Screening of 200 Core SNPs and the Construction of a Systematic SNP-DNA Standard Fingerprint Database with More Than 20,000 Maize Varieties

Hongli Tian †, Yang Yang †, Rui Wang †, Yaming Fan, Hongmei Yi, Bin Jiang, Lu Wang, Jie Ren, Liwen Xu, Yunlong Zhang, Jianrong Ge, Yawei Liu, Fengge Wang * and Jiuran Zhao *

Citation: Tian, H.; Yang, Y.; Wang, R.; Fan, Y.; Yi, H.; Jiang, B.; Wang, L.; Ren, J.; Xu, L.; Zhang, Y.; et al. Screening of 200 Core SNPs and the Construction of a Systematic SNP-DNA Standard Fingerprint Database with More Than 20,000 Maize Varieties. Agriculture 2021, 11, 597. https://doi.org/10.3390/agriculture11070597

1. Introduction

Maize is the crop with the highest total output and the largest market value in the seed industry in China, which plays an important role in the structure of China’s agricultural economy. With the recent explosion of new varieties and the rapid development of molecular breeding technology, maize molecular identification faces significant obstacles. As of 2018, more than 12,000 maize varieties have been approved in China. As of 2019, applications for rights to 7570 new plant varieties have been filed since 1999, and more than 5000 hybridized combinations are tested each year ([1], http://202.127.42.145/bigdataNew/, http://www.nybkjfzzx.cn/p_pzbh/pzbh.aspx (accessed on 1 December 2020)). Although the number of varieties is increasing sharply, most are the products of imitation breeding, and few are original innovations. At the same time, the introduction of new varieties...
produced by the application of technologies such as molecular biology and whole-genome selection have brought new challenges to variety identification. The fundamental solution to these problems is efficient and accurate identification, monitoring, and protection of maize varieties and materials. Variety identification based on field plantings is difficult because of limitations such as a long cycle time, high cost, environmental impact, and difficulty in forming standard data. The use of molecular technology to assign each variety a clear, effective, and easily identifiable molecular ID card is thus urgently required.

The development of molecular marker identification technology for crop varieties has gone through three important stages, i.e., first-generation marker technologies, such as those based on restriction fragment length polymorphisms (RFLP) and random amplified polymorphic DNA (RAPD) in the 1990s; second-generation marker technology involving simple sequence repeats (SSR); and third-generation marker technologies launched in recent years that rely on single nucleotide polymorphisms (SNP) and insertion–deletion polymorphisms (InDel) [2–5]. Among these markers, SSR and SNP are optimal markers for variety identification which are both co-dominantly inherited, test the determine sequence difference and facilitate high-throughput screening [6]. The industrial standard for the identification of maize varieties using SSR molecular markers was first issued in 2007; as of 2020, technical protocols for variety identification based on SSR markers have been developed for 20 crop species [7–13]. Construction of standard DNA fingerprint databases using the above-mentioned protocols, has been initiated for maize, rice, and wheat. An SSR fingerprint database of maize varieties containing 3998 samples was constructed in 2017, and more than 50,000 samples have since been incorporated [14]. Standard fingerprint databases for rice and wheat now include data for more than 5000 and 10,000 varieties, respectively. Crop-variety SSR-DNA fingerprint constructed outside of China include 1537 maize varieties from France [15] and 502 European wheat varieties in Germany [16]. Although SSR marker-based identification technology has been widely applied to differentiate crop varieties, further increases in detection throughput are likely limited. Given the explosive growth in the number of varieties and continuous changes in detection requirements, the need to integrate and share standard fingerprint databases has further increased. Because of their high distribution density on the genome, bi-allelic variation, easy shareability, and high-throughput detection, SNP markers have attracted the attention of researchers. The development of more efficient and accurate SNP molecular identification technologies and establishment of a maize SNP-DNA fingerprint database is thus urgently needed.

Preliminary attempts to construct SNP-DNA fingerprints for various crops have taken place during the past 5 years in China. For instance, molecular barcodes of 429 well-known Chinese wheat varieties have been constructed using microarray chips [17]. As another example, the number of SNP markers suitable for the construction of soybean fingerprints has been investigated, and hundreds of fingerprints of soybean germplasm resources and varieties have been generated [18–20]. In addition, 393 core SNP marker fingerprints covering 719 cotton resources have been established using high-density array, and a set of SNPs suitable for the construction of cotton accessions has been assessed and screened [21,22]. Moreover, a DNA fingerprint of 220 tested varieties of rapeseed has been generated [23]. From the studies of maize varieties/materials based on SNP markers, mainly focused on genetic assessment of germplasm resources, in addition to the 335 nationally approved hybrids, fingerprints have been constructed [24–28].

Along with the continuous advancement of modern molecular biology, such as the rise of high-throughput genotyping technology, sequencing of multiple reference genomes in maize, and the release of numerous SNP arrays have greatly promoted the development of maize SNP identification technology and standard fingerprint database construction [29–39]. Kompetitive allele-specific polymerase chain reaction (KASP) technology is a competitive allele-specific PCR-based system that is compatible with 96-, 384-, and 1536-well plates [33]. This system, which has the features of simplicity, high sample throughput, and flexible test design, is especially suitable for SNP genotyping [33].
In this study, in response to the diversified needs of maize variety management, market monitoring, and intellectual property protection, we generated a set of core SNPs for fingerprint database construction of maize varieties, constructed a comprehensive SNP-DNA fingerprint database of 20,075 materials encompassing maize hybrid lines, hybrid combinations, and inbred lines based on KASP technology. The maize SNP-DNA fingerprint database established in this study can play an important role in variety authentication, purity determination and the protection of variety rights, thereby providing reliable and comprehensive data support for developmental use in the seed industry.

