UPOV
International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Testing Methods and Techniques **TWM/2/16**

Second Session
Virtual meeting, April 8 to 11, 2024

Original: English
Date: April 3, 2024

MAIZE6H-60K: A GENOME-WIDE SINGLE NUCLEOTIDE POLYMORPHISM ARRAY AND ITS APPLICATION

Document prepared by an expert from China

Disclaimer: this document does not represent UPOV policies or guidance

The annex to this document contains a copy of a presentation “Maize6H-60K: A genome-wide single nucleotide polymorphism array and its application”, to be made by an expert from China, at the second session of the Technical Working Party on Testing Methods and Techniques (TWM).

[Annex follows]

Maize6H-60K: A genome-wide single nucleotide polymorphism array and its application

Presented by Hongli Tian

Maize Research Institute,

Beijing Academy of Agriculture and Forestry Sciences

UPOV/TWM/2, April 8 to 12, 2024

1

1. Research background

2

1.1 Maize, an important crop widely cultivated around the world

- **Maize is one of the most important crops grown worldwide for food, animal feed and fuel production.**
- **Because of its high recombination rate and rich genetic diversity as well as the large body of maize genomic research, maize is a significant model plant for use in genetic studies.**

3

3

1.2 The urgency of developing high-density and quality molecular marker set for maize

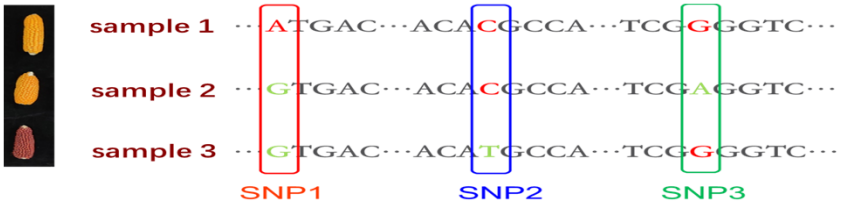
- **Since the 1990s, the development and utilization of molecular markers in plant genetics has greatly advanced.**
- **Beginning with RAPDs (random amplified polymorphic DNA) and ISSRs (inter-simple sequence repeats) and then expanding to SSRs (simple sequence repeats) and SNPs (single nucleotide polymorphisms), developed markers have gradually become more accurate, dense, and uniformly distributed.**
- **Given demand for improving the throughput of detection samples, increasing the density of identification markers and strengthening fingerprint data sharing in maize germplasm resource evaluation, variety identification, and MAS.**
- **For maize, there is an urgent need to develop a set of high-density and data easily integrated and shared marker set.**

4

4

1.3 Characteristics of SNP markers

- **Widely distributed throughout the plant genome, 44–75 bp /SNP (in the maize genome).**
- **Bi-allelic markers, the data can be easily analyzed, integrated and compared.**
- **Be located in intragenic regions, provides a greater chance of association with a functional gene.**
- **Relatively high genetic stability, low mutation frequency.**
- **Easy to achieve high-throughput detection.**
- **These advantages of SNP markers have led to its widespread application in genetic studies.**



5

5

1.4 Develop a high-density and high-quality SNP array

- **The reported SNP arrays for maize are mainly based on early sequencing data.**
- **With the improvement of sequencing quality and the reduction of costs, it is necessary to start mining, developing, and evaluating SNP loci based on whole genome sequencing data.**
- **Therefore, we have developed the Maize6H-60K SNP array based on the whole-genome sequence data of 388 widely representative inbred lines.**

