

SLAF-seq Efficiently Identifies SNP Markers for Wheat (*Triticum aestivum* L.) Development

Dongsheng Yang

Hetao College

Hao Liang

Hetao College

Shijun Sun

Hetao College

Haiwei Wang

Hetao College

Chao Cui

Hetao College

Ruinian Xu

Hetao College

Yulei Liu

Hetao College

Che Liu

Hetao College

Lei Wang

Hetao College

Shuiyuan Hao

haoshuiyuan@126.com

Hetao College

Xu Gao

Inner Mongolia Tianfu Hetao Germplasm Science and Technology Development Co., Ltd.

Research Article

Keywords: Wheat, SLAF-seq, SNP, Genetic diversity, Fingerprint map

Posted Date: February 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8808792/v1>

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomics on May 21st, 2026. See the published version at <https://doi.org/10.1186/s12864-026-12921-6>.

Abstract

Molecular markers are indispensable tools for identifying genetic variation among plant individuals and enhancing breeding efficiency. In this study, we developed SNP markers, assessed genetic diversity, and established fingerprint maps for 306 wheat germplasm accessions from China using SLAF-seq technology. We obtained 4978.16 Mb of clean reads after quality control of individual sample sequencing data. The number of SNP markers detected per sample ranged from 7.03 to 35.92 million. A total of 554,315 SLAF tags were identified, including 356,643 polymorphic tags. After population-level SNP filtering, 52,228 highly consistent and effective SNP markers were retained. Genetic diversity analysis revealed relatively close genetic relationships among the wheat varieties, with an average observed heterozygosity of 0.090 and a mean polymorphism information content (PIC) of 0.251. Population structure analysis ($K = 4$) indicated that most accessions shared close ancestral relationships, with evidence of admixture. Cluster analysis grouped the 306 wheat germplasm resources into four distinct clusters. Further filtering identified 114 core SNP markers, enabling the successful construction of a fingerprint database encompassing all 306 accessions. This study demonstrates that SLAF-seq is a cost-effective and efficient method for high-throughput SNP marker development and a powerful tool for wheat germplasm genetic analysis. The SNP markers identified here can facilitate germplasm identification, varietal improvement, protection, utilization, and QTL mapping of important traits with yield and quality, significantly advancing molecular breeding efforts in wheat.

1. INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the world's major food crops and the second-largest staple crop in China, providing approximately 20% of protein intake and 21% of caloric intake for humans. By 2050, the global population is projected to exceed 9 billion (<https://population.un.org/wpp/>); wheat production will need to be sufficient to ensure food security [1]. Breeders have been gradually developing wheat varieties adapted to various cultivation environments through artificial selection based on landraces [2, 3]. Researchers often focus on identifying and utilizing genes closely associated with important and complex agronomic traits, as this targeted trait-based selective breeding approach yields higher efficiency [4, 5]. Traditional linkage analysis and association mapping are effective methods for identifying genetic locus variations, but they still exhibit limitations in detecting genetic variations related to domestication and improvement [6, 7]. Studying genetic variation regions across different populations or within populations can help identify loci under selection [8, 9].

The cultivation of excellent varieties is crucial for ensuring high and stable wheat yields in China. However, when breeding for wheat varieties with superior agronomic traits, breeders have a tendency to overemphasize the utilization of a limited number of elite wheat germplasms. This has led to an increasingly narrow genetic base in the wheat germplasms that have been bred, which in turn has resulted in slow progress and limited advancement in wheat breeding. Consequently, research on germplasm resources, as a fundamental aspect of developing new wheat varieties, has garnered increasing attention among researchers. Analyzing the genetic diversity of germplasms, understanding the genetic basis and kinship of existing germplasms, and identifying and utilizing superior germplasm materials are of great significance for broadening the genetic base of wheat breeding, accelerating the breeding process, and enhancing breeding efficiency.

Wheat has a large and complex genome with a high proportion of repetitive sequences or transposons [10]. As such, high-throughput, high-density genotyping strategies are often used to reduce research costs [11]. Single nucleotide polymorphisms (SNPs), which are third-generation DNA molecular markers, are highly accurate, offer high throughput, and are low-cost. SNPs are widely used in plant molecular biology research [12–14]. Many

technical methods for SNP marker development exist, including GBS, SLAF, and RAD [15–17]. Specific-locus amplified fragment sequencing (SLAF-seq) is a reduced-representation genome sequencing technology that combines high-throughput sequencing to screen representative SLAF fragments [18]. First, in-silico enzyme digestion is predicted based on the reference genome of the species or a closely-related one. Then, two optimal restriction enzymes are chosen for double digestion. Finally, 300–500-bp-long enzyme-digested fragments are identified as SLAF tags. The advantages of SLAF-seq are applicability to species with or without a reference genome, uniform tag distribution across the genome, suitably-sized enzyme-digested fragments, minimal sequencing data redundancy, and high data utilization efficiency [19]. SLAF-seq-based SNP marker development has been widely used in major crops such as soybean (*Glycine max* (L.) Merr.), maize (*Zea mays* L.), and rice (*Oryza sativa* L.), for germplasm resource kinship identification, genetic diversity analysis, high-density genetic map construction, and gene mapping [20–22]. The wheat genome is extremely large and complex, with significant heterozygosity. Reduced-representation genome sequencing can overcome these issues, creating favorable conditions for large-scale SNP development.

In this study, we used SLAF-seq to develop wheat genome-wide SNP markers and analyzed the genetics of 306 wheat samples. Sequences were examined against the wheat reference genome. Using 17,233 novel genome-wide SNPs, we gauged genetic diversity and population structure. We provide key data for wheat-related research such as genome-wide association studies and germplasm evaluation. Our results could aid in wheat resource conservation and utilization, as well as genetic mapping, important-trait association analysis, and marker-assisted breeding.

2. MATERIALS AND METHODS

2.1 Test Materials

The experimental materials consisted of 306 wheat germplasm resources with spring and weak winter characteristics collected from various regions in China, all of which could grow normally in the Bayannur area (S1 Table). They were planted at the Hetao College Crop Experimental Field in April 2024.

2.2 DNA Extraction and Detection

From each sample, around 150 mg of tender leaf tissue was gathered, swiftly frozen in liquid nitrogen, and stored at – 80°C in an ultra-cold freezer for later use. Genomic DNA of wheat was extracted via the CTAB method. An ultra-micro spectrophotometer, the NanoDrop ND 2000, was employed to assay the DNA concentration and 1.5% agarose gel electrophoresis was used to check that the DNA quality was sufficient for library preparation and sequencing.

2.3 Library Construction

The Wheat_KN9204 genome (<https://ngdc.cncb.ac.cn/gwh/Assembly/25997/show>) was selected as the wheat reference genome for restriction enzyme digestion prediction, with a genome size of 14.47 Gb and a GC content of 46.16%. The optimal restriction enzyme digestion strategy was developed to ensure the following: 1) a minimal proportion of restriction fragments located in repetitive sequences; 2) a relatively uniform distribution of restriction fragments across the genome; 3) fragment lengths that matched with the specific experimental system [23]; and 4) a sufficient number of restriction fragments (SLAF tags) to meet the expected tag count. Qualified genomic DNA samples were individually digested and the resulting restriction fragments (SLAF tags) underwent 3'-end A-

tailing, dual-index adapter ligation, PCR amplification, purification, pooling, gel excision, and target fragment selection [24].

2.4 Illumina HiSeq Sequencing and Data Quality Assessment

After passing the library quality inspection, high-throughput sequencing was performed using Illumina HiSeq. Paired-end sequencing was employed to identify raw data and obtain the read count for each sample. Complex structural regions on the genome (such as loop domains and consecutive restriction sites), low individual purity of genomic DNA, insufficient digestion time of restriction enzymes, and other factors may affect the activity of restriction enzymes, leading to some restriction sites remaining uncut. The sequencing quality score (Q) is a crucial metric for evaluating the single-base error rate in high-throughput sequencing. A higher sequencing quality score corresponds to a lower base-calling error rate. If the probability of a base-calling error is 0.001, the quality score of that base should be Q30.