2. Materials and Methods

2.1. Plant Materials and DNA Extraction

To construct the maize SNP-DNA fingerprint database, we used 20,075 samples derived from three sources, namely, (1) 6755 nationally and provincially approved hybrids obtained between 1984 and 2019, which were selected and developed by breeding institutions or companies in China; (2) 7837 hybridized combinations of regional test samples in 2019; and (3) germplasm resource materials, including 3478 inbred lines with new plant variety rights and 2005 samples from three breeding populations. Two populations were developed from recombinant inbred line (RIL) populations constructed from Jing 2416 and Jing 92, Jing 2416 and Jing 2418. The third population was DH lines constructed from the single-cross Xianyu335. Total DNA of each material was extracted from a pooled sample of at least 30 individual plants by the modified CTAB (hexadyltrimethyl ammonium bromide) method [8]. An ultraviolet spectrophotometer (Nanodrop 2000) was used to determine the concentration and quality of extracted DNA. DNA was considered to be of sufficient quality according to the following criteria: OD\(_{260}/280\) and OD\(_{260}/230\) values of 1.5–2.0. After quantification, the concentration of each DNA working solution was uniformly adjusted to 20 ng µL\(^{-1}\).

2.2. Selection and Validation of Core Single Nucleotide Polymorphism (SNP) Markers for Fingerprint Database Construction

A three-step process was performed to select core SNPs using the 329 representative inbred lines and 221 main popularized hybrid lines [39]. First, we obtained a set of candidate SNPs based on 61,214 SNPs contained in maize 6H-60K array [38,39]. The filter criteria were: high-quality loci (highly repeatable and stable markers, which were classified as poly high resolution); the 60 bp flanking sequence was highly conserved, with no insertion-deletion polymorphisms (InDels) and no more than three SNPs; high polymorphism; and completely in accordance with Mendelian inheritance. Second, KASP primers for candidate SNPs were designed. The high quality and compatible SNPs were obtained based on the KASP genotyping system. The screening indicators are as follows: perfect genotyping loci (their AA, BB and AB three genotypes fell into clearly separated clusters); the primers have good stability and repeatability; high variety differentiation efficiency (minor allele frequencies (MAF) ≥2); and are compatible with the main SNP genotyping platform. Finally, a set of 200 core SNP markers was obtained for the construction of a maize DNA fingerprint database. The principle of determination is that the loci are relatively evenly distributed on 10 pairs of chromosomes, and have high cumulative variety differentiation ability.

2.3. Construction of SNP-DNA Fingerprinting

SNP genotyping of maize materials based on the 200 core SNPs was carried out using the KASP technology system on the SNPline platform (LGC, UK) (Table S1). Polymerase chain reaction (PCR) amplifications were performed in 1536-microwell plates with 1 µL reaction volumes containing template DNA (1.5 µL of working solution, dehydrated), 0.5 µL of 2 × PCR master mix, 0.014 µL of primer working solution, and 0.486 µL of deionized water. The amplification protocol was as follows: 94 °C for 15 min; 10 cycles of 94 °C for 20 s and 61–55 °C (with a decrease of 0.6 °C per cycle) for 60 s; and 26 cycles of
94 °C for 20 s and 55 °C for 60 s. Five out of every 384 samples, two replicate samples, two reference samples, and one blank control were run. In this study, the maize inbred lines Jing724 and Jing92 were used as reference samples.

Fluorescence signals were scanned on a Phearstar plate reader (LGC Biosearch Technologies, Hoddesdon, UK). The scanned original signal data were imported into Kraken (LGC Biosearch Technologies, Hoddesdon, UK) and subjected to a strict genotype-calling scheme. Genotype data of the 200 SNPs and original signal values of all samples were imported into the Plant Variety SNP Fingerprint Database Management System of the Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences (registration number: 2018SR043088).

2.4. Data Analysis

A diagram physical distribution of 200 SNP loci on maize genome was generated using the Python3 language and the Biopython graphics library (https://biopython.org (accessed on 17 August 2020)). MAF and polymorphism information content (PIC) values of the 200 SNPs based on the genotype data of maize hybrids, hybridized combinations, and inbred lines were analyzed using the SNP Comparison Statistical Tool (v1.1, registration number: 2018SR026743, Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China). To determine the distinguishing ability of SNP set, the VDP (variety discrimination power) values of 200 core SNPs were analyzed based on DNA fingerprinting data of 538 national approved maize hybrids by VDPtools (https://github.com/caurwx1/VDPtools.git (accessed on 22 May 2020)) [40].

