Wu, Q., Ge, S., et al. (2019). Transposable elements drive rapid phenotypic variation in Capsella rubella. Proc. Natl. Acad. Sci. USA **116**:6908–6913. https://doi.org/10.1073/pnas.1811498116.

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. **176**:1410–1422. https://doi.org/10.1104/pp.17.01310.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295. https://doi.org/10.1038/nbt.3122.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**:559–575. https://doi.org/10.1086/519795.

Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell **184**:3542–3558.e16. https://doi.org/10.1016/j.cell.2021.04.046.

Rao, M.J., Xu, Y., Huang, Y., Tang, X., Deng, X., and Xu, Q. (2019). Ectopic expression of citrus UDP-GLUCOSYL TRANSFERASE gene enhances anthocyanin and proanthocyanidins contents and confers high light tolerance in *Arabidopsis*. BMC Plant Biol. **19**:603. https://doi.org/10.1186/s12870-019-2212-1.

Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. **21**:245. https://doi.org/10.1186/s13059-020-02134-9.

Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinf. **19**:460. https://doi.org/10.1186/s12859-018-2485-7.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572–1574. https://doi.org/10.1093/bioinformatics/btg180.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. **16**:259. https://doi.org/10.1186/s13059-015-0831-x.

She, R., Chu, J.S.C., Wang, K., Pei, J., and Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. **19**:143–149. https://doi.org/10.1101/gr.082081.108.

Shrestha, N., Hu, H., Shrestha, K., and Doust, A.N. (2023). Pearl millet response to drought: A review. Front. Plant Sci. **14**:1059574. https://doi.org/10.3389/fpls.2023.1059574.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Srivastava, R.K., Yadav, O.P., Kaliamoorthy, S., Gupta, S.K., Serba, D.D., Choudhary, S., Govindaraj, M., Kholová, J., Murugesan, T., Satyavathi, C.T., et al. (2022). Breeding Drought-Tolerant Pearl Millet Using Conventional and Genomic Approaches: Achievements and Prospects. Front. Plant Sci. **13**:781524. https://doi.org/10.3389/fpls.2022.781524.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **25**:10.1-4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

Valencia, J.D., and Girgis, H.Z. (2019). LtrDetector: A tool-suite for detecting long terminal repeat retrotransposons de-novo. BMC Genom. **20**:450. https://doi.org/10.1186/s12864-019-5796-9.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics **43**:11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.

Wang, K., Hua, G., Li, J., Yang, Y., Zhang, C., Yang, L., Hu, X., Scheben, A., Wu, Y., Gong, P., et al. (2024). Duck pan-genome reveals two transposon insertions caused bodyweight enlarging and white plumage phenotype formation during evolution. iMeta **3**:e154. https://doi.org/10.1002/imt2.154.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. **38**:e164. https://doi.org/10.1093/nar/gkq603.

Wang, Y.W., Li, X.R., Fan, X.Q., Zhao, Z.H., Liu, Y.H., Fan, G.Y., and Wang, X.M. (2024). Genome-wide association analysis of maturity traits in *Setaria italica* using 96K SNP microarrays. Seed **43**:37–50. https://doi.org/10.16590/j.cnki.1001-4705.2024.04.037.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**:W265–W268. https://doi.org/10.1093/nar/gkm286.

Yan, J., and Wang, X. (2023). Machine learning bridges omics sciences and plant breeding. Trends Plant Sci. **28**:199–210. https://doi.org/10.1016/j.tplants.2022.08.018.

Yang, C., Yan, J., Jiang, S., Li, X., Min, H., Wang, X., and Hao, D. (2022). Resequencing 250 soybean accessions: new insights into genes associated with agronomic traits and genetic networks. Genom. Proteom. Bioinform. **20**:29–41. https://doi.org/10.1016/j.gpb.2021.02.009.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591. https://doi.org/10.1093/molbev/msm088.

Yuan, Y., Bayer, P.E., Batley, J., and Edwards, D. (2021). Current status of structural variation studies in plants. Plant Biotechnol. J. **19**:2153–2163. https://doi.org/10.1111/pbi.13646.

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z., Wang, W., et al. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat. Biotechnol. **30**:549–554. https://doi.org/10.1038/nbt.2195.

Zheng, Z., Zhu, M., Zhang, J., Liu, X., Hou, L., Liu, W., Yuan, S., Luo, C., Yao, X., Liu, J., and Yang, Y. (2024). A sequence-aware merger of genomic structural variations at population scale. Nat. Commun. **15**:960. https://doi.org/10.1038/s41467-024-45244-9.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **44**:821–824. https://doi.org/10.1038/ng.2310.

Plant Communications, Volume 7

Supplemental information

Integrating pan-genome analysis, GWAS, and interpretable machine learning to prioritize trait-associated structural variations in *Setaria italica*

Wenying Wang, Tianhao Wu, Guangyu Fan, Shuai Zhang, Songyu Liu, Shuqin Jiang, Qian Cheng, Meiqi Shang, Yanfen Xu, Wenlin Zhang, Jianan Zhang, Xiangfeng Wang, Zhihai Zhao, and Jun Yan

Supplemental Information for

Integrating Pan-genome, GWAS, and Interpretable Machine Learning to Prioritize Trait-Associated Structural Variations in *Setaria italica*

Wenying Wang[a,#], Tianhao Wu[a,#], Guangyu Fan[b,#], Shuai Zhang[b], Songyu Liu[a], Shuqin Jiang[c], Qian Cheng[a], Meiqi Shang[a], Yanfen Xu[c], Wenlin Zhang[c], Jianan Zhang[c], Xiangfeng Wang[a,*], Zhihai Zhao[b,*], and Jun Yan[a,*]

[a]State Key Laboratory of Maize Bio-breeding, National Maize Improvement Center, Frontiers Science Center for Molecular Design Breeding, College of Agronomy and Biotechnology, China Agricultural University, Beijing, 100094, China

[b]Institute of Millet, Zhangjiakou Academy of Agricultural Science, Zhangjiakou, Hebei Province, 075000, China

[c]Molbreeding Biotechnology Co., Ltd, Shijiazhuang, Hebei Province, 051430, China

[#]These authors contributed equally to the article.

[*]Corresponding authors

***Email addresses*:**

yanjun@cau.edu.cn (J. Yan)

zhaozhihai58@163.com (Z. Zhao)

xwang@cau.edu.cn (X. Wang)

Supplementary Figures

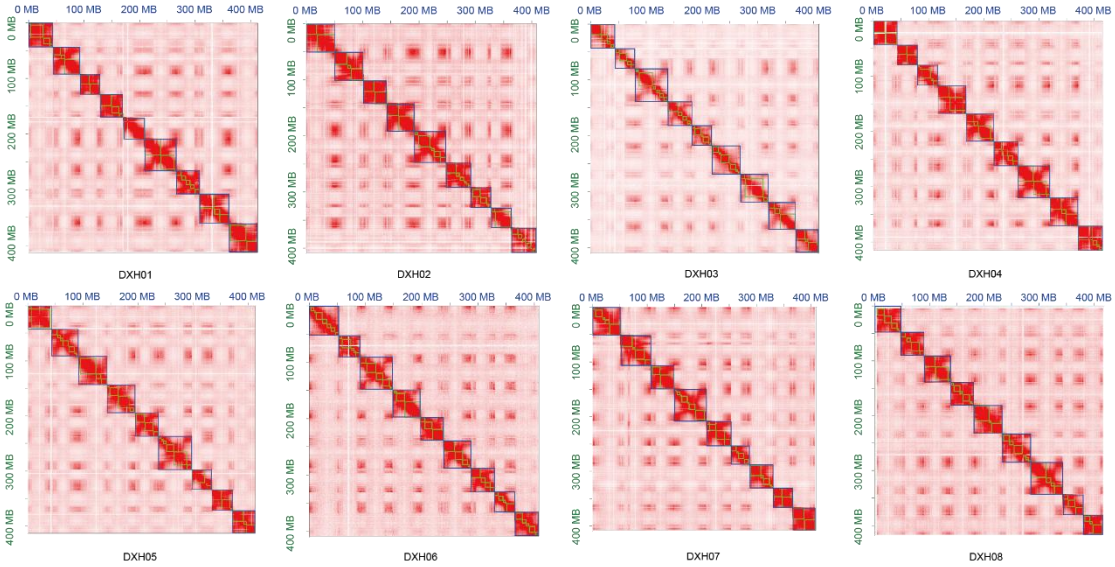


Figure S1. Hi-C contact heatmap and analysis of chromosomal interactions in eight millet genome assemblies, DXH01–DXH08.

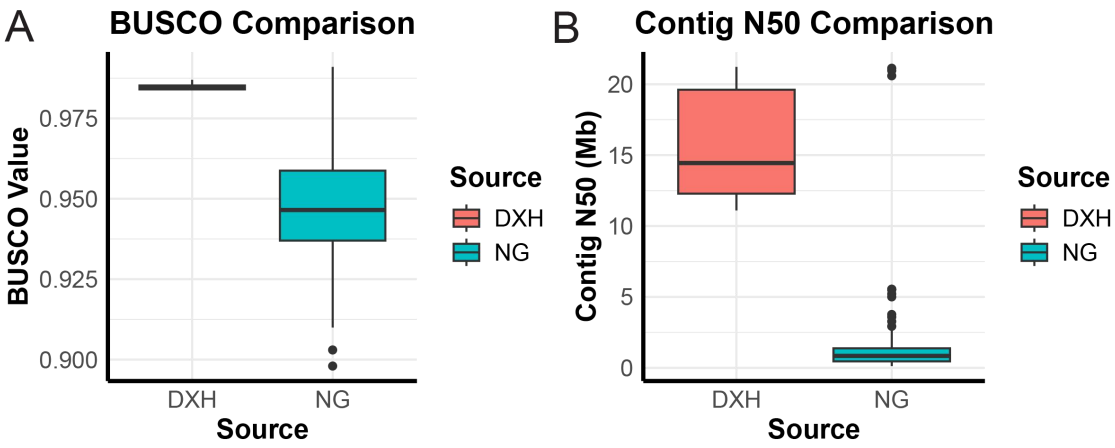


Figure S2. Comparison of pan-genome assembly metrics between this study and He *et al.* (2023): BUSCO (**A**) and Contig N50 (**B**).

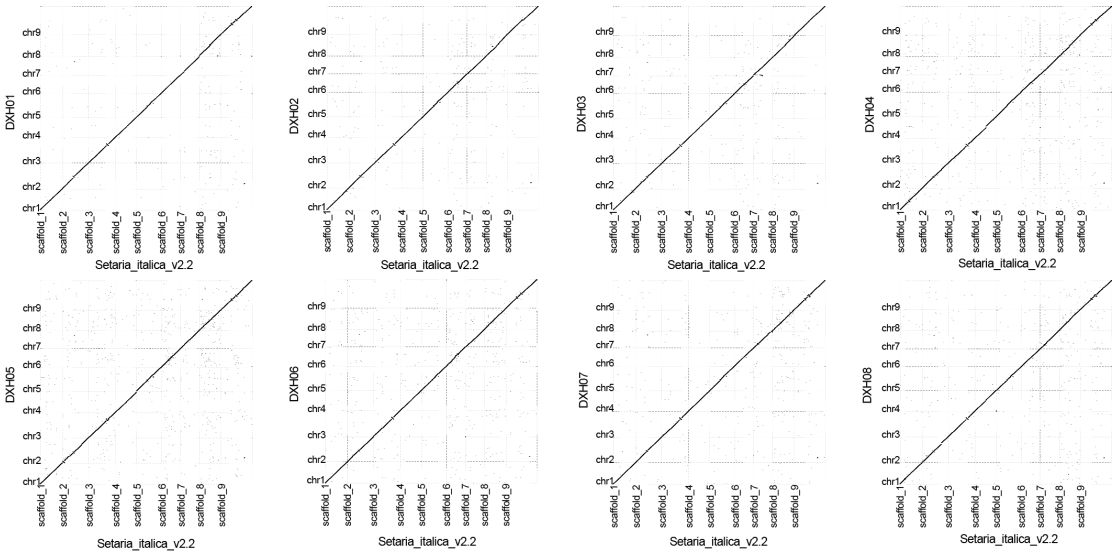


Figure S3. Genome-wide collinearity among the eight new millet assemblies and the *S. italica* reference genome.

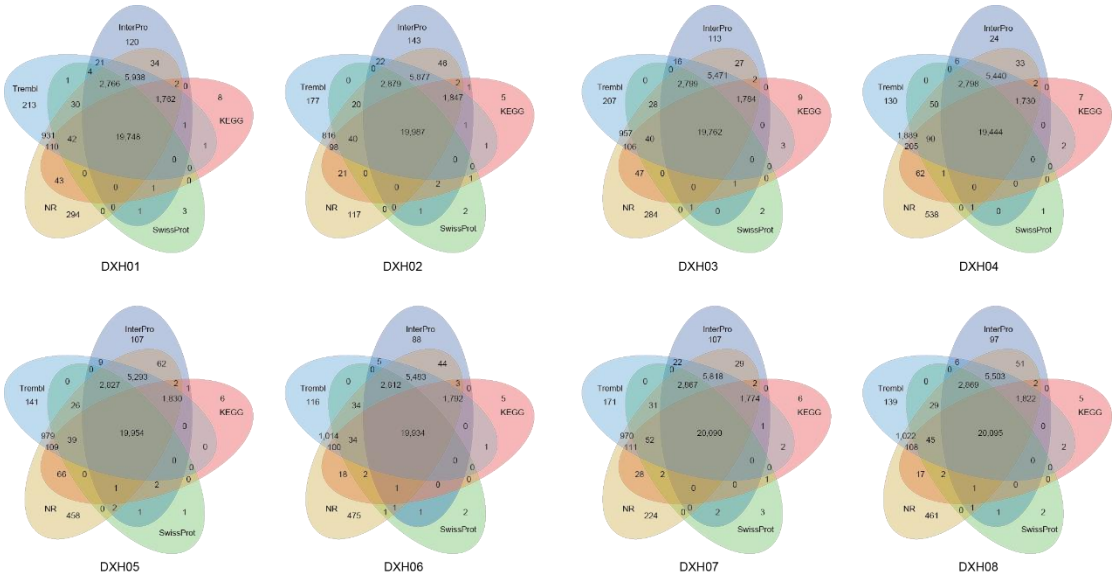


Figure S4. Venn diagrams showing the numbers of shared genes among different sets of functionally annotated protein-coding genes identified using the Swissprot, KEGG, TrEMBL, and InterPro databases.

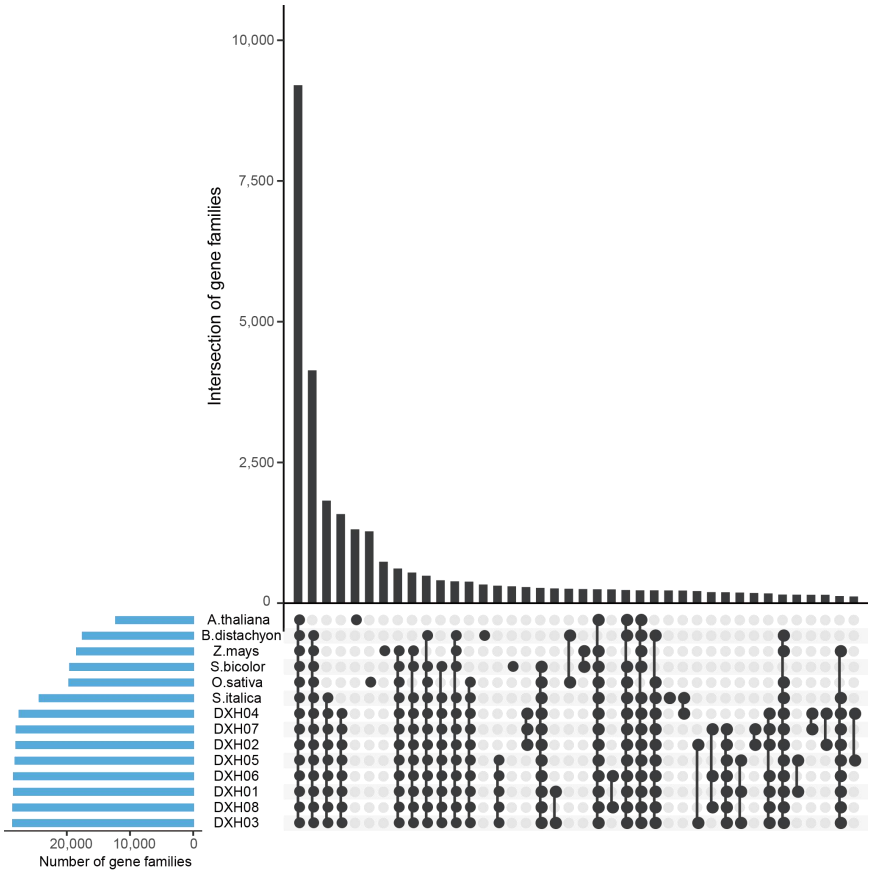


Figure S5. Intersections of clustered gene families among the eight newly assembled millets (DXH01–DXH08), the *S. italica* reference genome, *S. bicolor*, *Z. mays*, *O. sativa*, *B. distachyon*, and *A. thaliana*, plotted using UpSetR. The rows represent gene families, and the columns represent their intersections. For each set that is part of a given intersection, a black filled circle is placed in the corresponding matrix cell. If a set is not part of the intersection, a light gray circle is shown. A vertical black line connects the topmost black circle with the bottommost black circle in each column to emphasize the column-based relationships. The size of the intersections is shown as a bar chart on top of the matrix, such that each column lines up with exactly one bar. A second bar chart showing the size of each set is shown to the left of the matrix.

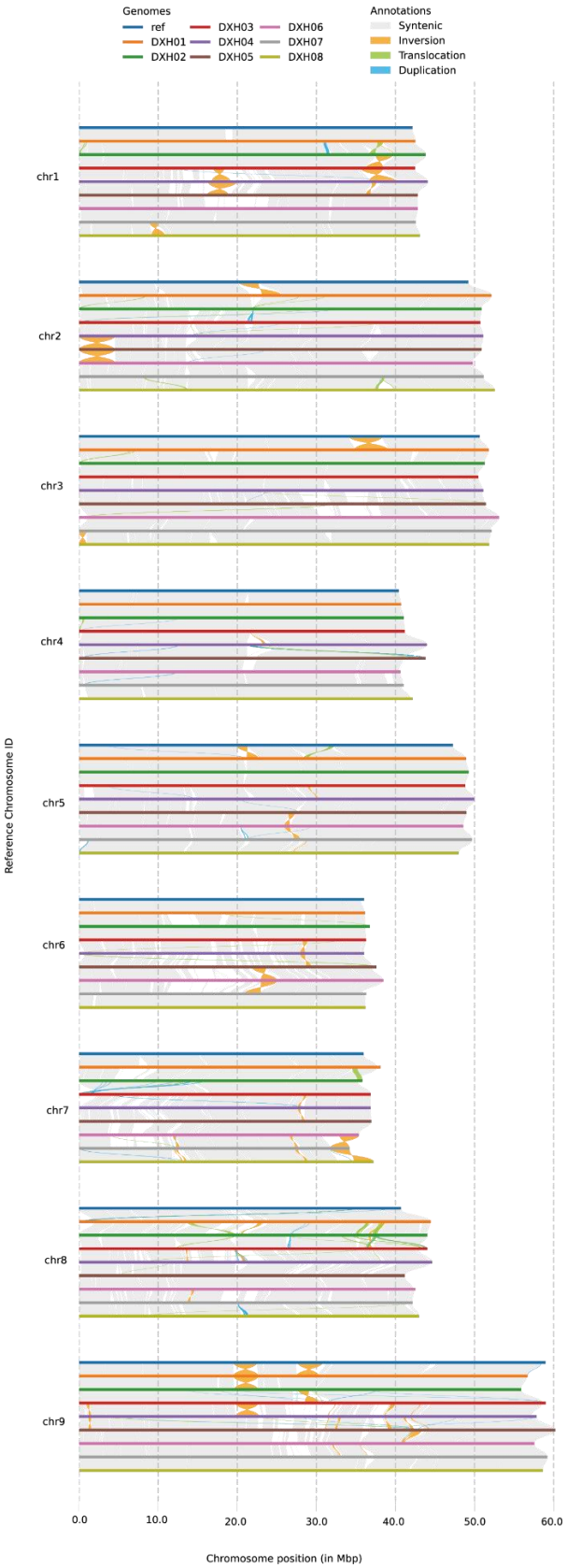


Figure S6. Chromosome-wide illustration of the three types of major SVs (translocations, inversions, and duplications) in nine millet accessions.

