UPOV

International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Testing Methods and Techniques **TWM/3/11**

Third Session
Beijing, China, April 28 to May 1, 2025

Original: English
Date: April 15, 2025

EXPLORATION OF IDENTIFICATION TECHNIQUES BASED SNP MARKERS FOR ESSENTIALLY DERIVED VARIETIES OF WHEAT

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 “Exploration of Identification Techniques based SNP markers for Essentially Derived Varieties of Wheat”, to be made by an expert from China, at the third session of the TWM.

[Annex follows]

Exploration of Identification Techniques based SNP markers for Essentially Derived Varieties of Wheat

Binshuang Pang

Hybrid Wheat Research Institute
Beijing Academy of Agriculture and Forestry Sciences, in China

1

Contents

- research Background
- Criteria for SNP Loci Selection
- Genetic Similarity Threshold for EDVs
- Threshold for Minimum SNP Number
- Validation of the EDV threshold

2

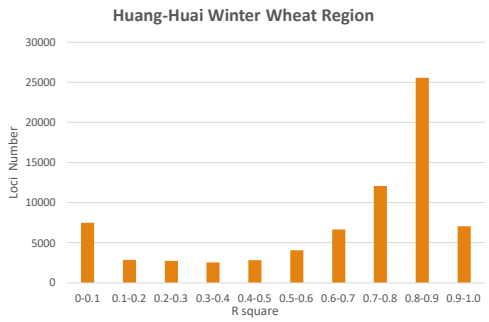
Research Background

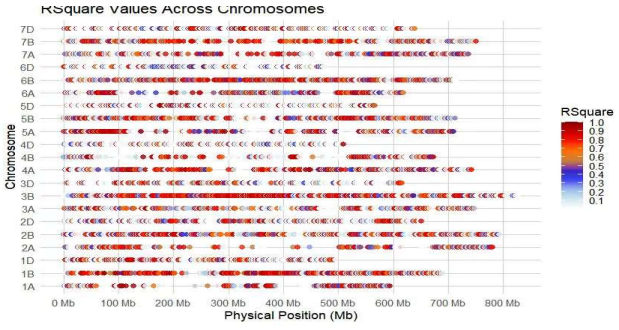
- How to protect and identify Essential Derived Varieties (EDVs) has become a topic of common concern among UPOV members.
- Estimates based on molecular markers are a preferred approach for assessing genetic conformity between putative Essential Derived Varieties (EDVs) and their Initial Varieties (IVs).
- The ISF has provided a list of 3072 SNP markers used in the guidelines for essential derivation in maize
- The results based on 146 SSR markers (from my lab) indicated that pairs of varieties with genetic similarity (GS) ≤ 90% were distinct (exhibiting more traits). When the GS > 90%, approximately 35% of the variety pairs were not distinct, while the remaining 65% were distinct in 1-2 traits (EDVs) .
- Currently, there is no internationally recognized rapid identification method for EDVs based on high-density SNP markers in wheat

3

Criteria for SNP Loci Selection

The selection of SNP loci was based on their distribution across the genome and their polymorphism information content (PIC). These loci were diploidized with unique positions, simply clustered, and characterized by low heterozygosity (less than 10%), $R^2 < 0.8$, and a low missing rate, making them suitable as candidate loci. We developed a BAAFS AFFY wheat 90K SNP array containing 84,662 SNP loci, which is highly effective for distinguishing hexaploid wheat accessions in China.





4

Genetic Similarity Threshold for EDVs

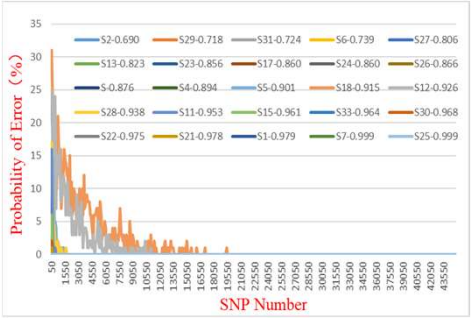
We constructed a fingerprint database using both SSR and SNP markers simultaneously. The results indicated that when the genetic similarity (GS) threshold based on SSR markers for similar varieties (SV) is above 90.0%, the GS threshold for Essential Derived Varieties (EDVs) is above 92.0%. If the GS between an Initial Variety (IV) and a putative EDV exceeds 92.0%, the latter may be classified as an EDV.