The heterozygous genotype frequencies and percentage of differential loci of pairwise comparisons samples within each group of hybrid lines, hybridized combinations, and inbred lines were analyzed using the SNP Comparison Statistical Tool (v1.1). In order to analyze the results consistency between 200 and 60 K SNPs in variety identification, the Pearson correlation coefficients between the two sets of SNPs were analyzed using the pairwise genetic distances of 329 representative inbred lines and 221 main popularized hybrid lines by SNP Comparison Statistical Tool (v1.1) [39]. For comparison the results consistency between 200 and 60 K SNPs in germplasm resources evaluation, UPGMA (unweighted pair group method with arithmetic mean) clustering analysis of 329 inbred lines was carried out in the PowerMarker v3.25 and MEGA7 programs [41,42]. The comparison analysis of the clustering trees formed by the two sets of SNPs was analyzed using Dendroscope v3.5.8 software [43].

3. Results

3.1. Detection and Selection of Core SNPs for Fingerprint Database Construction

Using 329 inbred and 221 hybrid lines, 384 candidate SNPs were obtained from 61,214 SNPs contained in maize6H-60K array based on the principles of flanking sequence conservation and good genotyping effect on chip platform [39]. Based on the high-throughput KASP technology system, 300 high-quality and compatible SNPs were obtained. According to the cumulative variety differentiation ability and relatively uniform distribution of SNP locus combination, 200 core SNPs with high discrimination, high accuracy, high stability, compatibility with multiple platforms and suitable for automatic genotyping were finally determined for the maize SNP-DNA fingerprints construction (Table S1, Figures 1 and 2).
Figure 1. Genotyping patterns of MSNP133 and MSNP164 markers on competitive allele-specific polymerase chain reaction (KASP) assay. Each solid dot is an actual genotype of a sample at the SNP marker. The coordinate position of each dot is obtained by calculating the fluorescence signal values of the two alleles. The data points of the same genotype gather together, and the heterozygous genotype points are between the two homozygous genotypes. The green points represent heterozygous genotype, the red and blue points represent two homozygous genotypes.

Figure 2. Distribution of 200 SNPs on 10 chromosomes of maize genome.

3.2. Characteristics of the 200 SNP Markers Used for Maize DNA Fingerprint Database Construction

KASP genotyping with the 200 SNP markers gave ideal results: the statuses of SNP genotypes assigned by Kraken fell into three clearly separated clusters (Figure 1). As revealed by plotting the physical locations of SNPs on the maize nuclear genome, the 200 SNPs were relatively uniformly distributed on the 10 maize chromosomes (Figure 2). The number of SNPs on each chromosome was positively correlated with chromosome...
length; the number of SNPs in centromeric regions was generally small, as was the number distributed on the short arm of chromosome 6. All 200 SNPs were located in intragenic regions: 106 (53.0%) in exons, 35 (17.5%) in 3' untranslated regions (UTRs), 34 (17.0%) in promoter regions, 18 (9.0%) in 5' UTRs, and 7 (3.5%) in introns (Table S1).

As revealed by MAF and PIC values based on the data from 6755 maize hybrids, 7837 hybridized combinations, and 3478 inbred lines, the 200 SNPs were highly polymorphic and had good variety-discrimination ability (Table S1, Figure 3). MAFs of the 200 SNPs in maize hybrids, hybridized combinations, and inbred lines ranged from 0.184–0.500, 0.100–0.499, and 0.154–0.499, respectively, with corresponding averages of 0.385, 0.350, and 0.378, while PIC values were 0.255–0.375 (average, 0.354), 0.164–0.375 (0.335), and 0.226–0.375 (0.351), respectively. According to these values, approved hybrids had the highest levels of polymorphism, followed by inbred lines and then hybridized combinations. More than 99.0%, 98.5%, and 86.0% of the 200 SNPs in hybrids, inbred lines, and hybridized combinations, respectively, had MAFs greater than 0.20, while more than 84.5%, 79.5%, and 67.5% of these SNPs had MAFs above 0.30. PIC values based on hybrids, inbred lines, and hybridized combinations were higher than 0.30 for more than 96.5%, 92.0% and 79.0% of the loci. We selected 538 national approved maize varieties to test the validity of 200 core SNPs. The results showed that it has higher ability to distinguish varieties, and the VDP value was 0.98. The samples that could not be discriminated were highly similar varieties.

![Figure 3](image-url)

**Figure 3.** Distribution of minor allele frequencies (MAF) and polymorphism information content (PIC) values of 200 SNPs based on maize hybrid lines, hybridized combinations and inbred lines respectively. The No. of SNP markers in abscissa was ranked from small to large according to MAF or PIC values evaluated using hybrid lines.
3.3. Construction and Analysis of Maize SNP-DNA Fingerprints

The genotype data of the 20,075 samples imported into the SNP-DNA fingerprint database management system could be searched and compared (Table S2). To ensure the accuracy of the genotype data imported into the fingerprint database, a strict analysis scheme was adopted when using the Kraken (LGC Biosearch Technologies, Hoddesdon, UK) software to analyze all samples. In particular, any data falling outside of the clusters were eliminated, an approach that could improve the accuracy of the data, however, it will lead to an increase in the percentage of missing data, especially that of hybrids. The overall missing data rate based on the three types of material, hybrids, hybridized combinations, and inbred lines was 7.5%, 8.8%, and 0.49%, respectively. The missing data of nationally approved hybrids (538 samples) was 3.7%. The frequency of heterozygous genotypes was 0.11–0.72 and 0.12–0.80, respectively, in 6755 maize hybrids and 7837 hybridized combinations, with average values of 0.48 and 0.47, respectively. More than 97% of hybrids and hybridized combinations had heterozygous genotype frequencies between 0.31 and 0.60. Heterozygous genotype frequencies of the 3478 inbred lines were 0.000–0.097, with an average of 0.012, and were lower than 0.060 in 96.09% of inbred lines (Figure 4).