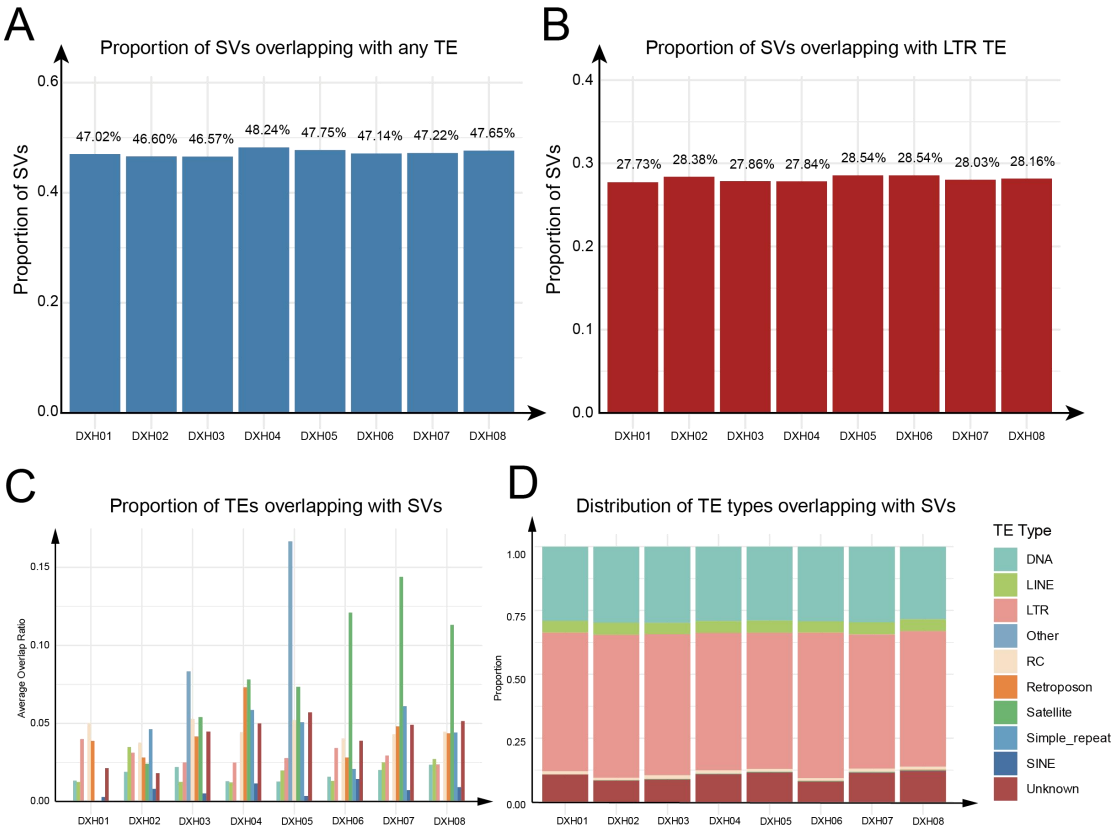


Figure S7. Associations between structural variants (SVs) and transposable elements (TEs). **A.** Proportion of SVs overlapping with all TEs across eight accessions (DXH01–DXH08). **B.** Proportion of SVs overlapping with LTR-type TEs across accessions. **C.** Overlap ratios of nine major TE types associated with SVs. **D.** Proportion of different TE categories, highlighting LTR elements as the predominant contributors to TE-associated SVs.

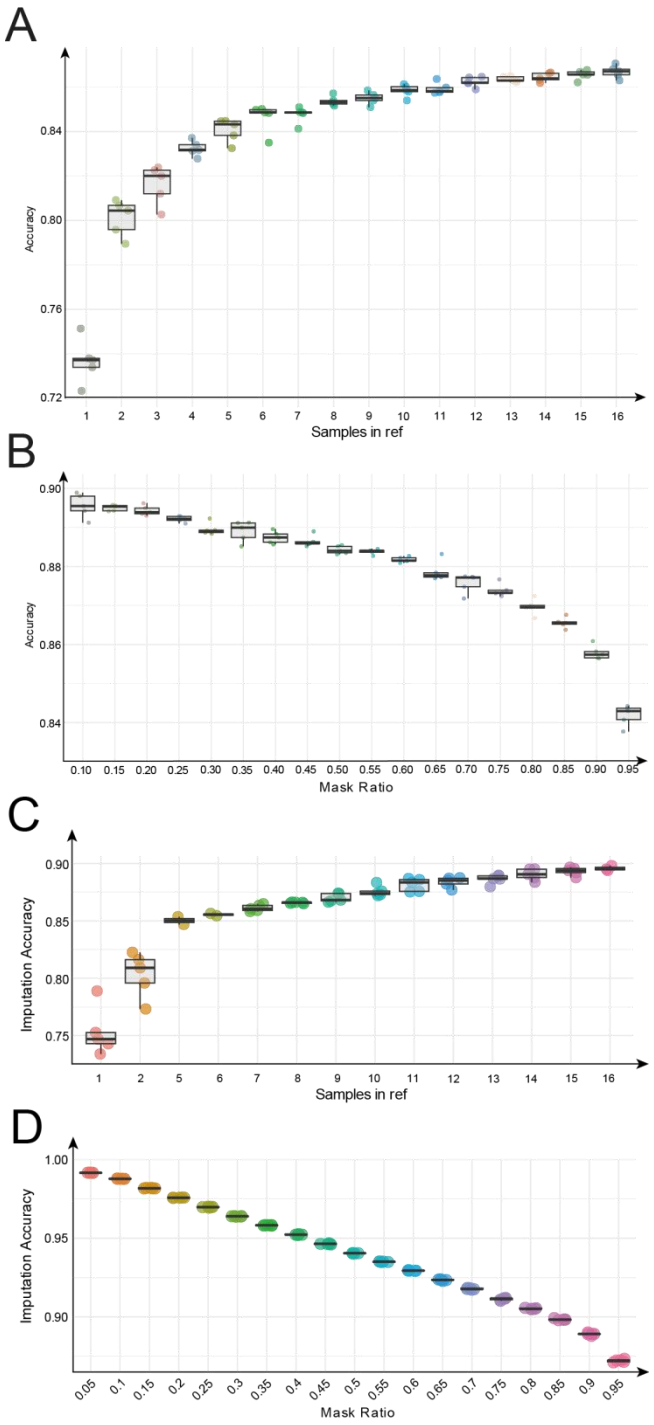
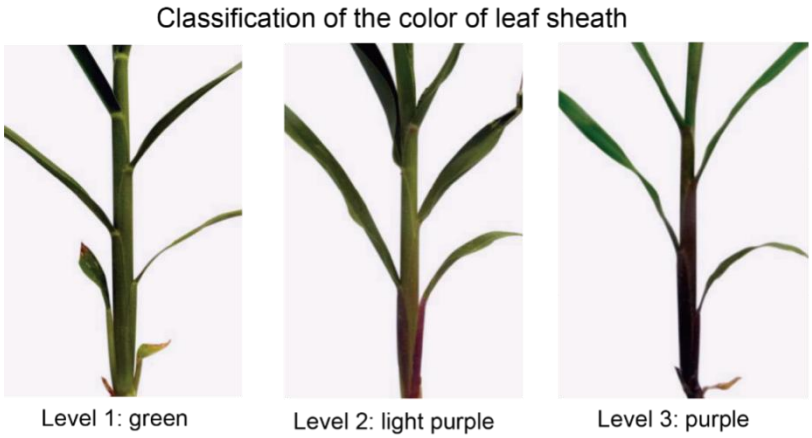
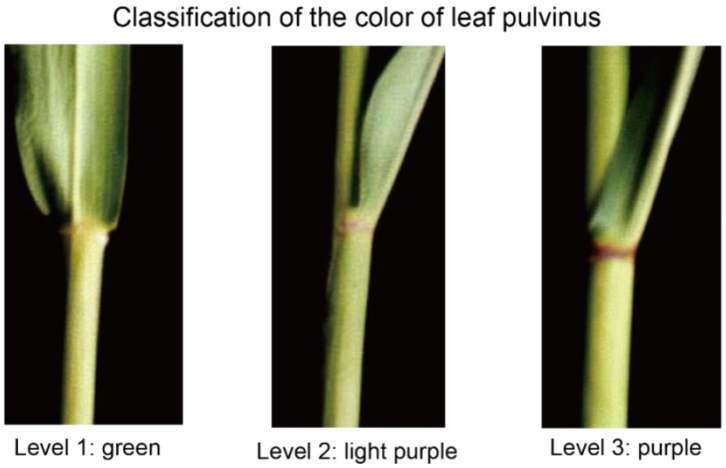


Figure S8. Evaluation of genotype imputation accuracy. **A.** Using SNP data under different reference panel sizes. **B.** Using SNP data under varying proportions of missing genotypes. **C.** Using PAV data under different reference panel sizes. **D.** Using PAV data under varying proportions of missing genotypes.





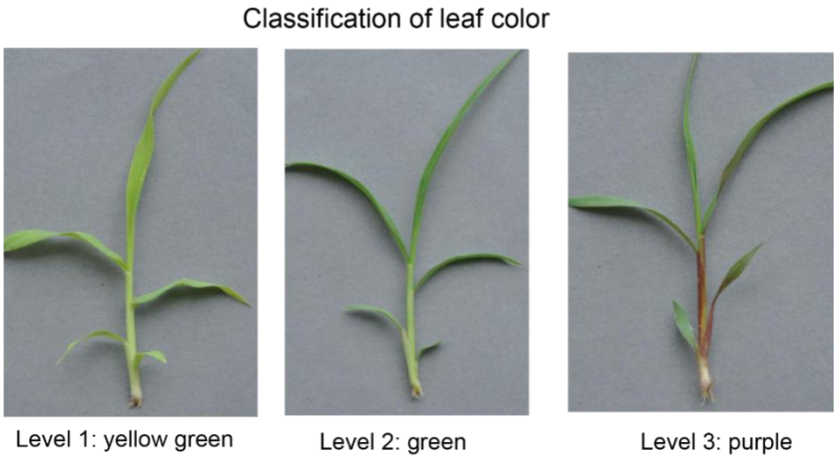


Figure S9. Classification standards for green, light purple, and deep purple colors of the leaf, leaf sheath, and leaf pulvinus.

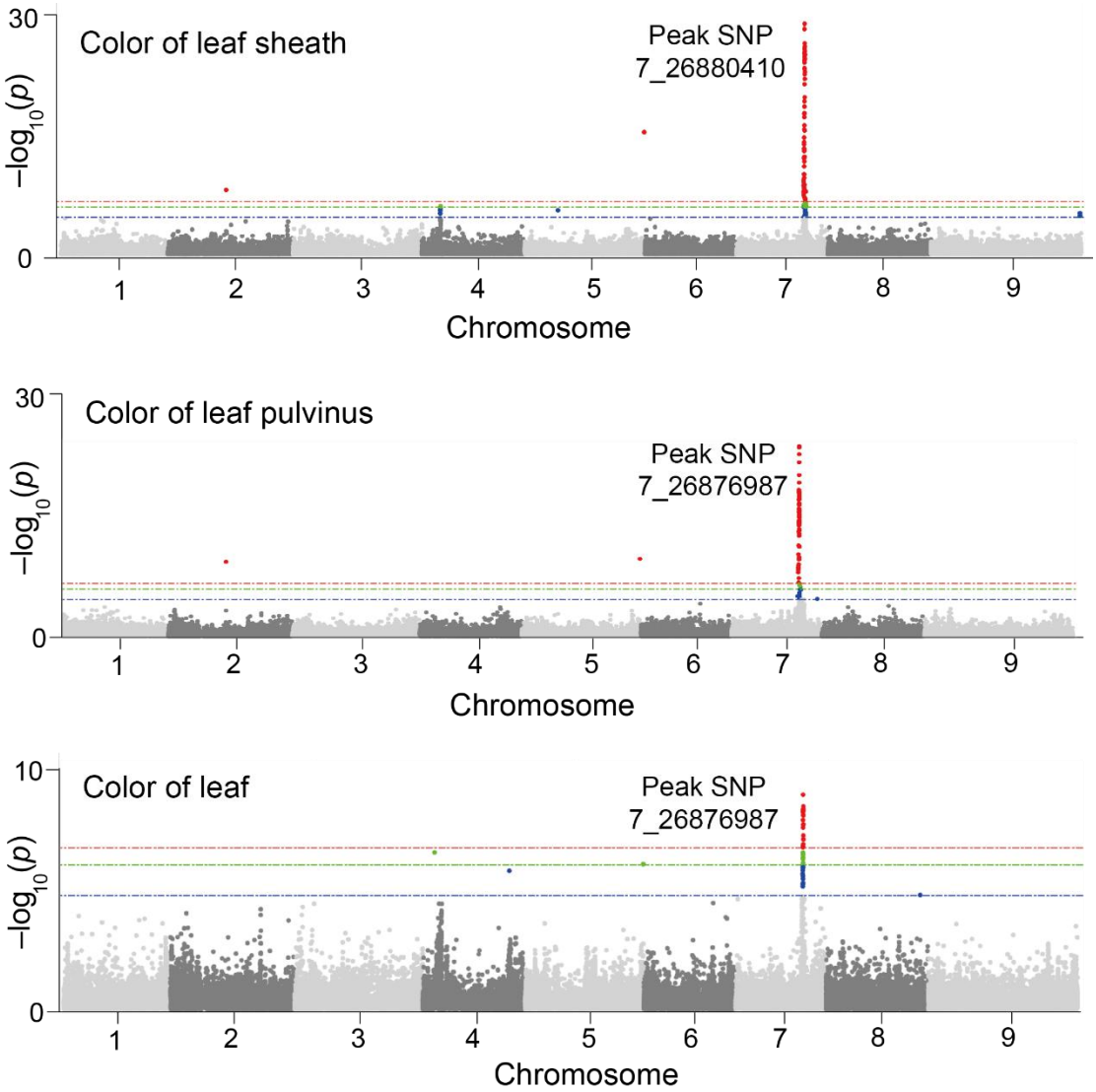


Figure S10. GWAS analysis of leaf, leaf sheath, and pulvinus color using the 96K genotype data.

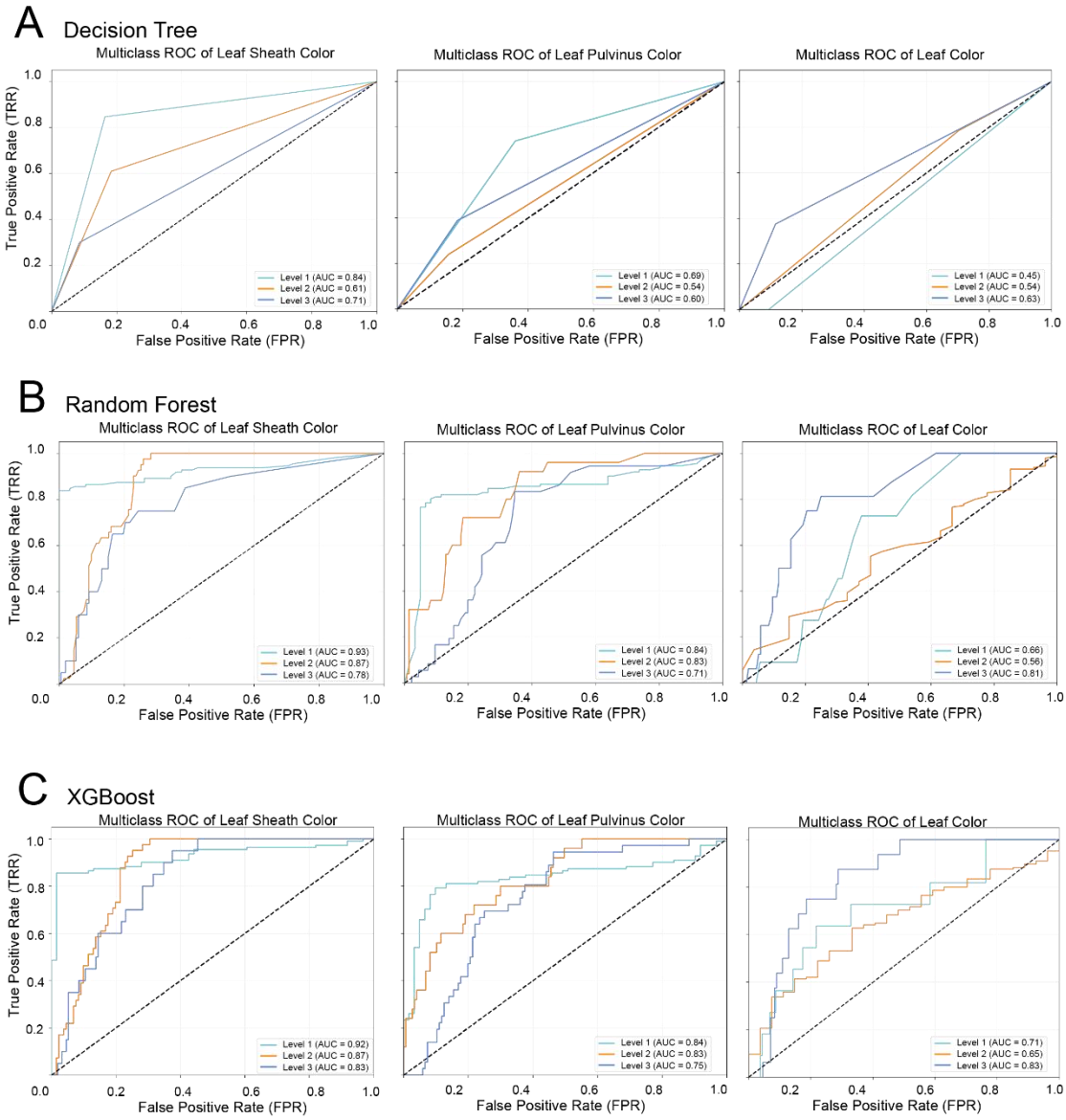


Figure S11. Multiclass ROC curves for prediction of three traits with the decision tree **(A)**, random forest **(B)**, and XGBoost **(C)** models.

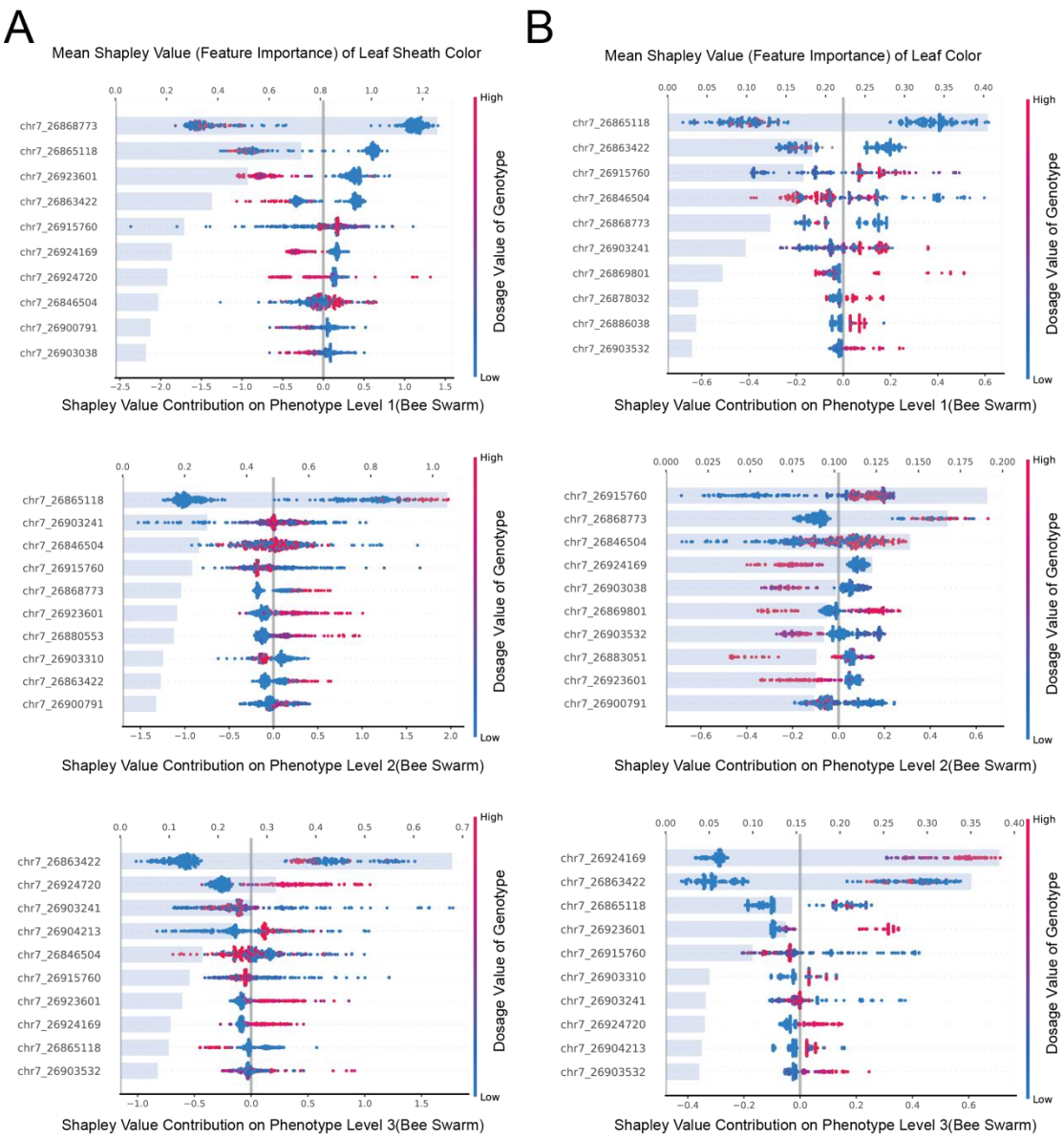


Figure S12. Bee swarm plot of Shapley values with a bar plot of feature importance for the top 10 features related to the three levels of leaf sheath color **(A)** and leaf color **(B)**.

Figure S13. Transposable elements identified in the syntenic regions surrounding the PPLS1 gene across representative genomes. Syntenic regions containing the PPLS1 gene were compared across eight representative genomes from this study (DXH01-DXH08) and selected foxtail millet genomes from He *et al*. (2023). Transposable elements (TEs) were annotated and classified as DNA transposons (*Harbinger*, *hAT*, *MuDR*, *Tc1/Mariner*) or retrotransposons (*Copia*, *Gypsy*). The structure and composition of TEs varied across accessions, with insertions mainly flanking or overlapping PPLS1. The visualization expands on Figure 6A by incorporating additional accessions, enabling a broader understanding of TE variation at this locus among wild, landrace, and cultivated lines.