6

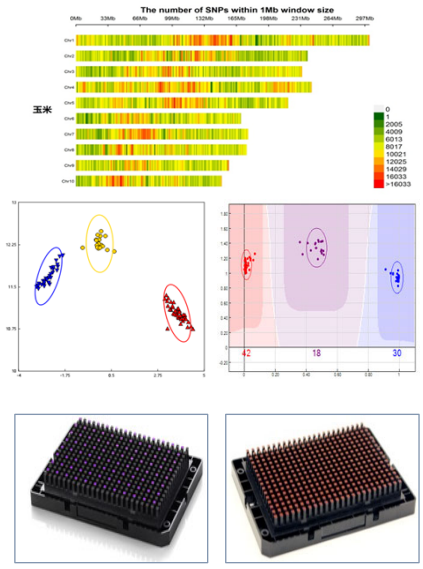
6

2. Development of maize6H-60K array

7

2.1 Pipeline of SNP detection and selection for the Maize6H-60K array

raw SNPs set information 160,912,141 (160M)	based on the whole genome sequencing data of 388 representative maize inbred lines, SNPs were mined and located by BWA and GATK
superior SNPs pool 25,697,116 (25.6M)	data coverage ≥ 50%, MAF ≥ 0.05, filtering for high SNP quality score, high-quality SNPs released by reported maize arrays and 100 chloroplast loci were added to the SNPs pool
screening array SNPs （199,875） 200K	flanking sequence is conserved, heterozygous rate ≤ 0.05, P convert value ≥0.4 (Affymetrix Axiom), final score ≥0.6 (Illumina Infinium); 199,875 loci were selected evenly in the genome including 70 chloroplast loci
Maize6H-60K （61,282） 60K	P convert value ≥0.6, final score ≥0.8; be Poly High Resolution category, missing data<5%, MAF≥0.1; be consistent with Mendelian inheritance in trios samples; total of 61, 282 variants were selected evenly including 68 chloroplast loci



8

2.2 Materials: comprehensive evaluation of the Maize6H-60K array

- 34 sets of parent–offspring trios: varieties with large planting, control varieties for trial test, and some specialized hybrids.
- 329 representative inbred lines: covered a broad set of germplasm resources and have been frequently used to develop maize.
- 221 nationally approved hybrids: basically represented the main popularized varieties in China.
- breeding populations: RIL and DH populations.

9

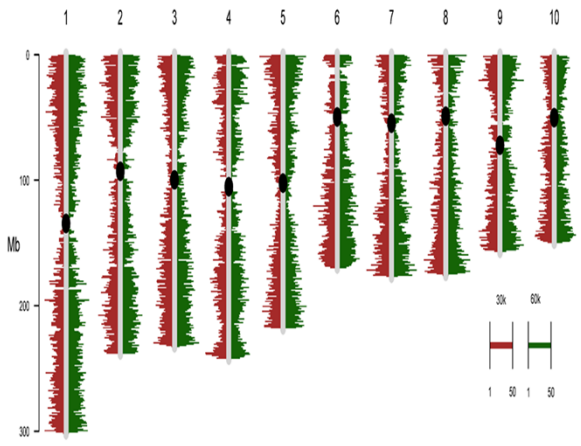
9

3. Characteristics of the Maize6H-60K array and its application

10

10

3.1 Characteristics of the Maize6H-60K array

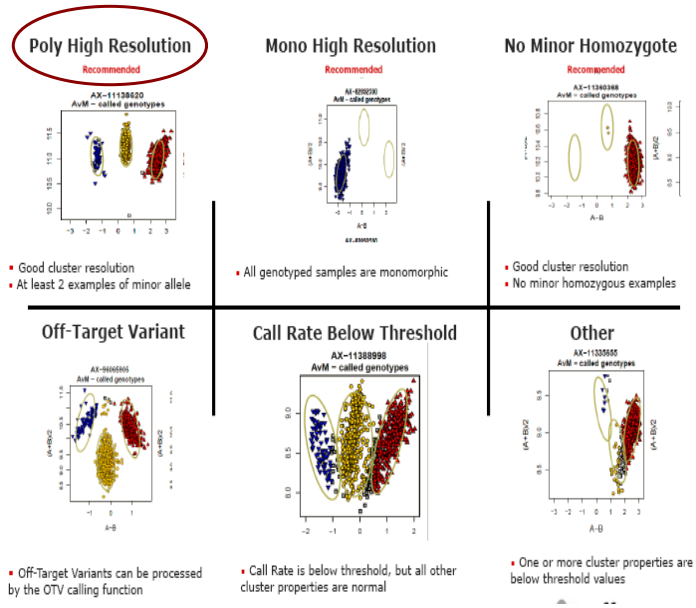


Distribution of 61,214 (green) and 30,171 (red) SNPs on the 10 maize chromosomes. The window size is 1,000 kb.