Figure 4. The distribution of heterozygous genotype rate of maize hybrid lines (A), hybridized combinations (A) and inbred lines (B).
3.4. Assessment the Efficiency of SNP Panels in Identification of Maize Hybrid and Inbred Lines

In pairwise comparisons of maize hybrid lines, hybridized combinations, and inbred lines, the number of different SNPs ranged from 1 to 164, 1 to 160, and 1 to 140, respectively. We detected 80–125, 70–115, and 80–110 different SNPs in 83.28%, 80.06%, and 82.51% of pairwise comparisons within these three respective groups; the percentage of different SNPs within 5% among pairwise comparisons (the number of difference SNPs is less than 10) accounted for 0.013%, 0.011%, and 0.030%, respectively (Figure 5). We calculated pairwise Nei’s (1973) genetic distance values of 329 representative inbred lines and 221 main popularized based on the 200 core SNPs and 60 K SNPs [39,44]. Genetic distances based on 200 SNPs were highly correlated with those calculated from the 60 K SNP data (Pearson correlation coefficient is 0.82 and 0.87 in inbred and hybrid lines respectively) (Figure 6). In the graph shown in Figure 6, the data points exhibited a concentrated distribution and displayed a linear relationship. To compare the consistency of 200 and 60 K SNPs in germplasm resources evaluation, UPGMA clustering analyzed were carried out using 329 inbred lines [39]. It showed that the evaluation results of the two sets of loci had high consistency (Figure 7). As the same as those reported in Tian et al., 2021, the 329 inbred lines could be classified into nine main groups: BSSS (mainly American materials), Lancaster (LAN; mainly American materials), Tang-Si-Ping-Tou (TSPT), PA, PB, Lvda red cob (LRC), X, Iodent (IDT), and Landrace [39]. There were 22 samples in the PA and PB groups, and a few sporadic samples had different clustering results (Figure 7).

![Figure 5](image_url)  
**Figure 5.** The distribution of the number of different SNPs obtained by pairwise analysis of maize hybrid lines, hybridized combinations and inbred lines respectively.
Figure 6. The correlation analysis of 200 SNPs and 60 K SNPs data based on pairwise Nei 1973 genetic distances (GD) of 329 inbred lines (A) and 221 hybrid lines (B). P is Pearson correlation coefficient.

Figure 7. Dendrogram for 329 representative inbred lines based on 200 (the right) and 60 K SNPs (the left).
4. Discussion

4.1. Selection and Verification of a High-Efficiency Core-SNP Marker Combination for Maize Fingerprint Database Construction

Selection of a set of high-efficiency core SNPs suitable for the identification of maize varieties is essential for the construction of a standard DNA fingerprint database. The criteria used for marker screening differ depending on the purpose of variety identification, that is, discrimination of maize varieties vs. confirmation of intellectual property rights to breeding material. Assuming the combination of loci is sufficient for the accuracy and reproducibility of detection, the first goal mentioned, discrimination of varieties, relies on criteria such as the ability to distinguish varieties and find differences between samples; several hundred loci are typically required, and the determination parameter is generally the number or percentage of different loci. In the second case—seeking to confirm intellectual property rights—the purpose is verifying the degree of similarity between samples, generally on the basis of genetic similarity, and a uniform genomic distribution of loci, typically several thousand, is desirable [24,45]. The application of the maize SNP-DNA fingerprint database constructed in this study is authenticity and distinctness identification of varieties. The analysis scheme is based on fingerprint comparison as well as screening and comparison across the whole database. Taking into account the scale of the constructed database, efficiency of comparison, and practical application requirements, we adopted a subset of SNPs suitable for variety discrimination as the core SNP combination for database construction. In this study, the combination of 200 SNPs reported here have high accuracy, highly discrimination power, and excellent practical application value in maize varieties identification (Figures 3, 5 and 6).

The evaluation of a set of high-efficiency SNP markers suitable for maize variety identification involves two stages: selection of individual core SNPs, and selection of the core SNP combination. Screening criteria for individual core SNPs are as follows: highly conserved SNP flanking sequences; stable primers; good genotyping effects; conformance to Mendelian inheritance; high polymorphism; and compatibility with multiple genotyping platforms. For the core SNP combination, screening criteria include a strong ability to distinguish varieties and a relatively uniform genomic distribution without close linkage. The distribution of genetic recombination in maize genome is not uniform, the recombination rate near telomere is high, and there is almost no cross-over around centromere [46]. In some regions, such as the short arm of chromosome 6, the recombination rate is low even near the telomere [46]. Therefore, the selected SNP sets should only be relatively evenly distributed in the genome, and it is difficult to achieve complete uniform distribution. In addition, factors such as genetic differences between maize varieties, genome size, and genetic recombination rates of genomes [6,45] are comprehensively considered when determining the number of SNPs in the core marker combination.