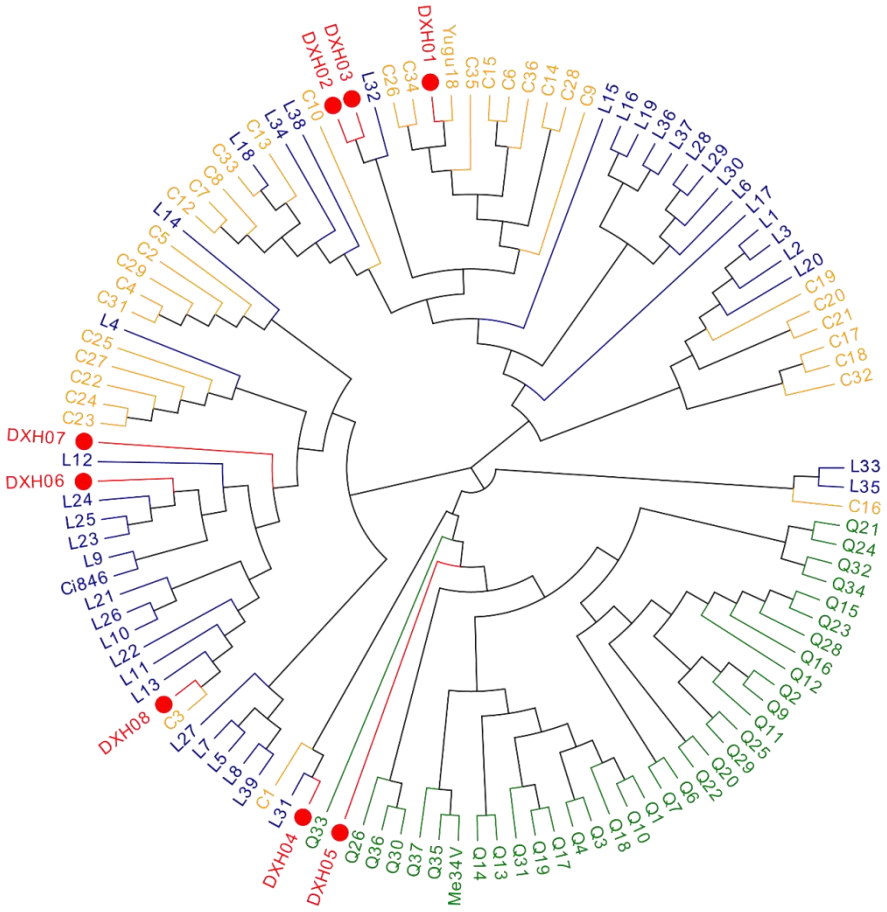


Figure S14. Phylogenetic tree of the eight millet varieties in this study compared to the 110 millet varieties from He *et al.* (2023). Genome sequences of 110 foxtail millet accessions from He *et al.* (2023) were combined with genome assemblies of the eight foxtail millet accessions (DXH01-DXH08) from this study to construct a phylogenetic tree. In the resulting phylogenetic tree, yellow labels marked with "C" indicate cultivated varieties, blue labels marked with "L" represent landraces, and green labels marked with "Q" denote wild varieties. The 110 published accessions clustered into three major subgroups. The eight DXH accessions were distributed across all three subgroups, demonstrating the representative genetic diversity of the samples included in this study.

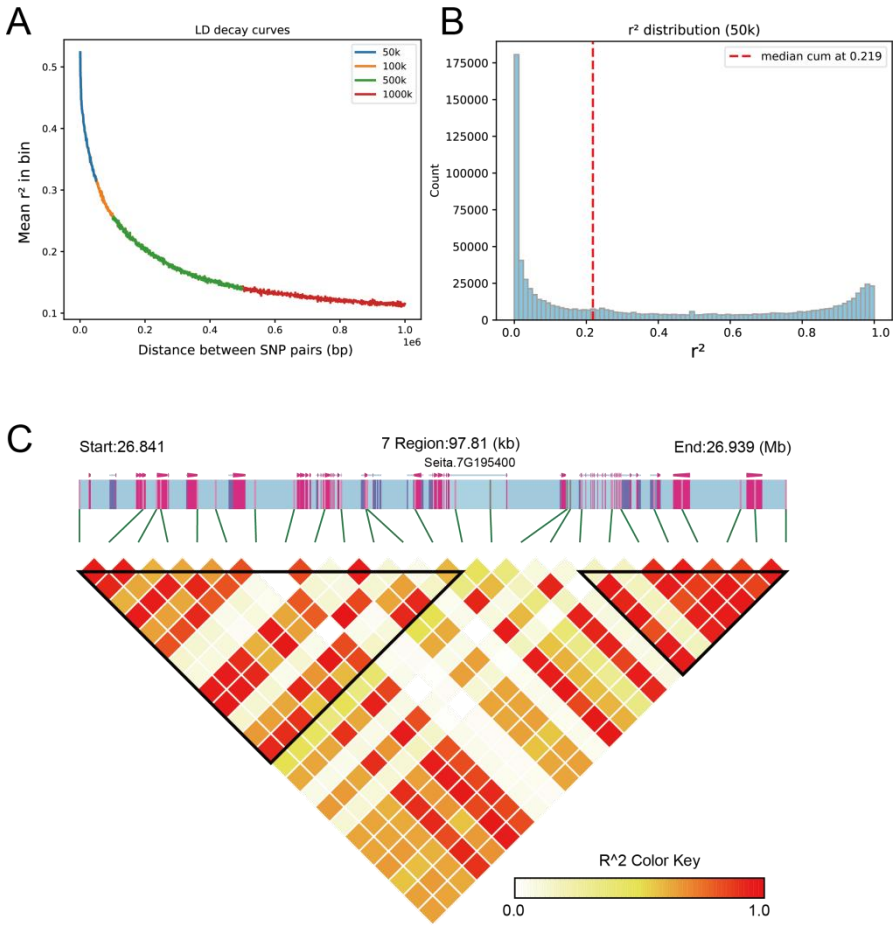


Figure S15. Linkage disequilibrium (LD) landscape in foxtail millet. **A.** Genome-wide LD decay curve showing the relationship between r^2 and physical distance. **B.** Genome-wide histogram of r^2 values calculated in 50-kb windows. **C.** LD heatmap of the genomic interval Chr7: 26.84–26.94 Mb, showing variation in LD strength across the region.

Supplementary Tables

Table S1. Information of the eight pan-genome accessions

Accession ID	Accession name	Chinese name	Origin	Type
DXH01	Yugu18	豫谷 18	China	Cultivar
DXH02	HunanWN	湘西糯	China	Landrace
DXH03	JP-60Days	日本 60 日	Japan	Cultivar
DXH04	IDSE1541	IDSE1541	India	Landrace
DXH05	IDSE120	IDSE120	India	Landrace
DXH06	HSY-2	二黑蛇妖	China	Landrace
DXH07	SPB-ZY2151	蒜皮白 ZY2151	China	Landrace
DXH08	SPB-ZY2152	蒜皮白 ZY2152	China	Landrace

Table S2. Summary statistics for the assembly and annotation of eight foxtail millet genomes

	DXH01	DXH02	DXH03	DXH04	DXH05	DXH06	DXH07	DXH08
Statistics of the eight genome assemblies								
Total length of scaffolds length (bp)	438,770,360	433,674,633	434,320,254	427,530,610	426,341,061	428,629,399	436,206,849	435,642,271
Scaffold N50 (bp)	44,455,500	44,030,813	44,019,000	44,627,284	43,805,500	42,799,932	42,551,000	43,081,500
Number of scaffold N50	5	5	5	5	5	5	5	5
Total length of contigs (bp)	438,738,860	433,635,633	434,275,754	427,511,610	426,307,061	428,600,899	436,121,349	435,587,271
Contig N50 (bp)	19,892,000	19,515,112	11,102,949	21,221,235	16,027,446	11,386,525	12,585,477	12,864,201
Number of contig N50	9	10	12	10	10	14	12	12
Sequences anchored to chromosome (%)	93.77	94.27	94.35	97.19	97.05	95.37	93.61	94.74
Gap counts	63	78	89	38	68	57	171	110
Complete BUSCOs (%)	98.4	98.6	98.4	98.7	98.4	98.5	98.4	98.5
Single copy (%)	96.3	95.8	95.6	96.7	96.5	96.5	96	96
Duplicated copy (%)	2.1	2.8	2.8	2	1.9	2	2.4	2.5
Fragmented (%)	0.7	0.7	0.7	0.7	0.8	0.9	0.8	0.7
Missing (%)	0.9	0.7	0.9	0.6	0.8	0.6	0.8	0.8
Mapping rate TGS(%)	99.8	99.8	99.6	99.7	99.0	99.3	99.8	99.6
QV	62.6	60.5	56.5	68.4	62.0	59.5	61.6	57.3
LTR assembly index(LAI)	17.2	18.1	18.1	17.8	17.3	17.8	18.3	15.6
Statistics of the transposable elements annotation								
TE ratio(%)	49.18	49.03	46.32	47.36	47.28	46.81	46.59	47.28
Percentage of repeat sequences by GCE (%)	34.67	33.64	37.73	33.43	34.41	34.5	34.87	34.37
Percentage of repeat sequences by GenomeScope (%)	27.8	29.2	31.5	28	28.5	28.1	28.9	29
Statistics of protein-coding gene prediction and annotation								
Number of genes	34,841	34,222	34,574	34,434	34,511	34,307	34,628	34,586
Average gene length (bp)	2,566	2,580	2,591	2,556	2,606	2,717	2,658	2,700
Average mRNA length (bp)	1,144	1,144	1,152	1,119	1,156	1,170	1,151	1,170
Average exons per gene	4.31	4.35	4.35	4.29	4.36	4.44	4.38	4.44
Complete BUSCOs*(%)	99.1	99.1	99	99.2	98.9	99.2	99	99.2

Table S3. Summary statistics for transposable elements (TEs) in the eight genome assemblies

DXH01								
	Repbase TEs		TE protiens		De novo		Combined TEs	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	74,227,493	16.918	59,003,444	13.448	119,993,569	27.35	137,195,936	31.271
LTR	68,124,315	15.527	51,897,707	11.829	113,865,446	25.953	127,823,523	29.134
Superfamily: Gypsy	43,834,252	9.991	36,857,840	8.401	90,346,900	20.592	96,305,654	21.951
Superfamily: Copia	24,053,850	5.482	14,867,047	3.389	22,323,851	5.088	30,833,442	7.028
LINE	6,063,863	1.382	7,128,927	1.625	6,010,504	1.37	9,354,967	2.132
SINE	66,695	0.015	0	0	130,646	0.03	192,946	0.044
ClassII:DNA	28,066,390	6.397	15,086,194	3.439	45,394,276	10.347	51,469,448	11.731
Unclassified	3,715	0.001	7,243	0.002	39,344,213	8.968	39,355,008	8.97
Total	102,172,937	23.288	73,840,848	16.83	203,329,407	46.344	215,753,895	49.176
DXH02								
	Repbase TEs		TE protiens		De novo		Combined TEs	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	74,243,178	17.121	59,304,454	13.676	125,865,435	29.026	138,650,157	31.974
LTR	68,144,338	15.715	51,987,481	11.989	119,309,852	27.514	129,188,647	29.792
Superfamily: Gypsy	43,978,119	10.142	36,883,905	8.506	92,570,182	21.347	97,124,477	22.398
Superfamily: Copia	23,934,542	5.52	14,911,764	3.439	26,066,087	6.011	31,827,572	7.34
LINE	6,061,659	1.398	7,339,898	1.693	6,457,403	1.489	9,461,948	2.182
SINE	64,979	0.015	0	0	103,882	0.024	167,552	0.039
ClassII:DNA	28,372,244	6.543	15,356,244	3.541	45,587,628	10.513	50,933,235	11.746
Unclassified	4,199	0.001	6,313	0.001	31,107,542	7.174	31,118,054	7.176
Total	102,488,332	23.635	74,420,457	17.162	200,966,331	46.345	212,589,799	49.025
DXH03								
	Repbase TEs		TE protiens		De novo		Combined TEs	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	74,502,742	17.156	59,603,777	13.725	120,275,194	27.696	135,286,870	31.152
LTR	68,416,795	15.754	52,267,376	12.036	114,371,827	26.336	125,880,581	28.986
Superfamily: Gypsy	44,267,457	10.193	37,122,932	8.548	89,547,703	20.62	95,351,291	21.956
Superfamily: Copia	23,973,890	5.52	14,963,586	3.446	24,140,907	5.559	30,441,779	7.01
LINE	6,053,207	1.394	7,360,937	1.695	5,475,495	1.261	9,077,221	2.09

SINE	59,042	0.014	0	0	451,213	0.104	502,866	0.116
ClassII:DNA	28,235,147	6.502	15,190,534	3.498	46,746,015	10.764	52,300,363	12.043
Unclassified	3,763	0.001	7,601	0.002	23,927,055	5.51	23,937,668	5.512
Total	102,599,192	23.625	74,568,298	17.171	189,272,519	43.583	201,152,112	46.319
DXH04								
	Repbase TEs		**TE protiens**		**De novo**		**Combined TEs**	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	74,518,347	17.431	59,427,795	13.901	117,000,527	27.368	134,036,536	31.353
LTR	68,518,600	16.027	52,286,138	12.23	110,319,540	25.805	124,501,015	29.122
Superfamily: Gypsy	44,581,991	10.428	37,416,527	8.752	84,939,725	19.868	94,184,117	22.031
Superfamily: Copia	23,684,270	5.54	14,694,583	3.437	24,645,175	5.765	30,317,508	7.092
LINE	5,968,798	1.396	7,169,017	1.677	6,660,305	1.558	9,638,309	2.255
SINE	59,579	0.014	0	0	29,252	0.007	87,491	0.02
ClassII:DNA	28,143,065	6.583	15,038,181	3.518	44,915,268	10.506	52,121,675	12.192
Unclassified	3,191	0.001	7,039	0.002	31,791,042	7.436	31,801,272	7.439
Total	102,517,017	23.98	74,241,185	17.366	191,256,685	44.737	202,456,673	47.357
DXH05								
	Repbase TEs		**TE protiens**		**De novo**		**Combined TEs**	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	75,682,137	17.753	56,122,774	13.165	110,758,488	25.981	135,938,417	31.887
LTR	69,706,603	16.351	49,535,065	11.62	104,005,858	24.397	126,531,501	29.681
Superfamily: Gypsy	44,813,047	10.512	35,198,364	8.257	88,598,714	20.783	95,927,908	22.502
Superfamily: Copia	24,693,738	5.792	14,179,264	3.326	14,704,753	3.449	30,469,161	7.147
LINE	5,947,706	1.395	6,605,251	1.549	6,720,210	1.576	9,464,723	2.22
SINE	54,866	0.013	0	0	36,525	0.009	91,391	0.021
ClassII:DNA	27,719,589	6.502	13,550,442	3.179	45,697,755	10.719	51,031,463	11.971
Unclassified	4,132	0.001	7,375	0.002	35,847,182	8.409	35,858,600	8.411
Total	103,250,493	24.22	69,450,030	16.291	190,064,353	44.584	201,544,648	47.277
DXH06								
	Repbase TEs		**TE protiens**		**De novo**		**Combined TEs**	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	74,626,055	17.412	59,207,345	13.814	124,680,246	29.09	135,951,640	31.72
LTR	68,646,248	16.016	52,011,042	12.135	118,459,085	27.639	126,833,537	29.592
Superfamily: Gypsy	43,034,748	10.041	36,105,513	8.424	91,220,510	21.283	95,048,471	22.176
Superfamily: Copia	25,396,318	5.925	15,751,815	3.675	26,490,541	6.181	31,540,552	7.359

LINE	5,952,523	1.389	7,218,000	1.684	6,118,930	1.428	9,098,273	2.123
SINE	52,790	0.012	0	0	106,372	0.025	154,094	0.036
ClassII:DNA	27,840,549	6.496	14,971,998	3.493	45,950,867	10.721	51,360,532	11.983
Unclassified	2,914	0.001	8,664	0.002	20,090,964	4.688	20,102,542	4.69
Total	102,305,072	23.87	73,948,029	17.253	189,170,555	44.137	200,613,262	46.807
DXH07								
	Repbase TEs		**TE protiens**		**De novo**		**Combined TEs**	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	75,798,620	17.38	60,201,863	13.804	111,519,803	25.571	137,428,648	31.512
LTR	69,646,928	15.97	52,920,246	12.134	104,424,464	23.944	127,357,893	29.202
Superfamily: Gypsy	44,409,299	10.183	37,136,395	8.515	87,127,782	19.978	96,292,757	22.079
Superfamily: Copia	24,980,845	5.728	15,594,932	3.576	16,602,432	3.807	30,923,361	7.091
LINE	6,124,790	1.404	7,303,690	1.675	6,932,048	1.589	10,008,742	2.295
SINE	54,402	0.012	0	0	168,755	0.039	217,895	0.05
ClassII:DNA	28,319,849	6.494	15,042,383	3.449	45,133,616	10.349	51,113,194	11.72
Unclassified	4,199	0.001	7,925	0.002	36,197,725	8.3	36,208,198	8.302
Total	103,971,517	23.84	75,020,687	17.202	191,197,446	43.84	203,199,886	46.593
DXH08								
	Repbase TEs		**TE protiens**		**De novo**		**Combined TEs**	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
ClassI	76,769,013	17.624	61,015,720	14.008	113,375,979	26.028	136,338,265	31.3
LTR	70,778,467	16.249	53,844,617	12.361	107,037,297	24.573	126,976,208	29.151
Superfamily: Gypsy	45,085,912	10.351	37,872,562	8.695	83,079,223	19.073	94,985,152	21.806
Superfamily: Copia	25,481,096	5.85	15,791,887	3.625	23,336,264	5.357	31,995,808	7.345
LINE	5,966,577	1.37	7,191,335	1.651	6,185,610	1.42	9,284,681	2.132
SINE	53,955	0.012	0	0	203,704	0.047	257,442	0.059
ClassII:DNA	28,322,595	6.502	15,263,298	3.504	45,756,428	10.505	51,431,406	11.807
Unclassified	3,566	0.001	6,355	0.001	37,090,290	8.515	37,100,048	8.517
Total	104,922,422	24.088	76,058,920	17.461	194,010,147	44.54	205,946,302	47.28

Table S4. Summary statistics for non-coding RNAs in the eight assemblies

DXH01				
Type	**Copy(w)**	**Average length (bp)**	**Total length (bp)**	**% of genome**
miRNA	161	133	21,424	0.0049
tRNA	2,198	75	164,748	0.0376
rRNA	7,509	376	2,823,460	0.6435
18S	1,157	1,664	1,925,255	0.4388
28S	4,246	144	612,537	0.1396
5.8S	1,058	158	167,310	0.0381
5S	1,048	113	118,358	0.0270
snRNA	438	116	50,730	0.0116
CD-box	289	102	29,579	0.0067
HACA-box	56	128	7,158	0.0016
splicing	93	150	13,993	0.0032
DXH02				
Type	**Copy(w)**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	162	133	21,621	0.0050
tRNA	1,775	75	132,941	0.0307
rRNA	4,773	359	1,712,024	0.3948
18S	683	1,672	1,141,700	0.2633
28S	2,547	144	367,652	0.0848
5.8S	632	158	100,100	0.0231
5S	911	113	102,572	0.0237
snRNA	449	117	52,382	0.0121
CD-box	296	104	30,702	0.0071
HACA-box	59	127	7,472	0.0017
splicing	94	151	14,208	0.0033
DXH03				
Type	**Copy(w)**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	161	133	21,423	0.0049
tRNA	1,186	75	88,923	0.0205
rRNA	3,729	381	1,419,246	0.3268
18S	571	1,698	969,379	0.2232
28S	2,151	144	310,622	0.0715
5.8S	536	158	84,919	0.0196
5S	471	115	54,326	0.0125
snRNA	436	117	50,920	0.0117
CD-box	285	103	29,362	0.0068
HACA-box	58	127	7,363	0.0017
splicing	93	153	14,195	0.0033
DXH04				
Type	**Copy(w)**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	163	134	21,795	0.0051
tRNA	667	75	50,035	0.0117

rRNA	5,217	359	1,873,440	0.4382
18S	732	1,708	1,249,952	0.2924
28S	2,750	145	397,669	0.0930
5.8S	691	159	109,714	0.0257
5S	1,044	111	116,105	0.0272
snRNA	451	117	52,754	0.0123
CD-box	300	104	31,107	0.0073
HACA-box	56	128	7,160	0.0017
splicing	95	152	14,487	0.0034
DXH05				
Type	**Copy(w)**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	160	133	21,341	0.0050
tRNA	661	75	49,719	0.0117
rRNA	4,354	369	1,608,756	0.3774
18S	630	1,724	1,085,857	0.2547
28S	2,410	145	348,474	0.0817
5.8S	599	158	94,789	0.0222
5S	715	111	79,636	0.0187
snRNA	437	117	51,034	0.0120
CD-box	290	103	29,925	0.0070
HACA-box	53	129	6,822	0.0016
splicing	94	152	14,287	0.0034
DXH06				
Type	**Copy(w)**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	159	133	21,182	0.0049
tRNA	684	75	51,302	0.0120
rRNA	4,379	363	1,588,682	0.3707
18S	616	1,729	1,065,254	0.2485
28S	2,318	145	335,031	0.0782
5.8S	587	158	92,966	0.0217
5S	858	111	95,431	0.0223
snRNA	463	116	53,885	0.0126
CD-box	305	103	31,500	0.0073
HACA-box	61	127	7,719	0.0018
splicing	97	151	14,666	0.0034
DXH07				
Type	**Copy(w)**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	165	132	21,841	0.0050
tRNA	699	75	52,352	0.0120
rRNA	6,244	374	2,336,794	0.5358
18S	939	1,691	1,588,019	0.3641
28S	3,489	145	504,676	0.1157
5.8S	876	158	138,688	0.0318
5S	940	112	105,411	0.0242

snRNA	456	116	53,096	0.0122
CD-box	303	103	31,261	0.0072
HACA-box	56	128	7,160	0.0016
splicing	97	151	14,675	0.0034
DXH08				
Type	**Copy(w)**	**Average length(bp)**	**Total length(bp)**	**% of genome**
miRNA	163	134	21,761	0.0050
tRNA	687	75	51,599	0.0118
rRNA	4,762	353	1,679,029	0.3855
18S	671	1,656	1,111,304	0.2551
28S	2,493	145	360,409	0.0827
5.8S	610	158	96,331	0.0221
5S	988	112	110,985	0.0255
snRNA	440	116	51,212	0.0118
CD-box	292	103	30,021	0.0069
HACA-box	56	128	7,161	0.0016
splicing	92	152	14,030	0.0032

Note: '% of genome' was calculated by the non-gap genome size 435,587,271 bp.