- **The 61,214 nuclear SNPs were evenly distributed across the whole maize genome.**
- **Among the 60K loci, 21,460 SNPs (35%) were located in genic regions, 12,835 SNPs (21%) were found in coding regions, and 8,625 SNPs (14%) were situated in in-tronic regions.**
- **30,171 SNPs (50%) were classified into the PHR category (Poly High Resolution) in all projects, the distributional characteristics of 30,171 SNPs were also evenly distributed across the genome.**

11

11



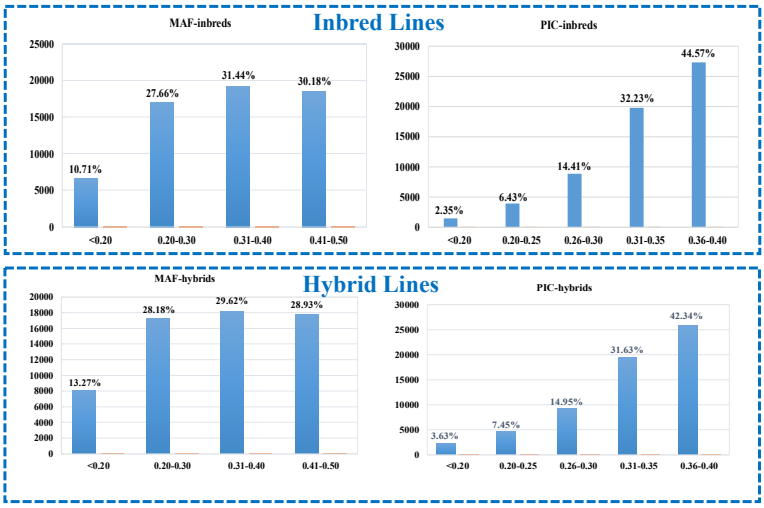
All SNPs were divided into six categories

- **PHR: Poly High Resolution**
- **Good cluster resolution, and their three genotype groups fell into clearly separated clusters.**
- **As high-quality candidate loci.**

12

12

3.1 Characteristics of the Maize6H-60K array



Distribution of MAF and PIC values of 61,214 SNPs based on 329 inbred lines and 221 hybrid lines

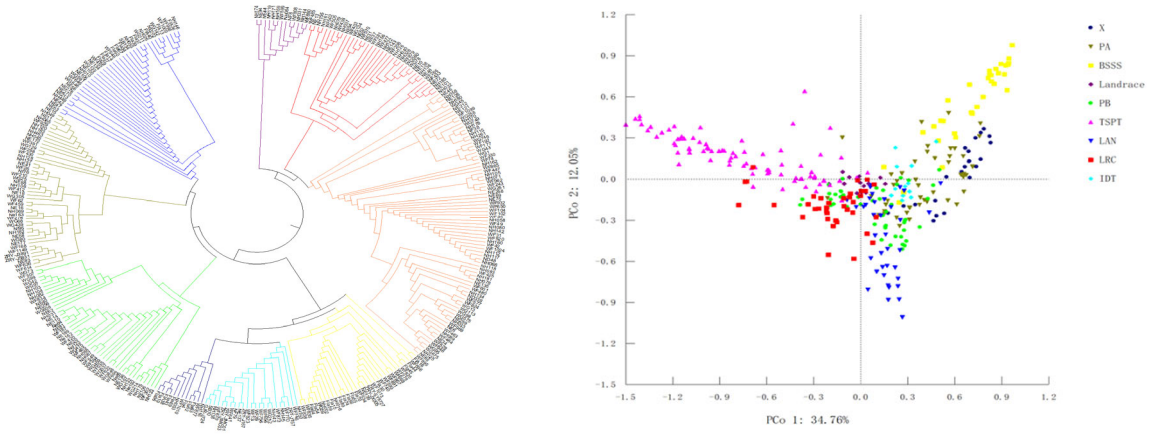
- **329 inbred lines:**
  - ✓ **89.29% and 61.63% SNPs with MAF values ≥ 0.20 and > 0.3;**
  - ✓ **97.65% and 76.81% SNPs with PIC values ≥ 0.20 and > 0.3.**
- **221 hybrid lines:**
  - ✓ **86.73% and 58.55% SNPs with MAF values ≥ 0.20 and > 0.3;**
  - ✓ **96.37% and 73.97% SNPs with PIC values ≥ 0.20 and > 0.3.**