4.2. Difficulties and Key Considerations when Establishing a SNP-DNA Standard Fingerprint Database of Maize Varieties

To ensure that a maize standard fingerprint database is representative, accurate, and shareable, several key issues must be resolved during construction. The first consideration is DNA preparation to properly represent the genomic information of a maize variety. The second issue concerns the formatting of fingerprint data for compatibility between different platforms and laboratories. A third point is ensuring the accuracy of data imported into the fingerprint database. Finally, a mature fingerprint database management system needs to be developed.

A maize variety comprises a large population of plants with relatively stable genetic traits and identical characteristics and economic value. For various reasons, such as a lack of homozygosity in the inbred parent or mixing during seed production, a certain proportion of heterogeneous plants are present within a variety; consequently, the uniformity of the variety is less than 100%. If only one to three individual plants are used as the source of DNA, the variation in the genome of the variety will be poorly represented, and the
information will be biased. To truly reflect the main genotypes of the variety (while taking into account experimental cost and efficiency vs. the quality of the generated data), pooling the leaves of at least 30 individual plants for DNA extraction is preferred to ensure that the obtained DNA fingerprint represents the population characteristics of the variety [8,45].

Although SNP markers are mainly bi-allelic variants, a unified data description format must be used when entering fingerprint data into a database to facilitate data sharing; this is because different alleles are defined in the primer or probe design of different genotyping platforms and also because genotype data can be exported from the same genotyping platform in various formats. We recommend the use of the A/T/C/G base format for maize SNP-DNA fingerprint data, with the fingerprint information of at least two reference samples also included. If the KASP typing platform is not used for the fingerprint data collection, a conversion based on the fingerprint data of a reference sample provided by the institution constructing the database is required. Because the major purpose of the fingerprint database is allowing comparisons to be performed directly using the data without conducting parallel comparison tests, the accuracy of the fingerprint data is a key issue. To ensure the accuracy of fingerprints entered into database management system, the method used to collect the genotype data—in addition to the design of various repeated tests—is critical. When collecting genotype data, the SNP genotyping system first scans and collects two fluorescence signals and then standardizes the fluorescence signals to obtain the coordinate values of the data points. The data points of a group of samples are divided into three clusters representing the two homozygous genotypes and one heterozygous genotype. In this study, a strict analysis scheme was adopted when the genotype data clusters were divided, namely, only those data points clearly within the cluster circle were collected. Although this approach increases the probability of missing data, it guarantees the accuracy of the data entered into the standard fingerprint database. Finally, the popularization and application of the shared standard fingerprint database is an inseparable component of a mature management system. Incorporation of the maize SNP-DNA fingerprint data into our Plant Variety SNP Fingerprint Database Management System already integrated with a SSR fingerprint database allows easy toggling between the fingerprint databases of the two types of marker; as a consequence, these data can be readily integrated, compared, and further analyzed [14].

4.3. Extensibility and Application of the Maize SNP-DNA Fingerprint Database

With the development of molecular marker genotyping technology, research related to molecular fingerprint identification of major crops has begun to focus on SNP markers. SSR and SNP marker technologies are both ideal methods for molecular identification, they have different advantages and disadvantages and can complement each other. Compared with SNP markers, SSR markers are highly polymorphic at individual loci and suitable for routine laboratory use, but their high-throughput detection is difficult to achieve on a capillary electrophoresis platform. Although SSR-DNA fingerprint data can be shared, multiple standardization and consolidation steps are required. The advantages of SNP marker fingerprint identification are exactly opposite those of SSR markers. First, high-throughput detection of SNPs is easy to achieve, with a detection throughput of thousands or tens of thousands of sites. Second, statistical analysis of SNP data is simple and accurate, and comparison and integration of data from different sample batches or laboratories is easily achievable [47]. Therefore, we should continue to play the role of SSR-DNA fingerprint database, and actively promote the application of the SNP-DNA fingerprint database in the advantageous fields.

The maize SNP-DNA standard fingerprint database can be directly applied for the authentication, specification, and intellectual property-right protection of varieties or seeds and can be indirectly exploited in variety selection and breeding. The database can be used for variety authentication in three main ways: (1) to detect whether samples from different years or different groups of hybrid combinations have been replaced in variety regional trials; (2) to test the authenticity of random samples for market monitoring; and (3) to
verify seed quality for seed enterprises. With respect to variety specification, the fingerprint database is primarily useful for variety regional trial verification testing. To determine the uniqueness of a new hybrid combination, the trialed variety needs to be compared with fingerprints of all known varieties in the database. In regards to protection of new plant variety rights, the uniqueness of submitted application materials can be confirmed by comparison against known varieties in the database. Another indirect application of the SNP-DNA standard fingerprint database of maize varieties is to provide important breeding reference information for variety selection. Understanding the characteristics and developmental trends of previously bred maize varieties through a multi-faceted statistical analysis of the DNA fingerprint data has important reference value for the preparation of maize breeding schemes.

In summary, the maize DNA fingerprint database constructed in this study includes different type materials and a large number of varieties. The generated SNP fingerprint were imported into a unified management system to ultimately yield a joint construction, shareable, standard fingerprint database. The database can be applied to numerous fields of maize variety research, such as regional trials, verification, market monitoring, and variety rights protection, thereby providing reliable, comprehensive data support for use in the seed industry.