Table S5. Statistics for core genomes and pan-genomes of millet, along with an additional genome assembly

Assemblies	Number of core-genes	Number of genes in pan-genome	Number of core-genes	Number of genes in pan-genome	Size of pan-genome (bp)
DXH01	34,841	34,841	27,384	27,384	438,770,360
DXH01+2	32,123	35,426	23,967	29,691	463,620,574
DXH01+2+3	31,710	35,616	23,482	30,632	469,581,873
DXH01+2+3+4	30,205	36,353	22,710	32,262	500,019,351
DXH01+2+3+4+5	29,361	37,037	22,259	33,286	528,680,019
DXH01+2+3+4+5+6	28,976	37,459	21,943	34,227	548,030,641
DXH01+2+3+4+5+6+7	28,781	37,671	21,748	34,824	558,394,086
DXH01+2+3+4+5+6+7+8	28,592	37,862	21,645	35,240	565,557,199

Table S6. Summary statistics for clustered gene families of the eight new millet assemblies, the millet reference assembly, and five model plant species

Species	Genes number	Genes in families	Unclustered genes	Family number	Unique families	Average genes per family
DXH01	34,841	33,147	1,694	28,378	9	1.17
DXH02	34,222	32,868	1,354	27,998	14	1.17
DXH03	34,574	33,424	1,150	28,507	25	1.17
DXH04	34,434	32,210	2,224	27,485	46	1.17
DXH05	34,511	32,992	1,519	28,130	38	1.17
DXH06	34,307	33,145	1,162	28,374	9	1.17
DXH07	34,628	32,975	1,653	27,950	12	1.18
DXH08	34,586	33,748	838	28,497	7	1.18
S. italica	34,584	30,339	4,245	24,306	225	1.25
S. bicolor	28,244	25,725	2,519	19,477	299	1.32
B. distachyon	25,537	22,788	2,749	17,476	331	1.3
O. sativa	42,189	30,333	11,856	19,612	1,275	1.55
Z. mays	34,351	29,458	4,893	18,395	735	1.6
A. thaliana	27,416	22,789	4,627	12,203	1,311	1.87

Note: Unclustered genes refer to special gene of corresponding species; Unique families refer to special gene families of corresponding species.

Table S7. GO enrichment of expanded and contracted genes in the lineage of the most recent common ancestor of nine millets

GO ID	Categories	Qvalue	GO description
Expanded genes			
GO:0016705	MF	9.58E-08	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
GO:0005506	MF	1.29E-07	iron ion binding
GO:0020037	MF	1.57E-06	heme binding
GO:0030170	MF	6.77E-05	pyridoxal phosphate binding
GO:0043531	MF	3.48E-04	ADP binding
GO:0055114	BP	4.85E-03	oxidation-reduction process
Contracted genes			
GO:0004672	MF	2.48E-104	protein kinase activity
GO:0006468	BP	2.48E-104	protein phosphorylation
GO:0005524	MF	1.64E-75	ATP binding
GO:0048544	BP	1.11E-68	recognition of pollen
GO:0015299	MF	1.82E-08	solute:proton antiporter activity
GO:0006812	BP	2.33E-07	cation transport
GO:0005515	MF	3.43E-07	protein binding
GO:0008107	MF	1.23E-05	galactoside 2-alpha-L-fucosyltransferase activity
GO:0042546	BP	1.23E-05	cell wall biogenesis
GO:0004674	MF	1.32E-03	protein serine/threonine kinase activity
GO:0016705	MF	1.36E-03	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
GO:0005506	MF	2.71E-03	iron ion binding
GO:0055085	BP	1.03E-02	transmembrane transport
GO:0016021	CC	1.61E-02	integral component of membrane
GO:0020037	MF	1.61E-02	heme binding

Table S8. GO annotations of 22 positively selected genes (PSGs)

GO ID	categories	# of Genes	Qvalue	GO description	Gene IDs
GO:0004825	MF	2	0.721559	methionine-tRNA ligase activity	DXH0124065.1, DXH0132985.1
GO:0017150	MF	1	0.721559	tRNA dihydrouridine synthase activity	DXH0124065.1
GO:0032977	MF	1	0.721559	membrane insertase activity	DXH0124065.1
GO:0106413	MF	1	0.721559	dihydrouridine synthase activity	DXH0132985.1
GO:0140101	MF	1	0.721559	catalytic activity, acting on a tRNA	DXH0131605.1
GO:0140597	MF	1	0.721559	protein carrier chaperone	DXH0131605.1
GO:0003950	MF	1	0.900863	NAD+ ADP-ribosyltransferase activity	DXH0131605.1
GO:0140104	MF	1	0.900863	molecular carrier activity	DXH0104890.1
GO:0000166	MF	1	1	nucleotide binding	DXH0104099.1
GO:0003674	MF	1	1	molecular_function	DXH0104099.1
GO:0003676	MF	2	1	nucleic acid binding	DXH0127830.1, DXH0131605.1
GO:0003677	MF	2	1	DNA binding	DXH0127830.1, DXH0131605.1
GO:0003700	MF	1	1	DNA-binding transcription factor activity	DXH0104099.1
GO:0003824	MF	4	1	catalytic activity	DXH0100536.1, DXH0104099.1, DXH0127830.1, DXH0131605.1
GO:0004812	MF	1	1	aminoacyl-tRNA ligase activity	DXH0100536.1
GO:0005488	MF	4	1	binding	DXH0100536.1, DXH0104099.1, DXH0127830.1, DXH0131605.1
GO:0005515	MF	1	1	protein binding	DXH0100536.1
GO:0005575	CC	1	1	cellular_component	DXH0100536.1
GO:0005634	CC	1	1	nucleus	DXH0104099.1
GO:0006082	BP	1	1	organic acid metabolic process	DXH0100536.1
GO:0006139	BP	1	1	nucleobase-containing compound metabolic process	DXH0100536.1
GO:0006355	BP	1	1	regulation of DNA-templated transcription	DXH0104099.1

GO:0006396	BP	2	1	RNA processing	DXH0124065.1, DXH0132985.1
GO:0006397	BP	1	1	mRNA processing	DXH0104890.1
GO:0006399	BP	2	1	tRNA metabolic process	DXH0124065.1, DXH0132985.1
GO:0006418	BP	1	1	tRNA aminoacylation for protein translation	DXH0124065.1
GO:0006431	BP	1	1	methionyl-tRNA aminoacylation	DXH0132985.1
GO:0006471	BP	1	1	protein ADP-ribosylation	DXH0132985.1
GO:0006520	BP	2	1	cellular amino acid metabolic process	DXH0102447.1, DXH0104890.1
GO:0006725	BP	1	1	cellular aromatic compound metabolic process	DXH0124065.1
GO:0006807	BP	1	1	nitrogen compound metabolic process	DXH0121385.1
GO:0006810	BP	1	1	transport	DXH0132985.1
GO:0008033	BP	9	1	tRNA processing	DXH0100536.1, DXH0102447.1, DXH0104890.1, DXH0112303.1, DXH0119851.1, DXH0120220.1, DXH0121385.1, DXH0124065.1, DXH0132985.1
GO:0008150	BP	3	1	biological_process	DXH0104890.1, DXH0119851.1, DXH0120220.1
GO:0008152	BP	2	1	metabolic process	DXH0104890.1, DXH0119851.1
GO:0008270	MF	1	1	zinc ion binding	DXH0119851.1
GO:0008289	MF	2	1	lipid binding	DXH0102447.1, DXH0104890.1
GO:0009521	CC	2	1	photosystem	DXH0102447.1, DXH0124065.1
GO:0009522	CC	1	1	photosystem I	DXH0119851.1
GO:0009889	BP	1	1	regulation of biosynthetic process	DXH0113274.1
GO:0009987	BP	1	1	cellular process	DXH0119851.1
GO:0010468	BP	1	1	regulation of gene expression	DXH0104890.1
GO:0010556	BP	2	1	regulation of macromolecule biosynthetic process	DXH0102447.1, DXH0104890.1
GO:0015979	BP	2	1	photosynthesis	DXH0102447.1, DXH0104890.1
GO:0016020	CC	5	1	membrane	DXH0104890.1, DXH0119851.1, DXH0120220.1, DXH0124065.1, DXH0132985.1

GO:0016021	CC	5	1	integral component of membrane	DXH0104890.1, DXH0119851.1, DXH0120220.1, DXH0124065.1, DXH0132985.1
GO:0016070	BP	4	1	RNA metabolic process	DXH0102447.1, DXH0104890.1, DXH0124065.1, DXH0132985.1
GO:0016071	BP	11	1	mRNA metabolic process	DXH0100536.1, DXH0102447.1, DXH0104890.1, DXH0112303.1, DXH0113274.1, DXH0119851.1, DXH0120220.1, DXH0121385.1, DXH0124065.1, DXH0131605.1, DXH0132985.1
GO:0016491	MF	2	1	oxidoreductase activity	DXH0124065.1, DXH0132985.1
GO:0016627	MF	2	1	oxidoreductase activity, acting on the CH-CH group of donors	DXH0124065.1, DXH0132985.1
GO:0016757	MF	2	1	glycosyltransferase activity	DXH0124065.1, DXH0132985.1
GO:0016763	MF	2	1	pentosyltransferase activity	DXH0124065.1, DXH0132985.1
GO:0016788	MF	2	1	hydrolase activity, acting on ester bonds	DXH0100536.1, DXH0112303.1
GO:0016874	MF	5	1	ligase activity	DXH0102447.1, DXH0104890.1, DXH0113274.1, DXH0124065.1, DXH0132985.1
GO:0016875	MF	2	1	ligase activity, forming carbon-oxygen bonds	DXH0124065.1, DXH0132985.1
GO:0019219	BP	1	1	regulation of nucleobase-containing compound metabolic process	DXH0132985.1
GO:0019222	BP	1	1	regulation of metabolic process	DXH0104890.1
GO:0019752	BP	1	1	carboxylic acid metabolic process	DXH0100317.1
GO:0031123	BP	2	1	RNA 3'-end processing	DXH0124065.1, DXH0132985.1
GO:0031124	BP	3	1	mRNA 3'-end processing	DXH0100317.1, DXH0124065.1, DXH0132985.1
GO:0031224	CC	1	1	intrinsic component of membrane	DXH0100317.1
GO:0031323	BP	2	1	regulation of cellular metabolic process	DXH0100317.1, DXH0124065.1
GO:0031326	BP	3	1	regulation of cellular biosynthetic process	DXH0100317.1, DXH0124065.1, DXH0132985.1
GO:0032991	CC	3	1	protein-containing complex	DXH0100317.1, DXH0124065.1, DXH0132985.1
GO:0034470	BP	1	1	ncRNA processing	DXH0104099.1
GO:0034641	BP	3	1	cellular nitrogen compound metabolic process	DXH0100317.1, DXH0124065.1, DXH0132985.1

GO:0034660	BP	3	1	ncRNA metabolic process	DXH0100317.1, DXH0124065.1, DXH0132985.1
GO:0036094	MF	1	1	small molecule binding	DXH0124065.1
GO:0043038	BP	3	1	amino acid activation	DXH0100317.1, DXH0124065.1, DXH0132985.1
GO:0043039	BP	1	1	tRNA aminoacylation	DXH0132985.1
GO:0043167	MF	1	1	ion binding	DXH0132985.1
GO:0043168	MF	1	1	anion binding	DXH0132985.1
GO:0043169	MF	1	1	cation binding	DXH0100317.1
GO:0043170	BP	1	1	macromolecule metabolic process	DXH0100317.1
GO:0043226	CC	1	1	organelle	DXH0124065.1
GO:0043227	CC	3	1	membrane-bounded organelle	DXH0100317.1, DXH0124065.1, DXH0132985.1
GO:0043229	CC	1	1	intracellular organelle	DXH0132985.1
GO:0043231	CC	1	1	intracellular membrane-bounded organelle	DXH0132985.1
GO:0043436	BP	1	1	oxoacid metabolic process	DXH0132985.1
GO:0043565	MF	1	1	sequence-specific DNA binding	DXH0132985.1
GO:0044237	BP	4	1	cellular metabolic process	DXH0100317.1, DXH0104890.1, DXH0124065.1, DXH0132985.1
GO:0044238	BP	4	1	primary metabolic process	DXH0100317.1, DXH0104099.1, DXH0124065.1, DXH0132985.1
GO:0044281	BP	4	1	small molecule metabolic process	DXH0100317.1, DXH0104890.1, DXH0124065.1, DXH0132985.1
GO:0046483	BP	1	1	heterocycle metabolic process	DXH0132985.1
GO:0046872	MF	5	1	metal ion binding	DXH0100317.1, DXH0104099.1, DXH0124065.1, DXH0127830.1, DXH0132985.1
GO:0046914	MF	1	1	transition metal ion binding	DXH0127830.1
GO:0050660	MF	1	1	flavin adenine dinucleotide binding	DXH0119851.1
GO:0051171	BP	1	1	regulation of nitrogen compound metabolic process	DXH0119851.1
GO:0051179	BP	1	1	localization	DXH0119851.1
GO:0051234	BP	1	1	establishment of localization	DXH0119851.1
GO:0051252	BP	1	1	regulation of RNA metabolic process	DXH0119851.1

GO:0055085	BP	1	1	transmembrane transport	DXH0119851.1
GO:0060255	BP	1	1	regulation of macromolecule metabolic process	DXH0119851.1
GO:0071704	BP	1	1	organic substance metabolic process	DXH0119851.1
GO:0080090	BP	4	1	regulation of primary metabolic process	DXH0100317.1, DXH0104890.1, DXH0124065.1, DXH0132985.1
GO:0090304	BP	1	1	nucleic acid metabolic process	DXH0119851.1
GO:0097159	MF	1	1	organic cyclic compound binding	DXH0119851.1
GO:0098796	CC	1	1	membrane protein complex	DXH0119851.1
GO:0110165	CC	1	1	cellular anatomical entity	DXH0119851.1
GO:0140098	MF	4	1	catalytic activity, acting on RNA	DXH0100317.1, DXH0104890.1, DXH0124065.1, DXH0132985.1
GO:0140110	MF	1	1	transcription regulator activity	DXH0119851.1
GO:0140640	MF	1	1	catalytic activity, acting on a nucleic acid	DXH0119851.1
GO:1901265	MF	5	1	nucleoside phosphate binding	DXH0100317.1, DXH0104099.1, DXH0104890.1, DXH0124065.1, DXH0132985.1
GO:1901360	BP	1	1	organic cyclic compound metabolic process	DXH0127830.1
GO:1901363	MF	1	1	heterocyclic compound binding	DXH0127830.1
GO:1903506	BP	7	1	regulation of nucleic acid-templated transcription	DXH0100317.1, DXH0104099.1, DXH0104890.1, DXH0119851.1, DXH0124065.1, DXH0127830.1, DXH0132985.1
GO:2001141	BP	1	1	regulation of RNA biosynthetic process	DXH0127830.1

Table S9. Functional descriptions of 22 PSGs

S.italica_v2.2	DXH01	DXH02	DXH03	DXH04	DXH05	DXH06	DXH07	DXH08	Function description
Seita.1G032400.1	DXH0100317.1	DXH0200354.1	DXH0300312.1	DXH0400318.1	DXH0500316.1	DXH0600324.1	DXH0700331.1	DXH0800317.1	protein CLP1 homolog
Seita.1G054100.1	DXH0100536.1	DXH0200571.1	DXH0300524.1	DXH0400543.1	DXH0500531.1	DXH0600540.1	DXH0700559.1	DXH0800536.1	methyl-CpG-binding domain-containing protein 9 isoform X2
Seita.1G097800.1	DXH0100954.1	DXH0201003.1	DXH0300955.1	DXH0400982.1	DXH0500968.1	DXH0600973.1	DXH0700988.1	DXH0800985.1	titin homolog
Seita.1G247500.1	DXH0102447.1	DXH0202491.1	DXH0302454.1	DXH0402446.1	DXH0502520.1	DXH0602480.1	DXH0702424.1	DXH0802491.1	alcohol dehydrogenase-like 6
Seita.1G353500.1	DXH0103502.1	DXH0203579.1	DXH0303519.1	DXH0403510.1	DXH0503580.1	DXH0603542.1	DXH0703458.1	DXH0803545.1	uncharacterized protein LOC101759919 isoform X2
Seita.2G031300.1	DXH0104099.1	DXH0204222.1	DXH0304115.1	DXH0404131.1	DXH0504022.1	DXH0604148.1	DXH0704082.1	DXH0804149.1	photosystem I reaction center subunit psaK, chloroplastic
Seita.2G112900.1	DXH0104890.1	DXH0205043.1	DXH0304898.1	DXH0404967.1	DXH0505034.1	DXH0604942.1	DXH0704922.1	DXH0804933.1	hypothetical protein PAHAL_B01547
Seita.3G032600.1	DXH0108499.1	DXH0208623.1	DXH0308457.1	DXH0408557.1	DXH0508702.1	DXH0608664.1	DXH0708516.1	DXH0808708.1	uncharacterized protein
Seita.3G120400.1	DXH0109346.1	DXH0209463.1	DXH0309315.1	DXH0409418.1	DXH0509556.1	DXH0609523.1	DXH0709355.1	DXH0809549.1	uncharacterized protein
Seita.3G193400.1	DXH0110093.1	DXH0210188.1	DXH0310055.1	DXH0410157.1	DXH0510297.1	DXH0610245.1	DXH0710076.1	DXH0810255.1	zinc finger protein 830
Seita.3G271200.1	DXH0110899.1	DXH0210920.1	DXH0310838.1	DXH0410896.1	DXH0511059.1	DXH0610999.1	DXH0710833.1	DXH0811084.1	mitochondrial zinc maintenance protein 1, mitochondrial
Seita.4G001000.1	DXH0112303.1	DXH0212268.1	DXH0312288.1	DXH0412215.1	DXH0512483.1	DXH0612409.1	DXH0712172.1	DXH0834336.1	BTB/POZ domain-containing protein
Seita.4G099600.1	DXH0113274.1	DXH0213248.1	DXH0313299.1	DXH0413208.1	DXH0513461.1	DXH0613380.1	DXH0713145.1	DXH0813521.1	GDSL esterase/lipase
Seita.5G353300.1	DXH0118659.1	DXH0218522.1	DXH0318634.1	DXH0418690.1	DXH0518963.1	DXH0618790.1	DXH0718466.1	DXH0818792.1	MIZU-KUSSEI 1
Seita.5G468400.1	DXH0119851.1	DXH0219691.1	DXH0319794.1	DXH0419868.1	DXH0520145.1	DXH0619953.1	DXH0719617.1	DXH0819946.1	probable WRKY transcription factor 35
Seita.6G035600.1	DXH0120220.1	DXH0220096.1	DXH0320177.1	DXH0420305.1	DXH0520518.1	DXH0620334.1	DXH0720024.1	DXH0820331.1	nucleolin 1
Seita.6G161800.1	DXH0121385.1	DXH0221204.1	DXH0321387.1	DXH0421438.1	DXH0521742.1	DXH0621542.1	DXH0721176.1	DXH0821530.1	ENHANCED DISEASE RESISTANCE 2-like isoform X2
Seita.7G178400.1	DXH0124065.1	DXH0223821.1	DXH0324108.1	DXH0424047.1	DXH0524574.1	DXH0624153.1	DXH0723792.1	DXH0824254.1	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like
Seita.7G187600.1	DXH0124176.1	DXH0223941.1	DXH0324222.1	DXH0424163.1	DXH0524688.1	DXH0624268.1	DXH0723906.1	DXH0824367.1	uncharacterized protein LOC101759402
Seita.8G222300.1	DXH0127830.1	DXH0227618.1	DXH0327890.1	DXH0427688.1	DXH0528311.1	DXH0627780.1	DXH0727112.1	DXH0827870.1	magnesium/proton exchanger 1
Seita.9G350300.1	DXH0131605.1	DXH0231292.1	DXH0331765.1	DXH0431417.1	DXH0532228.1	DXH0631565.1	DXH0730901.1	DXH0831663.1	mitochondrial inner membrane protein OXA1
Seita.9G496100.1	DXH0132985.1	DXH0232706.1	DXH0333150.1	DXH0432835.1	DXH0533656.1	DXH0632958.1	DXH0732298.1	DXH0833054.1	methionine--tRNA ligase, chloroplastic/mitochondrial

Table S10. Three hundred and fifty completely present/absent genes identified in the eight millet accessions