2.5 Development of SNP Markers

Reads were sequenced from restriction fragments produced by the same restriction enzyme acting on different samples. Reads derived from SLAF-seq were clustered based on sequence similarity. When SLAF tag sequences differed between samples, these were considered polymorphic SLAF tags. SNP markers were developed based on the reference sequence with the highest depth within each SLAF tag. Sequencing reads were aligned to the reference genome using BWA [25], and SNPs were called using GATK [26] and SAMtools [27]. The intersection of SNP markers obtained from both methods was considered a reliable SNP marker dataset. Finally, SNPs with a minor allele frequency (MAF) > 0.05 and completeness > 0.8 were selected for subsequent analysis.

2.6 Phylogenetic Tree Construction

The developed SNP markers were preliminarily filtered according to a criteria of completeness > 0.5 and MAF > 0.05. Based on the initially screened population-consistent SNPs, a phylogenetic tree of all samples was constructed using MEGA X [28] software with the neighbor-joining method under the Kimura 2-parameter model, with 1,000 bootstrap replicates.

2.7 Analysis of Genetic Diversity

SNPs developed from populations composed of non-genetic group materials should possess characteristics such as high marker quality, strong representativeness, high discriminative power among materials, uniform genomic distribution, and strong specificity. Based on these principles, markers should be uniformly distributed across the genome and markers should have no missing loci (i.e., a locus completeness of 100%). Loci were discarded if any of the following were true: MAF < 20%, PIC < 0.35, and P-value from Hardy-Weinberg equilibrium testing < 0.01. Also, if there were no other locus mutations within 100 bp upstream and downstream of the selected markers, the marker was discarded.

The PowerMarker V3.25 software was used to calculate the MAF, expected number of alleles, observed number of alleles, expected heterozygosity (H_e), observed heterozygosity (H_o), Nei's diversity index (H), PIC, and Shannon-Wiener index (I).

2.8 Population Genetic Structure and Principal Component Analysis

Population structure analysis was conducted using Admixture (v1.22) [29]. The number of subgroups (*K* value) was preset from 1 to 10 for clustering, and the clustering results were cross-validated. The optimal number of clusters was determined based on the trough of the cross-validation error rate [30]. The R package Pophelper was used to generate stacked Q-matrix plots for each *K* value (<http://royfrancis.github.io/pophelper>). The smartPCA program in the Eigensoft (v6.0) [31] package was employed to perform principal component analysis (PCA) based on SNP data, revealing the clustering patterns of the samples. Principal component analysis was utilized to transform multiple indicators into several comprehensive indicators for analyzing genetic relationships among individuals.

2.9 Construction of SNP Fingerprint Map

Based on the high-quality core SNP markers selected, genotyping results were obtained for all samples. SNP loci with consistent genotypes among replicate samples were identified and converted into binary-encoded data to construct the SNP fingerprint map of 306 germplasm resources containing wheat varieties and high-generation wheat lines.

3. RESULTS

3.1 SLAF Library Quality Assessment

SLAF-seq reads are genomic DNA restriction fragments with a base distribution influenced by restriction enzyme sites and PCR amplification. The first two bases of the sequencing reads exhibited base separation consistent with the restriction sites, while subsequent bases showed varying degrees of fluctuation (Fig. 1A). Using SLAF-predict software to perform *in silico* restriction digestion prediction on the wheat reference genome, the restriction enzyme combination was determined to be Rsa I + Hae III. Digestion efficiency is a key indicator for evaluating the success of SLAF experiments. In this study, the paired-end mapping efficiency was 88.72%, with residual restriction sites accounting for 11.28%, indicating normal SLAF tag digestion. Sequences with fragment lengths between 464 and 494 bp were defined as SLAF tags, and 560,935 SLAF tags were predicted, which were generally evenly distributed across the genome. To assess the normality of the SLAF experimental process and the effectiveness of the restriction digestion protocol, the SOAP software was used to align sample reads to the reference genome. The results demonstrated a paired-end mapping efficiency of 91.66%, single-end mapping efficiency of 4.23%, and unmapped efficiency of 4.11%. Additionally, the insert size distribution fell within the expected range (Fig. 1B), confirming that the SLAF library construction was normal and effective.

3.2 Sequencing Data Statistics and Evaluation

To ensure analytical quality, 126 bp × 2 read lengths were used for subsequent data evaluation and analysis. Sequencing of 306 wheat materials yielded read counts ranging from 7,032,048 to 35,918,412, totaling 4978.16 Mb reads, with sample W33 exhibiting the longest sequences. The sequencing Q30 values ranged between 85.03% and 92.27%, with an average of 90.65%. The GC content varied from 47.38% to 51.53%, averaging 48.55%, indicating high sequencing quality and reliable results. Furthermore, alignment with the reference genome showed that the percentage of clean reads mapped to the reference genome ranged from 92.82% to 99.70%, with an average of 99.11%. The proportion of paired-end sequences properly aligned to the reference genome with insert sizes matching the expected fragment length distribution ranged from 77.32% to 96.58%, averaging 86.53%.

3.3 Development of SLAF Tags and SNP Markers

A total of 554,315 SLAF tags were obtained from the genome sequencing of 306 wheat samples, with an average sequencing depth of 10.97×, among which 356,643 SLAF tags were polymorphic. The SLAF tags were mapped to the reference genome using BWA software, and their distribution across chromosomes was visualized (Fig. 2A). A total of 5,232,218 population SNPs were developed, with the number of SNP markers per sample ranging between 1,862,063 and 3,095,277. The completeness of these SNP markers was 35.59%–59.16%, and the heterozygosity rate was 3.63%–14.50%. The distribution of SNPs across chromosomes was also plotted (Fig. 2B). The developed SLAF tags and SNPs covered nearly all chromosomes, with the highest SNP density observed on chromosomes GWHBJWI00000004, GWHBJWI00000005, GWHBJWI00000008, and GWHBJWI00000020. After filtering the population SNPs, 52,228 highly consistent and effective SNP markers were obtained for subsequent germplasm genetic evolution analysis.

3.4 Genetic Diversity and Principal Component Analysis

Through rigorous screening of indicators, such as the heterozygosity rate, the missing rate, and polymorphism, high-quality SNP markers were obtained and used to estimate genetic diversity parameters for 306 wheat samples. The expected allele number was 1.105–2.000, with an average of 1.509, and the observed allele number was 2.000 for all markers. The expected heterozygous number was 0.095–0.500, with an average of 0.308, and the observed heterozygosity number ranged from 0.003 to 1.000, with an average value of 0.090. The Nei diversity index ranged from 0.095 to 0.502, with an average of 0.309. The PIC ranged from 0.090 to 0.375, with an average of 0.251. The Shannon Wiener index ranged from 0.199 to 0.693, with an average of 0.475 (Table 1). These results indicate that the dataset is suitable for genotyping and fingerprint construction of the 306 wheat samples.

Table 1
Genetic diversity index of wheat population samples

Index	Expected allele number	Expected heterozygous number	Observed allele number	Observed heterozygous number	Nei diversity index	Polymorphism information content	Shannon Wiener index
Range	1.105–2.000	0.095–0.500	2.000–2.000	0.003–1.000	0.095–0.502	0.090–0.375	0.199–0.693
Average	1.509	0.308	2.000	0.090	0.309	0.251	0.475

Using the neighbor-joining algorithm of the MAGA software, a phylogenetic tree of 306 wheat samples was constructed, with the 306 samples divided into four groups (Fig. 3). Group 1 contained the largest number of samples at 159, which were further divided into two subgroups: subgroup 1 comprised 14 samples, while subgroup 2 was further divided into three smaller groups containing 25, 25, and 95 samples respectively. Group 2 contained 50 samples; Group 3 contained 56 samples; and Group 4 contained 41 samples. PCA was conducted based on high-quality group SNPs, and the clustering of individuals is shown in Fig. 4. Principal component PC1, PC2, and PC3, accounted for 9.59%, 4.29%, and 3.01% of wheat individual genetic variation, respectively. The wheat individuals were divided into four main clusters, with Cluster A containing the largest number of individuals, indicating significant differences in genetic structure among individuals.