5. Conclusions

In this study, we first evaluated and obtained a set of core SNP combinations including 200 loci. Based on the 200 core SNPs, an expanding systematic SNP-DNA standard fingerprint database with more than 20,000 maize materials covering approved maize hybrid lines, hybridized combinations, and inbred lines was constructed. The evaluation results based on the samples of the above three groups showed that 200 SNPs had high ability to distinguish varieties, with MAF and PIC values greater than 0.30; the maximum number of different locus of pairwise comparison samples was 164; the percentage of different SNPs within 5% accounted for 0.013%, 0.011%, and 0.030% among pairwise comparisons of samples within hybrid lines, hybridized combinations and inbred lines, respectively. The results also showed that heterosis was well used in maize variety breeding in China, and the average frequency of heterozygous genotype of maize hybrids reached 0.48. The homozygous degree of inbred lines was higher, and the average frequency of homozygous genotype reached 0.988. Genetic distances between samples based on the 200 core SNPs were highly correlated with those obtained using 60 K SNPs. This SNP-DNA fingerprint database will provide basic data support for maize variety authenticity identification, purity identification, variety right protection, and molecular breeding.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/agriculture11070597/s1, Table S1: Detailed information of 200 core SNPs based on the B73 AGP_v3 reference; Table S2: Genotype data of 200 SNP loci in representative maize materials.

Author Contributions: J.Z. and F.W. conceived and designed the experiments and the article; H.T. and Y.Y. analyzed the data and wrote the manuscript; R.W. wrote the section of materials and methods; R.W., Y.F., H.Y., L.W., and J.R. performed the experiments; B.J. and J.G. helped complete the experiments; L.X. and Y.Z. helped complete the data analysis; Y.L. prepared materials; All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the 13th Five-Year National Key R & D Program of China (2017YFD0102001, 2017YFD0102005) and the Beijing Scholars Program (BSP041).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or Supplementary Material.

Conflicts of Interest: The authors declare that there are no conflict of interest.
References

1. Yang, Y.; Tian, H.L.; Yi, H.M.; Liu, Y.W.; Ren, J.; Wang, R.; Wang, L.; Zhao, J.R.; Wang, F.G. Analysis of the current status of protection of maize varieties in China. *Sci. Agric. Sin.* 2020, 53, 1095–1107.

2. Powell, W.; Morgante, M.; Andre, C.; Hanafey, M.; Vogel, J.; Tingey, S.; Rafalski, A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 1996, 2, 225–238. [CrossRef]

3. Rasheed, A.; Hao, Y.F.; Xia, X.C.; Khan, A.; Xu, Y.B.; Varshney, R.K.; He, Z.H. Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. *Mol. Plant.* 2017, 10, 1047–1064. [CrossRef]

4. Jamali, S.H.; James, C.; Hickey, I.T. Insights into deployment of DNA markers in plant variety protection and registration. *Theor. Appl. Genet.* 2019, 132, 1911–1929. [CrossRef] [PubMed]

5. Wang, F.G.; Tian, H.L.; Yi, H.M.; Zhao, H.; Huo, Y.X.; Kuang, M.; Zhang, L.K.; Lyu, Y.D.; Ding, M.Q.; Zhao, J.R. Principle and strategy of DNA fingerprint identification of plant variety. *Mol. Plant Breed.* 2018, 16, 4756–4766. [CrossRef]

6. UPOV (International Union for the Protection of New Varieties of Plants). *Possible Use of Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS)*; UPOV: Geneva, Switzerland, 2011.

7. Zhao, J.R.; Wang, F.G.; Guo, J.L.; Lyu, B.; Huo, C.Y.; Du, Y.Y. Maize variety identification molecular techniques. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 1432-2007; China Agriculture Press: Beijing, China, 2007.

8. Wang, F.G.; Yi, H.M.; Zhao, J.R.; Liu, P.; Zhang, X.M.; Tian, H.L.; Du, Y.Y. Protocol for the Identification of Maize Varieties-SSR Marker Method. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 1432-2014; China Agriculture Press: Beijing, China, 2014.

9. Zhuang, J.Y.; Shi, Y.F.; Lyu, B.; Chen, N.; Yang, K.; Ying, J.Z.; Zeng, R.Z. Identification of *Oryza sativa* varieties using microsatellite markers. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 1433-2007; China Agriculture Press: Beijing, China, 2007.

10. Xu, Q.; Wei, X.H.; Zhuang, J.Y.; Lyu, B.; Yuan, Y.P.; Liu, P.; Zhang, X.M.; Yu, H.Y.; Du, Y.Y. Protocol for identification of rice varieties-SSR marker method. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 2470-2013; China Agriculture Press: Beijing, China, 2013.

11. Li, R.Y.; Zhang, H.; Wang, D.J.; Sun, J.M.; Yao, F.X.; Zheng, Y.S.; Xu, J.F.; Duan, L.L.; Li, H. Protocol for the identification of wheat varieties-SSR marker method. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 2469-2013; China Agriculture Press: Beijing, China, 2013.