13

13

3.2 Application case: germplasm resource

UPGMA and PCA (Principal component analysis) plots

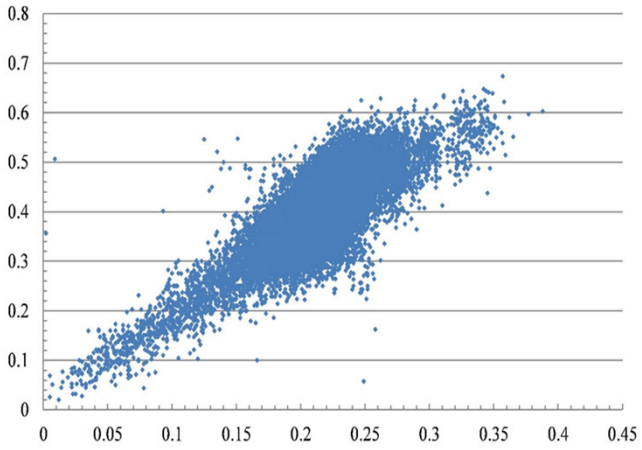


The 329 inbred lines were classified into nine groups, which are basically consistent with previously reported evaluation results and known pedigree relationships.

14

14

3.3 Application case: verification of maize varieties



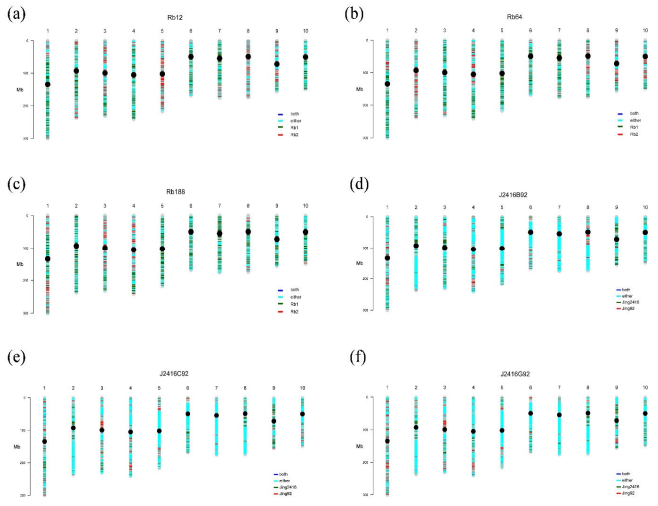
- **Genetic distances (Nei's 1973) based on SNPs were highly correlated with those calculated from the SSR data (Pearson correlation coefficient = 0.79);**
- **The data points exhibited a concentrated distribution and displayed a linear relationship.**

Correlation analysis of SNP and SSR data based on pairwise genetic distances of 221 hybrids. The abscissa and ordinate are genetic distances based on 30171 SNPs and 40 SSRs.

15

15

3.4 Application case: background analysis of molecular breeding population



- **The genetic background of six inbred progeny and ancestral materials based on 30,171 SNPs.**
- **Approximately 35% of SNPs were polymorphic among breeding population constructed from inter-group inbred lines.**

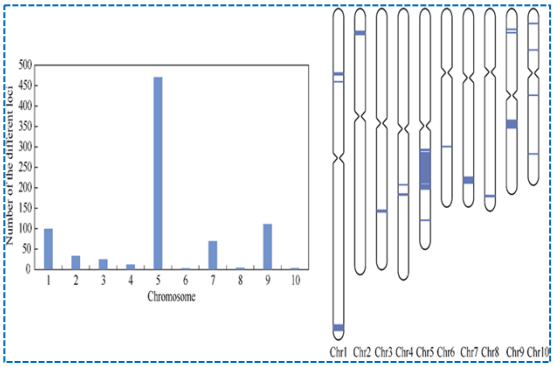
Genotype profiles of six inbred lines from two breeding