3.5 Population Structure Analysis

To further elucidate the genetic background relationships of wheat, the identified highly consistent SNP molecular markers were used to analyze the population genetic structure of 306 wheat germplasms using the Admixture

software. The results revealed a relatively distinct population structure among the 306 wheat germplasms, with the lowest cross-validation error rate corresponding to $K = 4$, thus determining the optimal cluster number as four, which was highly consistent with the phylogenetic tree classification results (Fig. 5). When $K = 4$, Group I comprised 75 samples dominated by ancestral Q1, Group II included 102 samples dominated by ancestral Q2, Group III consisted of 76 samples dominated by ancestral Q3, and Group IV contained 53 samples dominated by ancestral Q4. Most populations exhibited signs of admixture, indicating close relationships among individuals and the presence of certain genetic exchanges (Fig. 6).

Population structure analysis ($K = 4$) results indicated that the genetic heterozygosity proportion was greater than that of morphological analysis. According to traditional classification methods, wheat seeds are divided into hard wheat and soft wheat based on endosperm texture. Among the 306 materials, 121 were hard wheat with compact endosperm structure, translucent appearance, and a hardness index above 60. The remaining were soft wheat with hardness indices ranging from 33 to 42, loose endosperm structure, and gypsum-like appearance.

3.6 Core germplasm screening

Core germplasm represents the maximum range of genetic diversity of morphological characteristics, geographical distribution, genes and genotypes of a particular kind of plant or its wild relatives. It has significant academic and practical significance for promoting germplasm exchange, utilization and gene bank management. Core Hunter II software can extract diverse, representative and least redundant subsets from a large number of germplasm resources to construct core germplasm or micro-core germplasm. Based on the genetic variation marker (SNP) data combined with multiple evaluation measures (modified Rogers distance, Shannon's diversity index, etc.) for weighted processing, we determined that the screened materials have high diversity, high representativeness and high locus richness. Using Core Hunter II software combined with the weighted index modified Rogers distance (0.7) and Shannon's diversity index (0.3), screening was conducted according to the gradients of the total germplasm resource ratio of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. The gene coverage of the screened materials was evaluated, and finally 61 materials were determined as core germplasms.

3.7 SNP Fingerprint Construction

Based on phylogenetic analysis, population structure, and PCA, the genotypes of 114 SNP loci from 306 wheat varieties were determined to construct fingerprint profiles. The nucleotide bases at each locus exhibited four possible types: A, T, G, and C. According to SNP principles, each SNP locus may present four variant types: substitution, transversion, deletion, and insertion. The four homozygous loci types are A/A, G/G, C/C, and T/T, while the heterozygous loci include A/C, A/T, A/G, C/T, C/G, and T/G. Different colors were assigned to represent the four homozygous loci: yellow, green, blue, and purple, respectively, while white was used to denote all heterozygous loci (Fig. 7). Finally, individual and composite fingerprint profiles were constructed for the 306 wheat varieties using the 114 SNP loci. These profiles visually demonstrate genomic-level differences among common cultivated wheat varieties, providing a foundation for the identification, protection, and utilization of wheat germplasm resources.

4. DISCUSSION

4.1 Development of Molecular Markers

Interspecies genetic diversity and population evolution can be analyzed using various molecular markers, such as SSR, AFLP, SRAP, and SCot [32–34]. However, most of these molecular marker methods suffer from drawbacks such as cumbersome operation, poor stability, high cost, and insufficient accuracy. The use of different molecular markers for analyzing the same species can yield varying genetic diversity results, which is often directly related to the sensitivity and detection characteristics of the molecular markers themselves. SNP markers, as third-generation molecular markers based on genome sequencing, exhibit advantages such as high polymorphism, high density, large numbers, genetic stability, wide distribution, and high-throughput detection [35, 36]. They are ideal markers for studies on population genetic evolution, gene mapping, genetic linkage map construction, and genome-wide association analysis. The key to SLAF-seq technology lies in selecting the optimal enzyme digestion scheme to digest the species' genome, followed by sequencing to obtain genomic variation information of the target species. Unlike whole-genome sequencing, this approach avoids a large number of repetitive sequences, significantly reducing sequencing costs. It has played an important role in SNP mining for species such as hollyhock (*Alcea rosea* L.) [37], rose (*Rosa chinensis* Jacq.) [38], maize [39], rice [40], *Miscanthus* [41], and eggplant (*Solanum melongena*) [42]. In this study, SLAF-seq technology was employed to conduct reduced-representation genome sequencing on 306 wheat materials with an average sequencing depth of 10.97×. To improve marker quality, GATK and SAMtools were used for joint screening, resulting in 5,232,218 population SNP markers. After filtering, 52,228 high-quality SNP markers were obtained, fully meeting the requirements for specific locus analysis and differential genetic information discovery among the tested materials. The substantial number of SNP markers obtained can be utilized for genome-wide association studies (GWAS) on key traits related to wheat yield, quality, and stress resistance. And the results will provide a valuable reference for future molecular genetic research in wheat.

4.2 Analysis of Genetic Diversity

Genetic diversity is one of the core subjects in biological research and serves as the foundation for species survival and evolution [43]. Studying the genetic diversity of populations can provide deeper insights into a species' adaptability to the environment and its genetic structure, facilitating the conservation of germplasm resources for important species. Genetic diversity evaluation parameters reflect the levels of genetic variation among and within populations. In this study, the expected number of alleles (N_e) ranged from 1.105 to 2.000, with an average of 1.509, showing a certain discrepancy from the observed number of alleles, indicating uneven distribution of alleles within the population. MAF, which is commonly used to distinguish between common and rare variants in populations [44], was moderate at 0.224. PIC reflects the degree of variation at SNP loci. According to the theory proposed by Botstein [24], $PIC \geq 0.5$ indicates high polymorphism; PIC between 0.25 and 0.5 indicates moderate polymorphism; and $PIC \leq 0.25$ indicates low polymorphism. Nei's genetic diversity index reflects the level of genetic diversity within a population, with higher values indicating lower genetic uniformity and higher genetic diversity. SNP analysis results showed that the 306 wheat accessions tested, originating from 17 regions in China, exhibited a PIC ranging from 0.090 to 0.375, a Nei's diversity index between 0.095 and 0.502, and a total of 1,191,222 polymorphic markers, indicating a relatively high level of genetic diversity among the tested germplasms. Of the wheat accessions tested, 170 were spring wheat distributed in northeastern and northwestern China, including Heilongjiang, Inner Mongolia, Ningxia, Xinjiang, and Gansu, regions characterized by lower temperatures and shorter growing seasons. The remaining 136 wheat accessions were weak winter wheat, mainly distributed in Shandong, Henan, Hebei, Jiangsu, and Anhui, regions with milder climates and longer growing seasons. Generally, the richer the genetic diversity of a species, the stronger its adaptability to the environment [45]. In this study, all 136 weak winter wheat varieties were able to grow normally in Bayannur, a spring wheat cultivation region, further demonstrating that higher genetic diversity is the basis for better environmental

adaptation in species. In recent years, despite the development of numerous wheat varieties and increasingly abundant germplasm resources, the high-intensity artificial selection during breeding and the frequent use of the same parental lines have significantly reduced population genetic diversity, leading to a narrow genetic base among varieties and making it difficult to select new breakthrough cultivars. This may be related to breeders' converging selection criteria for agronomic traits, quality traits, and resistance, as well as the low diversity of climate types.

4.3 Phylogenetic Tree Construction and Population Structure Analysis

Phylogenetic trees can clearly reflect the genetic relationships among different groups but cannot reveal the genetic composition of individuals. In contrast, genetic structure analysis can display the genetic structure of test materials and the genetic composition of individuals at the DNA level, clarifying the ancestral origins of individuals and the exchange of genetic information among different genotypes. This serves as an effective method for analyzing the genetic backgrounds of diverse test materials and constitutes a fundamental task in studies on genetic diversity, varietal evolution, and association analysis of quantitative traits. It lays the foundation for understanding genetic evolution, resource conservation, and the utilization of species [46]. In this study, the classification results of the constructed evolutionary tree and the cross-validation error rate trough (K value) corresponding to the population structure both indicated that the optimal number of clusters for the 306 wheat samples was four groups. The constructed evolutionary tree showed that the 109 samples in Group 1 and the 50 samples in Group 2 were genetically close, suggesting a relatively close kinship between these two groups.