12. Dai, J.; Wang, X.S.; Ding, K.M.; Wang, Y.P.; Xu, P.; Feng, J.H.; Cheng, E.L. Protocol for identification of cotton variety-SSR marker method. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 2469-2013; China Agriculture Press: Beijing, China, 2013.

13. Li, D.M.; Liu, P.; Chen, L.J.; Tang, H.; Sun, L.F.; Chi, Y.Q.; Wang, X.Y.; Ma, N. Identification of soybean varieties-SSR marker method. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 2595-2014; China Agriculture Press: Beijing, China, 2014.

14. Wang, F.G.; Yang, Y.; Yi, H.M.; Zhao, J.R.; Ren, J.; Wang, L.; Ge, J.R.; Jiang, B.; Zhang, X.C.; Tian, H.L.; et al. Construction of an SSR-based standard fingerprint database for corn variety authorized in China. *Sci. Agric. Sin.* 2017, 50, 1–14.

15. Inghelandt, D.V.; Melchinger, A.E.; Lebreton, C.; Stich, B. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* 2010, 120, 1289–1299. [CrossRef]

16. Röder, M.; Wendehake, K.; Korzen, V.; Bredemeijer, G.; Laborie, D.; Bertrand, L.; Isaac, P.; Rendell, S.; Jackson, J.; Cooke, R.; et al. Construction and analysis of a microsatellite-based database of European wheat varieties. *Theor. Appl. Genet.* 2002, 106, 67–73. [CrossRef]

17. Gao, L.F.; Jia, J.Z.; Kong, X.Y. A SNP-based molecular barcode for characterization of common wheat. *PLoS ONE* 2016, 11, e0150947. [CrossRef] [PubMed]

18. Song, Q.J.; Hyten, D.L.; Jia, G.F.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, R.B. Fingerprinting soybean germplasm and its utility in genomic research. *G3 Genes Genomes Genet.* 2015, 5, 1999–2006. [CrossRef] [PubMed]

19. Liu, Z.X.; Li, J.; Fan, X.H.; Htwe, N.M.P.S.; Wang, S.M.; Huang, W.; Yang, J.Y.; Xing, L.L.; Chen, L.J.; Li, Y.H.; et al. Assessing the number of SNPs needed to establish molecular IDs and characterize the genetic diversity of soybean cultivars derived from *Tokachi nagaha*. *Crop J.* 2017, 5, 326–336. [CrossRef]

20. Wei, Z.Y.; Li, H.H.; Li, J.; Gamar, Y.A.; Ma, Y.S.; Qu, L.J. Accurate identification of varieties by nucleotide polymorphisms and establishment of scannable variety IDs for soybean germplasm. *Acta Agron. Sin.* 2018, 44, 315–323. [CrossRef]

21. Sun, Z.W.; Kuang, M.; Ma, Z.Y.; Wang, X.F. Construction of cotton variety fingerprints using *CottonSNP63K* Array. *Sci. Agric. Sin.* 2017, 50, 4692–4704.

22. Li, L.C.; Zhu, G.Z.; Su, X.J.; Guo, W.Z. Genome-wide screening and evaluation of SNP core loci for fingerprinting construction of cotton accessions (G. barbadense). *Acta Agron. Sin.* 2019, 45, 647–655. [CrossRef]

23. Zhao, R.X.; Li, S.Y.; Guo, R.X.; Zeng, X.H.; Wen, J.; Ma, C.Z.; Shen, J.X.; Tu, J.X.; Fu, T.D.; Yi, B. Construction of DNA fingerprinting for *Brassica napus* varieties based on SNP chip. *Acta Agron. Sin.* 2018, 44, 956–965. [CrossRef]

24. Tian, H.L.; Yang, Y.; Wang, L.; Wang, R.; Yi, H.M.; Xu, L.W.; Zhang, Y.L.; Ge, J.R.; Wang, F.G.; Zhao, J.R. Screening of compatible maizeSNP384 markers and the construction of DNA fingerprints of maize varieties. *Acta Agron. Sin.* 2020, 46, 1006–1015.
25. Lu, Y.L.; Yan, J.B.; Guimaraes, C.T.; Tabo, S.; Hao, Z.F.; Gao, S.B.; Chen, S.J.; Li, J.S.; Zhang, S.H.; Vivek, B.S.; et al. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor. Appl. Genet.* **2009**, *120*, 93–115. [CrossRef] [PubMed]

26. Romay, M.C.; Millard, M.; Glaubitz, J.C.; Peiffer, J.A.; Swarts, K.L.; Casstevens, T.M.; Elshire, R.J.; Acharya, C.B.; Mitchell, S.E.; Flint-Garcia, S.A.; et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **2013**, *14*, R55. [CrossRef]

27. Zhao, J.R.; Li, C.H.; Song, W.; Wang, Y.D.; Zhang, R.Y.; Wang, J.D.; Wang, F.G.; Tian, H.L.; Wang, R. Genetic diversity and population structure of important Chinese maize breeding germplasm revealed by SNP-chips. *Sci. Agric.* **2018**, *51*, 626–634.