Gene_ID	Chrom	Gene_Start	Gene_End	PAV_ID	PAV_Start	PAV_End	PAV_length	Alt	DXH01	DXH02	DXH03	DXH04	DXH05	DXH06	DXH07	DXH08
Seita.1G006000	1	412,761	414,791	1_412238	412,238	416,433	4,195	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.1G031200	1	2,919,611	2,920,674	1_2917348	2,917,348	2,921,730	4,382	*	Ref	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.1G042900	1	4,256,323	4,256,844	1_4255500	4,255,500	4,258,688	3,188	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.1G043000	1	4,257,015	4,257,911	1_4255500	4,255,500	4,258,688	3,188	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.1G043100	1	4,264,137	4,264,781	1_4263932	4,263,932	4,265,548	1,616	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.1G047300	1	4,627,153	4,632,208	1_4622637	4,622,637	4,635,389	12,752	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.1G047400	1	4,634,026	4,635,136	1_4622637	4,622,637	4,635,389	12,752	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.1G053400	1	5,146,435	5,147,109	1_5138558	5,138,558	5,147,762	9,204	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.1G070600	1	6,489,412	6,490,374	1_6484290	6,484,290	6,492,053	7,763	C	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.1G074400	1	6,761,216	6,762,112	1_6761004	6,761,004	6,762,301	1,297	*	Ref	Ref	Ref	A	A	Ref	A	A
Seita.1G074500	1	6,763,713	6,764,247	1_6763489	6,763,489	6,770,102	6,613	*	Ref	Ref	Ref	A	A	Ref	A	A
Seita.1G074600	1	6,768,659	6,769,303	1_6763489	6,763,489	6,770,102	6,613	*	Ref	Ref	Ref	A	A	Ref	A	A
Seita.1G083800	1	7,458,439	7,458,765	1_7458118	7,458,118	7,458,882	764	*	Ref	Ref	A	Ref	Ref	Ref	Ref	A
Seita.1G114400	1	9,944,945	9,946,209	1_9944490	9,944,490	9,946,490	2,000	C	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.1G119600	1	10,337,559	10,342,668	1_10336556	10,336,556	10,352,748	16,192	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.1G123100	1	10,715,712	10,717,484	1_10715269	10,715,269	10,719,581	4,312	*	Ref	Ref	Ref	A	A	A	Ref	A
Seita.1G126100	1	11,169,992	11,170,418	1_11169547	11,169,547	11,170,640	1,093	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.1G129100	1	11,707,078	11,708,163	1_11706134	11,706,134	11,709,121	2,987	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.1G131000	1	12,167,473	12,168,532	1_12155686	12,155,686	12,169,420	13,734	A	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.1G133600	1	12,765,043	12,765,851	1_12764336	12,764,336	12,767,052	2,716	C	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.1G133700	1	12,765,438	12,766,717	1_12764336	12,764,336	12,767,052	2,716	C	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.1G151500	1	20,296,327	20,296,858	1_20292327	20,292,327	20,302,203	9,876	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.1G157300	1	21,751,846	21,753,075	1_21748708	21,748,708	21,754,515	5,807	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.1G157900	1	22,008,204	22,008,681	1_22007658	22,007,658	22,010,330	2,672	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref

Seita.1G158000	1	22,009,496	22,010,105	1_22007658	22,007,658	22,010,330	2,672	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.1G163600	1	23,487,974	23,489,083	1_23485338	23,485,338	23,498,153	12,815	*	A	A	A	A	A	A	A	A
Seita.1G184400	1	26,524,443	26,525,655	1_26522647	26,522,647	26,526,227	3,580	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.1G188400	1	27,090,279	27,090,668	1_27089103	27,089,103	27,094,314	5,211	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.1G188500	1	27,091,704	27,093,739	1_27089103	27,089,103	27,094,314	5,211	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.1G191500	1	27,367,304	27,367,950	1_27367116	27,367,116	27,368,779	1,663	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.1G200800	1	28,066,356	28,067,645	1_28062030	28,062,030	28,068,282	6,252	T	Ref	Ref	Ref	A	A	A	A	Ref
Seita.1G225700	1	30,290,223	30,291,430	1_30289962	30,289,962	30,291,521	1,559	*	Ref	A	A	Ref	Ref	Ref	Ref	A
Seita.1G232900	1	30,935,591	30,936,842	1_30931650	30,931,650	30,941,078	9,428	A	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.1G233000	1	30,937,319	30,937,979	1_30931650	30,931,650	30,941,078	9,428	A	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.1G234200	1	31,286,822	31,288,740	1_31284945	31,284,945	31,296,000	11,055	*	Ref	A	Ref	Ref	A	A	Ref	Ref
Seita.1G234300	1	31,292,418	31,294,372	1_31284945	31,284,945	31,296,000	11,055	*	Ref	A	Ref	Ref	A	A	Ref	Ref
Seita.1G234600	1	31,322,802	31,325,991	1_31316555	31,316,555	31,326,060	9,505	*	Ref	Ref	Ref	A	A	A	A	A
Seita.1G243700	1	32,127,503	32,129,755	1_32127429	32,127,429	32,129,949	2,520	*	A	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.1G256500	1	33,149,836	33,150,415	1_33148001	33,148,001	33,152,330	4,329	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.1G320900	1	38,171,398	38,173,439	1_38170213	38,170,213	38,173,930	3,717	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.1G361600	1	41,120,917	41,121,725	1_41120837	41,120,837	41,121,991	1,154	*	Ref	A	Ref	A	A	A	Ref	Ref
Seita.2G010100	2	585,924	588,893	2_585902	585,902	589,104	3,202	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.2G012200	2	773,927	774,995	2_770837	770,837	775,439	4,602	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.2G013100	2	894,138	894,642	2_894045	894,045	895,255	1,210	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.2G015000	2	1,109,678	1,111,177	2_1107945	1,107,945	1,113,302	5,357	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.2G030100	2	2,439,575	2,440,246	2_2439380	2,439,380	2,441,780	2,400	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.2G052300	2	4,186,927	4,188,200	2_4186514	4,186,514	4,189,625	3,111	*	Ref	Ref	Ref	Ref	Ref	A	Ref	A
Seita.2G059500	2	4,952,562	4,953,111	2_4947201	4,947,201	4,953,332	6,131	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.2G065500	2	5,434,052	5,434,696	2_5433243	5,433,243	5,439,805	6,562	*	Ref	A	A	A	A	A	Ref	A
Seita.2G065600	2	5,439,109	5,439,643	2_5433243	5,433,243	5,439,805	6,562	*	Ref	A	A	A	A	A	Ref	A

Seita.2G065700	2	5,440,735	5,441,631	2_5439964	5,439,964	5,443,156	3,192	*	Ref	A	A	A	Ref	A	Ref	A
Seita.2G065800	2	5,441,811	5,442,332	2_5439964	5,439,964	5,443,156	3,192	*	Ref	A	A	A	Ref	A	Ref	A
Seita.2G067900	2	5,663,604	5,664,487	2_5663547	5,663,547	5,665,051	1,504	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.2G068000	2	5,664,543	5,664,906	2_5663547	5,663,547	5,665,051	1,504	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.2G086000	2	7,732,968	7,733,771	2_7732375	7,732,375	7,735,294	2,919	G	Ref	A	Ref	A	A	A	Ref	A
Seita.2G100200	2	9,029,275	9,029,809	2_9029052	9,029,052	9,036,016	6,964	*	Ref	Ref	A	A	Ref	Ref	Ref	Ref
Seita.2G114100	2	11,228,746	11,229,810	2_11225540	11,225,540	11,232,150	6,610	C	Ref	Ref	A	A	A	Ref	Ref	Ref
Seita.2G114200	2	11,231,078	11,232,130	2_11225540	11,225,540	11,232,150	6,610	C	Ref	Ref	A	A	A	Ref	Ref	Ref
Seita.2G118500	2	12,214,417	12,215,044	2_12213991	12,213,991	12,215,975	1,984	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.2G120500	2	12,703,261	12,705,153	2_12702741	12,702,741	12,705,372	2,631	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.2G149000	2	20,211,765	20,215,767	2_20211280	20,211,280	20,216,122	4,842	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.2G168300	2	25,097,532	25,097,825	2_25097261	25,097,261	25,097,876	615	A	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.2G171700	2	25,733,589	25,734,400	2_25712518	25,712,518	25,735,312	22,794	G	Ref	A	A	A	A	A	A	A
Seita.2G173500	2	26,192,214	26,193,143	2_26191922	26,191,922	26,193,444	1,522	*	Ref	A	A	A	Ref	A	A	A
Seita.2G196400	2	29,313,165	29,313,546	2_29312676	29,312,676	29,313,791	1,115	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.2G219100	2	32,117,325	32,119,641	2_32116709	32,116,709	32,121,454	4,745	C	Ref	Ref	Ref	A	Ref	Ref	Ref	A
Seita.2G233000	2	33,566,091	33,566,480	2_33565922	33,565,922	33,572,297	6,375	T	Ref	Ref	Ref	A	A	A	Ref	A
Seita.2G342800	2	42,499,078	42,500,589	2_42487703	42,487,703	42,502,285	14,582	*	A	Ref	Ref	A	A	A	A	Ref
Seita.2G424700	2	47,751,179	47,751,961	2_47751097	47,751,097	47,752,159	1,062	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.2G430100	2	48,120,875	48,121,339	2_48117887	48,117,887	48,122,308	4,421	*	Ref	A	Ref	A	A	Ref	Ref	Ref
Seita.2G430200	2	48,123,120	48,124,241	2_48122310	48,122,310	48,125,005	2,695	*	Ref	A	Ref	A	A	Ref	Ref	Ref
Seita.2G441000	2	48,943,209	48,943,436	2_48943189	48,943,189	48,943,536	347	*	A	A	A	A	A	A	A	A
Seita.3G013500	3	727,915	728,070	3_717643	717,643	734,757	17,114	*	Ref	A	A	Ref	Ref	Ref	Ref	A
Seita.3G013600	3	729,465	734,557	3_717643	717,643	734,757	17,114	*	Ref	A	A	Ref	Ref	Ref	Ref	A
Seita.3G022200	3	1,357,086	1,357,607	3_1356254	1,356,254	1,359,070	2,816	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.3G022300	3	1,357,743	1,358,686	3_1356254	1,356,254	1,359,070	2,816	*	Ref	Ref	Ref	A	A	A	Ref	Ref

Seita.3G022400	3	1,359,824	1,360,358	3_1359182	1,359,182	1,366,195	7,013	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.3G022500	3	1,364,781	1,365,434	3_1359182	1,359,182	1,366,195	7,013	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.3G087500	3	5,629,715	5,631,061	3_5629568	5,629,568	5,631,706	2,138	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.3G094200	3	6,161,841	6,162,362	3_6160997	6,160,997	6,163,343	2,346	*	Ref	Ref	Ref	Ref	A	A	Ref	A
Seita.3G094400	3	6,164,475	6,165,009	3_6164409	6,164,409	6,170,826	6,417	*	Ref	Ref	Ref	Ref	A	A	Ref	A
Seita.3G094500	3	6,169,422	6,170,066	3_6164409	6,164,409	6,170,826	6,417	*	Ref	Ref	Ref	Ref	A	A	Ref	A
Seita.3G106200	3	7,007,175	7,007,819	3_7006407	7,006,407	7,016,263	9,856	A	Ref	Ref	Ref	Ref	Ref	Ref	A	A
Seita.3G106300	3	7,013,839	7,014,735	3_7006407	7,006,407	7,016,263	9,856	A	Ref	Ref	Ref	Ref	Ref	Ref	A	A
Seita.3G106400	3	7,014,915	7,015,436	3_7006407	7,006,407	7,016,263	9,856	A	Ref	Ref	Ref	Ref	Ref	Ref	A	A
Seita.3G118300	3	7,966,487	7,971,449	3_7965783	7,965,783	7,971,465	5,682	*	Ref	Ref	Ref	A	A	Ref	Ref	A
Seita.3G118400	3	7,975,233	7,977,538	3_7974343	7,974,343	7,982,075	7,732	*	Ref	Ref	Ref	A	A	Ref	Ref	A
Seita.3G118500	3	7,978,373	7,980,347	3_7974343	7,974,343	7,982,075	7,732	*	Ref	Ref	Ref	A	A	Ref	Ref	A
Seita.3G141900	3	9,960,775	9,962,287	3_9959529	9,959,529	9,968,951	9,422	*	Ref	A	A	A	A	A	Ref	A
Seita.3G142000	3	9,965,151	9,966,055	3_9959529	9,959,529	9,968,951	9,422	*	Ref	A	A	A	A	A	Ref	A
Seita.3G144700	3	10,239,589	10,239,699	3_10239206	10,239,206	10,239,766	560	*	Ref	A	A	A	A	A	Ref	A
Seita.3G172600	3	12,772,726	12,772,938	3_12770826	12,770,826	12,773,364	2,538	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.3G234300	3	19,510,252	19,510,749	3_19508822	19,508,822	19,512,513	3,691	*	Ref	A	A	Ref	A	Ref	Ref	Ref
Seita.3G234500	3	19,520,448	19,522,947	3_19518517	19,518,517	19,525,908	7,391	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.3G234600	3	19,523,782	19,525,756	3_19518517	19,518,517	19,525,908	7,391	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.3G284800	3	26,806,748	26,808,875	3_26805439	26,805,439	26,809,880	4,441	*	Ref	A	A	A	A	A	A	A
Seita.3G285100	3	26,865,084	26,865,634	3_26862999	26,862,999	26,866,187	3,188	*	Ref	A	A	A	A	A	A	A
Seita.3G285700	3	26,980,654	26,983,143	3_26980325	26,980,325	26,987,474	7,149	*	A	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.3G310500	3	37,307,128	37,307,938	3_37305784	37,305,784	37,320,041	14,257	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.3G316200	3	39,080,925	39,081,569	3_39079536	39,079,536	39,086,003	6,467	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.3G316400	3	39,087,274	39,088,170	3_39086152	39,086,152	39,089,014	2,862	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.3G316500	3	39,088,291	39,088,812	3_39086152	39,086,152	39,089,014	2,862	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref

Seita.3G317200	3	39,223,376	39,224,402	3_39221185	39,221,185	39,225,334	4,149	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.3G323500	3	40,625,585	40,626,106	3_40624131	40,624,131	40,627,639	3,508	*	Ref	A	A	A	Ref	Ref	Ref	Ref
Seita.3G323600	3	40,626,286	40,627,182	3_40624131	40,624,131	40,627,639	3,508	*	Ref	A	A	A	Ref	Ref	Ref	Ref
Seita.3G323700	3	40,628,702	40,629,236	3_40627703	40,627,703	40,634,450	6,747	*	Ref	A	A	A	Ref	Ref	Ref	Ref
Seita.3G327200	3	41,138,269	41,140,101	3_41137736	41,137,736	41,142,094	4,358	G	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.3G327300	3	41,140,260	41,140,955	3_41137736	41,137,736	41,142,094	4,358	G	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.3G331400	3	42,384,431	42,384,886	3_42383329	42,383,329	42,387,052	3,723	G	A	Ref	Ref	A	A	Ref	Ref	Ref
Seita.3G337500	3	43,211,352	43,212,145	3_43210488	43,210,488	43,215,415	4,927	*	Ref	Ref	Ref	A	Ref	A	A	A
Seita.3G337600	3	43,213,206	43,214,460	3_43210488	43,210,488	43,215,415	4,927	*	Ref	Ref	Ref	A	Ref	A	A	A
Seita.3G339700	3	43,490,640	43,490,888	3_43488421	43,488,421	43,498,829	10,408	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.3G356400	3	45,707,644	45,708,101	3_45706852	45,706,852	45,709,050	2,198	*	Ref	Ref	Ref	Ref	A	A	A	A
Seita.3G366200	3	46,954,549	46,954,839	3_46953527	46,953,527	46,955,877	2,350	*	A	A	A	A	A	A	A	A
Seita.3G388500	3	48,995,855	48,997,036	3_48994143	48,994,143	48,997,040	2,897	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.3G393600	3	49,486,303	49,487,643	3_49481799	49,481,799	49,491,845	10,046	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.4G022200	4	1,471,928	1,472,248	4_1466372	1,466,372	1,473,987	7,615	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.4G037400	4	2,522,065	2,524,552	4_2522059	2,522,059	2,525,021	2,962	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.4G037500	4	2,536,980	2,538,233	4_2526598	2,526,598	2,540,349	13,751	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.4G073100	4	5,635,802	5,637,457	4_5633323	5,633,323	5,637,954	4,631	*	A	A	A	A	A	A	Ref	Ref
Seita.4G073200	4	5,648,131	5,648,668	4_5637955	5,637,955	5,654,495	16,540	*	A	A	A	A	A	A	A	Ref
Seita.4G073300	4	5,649,496	5,650,661	4_5637955	5,637,955	5,654,495	16,540	*	A	A	A	A	A	A	A	Ref
Seita.4G073400	4	5,650,927	5,651,742	4_5637955	5,637,955	5,654,495	16,540	*	A	A	A	A	A	A	A	Ref
Seita.4G073500	4	5,652,388	5,653,292	4_5637955	5,637,955	5,654,495	16,540	*	A	A	A	A	A	A	A	Ref
Seita.4G080800	4	6,321,229	6,325,482	4_6321026	6,321,026	6,326,982	5,956	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.4G085300	4	6,923,262	6,924,622	4_6922441	6,922,441	6,924,811	2,370	*	A	Ref	Ref	A	A	A	A	A
Seita.4G087000	4	7,144,355	7,144,739	4_7143374	7,143,374	7,145,894	2,520	*	A	Ref	Ref	A	A	Ref	Ref	Ref
Seita.4G088800	4	7,401,510	7,403,186	4_7401313	7,401,313	7,403,319	2,006	G	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref

Seita.4G123300	4	12,653,571	12,654,822	4_12648713	12,648,713	12,658,558	9,845	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.4G123400	4	12,654,244	12,657,704	4_12648713	12,648,713	12,658,558	9,845	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.4G166600	4	26,144,746	26,144,996	4_26144625	26,144,625	26,145,042	417	*	A	A	A	A	A	A	A	A
Seita.4G230700	4	35,210,442	35,211,229	4_35210312	35,210,312	35,211,648	1,336	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.4G231300	4	35,280,320	35,283,452	4_35280199	35,280,199	35,283,567	3,368	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.4G236800	4	36,066,043	36,066,707	4_36055212	36,055,212	36,067,893	12,681	*	Ref	Ref	Ref	Ref	A	Ref	A	Ref
Seita.4G269400	4	38,629,972	38,630,192	4_38629367	38,629,367	38,635,690	6,323	*	A	A	A	Ref	Ref	A	A	A
Seita.4G292500	4	40,379,841	40,382,270	4_40376402	40,376,402	40,382,758	6,356	*	Ref	Ref	Ref	A	A	A	A	A
Seita.5G016300	5	1,955,530	1,957,096	5_1954926	1,954,926	1,957,444	2,518	*	A	Ref	Ref	Ref	Ref	Ref	A	A
Seita.5G024300	5	2,455,782	2,457,676	5_2455590	2,455,590	2,458,609	3,019	*	A	A	A	Ref	A	Ref	A	A
Seita.5G035100	5	3,336,075	3,337,683	5_3335618	3,335,618	3,337,973	2,355	*	Ref	A	Ref	Ref	A	A	A	A
Seita.5G035400	5	3,347,888	3,349,140	5_3344958	3,344,958	3,349,525	4,567	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.5G050100	5	4,483,748	4,484,128	5_4483430	4,483,430	4,484,603	1,173	*	A	Ref	A	Ref	Ref	Ref	A	A
Seita.5G079800	5	6,803,264	6,806,599	5_6802407	6,802,407	6,806,918	4,511	T	Ref	A	A	A	A	A	A	A
Seita.5G090600	5	7,717,990	7,720,445	5_7716825	7,716,825	7,721,839	5,014	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.5G091500	5	7,811,236	7,811,757	5_7809578	7,809,578	7,813,076	3,498	*	Ref	Ref	Ref	Ref	A	Ref	Ref	A
Seita.5G091600	5	7,811,937	7,812,836	5_7809578	7,809,578	7,813,076	3,498	*	Ref	Ref	Ref	Ref	A	Ref	Ref	A
Seita.5G100800	5	8,560,801	8,561,493	5_8560250	8,560,250	8,562,847	2,597	*	A	Ref	Ref	Ref	Ref	A	A	A
Seita.5G111200	5	9,338,577	9,339,188	5_9334047	9,334,047	9,340,002	5,955	T	Ref	A	Ref	A	A	Ref	Ref	Ref
Seita.5G112200	5	9,388,404	9,389,798	5_9385839	9,385,839	9,389,848	4,009	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.5G131300	5	11,220,808	11,221,302	5_11220410	11,220,410	11,222,095	1,685	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.5G181700	5	22,686,606	22,687,497	5_22681587	22,681,587	22,691,244	9,657	*	A	A	A	A	A	Ref	A	A
Seita.5G181800	5	22,687,629	22,690,198	5_22681587	22,681,587	22,691,244	9,657	*	A	A	A	A	A	Ref	A	A
Seita.5G182000	5	22,692,961	22,693,535	5_22691300	22,691,300	22,695,647	4,347	*	A	A	A	A	A	Ref	A	A
Seita.5G182100	5	22,694,181	22,695,085	5_22691300	22,691,300	22,695,647	4,347	*	A	A	A	A	A	Ref	A	A
Seita.5G185800	5	23,674,183	23,677,989	5_23665046	23,665,046	23,680,745	15,699	*	A	A	A	A	A	Ref	A	A