The results demonstrated that genotypes could not be precisely clustered based on their geographic origins. On the one hand, some genotypes from different geographic origins were grouped into the same cluster, indicating a certain degree of correlation among them, possibly due to germplasm exchange or cross-border trade that facilitated the transfer of germplasm from one region to another. On the other hand, some genotypes from the same geographic origin clustered into different groups, suggesting a certain degree of genetic differentiation among populations. Genotype grouping that does not reflect geographic origins has also been observed in studies on the genetic diversity of cowpea germplasm [47].

5. CONCLUSIONS

This study utilized SLAF-seq technology to develop SNP markers for 306 spring and weak winter wheat samples adapted to the Hetao region, obtaining a total of 554,315 SLAF tags, among which 356,643 were polymorphic. A total of 5,232,218 population SNP markers were developed, and after population filtering, 52,228 highly consistent and effective SNP markers were obtained. Calculation of population genetic diversity indices indicated that wheat exhibits high genetic diversity at the population level. Using the screened SNP markers, a phylogenetic tree was constructed, classifying the 306 wheat sample resources into four groups. The findings of this study can provide a reference for the exploitation, utilization, and scientific conservation of wheat resources.

Authors' contributions

Dongsheng Yang and Shuiyuan Hao conceived and designed the research. Dongsheng Yang and Xu Gao performed the experiments. Dongsheng Yang and Hao Liang completed the writing of the article. Haiwei Wang, Shijun Sun, Chao Cui, Ruinian Xu, Yulei Liu, Che Liu and Lei Wang assisted in the completion of the experiments. All

authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Declarations

Authors' contributions

Dongsheng Yang and Shuiyuan Hao conceived and designed the research. Dongsheng Yang and Xu Gao performed the experiments. Dongsheng Yang and Hao Liang completed the writing of the article. Haiwei Wang, Shijun Sun, Chao Cui, Ruinian Xu, Yulei Liu, Che Liu and Lei Wang assisted in the completion of the experiments. All authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

Science and Technology Plan Project of Inner Mongolia Autonomous Region (2025YFHH0252); Science and Technology Plan Project of Bayannur (NMKJXM202408); Development of Inner Mongolia Through Science and Technology of China (NMKJXM202201); Development of Inner Mongolia Through Science and Technology of China (NMKJXM202302).

Data Availability

The raw sequencing data supporting the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1314945. These data are publicly available and can be accessed through the NCBI SRA database.

Ethics approval and consent to participate

The methods involved in the current study were carried out in obedience with intuitional and national regulations.

Consent for publication

Not applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

References

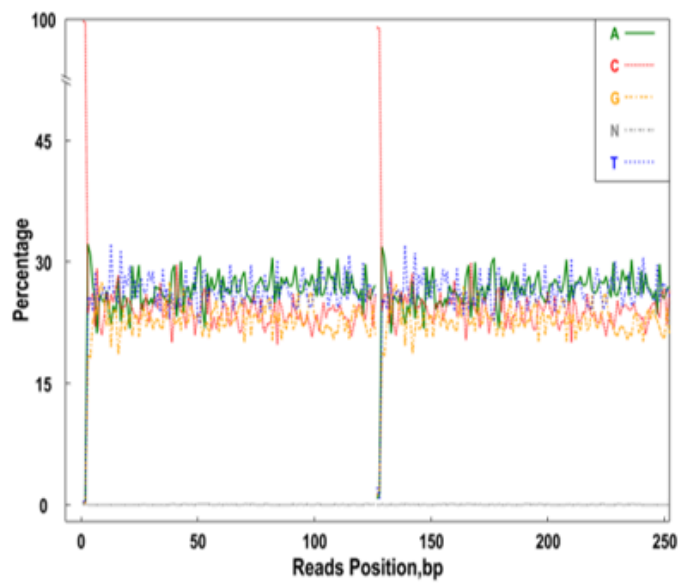
1. Afzal F, Li HH, Gul A, Subhani A, Ali A, Mujeeb-Kazi, A, Ogbonnaya F, Trethowan R, Xia XC, He ZH, Rasheed A: **Genome-wide analyses reveal footprints of divergent selection and drought adaptive traits in synthetic-derived wheats.** *Genes-Genomes-Genetics* 2019, **9**(6): 1957-1973.
2. Lopes MS, El-Basyoni I, Baenziger PS, Singh S, Royo C, Ozbek K, Aktas H, Ozer E, Ozdemir F, Manickavelu A, Ban T, Vikram P: **Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change.** *Journal of Experimental Botany* 2015, **66**(12): 3477-3486.
3. Rasheed A, Mujeeb-Kazi A, Ogbonnaya FC, He ZH, Rajaram S: **Wheat genetic resources in the post-genomics era: promise and challenges.** *Annals of botany* 2018, **121**(4): 603-616.

4. Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peebles M, Lepak N, Prim L, Romay MC, Ross-Ibarra J, Sanchez-Gonzalez JDJ, Schmidt C, Schuenemann VJ, Krause J, Matson RG, Weigel D, Buckler ES, Burbano HA: **Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America.** *Science* 2017, **357**(6350): 512-515.
5. Zhou Y, Chen ZX, Cheng MP, Chen J, Zhu TT, Wang R, Liu YX, Qi PF, Chen GY, Jiang QT, Wei YM, Luo MC, Nevo E, Allaby RG, Liu DC, Wang JR, Dvorák J, Zheng YL: **Uncovering the dispersion history, adaptive evolution and selection of wheat in China.** *Plant Biotechnology Journal* 2018, **16**(1): 280-291.
6. Heslot N, Jannink JL, Sorrells ME. **Perspectives for genomic selection applications and research in plants.** *Crop Science* 2015, **55**(1) :1-12.
7. Morrell PL, Buckler ES, Ross-Ibarra J: **Crop genomics: advances and applications.** *Nature Reviews Genetics* 2011, **13**(2): 85-96.
8. Wei DY, Cui YX, He YJ, Xiong Q, Qian LW, Tong CB, Lu GY, Ding YJ, Li JN, Jung C, Qian W: **A genome-wide survey with different rapeseed ecotypes uncovers footprints of domestication and breeding.** *Journal of Experimental Botany* 2017, **68**(17): 4791-4801.
9. Stephan W: **Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation.** *Molecular Ecology* 2016, **25**(1): 79-88
10. Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, Mayer KFX, Olsen OA: **Genome interplay in the grain transcriptome of hexaploid bread wheat.** *Science* 2014, **345**(6194): 1250091.
11. Allen AM, Winfield MO, BurrIDGE AJ, Downie RC, Benbow HR, Barker G, Wilkinson PA, Coghill JA, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor CA, Griffiths S, Bentley AR, Alda M, Jack P, Phillips AL, Edwards KJ: **Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*).** *Plant Biotechnology Journal* 2017, **15**(3): 390-401.
12. Eltaher S, Li J, Freeman B, Singh S, Ali GS: **A genome-wide association study identified SNP markers and candidate genes associated with morphometric fruit quality traits in mangoes.** *BMC genomics* 2025, **26**(1): 120.
13. Lippolis A, Hollebrands B, Acierno V, de Jong C, Pouvreau L, Paulo J, Gezan SA, Trindade LM: **GWAS identifies SNP markers and candidate genes for off-flavours and protein content in faba bean (*Vicia faba* L.).** *Plants* 2025, **14**(2): 193.
14. Yang FY, Lang T, Wu JY, Zhang C, Qu H.J, Pu ZG, Yang F, Yu M, Feng JY: **SNP loci identification and KASP marker development system for genetic diversity, population structure, and fingerprinting in sweetpotato (*Ipomoea batatas* L.).** *BMC Genomics* 2024, **25**(1): 1245.
15. Adedugba AA, Adeyemo OA, Adetumbi AJ, Ilesanmi OJ, Ogunkanmi LA: **Genetic diversity and population structure of some Nigerian and four African countries' sorghum landraces [*Sorghum bicolor* (L.) Moench] using Genotyping-By-Sequencing (GBS) SNP markers.** *South African Journal of Botany* 2023, **162**: 495-504.
16. Feng JY, Zhao S, Li M, Zhang C, Qu HJ, Li Q, Li JW, Lin Y, Pu ZG: **Genome-wide genetic diversity detection and population structure analysis in sweetpotato (*Ipomoea batatas*) using RAD-seq.** *Genomics* 2020, **112**(2): 1978-1987.
17. Wang XY, Wang WJ, Zhan DM, Ge SS, Tang LQ: **Genome-wide SNP markers based on SLAF-Seq uncover genetic diversity of *Saccharina* cultivars in Shandong, China.** *Frontiers in Marine Science* 2022, **9**: 849502.