16

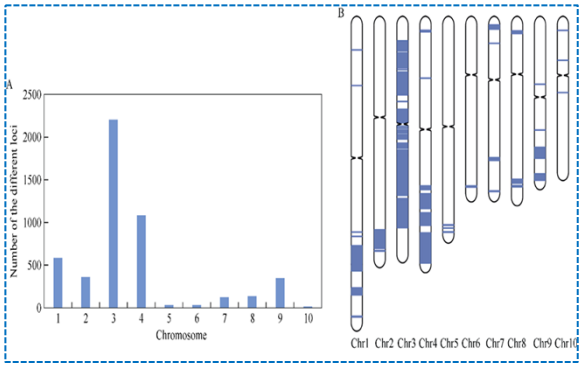
16

3.5 Application case: identification of maize similar inbred lines

Distribution of different loci between sample A and sample B (extremely similar inbreds) in the whole genome



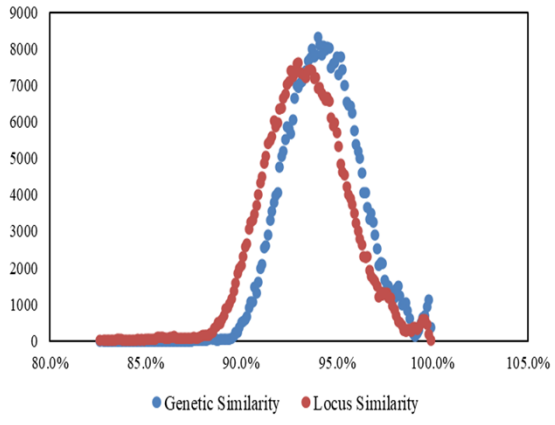
Distribution of different loci between sample C and sample D (similar inbreds) in the whole genome



The GS value between sample A and B was 98.7%, and 56.7% of the difference loci were concentrated on chromosome 5.
The GS value between sample C and D was 90.1%, and 44.8% of the difference loci were concentrated on chromosome 3.

17

17



Distribution of genetic similarity between two DH lines of 893 DH lines

- **893 DH lines of genetic population constructed by sample C and D.**
- **All DH lines could be clearly distinguished. The GS values of 893 DH lines in pairs ranged from 87.5% to 99.9%, with an average of 94.3%.**
- **Maize6H-60K SNP array could accurately evaluate the genetic background of maize similar or extremely similar inbred and DH lines, identify and distinguish all materials one by one, and had the potential to further lock the linkage markers of derived traits.**

18

4. conclusion

Advantages of Maize6H-60K SNP array:

- **high density** (61,282 loci evenly distributed across the entire genome);
- **high quality** (over 80% of SNPs were classified as PHR type);
- **high discrimination ability** (nearly 90% of SNPs with MAF ≥ 0.2);
- **high flexibility** (384, 96, and 24-well format plates);
- **high compatibility** (usable on Affymetrix Axiom, Illumina Infinium platforms and any genotype-by-targeted-sequencing platform);
- **high cost benefit** (lowest cost at the same loci);
- to pro-vide technical support for intellectual property protection and variety innovation of maize varieties.

19

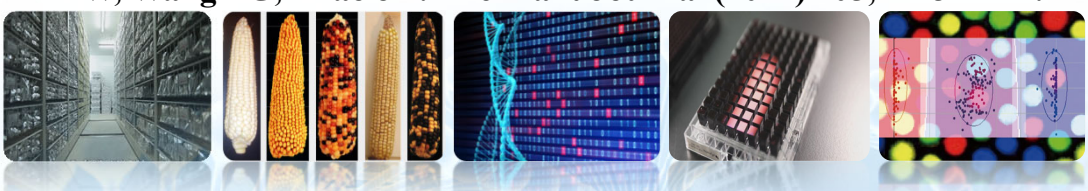
19

Thanks for your attention!

The article link is as follows

https://doi.org/10.1111/tpj.15089

Tian HL, Yang Y, Yi HM, Xu LW, He H, Fan YM, Wang L, Ge JR, Liu RW, Wang FG, Zhao JR. The Plant Journal (2021) 105, 1113–1122.



20

[End of Annex and of document]