Seita.5G185900	5	23,684,830	23,686,946	5_23682672	23,682,672	23,697,513	14,841	*	A	A	A	A	Ref	Ref	A	Ref
Seita.5G188300	5	24,048,645	24,049,419	5_24048038	24,048,038	24,050,283	2,245	*	A	A	A	A	A	A	A	A
Seita.5G197300	5	25,356,853	25,357,439	5_25354945	25,354,945	25,359,125	4,180	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.5G200800	5	25,822,666	25,823,466	5_25821075	25,821,075	25,825,423	4,348	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.5G200900	5	25,824,111	25,825,015	5_25821075	25,821,075	25,825,423	4,348	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.5G206100	5	26,455,184	26,456,286	5_26454697	26,454,697	26,456,364	1,667	*	Ref	Ref	Ref	Ref	A	Ref	Ref	A
Seita.5G206300	5	26,461,414	26,461,942	5_26456368	26,456,368	26,467,892	11,524	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.5G208000	5	26,747,621	26,748,161	5_26740912	26,740,912	26,751,207	10,295	*	Ref	A	A	Ref	A	Ref	Ref	Ref
Seita.5G208100	5	26,749,936	26,751,160	5_26740912	26,740,912	26,751,207	10,295	*	Ref	A	A	Ref	A	Ref	Ref	Ref
Seita.5G220900	5	28,061,383	28,061,822	5_28058956	28,058,956	28,062,451	3,495	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.5G227500	5	28,898,693	28,899,996	5_28896720	28,896,720	28,914,914	18,194	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.5G227600	5	28,901,212	28,902,267	5_28896720	28,896,720	28,914,914	18,194	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.5G227700	5	28,902,480	28,902,808	5_28896720	28,896,720	28,914,914	18,194	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.5G232700	5	29,497,776	29,498,818	5_29497331	29,497,331	29,501,190	3,859	*	Ref	A	A	Ref	Ref	A	A	A
Seita.5G244000	5	30,651,085	30,652,343	5_30650644	30,650,644	30,657,175	6,531	*	Ref	A	A	A	A	A	A	A
Seita.5G247600	5	30,897,010	30,899,956	5_30896741	30,896,741	30,904,575	7,834	*	Ref	A	A	A	Ref	A	A	A
Seita.5G276700	5	33,795,898	33,796,455	5_33793031	33,793,031	33,798,044	5,013	A	Ref	A	A	A	A	A	A	A
Seita.5G283900	5	34,318,674	34,319,210	5_34316742	34,316,742	34,319,265	2,523	C	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.5G298800	5	35,416,252	35,416,853	5_35412820	35,412,820	35,417,796	4,976	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.5G371000	5	40,744,709	40,748,693	5_40731535	40,731,535	40,751,152	19,617	*	A	A	A	A	A	A	A	A
Seita.5G412200	5	43,597,068	43,600,008	5_43595498	43,595,498	43,600,328	4,830	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.5G449400	5	46,104,308	46,107,247	5_46103605	46,103,605	46,109,867	6,262	*	Ref	A	A	A	A	Ref	Ref	Ref
Seita.5G459000	5	46,680,297	46,681,223	5_46678788	46,678,788	46,682,699	3,911	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.5G462700	5	46,857,694	46,858,353	5_46857287	46,857,287	46,862,856	5,569	*	A	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.6G011900	6	811,440	813,634	6_810328	810,328	814,137	3,809	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.6G012300	6	829,568	829,737	6_829560	829,560	829,795	235	C	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref

Seita.6G015300	6	1,037,250	1,038,417	6_1034141	1,034,141	1,038,873	4,732	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.6G017300	6	1,223,782	1,224,260	6_1223126	1,223,126	1,224,519	1,393	*	A	A	Ref	Ref	A	Ref	Ref	A
Seita.6G017400	6	1,228,433	1,228,834	6_1227788	1,227,788	1,229,173	1,385	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.6G020800	6	1,482,978	1,484,210	6_1482774	1,482,774	1,484,657	1,883	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.6G024300	6	1,740,299	1,741,914	6_1739953	1,739,953	1,743,793	3,840	A	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.6G028200	6	2,078,016	2,078,677	6_2077668	2,077,668	2,079,262	1,594	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.6G028300	6	2,080,207	2,080,665	6_2079319	2,079,319	2,080,904	1,585	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.6G028400	6	2,082,286	2,082,744	6_2081913	2,081,913	2,083,401	1,488	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G028500	6	2,084,413	2,084,871	6_2083811	2,083,811	2,085,109	1,298	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G034600	6	2,611,554	2,611,929	6_2609758	2,609,758	2,613,110	3,352	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G037600	6	2,875,703	2,876,506	6_2875459	2,875,459	2,887,135	11,676	*	A	A	A	A	A	A	A	A
Seita.6G037700	6	2,877,078	2,877,707	6_2875459	2,875,459	2,887,135	11,676	*	A	A	A	A	A	A	A	A
Seita.6G037800	6	2,880,380	2,880,917	6_2875459	2,875,459	2,887,135	11,676	*	A	A	A	A	A	A	A	A
Seita.6G037900	6	2,882,466	2,882,657	6_2875459	2,875,459	2,887,135	11,676	*	A	A	A	A	A	A	A	A
Seita.6G038200	6	2,906,011	2,906,889	6_2906011	2,906,011	2,907,207	1,196	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.6G038600	6	2,948,168	2,949,266	6_2947871	2,947,871	2,956,426	8,555	*	Ref	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.6G038900	6	2,965,278	2,966,175	6_2964703	2,964,703	2,966,415	1,712	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G039000	6	2,969,148	2,970,156	6_2968089	2,968,089	2,970,596	2,507	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G044000	6	3,394,122	3,396,846	6_3394119	3,394,119	3,397,829	3,710	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.6G044400	6	3,427,792	3,429,080	6_3426299	3,426,299	3,429,742	3,443	*	Ref	Ref	Ref	A	Ref	Ref	Ref	A
Seita.6G052600	6	4,120,675	4,121,196	6_4119840	4,119,840	4,123,311	3,471	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.6G052700	6	4,121,317	4,122,213	6_4119840	4,119,840	4,123,311	3,471	*	Ref	Ref	Ref	A	A	Ref	Ref	Ref
Seita.6G052800	6	4,123,612	4,124,146	6_4123547	4,123,547	4,131,926	8,379	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G052900	6	4,128,577	4,129,221	6_4123547	4,123,547	4,131,926	8,379	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G057000	6	4,571,249	4,573,131	6_4567772	4,567,772	4,578,985	11,213	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.6G065000	6	5,534,808	5,536,025	6_5534398	5,534,398	5,536,060	1,662	*	Ref	A	Ref	Ref	Ref	Ref	Ref	Ref

Seita.6G066500	6	5,683,329	5,683,865	6_5682678	5,682,678	5,683,878	1,200	*	Ref	A	A	Ref	Ref	A	Ref	A
Seita.6G076500	6	6,794,496	6,795,016	6_6793707	6,793,707	6,795,848	2,141	*	Ref	A	A	A	A	A	A	A
Seita.6G086600	6	7,695,175	7,699,130	6_7695170	7,695,170	7,703,196	8,026	*	Ref	A	Ref	Ref	Ref	A	A	A
Seita.6G086700	6	7,700,421	7,701,384	6_7695170	7,695,170	7,703,196	8,026	*	Ref	A	Ref	Ref	Ref	A	A	A
Seita.6G096100	6	10,096,714	10,097,834	6_10096137	10,096,137	10,099,736	3,599	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.6G102400	6	12,741,455	12,742,621	6_12740468	12,740,468	12,744,434	3,966	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G102500	6	12,742,838	12,743,467	6_12740468	12,740,468	12,744,434	3,966	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G102700	6	12,745,414	12,745,951	6_12744525	12,744,525	12,751,955	7,430	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G102800	6	12,746,904	12,749,060	6_12744525	12,744,525	12,751,955	7,430	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G102900	6	12,749,705	12,750,607	6_12744525	12,744,525	12,751,955	7,430	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G111200	6	17,264,483	17,265,634	6_17263121	17,263,121	17,270,160	7,039	*	Ref	Ref	Ref	A	Ref	A	Ref	Ref
Seita.6G134400	6	24,062,850	24,063,906	6_24062573	24,062,573	24,063,997	1,424	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G134500	6	24,068,138	24,068,638	6_24066570	24,066,570	24,068,805	2,235	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G145600	6	25,966,444	25,967,989	6_25964976	25,964,976	25,969,842	4,866	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.6G149100	6	26,476,788	26,483,742	6_26475967	26,475,967	26,486,279	10,312	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.6G152800	6	27,216,309	27,216,680	6_27209249	27,209,249	27,218,430	9,181	*	Ref	A	A	Ref	A	A	Ref	Ref
Seita.6G152900	6	27,226,851	27,227,860	6_27221869	27,221,869	27,229,835	7,966	*	Ref	A	A	Ref	A	A	Ref	Ref
Seita.6G153600	6	27,321,092	27,322,024	6_27318960	27,318,960	27,323,297	4,337	*	Ref	A	Ref	Ref	Ref	Ref	Ref	A
Seita.6G174600	6	29,833,596	29,833,893	6_29833557	29,833,557	29,833,904	347	*	A	A	A	A	A	A	A	A
Seita.6G199200	6	32,011,696	32,012,230	6_32011212	32,011,212	32,015,720	4,508	*	Ref	A	A	A	A	A	Ref	A
Seita.6G229300	6	34,253,501	34,256,821	6_34253037	34,253,037	34,257,791	4,754	*	A	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Seita.6G240200	6	35,062,244	35,062,765	6_35060661	35,060,661	35,063,204	2,543	*	A	A	A	A	A	A	A	Ref
Seita.6G240400	6	35,064,675	35,065,209	6_35064396	35,064,396	35,066,010	1,614	*	A	A	A	A	A	A	A	A
Seita.6G247300	6	35,491,213	35,492,553	6_35491160	35,491,160	35,493,467	2,307	*	Ref	A	A	A	A	A	A	Ref
Seita.7G000600	7	104,070	104,528	7_100964	100,964	105,083	4,119	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.7G007400	7	1,238,223	1,240,016	7_1237511	1,237,511	1,240,325	2,814	*	Ref	Ref	A	Ref	A	A	A	A

Seita.7G009700	7	1,699,741	1,700,211	7_1699722	1,699,722	1,700,324	602	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.7G013600	7	2,429,018	2,430,775	7_2428561	2,428,561	2,437,755	9,194	*	Ref	A	A	A	Ref	A	A	A
Seita.7G019100	7	3,663,749	3,665,017	7_3663696	3,663,696	3,667,860	4,164	*	Ref	A	A	Ref	Ref	Ref	Ref	A
Seita.7G029900	7	8,503,848	8,505,645	7_8502680	8,502,680	8,506,124	3,444	*	Ref	A	A	A	A	A	A	A
Seita.7G031300	7	9,470,132	9,470,481	7_9469298	9,469,298	9,476,431	7,133	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.7G035200	7	11,204,847	11,205,661	7_11204536	11,204,536	11,207,372	2,836	*	Ref	Ref	Ref	A	Ref	A	Ref	Ref
Seita.7G039600	7	12,119,332	12,120,860	7_12113665	12,113,665	12,121,468	7,803	*	Ref	A	A	A	A	A	A	A
Seita.7G041200	7	12,662,508	12,663,718	7_12660753	12,660,753	12,663,742	2,989	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.7G051700	7	14,740,518	14,741,996	7_14739899	14,739,899	14,742,280	2,381	*	Ref	A	A	Ref	Ref	A	A	Ref
Seita.7G051900	7	14,787,044	14,788,965	7_14785922	14,785,922	14,791,772	5,850	*	Ref	Ref	A	Ref	Ref	Ref	Ref	Ref
Seita.7G062000	7	16,412,878	16,413,412	7_16407036	16,407,036	16,417,299	10,263	*	Ref	Ref	Ref	A	A	Ref	Ref	A
Seita.7G062100	7	16,415,193	16,416,514	7_16407036	16,407,036	16,417,299	10,263	*	Ref	Ref	Ref	A	A	Ref	Ref	A
Seita.7G064300	7	16,679,889	16,680,509	7_16678582	16,678,582	16,682,508	3,926	*	Ref	Ref	Ref	A	Ref	A	A	A
Seita.7G064400	7	16,683,596	16,686,331	7_16683158	16,683,158	16,686,755	3,597	*	Ref	Ref	Ref	A	Ref	A	A	A
Seita.7G070500	7	17,429,676	17,430,305	7_17425200	17,425,200	17,438,256	13,056	*	Ref	Ref	Ref	A	A	A	A	A
Seita.7G070600	7	17,431,722	17,432,259	7_17425200	17,425,200	17,438,256	13,056	*	Ref	Ref	Ref	A	A	A	A	A
Seita.7G070700	7	17,433,086	17,434,252	7_17425200	17,425,200	17,438,256	13,056	*	Ref	Ref	Ref	A	A	A	A	A
Seita.7G070800	7	17,434,518	17,435,333	7_17425200	17,425,200	17,438,256	13,056	*	Ref	Ref	Ref	A	A	A	A	A
Seita.7G070900	7	17,435,979	17,436,883	7_17425200	17,425,200	17,438,256	13,056	*	Ref	Ref	Ref	A	A	A	A	A
Seita.7G082500	7	18,660,558	18,661,272	7_18660546	18,660,546	18,661,442	896	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.7G091000	7	19,463,143	19,463,677	7_19457304	19,457,304	19,464,137	6,833	*	Ref	Ref	Ref	A	Ref	A	A	A
Seita.7G102800	7	20,519,339	20,520,648	7_20517207	20,517,207	20,520,988	3,781	G	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.7G118300	7	21,542,604	21,543,664	7_21537836	21,537,836	21,545,294	7,458	*	Ref	Ref	A	A	Ref	Ref	A	A
Seita.7G125100	7	21,982,737	21,983,402	7_21982382	21,982,382	21,983,686	1,304	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.7G149300	7	23,675,874	23,679,586	7_23673660	23,673,660	23,684,615	10,955	T	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.7G186600	7	26,368,494	26,368,943	7_26368461	26,368,461	26,369,114	653	*	A	Ref	Ref	Ref	Ref	Ref	Ref	Ref

Seita.7G187100	7	26,389,779	26,390,423	7_26388969	26,388,969	26,398,933	9,964	*	A	A	Ref	A	A	Ref	Ref	Ref
Seita.7G187200	7	26,396,529	26,397,425	7_26388969	26,388,969	26,398,933	9,964	*	A	A	Ref	A	A	Ref	Ref	Ref
Seita.7G187300	7	26,397,546	26,398,067	7_26388969	26,388,969	26,398,933	9,964	*	A	A	Ref	A	A	Ref	Ref	Ref
Seita.7G189100	7	26,537,867	26,538,398	7_26537451	26,537,451	26,540,357	2,906	*	A	Ref	Ref	A	A	Ref	Ref	Ref
Seita.7G189200	7	26,542,712	26,543,356	7_26540359	26,540,359	26,544,145	3,786	*	A	Ref	Ref	A	A	Ref	Ref	Ref
Seita.7G191400	7	26,660,461	26,661,585	7_26660180	26,660,180	26,662,292	2,112	*	A	A	A	A	A	A	A	A
Seita.7G224100	7	28,967,288	28,967,769	7_28966754	28,966,754	28,969,216	2,462	*	A	A	Ref	Ref	Ref	A	A	A
Seita.7G234900	7	29,755,412	29,757,264	7_29755211	29,755,211	29,757,413	2,202	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.7G243200	7	30,375,616	30,375,975	7_30375448	30,375,448	30,377,635	2,187	T	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.7G249400	7	30,700,490	30,702,015	7_30697185	30,697,185	30,702,668	5,483	*	A	A	A	A	A	A	Ref	A
Seita.7G251600	7	30,842,713	30,844,342	7_30837779	30,837,779	30,847,076	9,297	*	Ref	Ref	Ref	A	Ref	A	Ref	A
Seita.7G253000	7	30,981,546	30,985,103	7_30979478	30,979,478	30,985,754	6,276	*	A	A	A	A	A	Ref	A	A
Seita.7G307100	7	34,479,954	34,480,970	7_34479893	34,479,893	34,481,525	1,632	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.8G007400	8	426,576	427,823	8_420524	420,524	427,846	7,322	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.8G010100	8	616,793	617,465	8_614580	614,580	617,819	3,239	*	Ref	Ref	Ref	Ref	A	A	A	Ref
Seita.8G010300	8	620,532	621,069	8_620423	620,423	621,393	970	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.8G010400	8	621,657	624,156	8_621612	621,612	626,624	5,012	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.8G010500	8	624,790	625,694	8_621612	621,612	626,624	5,012	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.8G021000	8	1,231,576	1,232,787	8_1229788	1,229,788	1,234,045	4,257	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.8G025200	8	1,545,508	1,546,622	8_1544937	1,544,937	1,547,171	2,234	*	Ref	Ref	Ref	Ref	A	A	A	A
Seita.8G046900	8	3,868,136	3,869,128	8_3867892	3,867,892	3,869,907	2,015	*	A	A	Ref	A	A	A	A	Ref
Seita.8G065800	8	6,497,259	6,497,561	8_6496611	6,496,611	6,501,369	4,758	*	Ref	A	Ref	A	Ref	Ref	A	Ref
Seita.8G065900	8	6,498,362	6,501,139	8_6496611	6,496,611	6,501,369	4,758	*	Ref	A	Ref	A	Ref	Ref	A	Ref
Seita.8G069700	8	7,108,263	7,109,978	8_7108070	7,108,070	7,113,232	5,162	*	Ref	A	Ref	Ref	Ref	A	Ref	Ref
Seita.8G072500	8	7,513,442	7,513,951	8_7512354	7,512,354	7,514,162	1,808	*	Ref	Ref	Ref	A	Ref	Ref	Ref	Ref
Seita.8G075700	8	8,204,740	8,205,970	8_8200228	8,200,228	8,209,556	9,328	G	A	Ref	Ref	A	A	A	A	A

Seita.8G084500	8	9,908,204	9,909,273	8_9908144	9,908,144	9,911,342	3,198	*	A	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.8G086000	8	10,343,377	10,349,142	8_10334374	10,334,374	10,352,181	17,807	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.8G086800	8	10,486,206	10,487,537	8_10485730	10,485,730	10,488,973	3,243	*	A	Ref	Ref	Ref	A	Ref	A	Ref
Seita.8G090500	8	11,044,941	11,046,071	8_11034774	11,034,774	11,046,135	11,361	*	A	Ref	Ref	Ref	Ref	A	Ref	A
Seita.8G100500	8	13,265,246	13,265,518	8_13265128	13,265,128	13,266,753	1,625	*	Ref	Ref	Ref	Ref	Ref	Ref	A	A
Seita.8G103800	8	13,926,295	13,927,538	8_13921710	13,921,710	13,939,161	17,451	*	Ref	Ref	Ref	A	A	A	A	A
Seita.8G103900	8	13,928,242	13,928,961	8_13921710	13,921,710	13,939,161	17,451	*	Ref	Ref	Ref	A	A	A	A	A
Seita.8G109200	8	16,761,773	16,762,669	8_16760876	16,760,876	16,763,824	2,948	*	A	Ref	Ref	Ref	A	Ref	A	Ref
Seita.8G109300	8	16,762,790	16,763,311	8_16760876	16,760,876	16,763,824	2,948	*	A	Ref	Ref	Ref	A	Ref	A	Ref
Seita.8G119600	8	22,590,612	22,591,024	8_22590369	22,590,369	22,591,426	1,057	*	A	A	Ref	A	A	A	A	A
Seita.8G137400	8	26,675,299	26,676,501	8_26662583	26,662,583	26,677,736	15,153	T	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.8G151900	8	29,020,756	29,022,649	8_29020733	29,020,733	29,025,225	4,492	*	Ref	Ref	Ref	A	A	Ref	A	Ref
Seita.8G154100	8	29,186,046	29,186,330	8_29184750	29,184,750	29,186,403	1,653	*	Ref	Ref	A	Ref	Ref	Ref	Ref	Ref
Seita.8G163300	8	30,545,021	30,545,826	8_30544654	30,544,654	30,546,723	2,069	*	Ref	Ref	A	Ref	A	A	Ref	A
Seita.8G165600	8	30,741,275	30,742,199	8_30740158	30,740,158	30,744,908	4,750	*	Ref	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.8G172900	8	31,589,026	31,590,871	8_31588009	31,588,009	31,594,414	6,405	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.8G173000	8	31,591,853	31,592,814	8_31588009	31,588,009	31,594,414	6,405	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.8G177700	8	32,033,838	32,034,015	8_32032091	32,032,091	32,035,443	3,352	*	Ref	Ref	Ref	A	A	Ref	A	Ref
Seita.8G186500	8	33,272,979	33,275,021	8_33267685	33,267,685	33,275,320	7,635	*	Ref	Ref	Ref	Ref	A	A	Ref	A
Seita.8G196900	8	34,377,140	34,380,592	8_34375940	34,375,940	34,384,129	8,189	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.8G221500	8	37,462,818	37,464,026	8_37461118	37,461,118	37,466,465	5,347	*	A	Ref	Ref	A	A	A	A	A
Seita.8G224800	8	37,708,587	37,709,008	8_37708456	37,708,456	37,709,160	704	*	A	Ref	Ref	A	A	A	A	A
Seita.8G231000	8	38,335,850	38,336,302	8_38335386	38,335,386	38,338,121	2,735	*	A	A	Ref	Ref	A	A	A	Ref
Seita.8G231400	8	38,389,493	38,389,889	8_38389279	38,389,279	38,389,969	690	*	Ref	A	Ref	Ref	Ref	Ref	Ref	Ref
Seita.8G239800	8	39,336,391	39,336,646	8_39335459	39,335,459	39,337,159	1,700	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.8G241800	8	39,512,486	39,513,374	8_39509992	39,509,992	39,514,486	4,494	*	Ref	Ref	Ref	A	Ref	Ref	Ref	A