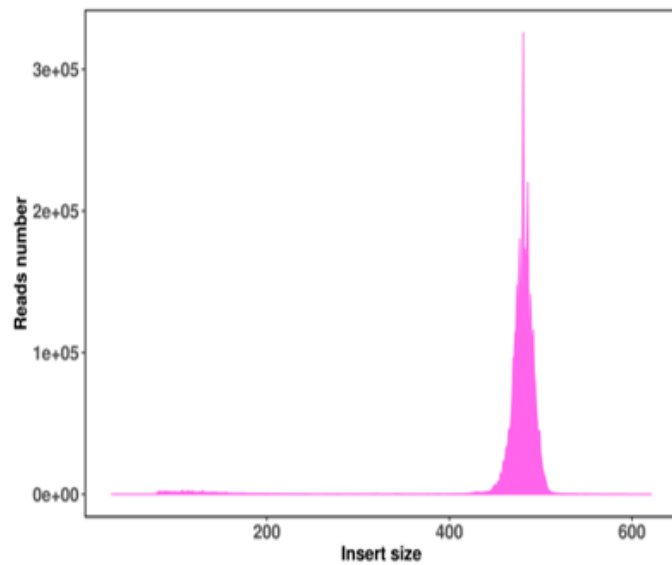
18. Sun XW, Liu DY, Zhang XF, Li WB, Liu H, Hong WG, Jiang CB, Guan N, Ma CX, Zeng HP, Xu CH, Song J, Huang L, Wang CM, Shi JJ, Wang R, Zheng XH, Lu CY, Wang XW, Zheng HK: **SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing.** *PLoS One* 2013, **8**(3): e58700.
19. Li J, Lin JY, Ou HB, Liu XS, Jiang Y, Liang RL: **Marker development and analysis of genetic diversity of phoebe bournei germplasms using SLAF-seq technology.** *Molecular Plant Breeding* 2021, **19**(13), 4517-4524
20. Ren HL, Han JN, Wang XR, Zhang B, Yu LL, Gao HW, Hong HL, Sun RJ, Tian Y, Qi XS, Liu ZX, Wu XX, Qiu LJ: **QTL mapping of drought tolerance traits in soybean with SLAF sequencing.** *The Crop Journal* 2020, **8**(6): 977-989.
21. Wen Y, Fang YX, Hu P, Tan YQ, Wang YY, Hou LL, Deng XM, Wu H, Zhu LX, Zhu L, Chen G, Zeng DL, Guo LB, Zhang GH, Gao ZY, Dong GJ, Ren DY, Shen L, Zhang Q, Xue DW, Qian Q, Hu J: **Construction of a high-density genetic map based on SLAF markers and QTL analysis of leaf size in rice.** *Frontiers in plant science* 2020, **11**: 1143.
22. Xia C, Chen LL, Rong TZ, Li R, Xiang Y, Wang P, Liu CH, Dong XQ, Liu B, Zhao D, Wei RJ, Lan H: **Identification of a new maize inflorescence meristem mutant and association analysis using SLAF-seq method.** *Euphytica* 2015, **202**(1): 35-44.
23. Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML: **Special features of RAD Sequencing data: implications for genotyping.** *Molecular ecology* 2013, **22**(11): 3151-3164.
24. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: **Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform.** *Applied and environmental microbiology* 2013, **79**(17): 5112-5120.
25. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**(14): 1754-1760.
26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: **The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data.** *Genome research* 2010, **20**(9): 1297-1303.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/map format and SAMtools.** *Bioinformatics* 2009, **25**(16): 2078-2079.
28. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: **Molecular evolutionary genetics analysis across computing platforms.** *Molecular Biology and Evolution* 2018, **35**(6): 1547-1549.
29. Alexander DH, Novembre J, Lange K: **Fast model-based estimation of ancestry in unrelated individuals.** *Genome research* 2009, **19**(9): 1655-1664.
30. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**(2): 945-959.
31. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nature Genetics* 2006, **38**(8): 904-909.
32. Deepika C, Venkatachalam SR, Yuvaraja A, Arutchenthil P, Indra N, Ravichandran V, Veeramani P, Kathirvelan P: **Comparison of genetic diversity using morphometric, molecular and digital imaging parameters in castor (*Ricinus communis* L.).** *Biocatalysis and Agricultural Biotechnology* 2025, **65**: 103525.
33. Hu SJ, Wang MZ, Yan XH, Cheng XM. **Genetic diversity and population structure of endangered orchid *Cypripedium flavum* in fragmented habitat using fluorescent AFLP markers.** *Plants* 2024, **13**(20): 2851.

34. Sinjare DYK, Abdulrahman SS, Khalid NS, Ismail RY, Naeem MY, Selamoglu Z, Issayev G, Ahmed AMN: **Molecular characterization and cuticular stomatal anatomy of *Punica Granatum* L. cultivars study in Dohuk governorate.** *Molecular Biology Reports* 2025, **52**(1): 287-287.
35. Sathishkumar R, Mohanrao MD, Geethanjali S, Prasad MSL, Senthilvel S: **A simple and cost-effective SNP genotyping assay for marker-assisted selection of wilt resistance in castor breeding.** *Industrial Crops and Products*, 2025, **226**: 120693.
36. Muturi P, Kyallo M, Gasura E, Yao N. **Diversity analysis and genome-wide association studies of seed weight trait in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) using diversity array technology sequence derived single nucleotide polymorphism markers.** *Euphytica* 2025, **221**(4): 34-34.
37. Deng WQ, Li YF, Chen X, Luo YZ, Pan YZ, Li X, Zhu ZS, Li FW, Liu XL Jia Y: **Development of single nucleotide polymorphism (SNP) markers and construction of DNA fingerprinting of *Alcea rosea* L. based on specific-locus amplified fragment sequencing (SLAF-seq) technology.** *Genetic Resources and Crop Evolution* 2024, **72**(2): 1-15.
38. Xia AN, Yang AA, Meng XS, Dong GZ, Tang XJ, Lei SM, Liu YG: **Development and application of rose (*Rosa chinensis* Jacq.) SNP markers based on SLAF-seq technology.** *Genetic Resources and Crop Evolution* 2022, **69**: 173-182.
39. Wen TT, Zhang XF, Zhu JJ, Zhang SS, Rhaman MS, Zeng W: **A SLAF-based high-density genetic map construction and genetic architecture of thermotolerant traits in maize (*Zea mays* L.).** *Frontiers in Plant Science* 2024, **15**: 1338086.
40. Liu X, Zhang N, Sun YR, Fu ZX, Han YH, Yang Y, Jia JC, Hou SY, Zhang BJ: **QTL mapping of downy mildew resistance in foxtail millet by SLAF-seq and BSR-seq analysis.** *Theoretical and Applied Genetics* 2024, **137**(7): 168.
41. Chen ZY, He YC, Iqbal Y, Shi YL, Huang HM, Yi ZL: **Investigation of genetic relationships within three *Miscanthus* species using SNP markers identified with SLAF-seq.** *BMC Genomics* 2022, **23**(1): 43.
42. Wei QZ, Wang WH, Hu TH, Hu HJ, Wang JL, Bao CL: **Construction of a SNP-based genetic map using SLAF-seq and QTL analysis of morphological traits in eggplant.** *Frontiers in Genetics* 2020, **11**: 178.
43. Souza IGB, Souza VAB, Lima PSC: **Molecular characterization of *Platonia insignis* Mart. ("bacurizeiro") using inter simple sequence repeat (ISSR) markers.** *Molecular biology reports* 2013, **40**(5): 3835-3845.
44. Dussault FM, Boulding EG: **Effect of minor allele frequency on the number of single nucleotide polymorphisms needed for accurate parentage assignment: A methodology illustrated using Atlantic salmon.** *Aquaculture Research* 2018, **49**: 1368-1372.
45. Pauls SU, Nowak C, Bálint M, Pfenninger M: **The impact of global climate change on genetic diversity within populations and species.** *Molecular ecology* 2013, **22**(4): 925-946.
46. Wei SP, Liu XF, Yang SX, Lv HY, Niu Y, Zhang YM: **Comparison of various clustering methods for population structure in Chinese cultivated soybean [*Glycine max* (L.) Merr.].** *Journal of Nanjing Agricultural University* 2011, **34**(2): 13-17.
47. Chipeta MM, Kafwambira J, Yohane, E: **Cowpea genetic diversity, population structure and genome-wide association studies in Malawi: insights for breeding programs.** *Frontiers in Plant Science* 2025, **15**: 1461631.