28. Wu, X.; Li, Y.X.; Shi, Y.S.; Song, Y.C.; Wang, T.Y.; Huang, Y.B.; Li, Y. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theor. Appl. Genet.* **2014**, *127*, 621–631. [CrossRef]

29. Jiao, Y.P.; Zhao, H.N.; Ren, L.H.; Song, W.B.; Zeng, B.; Guo, J.J.; Wang, B.B.; Liu, Z.P.; Chen, J.; Li, W.; et al. Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* **2012**, *44*, 812–817. [CrossRef]

30. Chia, J.M.; Song, C.; Bradbury, P.J.; Costich, D.; Leon, N.D.; Doebley, J.; Elshire, R.J.; Gaut, B.; Geller, L.; Glaubitz, J.C.; et al. Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* **2012**, *44*, 803–809. [CrossRef] [PubMed]

31. Jiao, Y.P.; Peluso, P.; Shi, J.H.; Liang, T.; Stitzer, M.C.; Wang, B.B.; Campbell, M.S.; Stein, J.C.; Wei, X.H.; Chin, C.S.; et al. Improved maize reference genome with single-molecular technologies. *Nature* **2017**. [CrossRef]

32. Bukowski, R.; Guo, X.S.; Lu, Y.L.; Zou, C.; He, B.; Rong, Z.Q.; Wang, B.; Xu, D.W.; Yang, B.C.; Xie, C.X.; et al. Construction of the third-generation *Zea mays* haplotype map. *GigaScience* **2018**, *7*, 1–12. [CrossRef] [PubMed]

33. Semagn, K.; Babu, R.; Hearne, S.; Olsen, M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol. Breed.* **2014**, *33*, 1–14. [CrossRef]

34. Ganal, M.W.; Durstewitz, G.; Polley, A.; Swarts, K.L.; Elshire, R.J.; Acharya, C.B.; Mitchell, S.E.; Flint-Garcia, S.A.; et al. A genome-wide *Zea mays* SNP genotyping array: Development and germplasm genotyping and genetic mapping to compare with the B73 reference genome. *PloS ONE* **2011**, *6*, e28334. [CrossRef]

35. Unterseer, S.; Bader, E.; Haberer, G.; Seidel, M.; Knaak, C.; Ouizunova, M.; Mettinger, T.; Strom, T.M.; Fries, R.; Pausch, H.; et al. A powerful tool for genome analysis in maize: Development and evaluation of the high density 60k SNP genotyping array. *BMC Genom.* **2014**, *15*, 823. [CrossRef] [PubMed]

36. Tian, H.L.; Wang, F.G.; Zhao, J.R.; Yi, H.M.; Wang, L.; Wang, R.; Yang, Y.; Song, W. Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. *Mol. Breed.* **2015**, *35*, 136. [CrossRef]

37. Xu, C.; Ren, Y.H.; Jian, Y.Q.; Guo, Z.F.; Zhang, Y.; Xie, C.X.; Fu, J.J.; Wang, H.W.; Wang, G.Y.; Xu, Y.B.; et al. Development of a maize 55K SNP array with improved genome coverage for molecular breeding. *Mol. Breed.* **2017**, *37*, 20. [CrossRef]

38. Wang, F.G.; Tian, H.L.; Zhao, J.R.; Yang, Y.; Yi, H.M.; Xu, L.W.; Wang, R.; Wang, L.; Ge, J.R.; Fan, Y.M. A Maize Genome Wide SNP Chip and Its Application. Chinese Invention Patent 201911186629.9, 28 November 2019.

39. Tian, H.L.; Yang, Y.; Yi, H.M.; Xu, L.W.; He, H.; Fan, Y.M.; Wang, L.; Ge, G.R.; Liu, Y.W.; Wang, F.G.; et al. New resources for genetic studies in maize (*Zea mays* L). A genome-wide Maize6H-60K SNP array and its application. *Plant J.* **2021**, *105*, 1113–1122. [CrossRef] [PubMed]

40. Yang, Y.; Tian, H.L.; Wang, R.; Wang, L.; Yi, H.M.; Liu, Y.W.; Fan, Y.M.; Zhao, J.R.; Wang, F.G. Variety Discrimination Power: An appraisal index for loci combination screening applied to plant variety discrimination. *Front. Plant Sci.* **2021**. [CrossRef]

41. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [CrossRef]

42. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]

43. Huson, D.H.; Scornavacca, C. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **2012**, *61*, 1061–1067. [CrossRef] [PubMed]

44. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3321–3323. [CrossRef]

45. Wang, F.G.; Yi, H.M.; Zhao, J.R.; Lyu, B.; Du, Y.Y.; Tian, H.L. General Guideline for Identification of Plant Varieties by DNA Fingerprinting. In *Agricultural Industry Standards of the People’s Republic of China*; NY/T 2594-2014; China Agriculture Press: Beijing, China, 2014.

46. Rodgers-Melnick, E.; Bradbury, P.J.; Elshire, R.J.; Glaubitz, J.C.; Acharya, C.B.; Mitchell, S.E.; Li, C.H.; Li, R.X.; Buckler, E.S. Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3823–3828. [CrossRef] [PubMed]

47. Guichoux, E.; Lagache, L.; Wagner, S.; Chaumeil, P.; Léger, P.; Lepais, O.; Lepoittevin, C.; Malasa, T.; Revardel, E.; Salin, F.; et al. Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* **2011**, *11*, 591–611. [CrossRef]