(WANG Lixin, JIA, 2015)

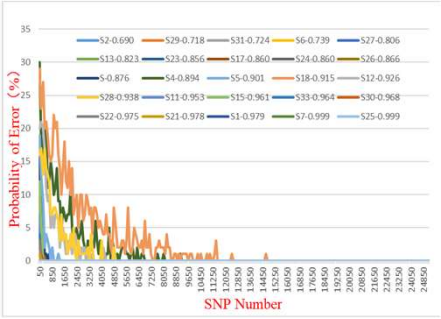
5

Threshold for Minimum SNP Number

The probability of misclassifying a variety as an EDV or non-EDV, whether using loci with an R^2 less than 0.8 or loci with an R^2 less than 1, decreases as the number of loci increases. However, the number of loci required for the former ($R^2 < 0.8$) is greater than that for the latter ($R^2 < 1$). When more than 20,000 loci are used for both types of markers, the error probability becomes extremely low.



Using loci with an R^2 less than 0.8



Using loci with an R^2 less than 1

In the figure, different colors represent pairs of varieties with varying levels of GS calculated using 50 000 loci. The colors indicate the probability of misclassifying a variety pair as an EDV or non-EDV under different numbers of markers."

The error probability when the threshold is 0.92

6

Threshold for Minimum SNP Number

The minimum number of markers required for accurate classification without error at different gradients of true GS (100 simulations) relative to the EDV threshold of 0.92. The linkage disequilibrium (LD) between markers can influence the number of markers required for accurate classification without error.

When the true GS is 0.915, at least 19550 markers with an $R^2 < 0.8$ are required to ensure accurate classification without error. Since, Identification is required for a range of GS values, a minimum of 20,000 loci is required.

0.92

GS / Loci / Sample	<0.86	0.86-0.88	0.88-0.9	0.915	0.926	0.94-0.96	0.96-0.98	>0.98
All Loci	250	550	1050	14750	4450	450	250	50
Loci of R^2<0.8	250	750	9250	19550	1150	1750	250	50

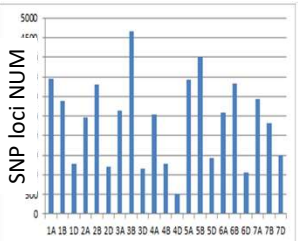
7

GS of EDVs and non-EDVs based on high-density SNP loci

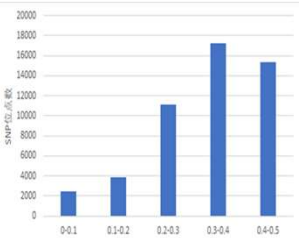
8

SNP array for the identification of wheat varieties domestically developed

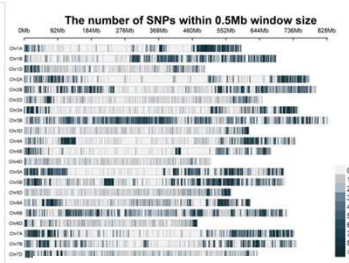
we developed a high quality national **BAAFS Laso wheat 65K SNP solid array** in 2024. The loci include the aforementioned 50,000 loci, functional gene markers (for traits such as yield, quality, and disease resistance), and 4,500 trait QTL loci.



Chromosome distribution



MAF value distribution



Chromosome distribution


Laso wheat 65K SNP solid array

9

Validation of the EDV threshold

The selection of a natural or induced mutant, or of a somaclonal variant, the selection of a variant individual from plants of the initial variety, backcrossing, or transformation by genetic engineering

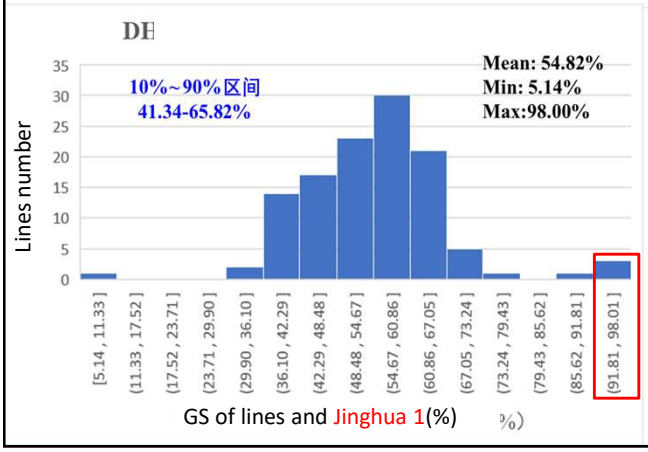
Original Variety	Putative EDVs	Loci similarity value	**Breeding method**
Bainong 207	Bai Han 207	99.9%	selection of a variant individual
Lunxuan 987	Zhongmai 123	98 ..5%	backcross
Xiaoyan 54	EDV 13-2	97.8 %	backcross（5） Backcross（5）
Yanzhan 1	EDV 14-2	97.9 %	backcross
Zhou Mai 18	Zhongmai 66	97.2 %	backcross
Jing 411	Y235	97.2 %	backcross
Zhoumai 16	**Tianmai 863**	95.6 %	**Conventional systematic breeding**
Zhoumai 16	Zheng Mai 618	93.6 %	backcross
Stone 4185	Shimai 14	92.3 %	**Conventional systematic breeding**
Xinmai 26	Kexing 3302	90.7 %	backcross
Jinmai 47	Linkang 11	82.6 %	backcross
Zhongke 1878	Zhongke 1878A	88.2%	selection of a variant individual
~~Bainong AK58~~	~~Bainong 418~~	~~72.3%~~	~~backcross~~