Figures



A



B

Figure 1

Distribution of insert fragments of control sequences

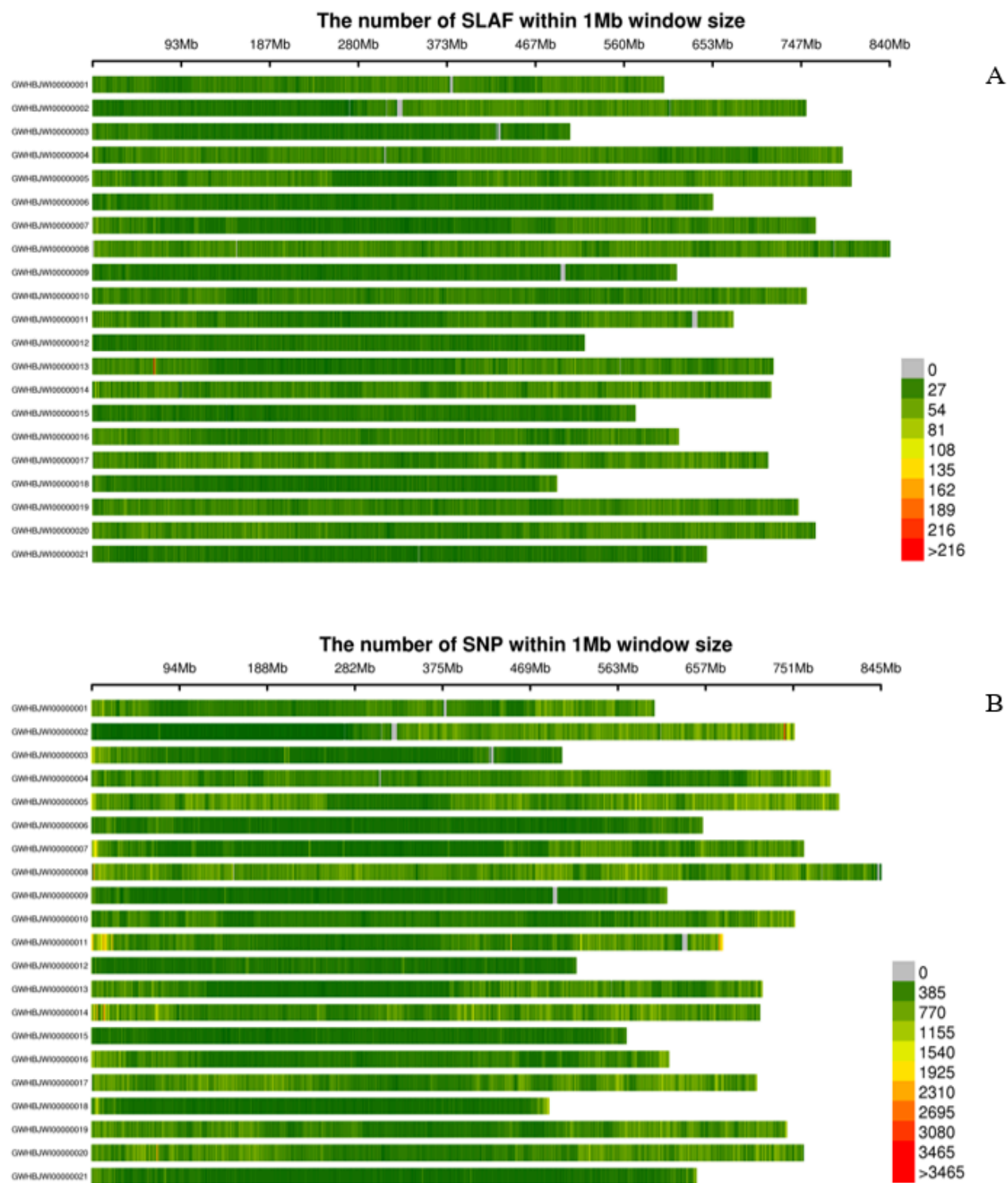


Figure 2

The distribution of SLAF tags and SNPs on chromosomes

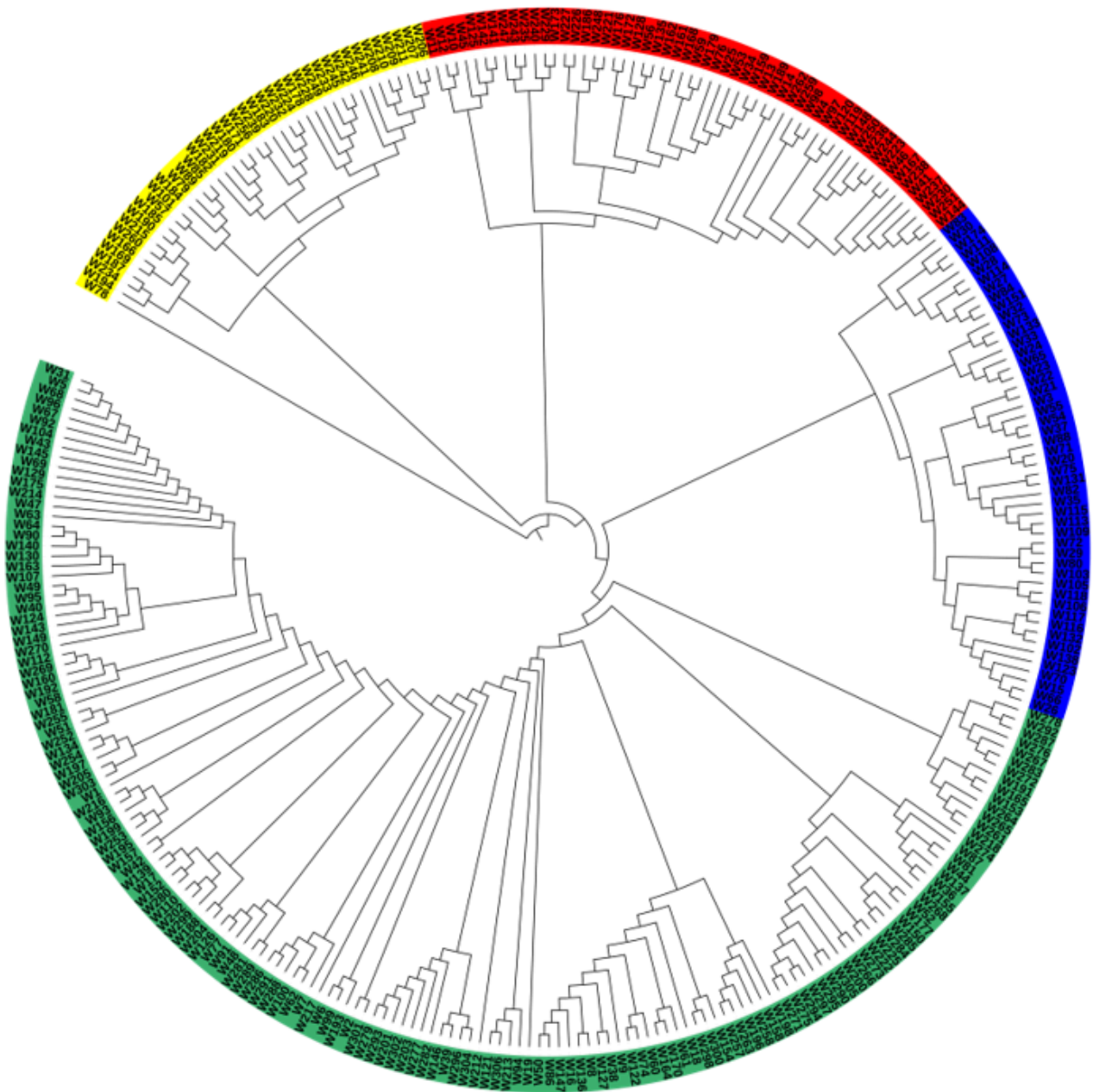


Figure 3

Phylogenetic tree of 306 wheat germplasm resources