Seita.8G243900	8	39,738,597	39,738,825	8_39737620	39,737,620	39,739,626	2,006	*	Ref	Ref	Ref	A	Ref	Ref	Ref	A
Seita.8G245500	8	39,918,468	39,919,286	8_39916701	39,916,701	39,925,782	9,081	*	A	A	A	A	A	A	A	A
Seita.8G250800	8	40,428,242	40,428,664	8_40427811	40,427,811	40,429,208	1,397	*	A	Ref	Ref	Ref	A	Ref	Ref	A
Seita.8G251400	8	40,460,647	40,461,515	8_40460306	40,460,306	40,461,877	1,571	*	Ref	Ref	Ref	Ref	Ref	A	A	A
Seita.9G000900	9	92,231	93,395	9_92048	92,048	97,090	5,042	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.9G016200	9	842,700	843,919	9_838426	838,426	844,466	6,040	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.9G016300	9	844,056	844,411	9_838426	838,426	844,466	6,040	*	Ref	Ref	Ref	A	A	A	Ref	Ref
Seita.9G045400	9	2,583,035	2,583,717	9_2568795	2,568,795	2,603,843	35,048	*	Ref	A	A	A	A	A	Ref	Ref
Seita.9G065300	9	3,778,836	3,779,767	9_3777875	3,777,875	3,783,012	5,137	*	A	Ref	A	Ref	Ref	Ref	Ref	Ref
Seita.9G102000	9	6,193,144	6,193,524	9_6192849	6,192,849	6,195,596	2,747	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.9G102100	9	6,193,891	6,195,424	9_6192849	6,192,849	6,195,596	2,747	*	Ref	Ref	Ref	Ref	Ref	Ref	Ref	A
Seita.9G137800	9	8,743,432	8,746,052	9_8740983	8,740,983	8,746,109	5,126	*	Ref	Ref	Ref	Ref	Ref	Ref	A	Ref
Seita.9G142400	9	9,109,884	9,110,212	9_9108326	9,108,326	9,111,478	3,152	*	Ref	Ref	Ref	A	Ref	A	Ref	Ref
Seita.9G143500	9	9,211,296	9,214,071	9_9210868	9,210,868	9,216,784	5,916	C	Ref	Ref	Ref	Ref	A	A	A	A
Seita.9G148900	9	9,687,426	9,690,668	9_9687401	9,687,401	9,694,801	7,400	*	Ref	Ref	Ref	Ref	A	A	Ref	A
Seita.9G210600	9	15,190,683	15,191,303	9_15190414	15,190,414	15,194,998	4,584	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.9G247800	9	19,495,687	19,496,208	9_19494612	19,494,612	19,497,526	2,914	*	A	Ref	Ref	A	Ref	A	A	A
Seita.9G247900	9	19,496,329	19,497,292	9_19494612	19,494,612	19,497,526	2,914	*	A	Ref	Ref	A	Ref	A	A	A
Seita.9G248000	9	19,498,184	19,498,718	9_19497532	19,497,532	19,503,737	6,205	*	A	Ref	Ref	A	Ref	A	A	A
Seita.9G250900	9	20,078,368	20,079,246	9_20078066	20,078,066	20,079,423	1,357	*	Ref	Ref	Ref	Ref	A	A	A	A
Seita.9G252600	9	20,231,315	20,231,849	9_20228938	20,228,938	20,235,480	6,542	*	A	Ref	Ref	A	A	A	A	A
Seita.9G252700	9	20,233,248	20,234,137	9_20228938	20,228,938	20,235,480	6,542	*	A	Ref	Ref	A	A	A	A	A
Seita.9G252800	9	20,234,317	20,234,838	9_20228938	20,228,938	20,235,480	6,542	*	A	Ref	Ref	A	A	A	A	A
Seita.9G285400	9	31,781,004	31,784,174	9_31779481	31,779,481	31,784,174	4,693	*	Ref	Ref	Ref	A	A	Ref	A	A
Seita.9G313400	9	36,136,156	36,136,677	9_36135340	36,135,340	36,137,832	2,492	*	Ref	Ref	Ref	Ref	A	Ref	A	Ref
Seita.9G313500	9	36,136,798	36,137,694	9_36135340	36,135,340	36,137,832	2,492	*	Ref	Ref	Ref	Ref	A	Ref	A	Ref

Seita.9G313600	9	36,139,110	36,139,644	9_36138363	36,138,363	36,140,480	2,117	*	Ref	Ref	Ref	Ref	A	Ref	A	Ref
Seita.9G313700	9	36,143,928	36,144,572	9_36143509	36,143,509	36,145,147	1,638	*	Ref	Ref	Ref	Ref	A	Ref	A	Ref
Seita.9G326400	9	37,921,579	37,923,124	9_37920817	37,920,817	37,923,162	2,345	*	A	Ref	Ref	A	A	A	A	A
Seita.9G338600	9	39,760,196	39,760,695	9_39759569	39,759,569	39,761,360	1,791	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.9G348200	9	40,813,028	40,816,782	9_40812602	40,812,602	40,817,132	4,530	*	A	Ref	Ref	A	Ref	Ref	Ref	A
Seita.9G359900	9	42,012,438	42,015,303	9_42009653	42,009,653	42,017,190	7,537	*	Ref	A	A	Ref	Ref	Ref	Ref	Ref
Seita.9G362400	9	42,320,686	42,320,988	9_42320246	42,320,246	42,321,241	995	*	A	A	A	A	A	Ref	Ref	A
Seita.9G381500	9	44,042,226	44,042,692	9_44040668	44,040,668	44,043,962	3,294	*	Ref	Ref	Ref	Ref	Ref	A	Ref	Ref
Seita.9G462700	9	50,789,554	50,794,316	9_50786657	50,786,657	50,796,225	9,568	*	A	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Seita.9G462800	9	50,799,221	50,803,120	9_50796227	50,796,227	50,803,783	7,556	*	A	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Seita.9G477800	9	51,828,058	51,828,672	9_51826045	51,826,045	51,829,165	3,120	*	Ref	Ref	Ref	Ref	A	A	A	A
Seita.9G513300	9	54,474,193	54,474,749	9_54473753	54,473,753	54,474,950	1,197	T	A	A	A	A	A	Ref	Ref	A
Seita.9G575500	9	58,531,473	58,532,548	9_58531246	58,531,246	58,535,856	4,610	*	Ref	Ref	Ref	Ref	A	Ref	Ref	Ref
Seita.9G578400	9	58,702,127	58,703,191	9_58701585	58,701,585	58,716,838	15,253	*	Ref	A	A	A	Ref	A	A	A

Table S11. Potentially functional variation sites in the 26.84 Mb to 26.94 Mb interval on chromosome 7

Var ID	Chr	PAV Start	PAV End	Var Type	Gene Annotation
7_26846504	chr7	26,846,504	26,847,566	pav_downstream	Seita.7G194500.v2.2(dist=649)
7_26848629	chr7	26,848,629	26,848,629	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon3:c.A1246G:p.S416G,
7_26848729	chr7	26,848,729	26,848,729	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon3:c.A1146G:p.I382M,
7_26848817	chr7	26,848,817	26,848,817	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon3:c.A1058G:p.D353G,
7_26849229	chr7	26,849,229	26,849,229	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon2:c.G671T:p.G224V,
7_26849235	chr7	26,849,235	26,849,235	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon2:c.G665C:p.W222S,
7_26849453	chr7	26,849,453	26,849,453	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon1:c.A470G:p.E157G,
7_26849573	chr7	26,849,573	26,849,573	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon1:c.T350G:p.V117G,
7_26849664	chr7	26,849,664	26,849,664	nonsynonymous_SNV	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon1:c.T259C:p.Y87H,
7_26851017	chr7	26,851,017	26,851,133	pav_downstream	Seita.7G194700.v2.2(dist=317)
7_26852007	chr7	26,852,007	26,852,007	nonsynonymous_SNV	Seita.7G194700.v2.2:Seita.7G194700.1.v2.2:exon2:c.G1015A:p.G339S,
7_26852382	chr7	26,852,382	26,852,382	nonsynonymous_SNV	Seita.7G194700.v2.2:Seita.7G194700.1.v2.2:exon2:c.G640A:p.V214I,
7_26852816	chr7	26,852,816	26,852,816	nonsynonymous_SNV	Seita.7G194700.v2.2:Seita.7G194700.1.v2.2:exon2:c.T206C:p.V69A,
7_26853053	chr7	26,853,053	26,853,053	nonsynonymous_SNV	Seita.7G194700.v2.2:Seita.7G194700.1.v2.2:exon1:c.T74C:p.V25A,
7_26858200	chr7	26,858,200	26,858,200	pav_intergenic	Seita.7G194800.v2.2(dist=1113),Seita.7G194900.v2.2(dist=3178)
7_26862091	chr7	26,862,091	26,862,091	nonsynonymous_SNV	Seita.7G194900.v2.2:Seita.7G194900.1.v2.2:exon1:c.A1619G:p.K540R,
7_26862734	chr7	26,862,734	26,862,734	nonsynonymous_SNV	Seita.7G194900.v2.2:Seita.7G194900.1.v2.2:exon1:c.T976A:p.S326T,
7_26862893	chr7	26,862,893	26,862,893	nonsynonymous_SNV	Seita.7G194900.v2.2:Seita.7G194900.1.v2.2:exon1:c.T817C:p.Y273H,
7_26863414	chr7	26,863,414	26,863,414	nonsynonymous_SNV	Seita.7G194900.v2.2:Seita.7G194900.1.v2.2:exon1:c.C296G:p.P99R,
7_26863422	chr7	26,863,422	26,863,422	frameshift_insertion	Seita.7G194900.v2.2:Seita.7G194900.1.v2.2:exon1:c.287_288insGATC:p.S96fs,
7_26863445	chr7	26,863,445	26,863,445	nonsynonymous_SNV	Seita.7G194900.v2.2:Seita.7G194900.1.v2.2:exon1:c.C265T:p.H89Y,
7_26863493	chr7	26,863,493	26,863,493	nonsynonymous_SNV	Seita.7G194900.v2.2:Seita.7G194900.1.v2.2:exon1:c.C217T:p.P73S,
7_26864648	chr7	26,864,648	26,864,919	pav_upstream	Seita.7G194900.v2.2(dist=910)
7_26865118	chr7	26,865,118	26,867,801	pav_intergenic	Seita.7G194900.v2.2(dist=1380),Seita.7G195000.v2.2(dist=3037)

7_26868773	chr7	26,868,773	26,869,378	pav_intergenic	Seita.7G194900.v2.2(dist=5035),Seita.7G195000.v2.2(dist=1460)
7_26869801	chr7	26,869,801	26,870,049	pav_downstream	Seita.7G195000.v2.2(dist=789)
7_26870059	chr7	26,870,059	26,870,059	pav_downstream	Seita.7G195000.v2.2(dist=779)
7_26870067	chr7	26,870,067	26,870,067	pav_downstream	Seita.7G195000.v2.2(dist=771)
7_26870863	chr7	26,870,863	26,870,863	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon4:c.G1565A:p.G522D,
7_26870927	chr7	26,870,927	26,870,927	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon4:c.G1501A:p.E501K,
7_26871184	chr7	26,871,184	26,871,184	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon4:c.A1244G:p.Q415R,
7_26871211	chr7	26,871,211	26,871,211	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon4:c.C1217A:p.P406Q,
7_26871407	chr7	26,871,407	26,871,407	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon3:c.G1051A:p.A351T,
7_26871427	chr7	26,871,427	26,871,427	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon3:c.G1031A:p.R344K,
7_26871570	chr7	26,871,570	26,871,570	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon3:c.G888C:p.R296S,
7_26871766	chr7	26,871,766	26,871,766	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon3:c.A692C:p.Q231P,
7_26871803	chr7	26,871,803	26,871,803	nonsynonymous_SNV	Seita.7G195000.v2.2:Seita.7G195000.1.v2.2:exon3:c.G655A:p.A219T,
7_26873145	chr7	26,873,145	26,873,145	pav_upstream;downstream	Seita.7G195000.v2.2(dist=359);Seita.7G195100.v2.2(dist=364)
7_26874972	chr7	26,874,972	26,874,972	nonsynonymous_SNV	Seita.7G195100.v2.2:Seita.7G195100.1.v2.2:exon6:c.T1076C:p.L359P,
7_26875141	chr7	26,875,141	26,875,141	nonsynonymous_SNV	Seita.7G195100.v2.2:Seita.7G195100.1.v2.2:exon6:c.G907T:p.A303S,
7_26875547	chr7	26,875,547	26,875,547	nonsynonymous_SNV	Seita.7G195100.v2.2:Seita.7G195100.1.v2.2:exon5:c.G588T:p.L196F,
7_26878032	chr7	26,878,032	26,878,084	pav_upstream	Seita.7G195100.v2.2(dist=982)
7_26880410	chr7	26,880,410	26,880,410	nonsynonymous_SNV	Seita.7G195200.v2.2:Seita.7G195200.1.v2.2:exon4:c.A250G:p.N84D,
7_26880458	chr7	26,880,458	26,880,458	frameshift_insertion	Seita.7G195200.v2.2:Seita.7G195200.1.v2.2:exon4:c.201_202insAG:p.S68fs,
7_26880509	chr7	26,880,509	26,880,509	nonsynonymous_SNV	Seita.7G195200.v2.2:Seita.7G195200.1.v2.2:exon4:c.G151T:p.D51Y,
7_26880553	chr7	26,880,553	26,880,553	frameshift_insertion	Seita.7G195200.v2.2:Seita.7G195200.1.v2.2:exon4:c.106_107insGATC:p.L36fs,
7_26880555	chr7	26,880,555	26,880,555	nonsynonymous_SNV	Seita.7G195200.v2.2:Seita.7G195200.1.v2.2:exon4:c.T105A:p.D35E,
7_26880580	chr7	26,880,580	26,880,580	nonsynonymous_SNV	Seita.7G195200.v2.2:Seita.7G195200.1.v2.2:exon4:c.A80G:p.E27G,
7_26881279	chr7	26,881,279	26,881,279	pav_intronic	Seita.7G195200.v2.2

7_26883051	chr7	26,883,051	26,883,051	pav_upstream	Seita.7G195200.v2.2(dist=524)
7_26886038	chr7	26,886,038	26,886,038	pav_upstream	Seita.7G195300.v2.2(dist=22)
7_26887194	chr7	26,887,194	26,887,194	nonsynonymous_SNV	Seita.7G195300.v2.2:Seita.7G195300.1.v2.2:exon2:c.T158C:p.L53P,
7_26887694	chr7	26,887,694	26,887,694	nonsynonymous_SNV	Seita.7G195300.v2.2:Seita.7G195300.1.v2.2:exon2:c.A658G:p.K220E,
7_26890988	chr7	26,890,988	26,890,988	pav_intronic	Seita.7G195400.v2.2
7_26895751	chr7	26,895,751	26,895,751	pav_intronic	Seita.7G195400.v2.2
7_26897142	chr7	26,897,142	26,897,142	pav_intronic	Seita.7G195400.v2.2
7_26900791	chr7	26,900,791	26,900,791	pav_upstream	Seita.7G195400.v2.2(dist=839)
7_26903038	chr7	26,903,038	26,903,038	pav_intergenic	Seita.7G195400.v2.2(dist=3086),Seita.7G195500.v2.2(dist=4184)
7_26903178	chr7	26,903,178	26,903,178	pav_intergenic	Seita.7G195400.v2.2(dist=3226),Seita.7G195500.v2.2(dist=4044)
7_26903212	chr7	26,903,212	26,903,212	pav_intergenic	Seita.7G195400.v2.2(dist=3260),Seita.7G195500.v2.2(dist=4010)
7_26903241	chr7	26,903,241	26,903,241	pav_intergenic	Seita.7G195400.v2.2(dist=3289),Seita.7G195500.v2.2(dist=3981)
7_26903249	chr7	26,903,249	26,903,249	pav_intergenic	Seita.7G195400.v2.2(dist=3297),Seita.7G195500.v2.2(dist=3973)
7_26903266	chr7	26,903,266	26,903,266	pav_intergenic	Seita.7G195400.v2.2(dist=3314),Seita.7G195500.v2.2(dist=3956)
7_26903310	chr7	26,903,310	26,903,310	pav_intergenic	Seita.7G195400.v2.2(dist=3358),Seita.7G195500.v2.2(dist=3912)
7_26903532	chr7	26,903,532	26,903,532	pav_intergenic	Seita.7G195400.v2.2(dist=3580),Seita.7G195500.v2.2(dist=3690)
7_26903725	chr7	26,903,725	26,903,725	pav_intergenic	Seita.7G195400.v2.2(dist=3773),Seita.7G195500.v2.2(dist=3497)
7_26903914	chr7	26,903,914	26,903,914	pav_intergenic	Seita.7G195400.v2.2(dist=3962),Seita.7G195500.v2.2(dist=3308)
7_26904213	chr7	26,904,213	26,904,213	pav_intergenic	Seita.7G195400.v2.2(dist=4261),Seita.7G195500.v2.2(dist=3009)
7_26907876	chr7	26,907,876	26,907,876	nonsynonymous_SNV	Seita.7G195500.v2.2:Seita.7G195500.1.v2.2:exon1:c.A178G:p.T60A,
7_26915698	chr7	26,915,698	26,915,698	nonsynonymous_SNV	Seita.7G195600.v2.2:Seita.7G195600.3.v2.2:exon17:c.C1259G:p.P420R,Seita.7G195600.v2.2:Seita.7G195600.2.v2.2:exon17:c.C1259G:p.P420R,Seita.7G195600.v2.2:Seita.7G195600.1.v2.2:exon16:c.C1340G:p.P447R,
7_26915760	chr7	26,915,760	26,916,035	pav_UTR3	Seita.7G195600.v2.2(Seita.7G195600.2.v2.2:c.*46_*321delTTAGGCCATGTTTAGTTACTCCCAACTTCCAACTTTGACACTATGCAAAAAGAAGATTCCCCATCACATCAAACTTGCGGTACATGTATGGAGTACTAAATGTAGATGAAATTAAAAAGTAATTGTACAGTTTTGTTGTACTTTACGAGACGAATCTTTTGAGCCTAATTAGTTAATATTTGGACAATAATTCACAAATACAAACGAAACGCTACAGTGTGCTACAGTGCTGTAACAGTAATTTGGCACCTCCCAAATTGGCCAACTAAACAAGGC,Seita.7G195600.3.v2.2:c.*44_*45delins-,Seita.7G195600.1.v2.2:c.*46_*321delTTAGGCCATGTTTAGTTACTCCCAACTTCCAACTTTGACACTATGCAAAAAGAAGATTCCCCATCACATCAAACTTGCGGTACAT

					GTATGGAGTACTAAATGTAGATGAAATTAAAAAGTAATTGTACAGTTTTGTTGTACTTTACGAGACGAATCTTTTGAGCCTAATTAGTTAATATTTGGACAATAATTCACAAAT
					ACAAACGAAACGCTACAGTGTGCTACAGTGCTGTAACAGTAATTTGGCACCTCCCAAATTGGCCAACTAAACAAGGC),Seita.7G195700.v2.2(Seita.7G195700.1.v2.2:c.*90
					8_*906delGCCTTGTTTAGTTGGCCAATTTGGGAGGTGCCAAATTACTGTTACAGCACTGTAGCACACTGTAGCGTTTCGTTTGTATTTGTGAATTATTGTCCAAATATTAACT
					AATTAGGCTCAAAAGATTCGTCTCGTAAAGTACAACAAAACTGTACAATTACTTTTTAATTTCATCTACATTTAGTACTCCATACATGTACCGCAAGTTTGATGTGATGGGGA
					ATCTTCTTTTTGCATAGTGTCAAAGTTGGAAGTTGGGAGTAACTAAACATGGCCTAA)
7_26917881	chr7	26,917,881	26,917,881	nonsynonymous_SNV	Seita.7G195700.v2.2:Seita.7G195700.1.v2.2:exon1:c.A184G:p.N62D,
7_26918461	chr7	26,918,461	26,918,461	pav_upstream	Seita.7G195700.v2.2(dist=4)
					Seita.7G195800.v2.2(Seita.7G195800.1.v2.2:c.*720_*721insTGGGCCAGTTTGGCAGGGTTCCGGCTCCGGCTCAGTGCGCTTACCGAAGCACGGAGGAGCCGGAGCCGC
7_26919917	chr7	26,919,917	26,919,917	pav_UTR3	ACCAAACGGCATTGACCGCGGAGCCATTTTTTAGCACATTTGGGAGGAGCCGAAGCCGTTTTAGGCCTGCATGGAGGAGCTCAAAAAAGTGGCTCCGCGGCTCCGGCTC
					CGGCTCCGCACGAGGAGCCCCTCGCGAGGAGCCCTGCCAAACTCGCCCTA)
7_26921744	chr7	26,921,744	26,921,744	pav_upstream	Seita.7G195800.v2.2(dist=544)
7_26923601	chr7	26,923,601	26,923,601	frameshift_insertion	Seita.7G195900.v2.2:Seita.7G195900.1.v2.2:exon1:c.691_692insCGAT:p.P231fs,
7_26924169	chr7	26,924,169	26,924,169	frameshift_insertion	Seita.7G195900.v2.2:Seita.7G195900.1.v2.2:exon1:c.1259_1260insGATC:p.K420fs,
7_26924720	chr7	26,924,720	26,924,720	frameshift_insertion	Seita.7G195900.v2.2:Seita.7G195900.1.v2.2:exon1:c.1810_1811insCGAT:p.P604fs,
7_26925111	chr7	26,925,111	26,925,111	nonsynonymous_SNV	Seita.7G195900.v2.2:Seita.7G195900.1.v2.2:exon1:c.T2201C:p.V734A,
7_26932181	chr7	26,932,181	26,932,181	pav_downstream	Seita.7G196000.v2.2(dist=884)
7_26934917	chr7	26,934,917	26,934,917	nonsynonymous_SNV	Seita.7G196000.v2.2:Seita.7G196000.1.v2.2:exon1:c.A353G:p.K118R,

Table S12. Performance metrics of machine learning models

Method	Pheno	Class	Precision	Recall	F1-score
DecisionTree	Leaf Color	Level 1	0.066667	0.090909	0.076923
		Level 2	0.84962406	0.779310345	0.81294964
		Level 3	0.217391304	0.3125	0.256410256
		Macro Avg	0.355671788	0.363936782	0.356453299
		Weighted Avg	0.736475288	0.686046512	0.709187569
		Overall Accuracy	0.686046512	-	-
	Leaf Sheath Color	Level 1	0.911764706	0.837837838	0.873239437
		Level 2	0.472727273	0.634146341	0.541666667
		Level 3	0.333333333	0.25	0.285714286
		Macro Avg	0.572608437	0.573994726	0.566873463
		Weighted Avg	0.739850972	0.720930233	0.725884863
		Overall Accuracy	0.720930233	-	-
	Leaf Pulvinus Color	Level 1	0.803921569	0.738738739	0.769953052
		Level 2	0.192307692	0.2	0.196078431
		Level 3	0.363636364	0.444444444	0.4
		Macro Avg	0.453288542	0.461061061	0.455343828
		Weighted Avg	0.622871486	0.598837209	0.609109009
		Overall Accuracy	0.598837209	-	-
RandomForest	Leaf Color	Level 1	0	0	0
		Level 2	0.845679012	0.944827586	0.892508143
		Level 3	0.222222222	0.125	0.16
		Macro Avg	0.355967078	0.356609195	0.350836048
		Weighted Avg	0.733598909	0.808139535	0.767288842
		Overall Accuracy	0.808139535	-	-
	Leaf Sheath Color	Level 1	0.95	0.855855856	0.900473934
		Level 2	0.553846154	0.87804878	0.679245283
		Level 3	0.428571429	0.15	0.222222222
		Macro Avg	0.644139194	0.627968212	0.600647146
		Weighted Avg	0.794936749	0.779069767	0.768872719
		Overall Accuracy	0.779069767	-	-
	Leaf Pulvinus Color	Level 1	0.837837838	0.837837838	0.837837838
		Level 2	0.285714286	0.24	0.260869565
		Level 3	0.325	0.361111111	0.342105263
		Macro Avg	0.482850708	0.47964965	0.480270889
		Weighted Avg	0.650249169	0.651162791	0.65021819
		Overall Accuracy	0.651162791	-	-
XGBoost	Leaf Color	Level 1	0	0	0
		Level 2	0.842105263	0.882758621	0.861952862
		Level 3	0.214285714	0.1875	0.2
		Macro Avg	0.352130326	0.356752874	0.353984287