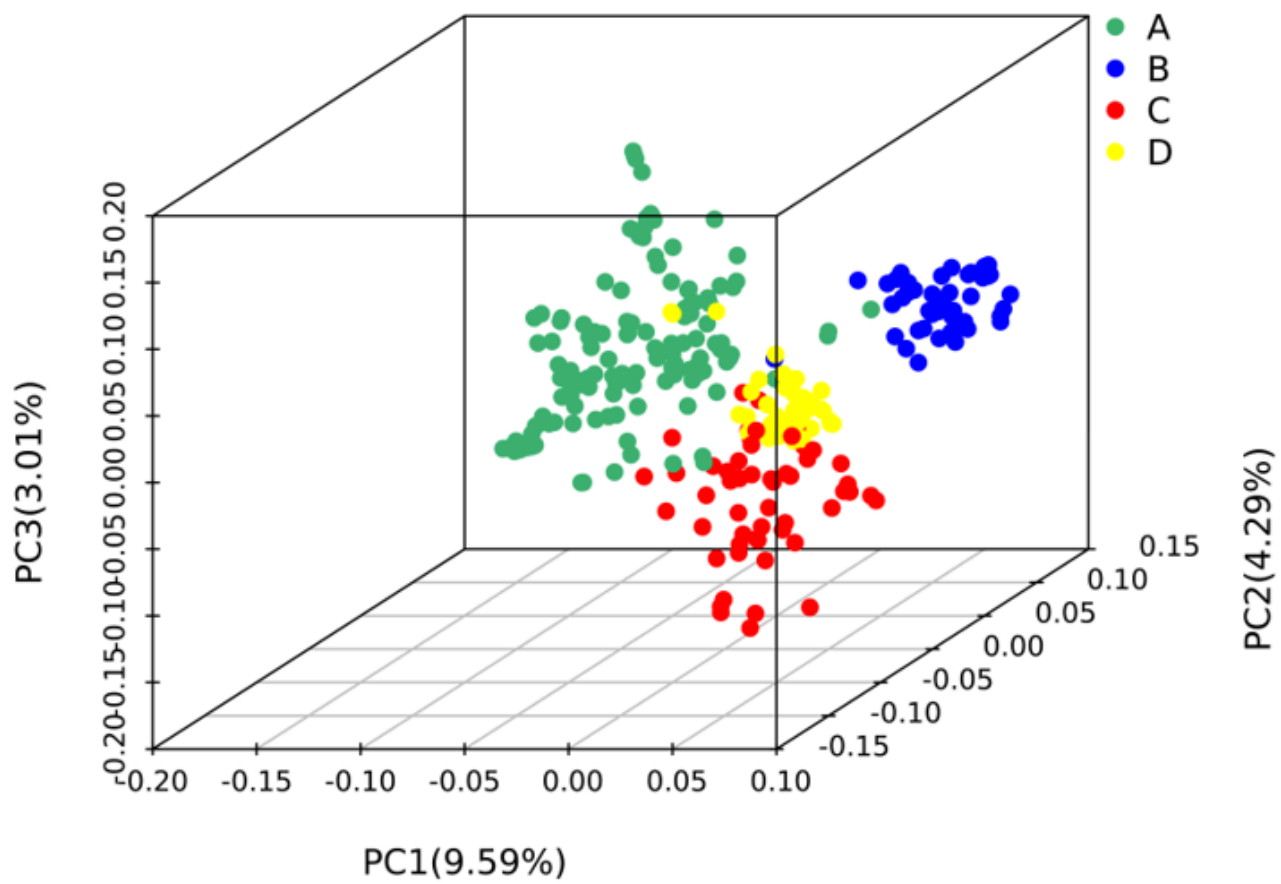


Figure 4

PCA plot of 306 wheat germplasm resources

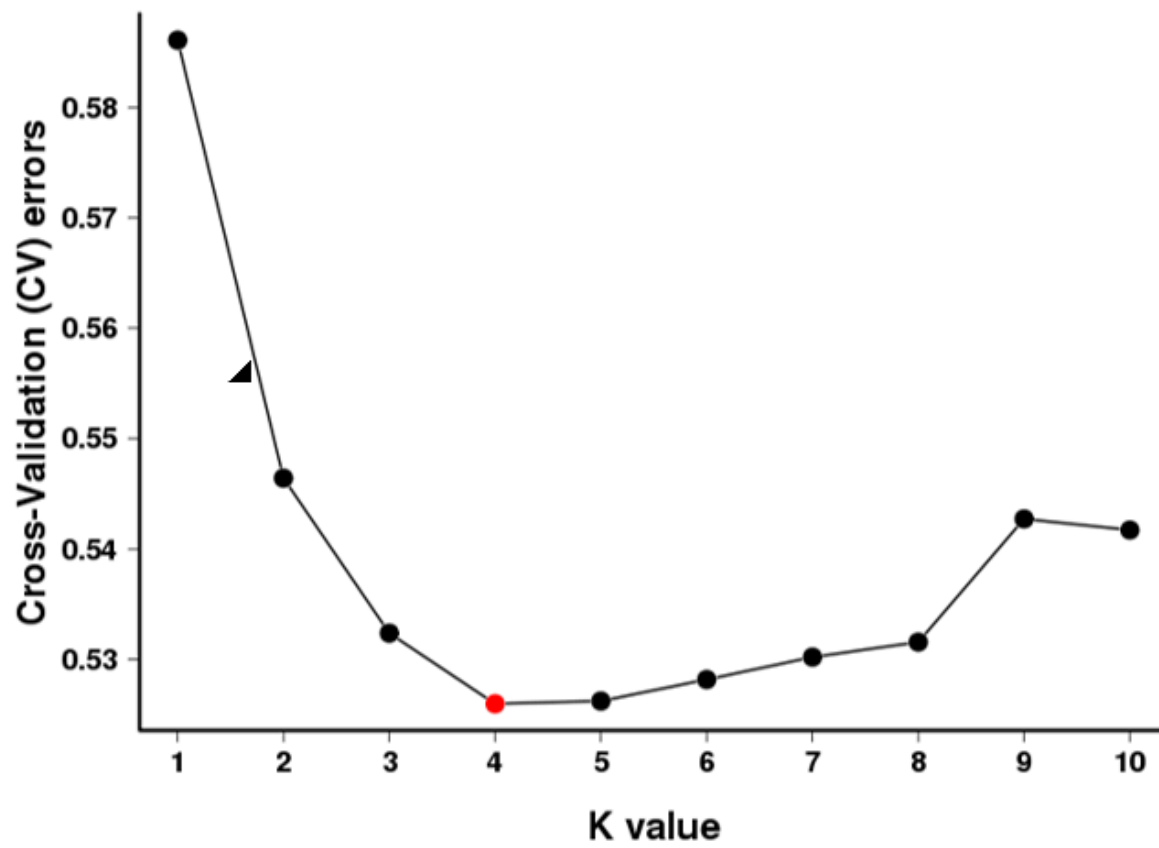


Figure 5

Cross-validation error rates corresponding to different K values

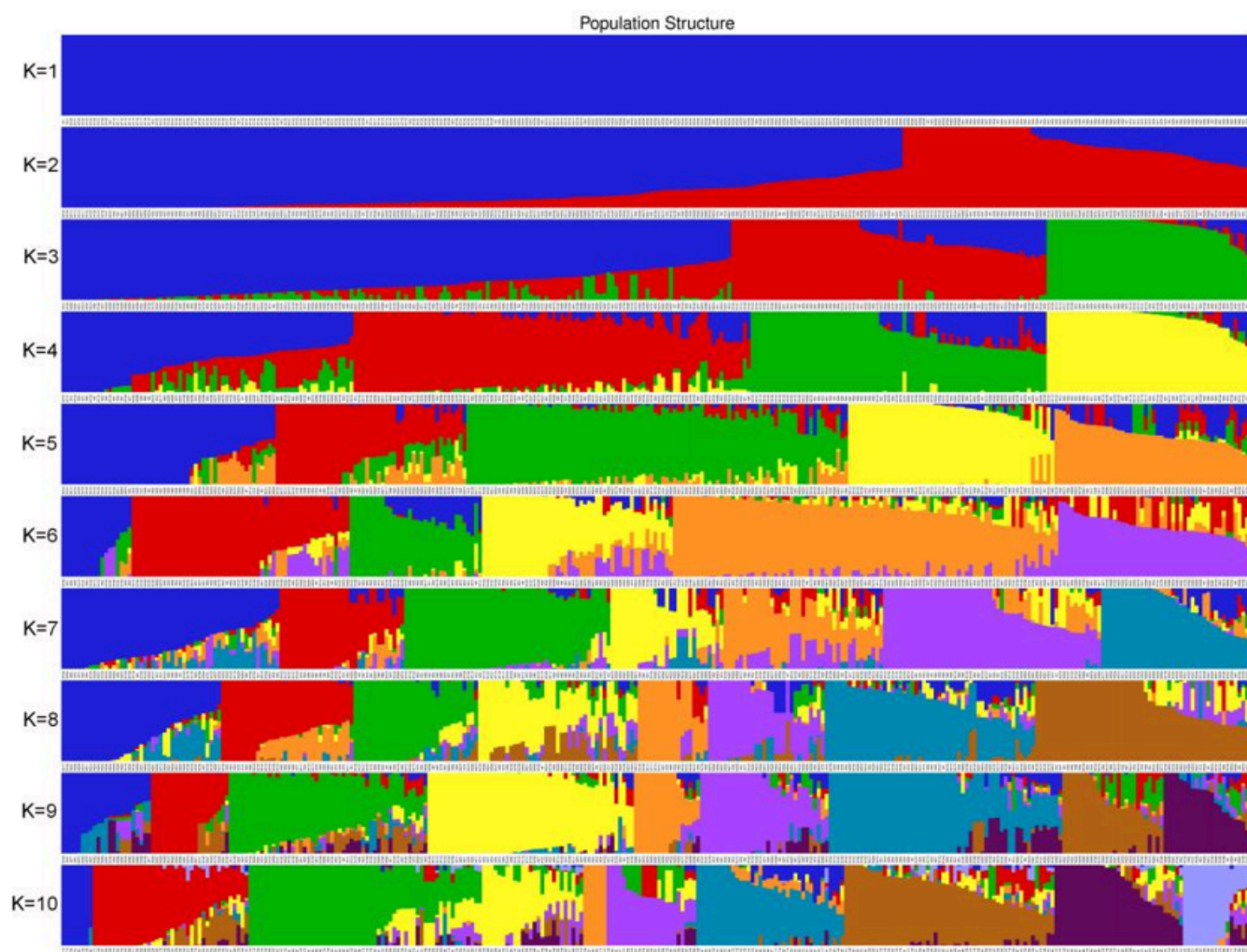


Figure 6

Admixture individual cluster values corresponding to each K value

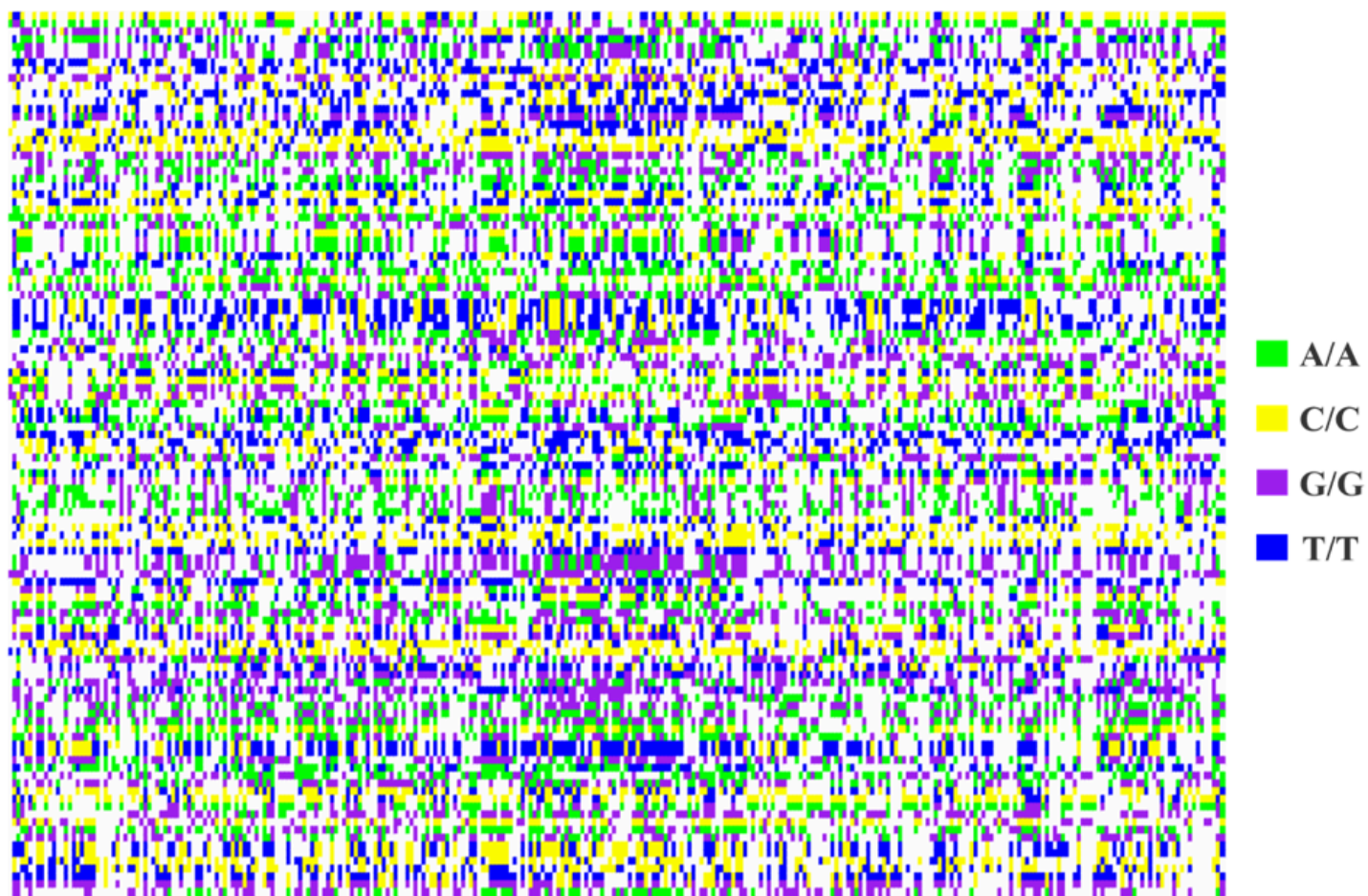


Figure 7

DNA fingerprints of 306 wheat samples. Each row represents the selected candidate loci, and each column represents a sample

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [STable.docx](#)