		Weighted Avg	0.729847876	0.761627907	0.745250959
		Overall Accuracy	0.761627907	-	-
	Leaf Sheath Color	Level 1	0.95959596	0.855855856	0.904761905
		Level 2	0.540983607	0.804878049	0.647058824
		Level 3	0.333333333	0.2	0.25
		Macro Avg	0.6113043	0.620244635	0.600606909
		Weighted Avg	0.786989221	0.76744186	0.767197577
		Overall Accuracy	0.76744186	-	-
	Leaf Pulvinus Color	Level 1	0.857142857	0.810810811	0.833333333
		Level 2	0.481481481	0.52	0.5
		Level 3	0.4	0.444444444	0.421052632
		Macro Avg	0.579541446	0.591751752	0.584795322
		Weighted Avg	0.70685985	0.691860465	0.698592411
		Overall Accuracy	0.691860465	-	-
LightGBM	**Leaf Color**	Level 1	0.222222222	0.25	0.235294118
		Level 2	0.865979381	0.622222222	0.724137931
		Level 3	0.350877193	0.952380952	0.512820513
		Macro Avg	0.479692932	0.608201058	0.490750854
		Weighted Avg	0.743204611	0.627906977	0.652863705
		Overall Accuracy	0.627906977	-	-
	Leaf Sheath Color	Level 1	0.98989899	0.882882883	0.933333333
		Level 2	0.6875	0.289473684	0.407407407
		Level 3	0.368421053	0.913043478	0.525
		Macro Avg	0.681940014	0.695133348	0.62191358
		Weighted Avg	0.839985303	0.755813953	0.762537683
		Overall Accuracy	0.755813953	-	-
	Leaf Pulvinus Color	Level 1	0.97	0.881818182	0.923809524
		Level 2	0.340425532	0.592592593	0.432432432
		Level 3	0.44	0.314285714	0.366666667
		Macro Avg	0.583475177	0.596232163	0.574302874
		Weighted Avg	0.763322613	0.720930233	0.733302655
		Overall Accuracy	0.720930233	-	-

Table S13. Mean Shapley values (feature importance) for the three traits

Features	Leaf Color				Leaf Sheath Color				Leaf Pulvinus Color			
	level1	level2	level3	sum_levels	level1	level2	level3	sum_levels	level1	level2	level3	sum_levels
chr7_26846504	0.1642	0.1450	0.0232	0.3324	0.1678	0.2464	0.1683	0.5825	0.1144	0.1351	0.1048	0.3544
chr7_26848629	0.0001	0.0057	0.0003	0.0061	0.0094	0.0281	0.0249	0.0624	0.0013	0.0018	0.0011	0.0042
chr7_26848729	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019	0.0019	0.0002	0.0000	0.0004	0.0007
chr7_26848817	0.0000	0.0000	0.0000	0.0000	0.0018	0.0026	0.0071	0.0114	0.0094	0.0114	0.0001	0.0209
chr7_26849229	0.0000	0.0000	0.0000	0.0000	0.0000	0.0017	0.0000	0.0017	0.0001	0.0000	0.0000	0.0001
chr7_26849235	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26849453	0.0000	0.0000	0.0133	0.0133	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	0.0000	0.0009
chr7_26849573	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0017	0.0011	0.0001	0.0029
chr7_26849664	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.0000	0.0000	0.0012
chr7_26852007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26852382	0.0000	0.0000	0.0000	0.0000	0.0136	0.0000	0.0068	0.0204	0.0005	0.0000	0.0018	0.0023
chr7_26852816	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26853053	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26858200	0.0001	0.0000	0.0185	0.0185	0.0061	0.0230	0.0050	0.0341	0.0318	0.0139	0.0042	0.0499
chr7_26862091	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26862734	0.0000	0.0000	0.0000	0.0000	0.0014	0.0000	0.0000	0.0014	0.0001	0.0002	0.0008	0.0012
chr7_26862893	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0003	0.0004
chr7_26863414	0.0000	0.0000	0.0000	0.0000	0.0024	0.0014	0.0010	0.0049	0.0009	0.0133	0.0004	0.0146
chr7_26863422	0.1836	0.0681	0.3509	0.6026	0.3770	0.1233	0.6787	1.1791	0.1750	0.0606	0.1684	0.4039
chr7_26863445	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0005	0.0001	0.0027	0.0001	0.0030
chr7_26863493	0.0077	0.0000	0.0000	0.0077	0.0035	0.0021	0.0219	0.0275	0.0031	0.0021	0.0013	0.0065
chr7_26864648	0.0027	0.0491	0.0088	0.0606	0.0136	0.0234	0.0255	0.0625	0.0186	0.0468	0.0162	0.0816
chr7_26865118	0.4055	0.0265	0.1453	0.5773	0.7257	1.0477	0.0995	1.8728	0.1421	0.2992	0.0411	0.4824

chr7_26868773	0.1302	0.1675	0.0000	0.2977	1.2568	0.1886	0.0615	1.5070	0.7170	0.0398	0.4332	1.1900
chr7_26869801	0.0696	0.0958	0.0000	0.1654	0.0460	0.0273	0.0159	0.0891	0.0103	0.0513	0.0298	0.0914
chr7_26870059	0.0084	0.0225	0.0000	0.0309	0.0126	0.0378	0.0325	0.0830	0.0097	0.0096	0.0055	0.0248
chr7_26870067	0.0000	0.0102	0.0000	0.0102	0.0005	0.0108	0.0080	0.0193	0.0016	0.0018	0.0011	0.0046
chr7_26870863	0.0000	0.0000	0.0118	0.0118	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0020	0.0020
chr7_26870927	0.0000	0.0000	0.0094	0.0094	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000	0.0008
chr7_26871184	0.0000	0.0064	0.0000	0.0064	0.0000	0.0006	0.0000	0.0006	0.0001	0.0017	0.0005	0.0023
chr7_26871211	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26871407	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0000	0.0003	0.0007
chr7_26871427	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26871570	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26871766	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26871803	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26873145	0.0000	0.0000	0.0000	0.0000	0.0018	0.0493	0.0517	0.1029	0.0030	0.0011	0.0008	0.0049
chr7_26874972	0.0168	0.0158	0.0000	0.0326	0.0056	0.0000	0.0000	0.0056	0.0000	0.0000	0.0000	0.0000
chr7_26875141	0.0000	0.0025	0.0000	0.0025	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0002
chr7_26875547	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26878032	0.0388	0.0314	0.0124	0.0827	0.0222	0.0511	0.0306	0.1039	0.0141	0.0952	0.0199	0.1292
chr7_26880410	0.0260	0.0000	0.0000	0.0260	0.0000	0.0000	0.0002	0.0002	0.0000	0.0000	0.0000	0.0000
chr7_26880458	0.0198	0.0054	0.0000	0.0252	0.0628	0.0289	0.0288	0.1205	0.0089	0.0663	0.0101	0.0854
chr7_26880509	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26880553	0.0006	0.0304	0.0000	0.0310	0.0505	0.1652	0.0492	0.2649	0.0208	0.0215	0.0303	0.0726
chr7_26880555	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26880580	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26881279	0.0013	0.0000	0.0000	0.0013	0.0151	0.0437	0.0215	0.0803	0.0117	0.0682	0.0167	0.0966
chr7_26883051	0.0195	0.0893	0.0008	0.1097	0.0166	0.0565	0.0595	0.1327	0.0396	0.0232	0.0575	0.1203

chr7_26886038	0.0361	0.0027	0.0001	0.0390	0.0107	0.0187	0.0196	0.0491	0.0088	0.0023	0.0028	0.0140
chr7_26887194	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26887694	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26897142	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26900791	0.0213	0.0881	0.0334	0.1428	0.1370	0.1082	0.0648	0.3099	0.0244	0.0737	0.0747	0.1728
chr7_26903038	0.0264	0.1010	0.0159	0.1434	0.1197	0.0696	0.0506	0.2399	0.0507	0.0257	0.0274	0.1037
chr7_26903241	0.0989	0.0827	0.0460	0.2276	0.0767	0.2731	0.2702	0.6199	0.0422	0.0441	0.0485	0.1348
chr7_26903249	0.0248	0.0187	0.0005	0.0441	0.0167	0.0338	0.0378	0.0883	0.0070	0.0202	0.0314	0.0586
chr7_26903266	0.0009	0.0271	0.0088	0.0367	0.0039	0.0107	0.0146	0.0292	0.0030	0.0016	0.0106	0.0152
chr7_26903310	0.0203	0.0429	0.0506	0.1138	0.0387	0.1306	0.0547	0.2240	0.0133	0.0130	0.0292	0.0555
chr7_26903532	0.0310	0.0939	0.0389	0.1638	0.0945	0.0746	0.0766	0.2457	0.0480	0.0268	0.0349	0.1097
chr7_26903914	0.0000	0.0000	0.0071	0.0071	0.0019	0.0111	0.0086	0.0216	0.0527	0.0070	0.0658	0.1256
chr7_26904213	0.0019	0.0345	0.0416	0.0780	0.0749	0.0923	0.2294	0.3966	0.0180	0.0748	0.0804	0.1732
chr7_26907876	0.0000	0.0116	0.0000	0.0116	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0003
chr7_26915698	0.0000	0.0000	0.0018	0.0018	0.0272	0.0158	0.0000	0.0430	0.0000	0.0002	0.0000	0.0002
chr7_26915760	0.1723	0.1910	0.0998	0.4631	0.2686	0.2240	0.1418	0.6345	0.0754	0.0521	0.0629	0.1904
chr7_26917881	0.0004	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26918461	0.0000	0.0000	0.0000	0.0000	0.0000	0.0221	0.0428	0.0649	0.0000	0.0007	0.0054	0.0061
chr7_26919917	0.0005	0.0000	0.0000	0.0005	0.0000	0.0016	0.0076	0.0092	0.0000	0.0000	0.0000	0.0000
chr7_26921744	0.0000	0.0327	0.0000	0.0327	0.0000	0.0024	0.0030	0.0054	0.0035	0.0000	0.0010	0.0046
chr7_26923601	0.0288	0.0891	0.1379	0.2558	0.5166	0.1762	0.1265	0.8193	0.1003	0.1341	0.0259	0.2603
chr7_26924169	0.0000	0.1227	0.3829	0.5056	0.2206	0.0587	0.1032	0.3825	0.0376	0.0184	0.0211	0.0771
chr7_26924720	0.0001	0.0231	0.0452	0.0684	0.2033	0.0655	0.3189	0.5877	0.0588	0.0450	0.0403	0.1441
chr7_26925111	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chr7_26932181	0.0019	0.0074	0.0106	0.0198	0.0057	0.0022	0.0030	0.0109	0.0080	0.0024	0.0043	0.0147
chr7_26934917	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table S14. Contribution statistics of the eight accessions in this study to the graph-based pan-genome

Sample	Number of Source Segments	Total Length of Source Segments (bp)	Proportion of Total Length (%)
DXH01	65	50793	0.00%
DXH02	3716	4716685	0.39%
DXH03	277	1201853	0.10%
DXH04	4529	7305194	0.61%
DXH05	14483	19616003	1.64%
DXH06	4087	7127677	0.60%
DXH07	1042	1683569	0.14%
DXH08	15256	14208308	1.19%
Statistics of the graph-based pan-genome			
Total number of segments	1160735		
Total segment length	1,197,866,326 bp		
Total length of DXH01–08-derived segments	55,910,082 bp (4.67%)		

Table S15. PAV and InDels in the 26.84 Mb to 26.94 Mb interval on chromosome 7 after integrating the 110-pangenome dataset

Var ID	Chr	PAV Start	PAV End	Var Type	Gene Annotation	110-pangenome specific
7_26840029	chr7	26840029	26840029	pav_upstream	Seita.7G194300.v2.2(dist=623)	YES
7_26840056	chr7	26840056	26840056	pav_upstream	Seita.7G194300.v2.2(dist=650)	YES
7_26842076	chr7	26842076	26842079	nonframeshift_substitution	Seita.7G194400.v2.2:Seita.7G194400.1.v2.2:exon1:c.225_228delinsGTCC,	YES
7_26845812	chr7	26845812	26845816	nonframeshift_substitution	Seita.7G194500.v2.2:Seita.7G194500.1.v2.2:exon1:c.86_90delinsCAGCC,	YES
7_26846501	chr7	26846501	26847565	pav_downstream	Seita.7G194500.v2.2(dist=646)	YES
7_26846504	chr7	26846504	26847566	pav_downstream	Seita.7G194500.v2.2(dist=649)	NO
7_26849664	chr7	26849664	26849665	nonframeshift_substitution	Seita.7G194600.v2.2:Seita.7G194600.1.v2.2:exon1:c.258_259delinsCC,	YES
7_26851016	chr7	26851016	26851134	pav_downstream	Seita.7G194700.v2.2(dist=316)	YES
7_26858200	chr7	26858200	26858200	pav_intergenic	Seita.7G194800.v2.2(dist=1113),Seita.7G194900.v2.2(dist=3178)	NO
7_26864648	chr7	26864648	26864919	pav_upstream	Seita.7G194900.v2.2(dist=910)	NO
7_26865118	chr7	26865118	26867801	pav_intergenic	Seita.7G194900.v2.2(dist=1380),Seita.7G195000.v2.2(dist=3037)	NO
7_26868773	chr7	26868773	26869378	pav_intergenic	Seita.7G194900.v2.2(dist=5035),Seita.7G195000.v2.2(dist=1460)	NO
7_26869801	chr7	26869801	26870049	pav_downstream	Seita.7G195000.v2.2(dist=789)	NO
7_26870059	chr7	26870059	26870059	pav_downstream	Seita.7G195000.v2.2(dist=779)	NO
7_26870067	chr7	26870067	26870067	pav_downstream	Seita.7G195000.v2.2(dist=771)	NO
7_26873145	chr7	26873145	26873145	pav_upstream;downstream	Seita.7G195000.v2.2(dist=359);Seita.7G195100.v2.2(dist=364)	NO
7_26878032	chr7	26878032	26878084	pav_upstream	Seita.7G195100.v2.2(dist=982)	NO
7_26880458	chr7	26880458	26880458	frameshift_insertion	Seita.7G195200.v2.2:Seita.7G195200.1.v2.2:exon4:c.201_202insAG:p.S68fs,	NO
7_26883051	chr7	26883051	26883051	pav_upstream	Seita.7G195200.v2.2(dist=524)	NO
7_26886038	chr7	26886038	26886038	pav_upstream	Seita.7G195300.v2.2(dist=22)	NO
7_26890988	chr7	26890988	26890988	pav_intronic	Seita.7G195400.v2.2	NO
7_26895751	chr7	26895751	26895751	pav_intronic	Seita.7G195400.v2.2	NO
7_26897142	chr7	26897142	26897142	pav_intronic	Seita.7G195400.v2.2	NO

7_26900791	chr7	26900791	26900791	pav_upstream	Seita.7G195400.v2.2(dist=839)	NO
7_26903038	chr7	26903038	26903038	pav_intergenic	Seita.7G195400.v2.2(dist=3086),Seita.7G195500.v2.2(dist=4184)	NO
7_26903178	chr7	26903178	26903178	pav_intergenic	Seita.7G195400.v2.2(dist=3226),Seita.7G195500.v2.2(dist=4044)	NO
7_26903212	chr7	26903212	26903212	pav_intergenic	Seita.7G195400.v2.2(dist=3260),Seita.7G195500.v2.2(dist=4010)	NO
7_26903241	chr7	26903241	26903241	pav_intergenic	Seita.7G195400.v2.2(dist=3289),Seita.7G195500.v2.2(dist=3981)	NO
7_26903249	chr7	26903249	26903249	pav_intergenic	Seita.7G195400.v2.2(dist=3297),Seita.7G195500.v2.2(dist=3973)	NO
7_26903266	chr7	26903266	26903266	pav_intergenic	Seita.7G195400.v2.2(dist=3314),Seita.7G195500.v2.2(dist=3956)	NO
7_26903310	chr7	26903310	26903310	pav_intergenic	Seita.7G195400.v2.2(dist=3358),Seita.7G195500.v2.2(dist=3912)	NO
7_26903532	chr7	26903532	26903532	pav_intergenic	Seita.7G195400.v2.2(dist=3580),Seita.7G195500.v2.2(dist=3690)	NO
7_26903720	chr7	26903720	26903720	pav_intergenic	Seita.7G195400.v2.2(dist=3768),Seita.7G195500.v2.2(dist=3502)	YES
7_26903725	chr7	26903725	26903725	pav_intergenic	Seita.7G195400.v2.2(dist=3773),Seita.7G195500.v2.2(dist=3497)	NO
7_26903914	chr7	26903914	26903914	pav_intergenic	Seita.7G195400.v2.2(dist=3962),Seita.7G195500.v2.2(dist=3308)	NO
7_26904213	chr7	26904213	26904213	pav_intergenic	Seita.7G195400.v2.2(dist=4261),Seita.7G195500.v2.2(dist=3009)	NO
7_26915760	chr7	26915760	26916035	pav_UTR3	Seita.7G195600.v2.2(Seita.7G195600.2.v2.2:c.*46_*321delTTAGGCCATGTTTAGTTACTCCCAACTTCCAACTTTGACACTATGCAAAAAGAAGATTCCCCATCACATCAAACTTGCGGTACATGTATGGAGTACTAAATGTAGATGAAATTAAAAAGTAATTGTACAGTTTTGTTGTACTTTACGAGACGAATCTTTTGAGCCTAATTAGTTAATATTTGGACAATAATTATTTGGCACCTCCCAAATTGGCCAACTAAACAAGGC,Seita.7G195600.3.v2.2:c.*44_*45delins-,Seita.7G195600.1.v2.2:c.*46_*321delTTAGGCCATGTTTAGTTACTCCCAACTTCCAACTTTGACACTATGCAAAAAGAAGATTCCCCATCACATCAAACTTGCGGTACATGTATGGAGTACTAAATGTAGATGAAATTAAAAAGTAATTGTACAGTTTTGTTGTACTTTACGAGACGAATCTTTTGAGCCTAATTAGTTAATATTTGGACAATAATTCACAAATACAAACGAAACGCTACAGTGTGCTACAGTGCTGTAACAGTAATTTGGCACCTCCCAAATTGGCCAACTAAACAAGGC),Seita.7G195700.v2.2(Seita.7G195700.1.v2.2:c.*908_*906delGCCTTGTTTAGTTGGCCAATTTGGGAGGTGCCAAATTACTGTTACAGCACTGTAGCACACTGTAGCGTTTCGTTTGTATTTGTGAATTATTGTCCAAATATTAACTAATTAGGCTCAAAAGATTCGTCTCGTAAAGTACAACAAAACTGTACAATTACTTTTTAATTTCATCTACATTTAGTACTCCATACATGTACCGCAAGTTTGATGTGATGGGGAATCTTCTTTTTGCATAGTGTCAAAGTTGGAAGTTGGGAGTAACTAAACATGGCCTAA)	NO
7_26918461	chr7	26918461	26918461	pav_upstream	Seita.7G195700.v2.2(dist=4)	NO
7_26918463	chr7	26918463	26918463	pav_upstream	Seita.7G195700.v2.2(dist=6)	YES

					Seita.7G195800.v2.2(Seita.7G195800.1.v2.2:c.*720_*721insTGGGCCAGTTTGGCAGGGTTCCGGCTCCGGCTCAGTGCGCTTACCGAAGCA	
7_26919917	chr7	26919917	26919917	pav_UTR3	CGGAGGAGCCGGAGCCGCACCAAACGGCATTGACCGCGGAGCCATTTTTTAGCACATTTGGGAGGAGCCGAAGCCGTTTTAGGCCTGCA	NO
					TGGAGGAGCTCAAAAAAGTGGCTCCGCGGCTCCGGCTCCGGCTCCGCACGAGGAGCCCCTCGCGAGGAGCCCTGCCAAACTCGCCCTA)	
					Seita.7G195800.v2.2(Seita.7G195800.1.v2.2:c.*718_*719insTATGGGCCAGTTTGGCAGGGTTCCGGCTCCGGCTCAGTGCGCTTACCGAAG	
7_26919919	chr7	26919919	26919919	pav_UTR3	CACGGAGGAGCCGGAGCCGCACCAAACGGCATTGACCGCGGAGCCATTTTTTAGCACATTTGGGAGGAGCCGAAGCCGTTTTAGGCCTG	YES
					CATGGAGGAGCTCAAAAAAGTGGCTCCGCGGCTCCGGCTCCGGCTCCGCACGAGGAGCCCCTCGCGAGGAGCCCTGCCAAACTCGCCC)	
7_26920879	chr7	26920879	26920882	nonframeshift_substitution	Seita.7G195800.v2.2:Seita.7G195800.1.v2.2:exon1:c.228_231delinsCGAC,	YES
7_26921744	chr7	26921744	26921744	pav_upstream	Seita.7G195800.v2.2(dist=544)	NO
7_26932181	chr7	26932181	26932181	pav_downstream	Seita.7G196000.v2.2(dist=884)	NO