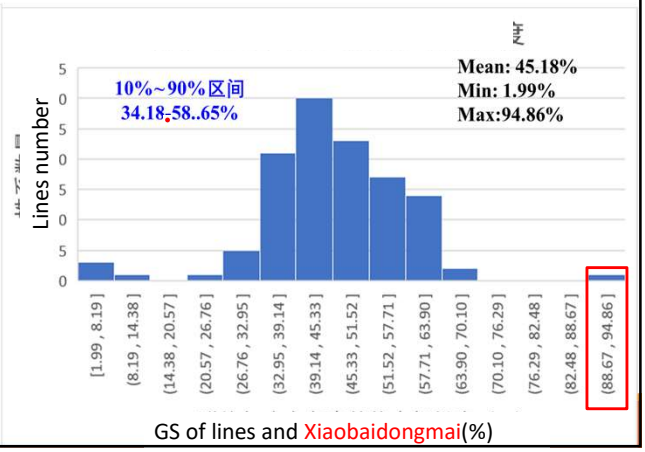
10

Analysis of EDV generation methods

doubled haploid (DH) breeding method

The genetic similarity (GS) between the parents is 0. The GS of DH offspring follows a normal distribution, with a small number of offspring having a GS greater than 0.92. This demonstrates that the DH method can also generate Essentially Derived Varieties (EDVs) of their parents





11

Analysis of EDV generation methods

Backcross Breeding Method

The generation of EDVs through the backcross breeding method is influenced by the GS between the parents and the number of backcross generations. When the GS between the parents is less than 50%, the likelihood of generating EDVs by the third backcross generation may be lower than expected.

Single parent	recurrent parent	BC	LS	Phenotypic differences	Single parent	LS	Phenotypic differences
DQM	ZhMai 366		52%	Plant height(PH), heading stage(HS), tillering stage(TS), flowering stage(FS), flag leaf(FL), Heading period(HP)			
P-EDV12	ZhMai 366	3	60%	PH, HS, TS, FS, FL	LGDQM	50 %	PH, HS, TS, FS, FL
P-EDV9	ZhMai 366	3	68%	HP, FP, FL	LGDQM	65%	PH, HS, TS, FS, FL
P-EDV1	ZhMai 366	3	75%	HP, FP	LGDQM	60%	PH, HS, TS, FS, FL
P-EDV3	ZhMai 366	3	77%	HS, TS, FS,	LGDQM	60%	PH, HS, TS, FS, FL
P-EDV4	ZhMai 366	3	77%	PH,HS,FS	LGDQM	60%	PH, HS, TS, FS, FL
P-EDV2	ZhMai 366	3	78 %	PH,HS,FS,FL	LGDQM	58%	HS, TS, FS, FL
P-EDV10	ZhMai 366	3	79%	HS,TS,FS,FL	LGDQM	61%	HS, TS, FS, FL
P-EDV5	ZhMai 366	3	81%	HS,FP	LGDQM	49%	HS, TS, FS, FL
P-EDV11	ZhMai 366	3	90%	HP, FP, FL	LGDQM	60 %	HS, TS, FS, FL
P-EDV7	ZhMai 366	3	91 %	HP, FP, FL	LGDQM	59 %	HS, TS, FS, FL
P-EDV6	ZhMai 366	3	91%	HP, FP, FL,HS	LGDQM	59 %	PH, HS, TS, FS, FL
P-EDV8	ZhMai 366	3	91 %	HP, FP, FL	LGDQM	59 %	PH, HS, TS, FS, FL

12

Summary

1. Quality control for SNP locus screening in the identification of allopolyploid hexaploid wheat EDVs— including locus uniqueness, heterozygosity, and missing rate — is crucial for accurate genotyping.
2. The determination of essential derivation is optimally based on genotypic data rather than solely on phenotypic comparisons.
3. Selecting over 20,000 SNP loci for wheat EDV identification is more reliable and secure.
4. Solid-phase wheat SNP chips are the most effective method for achieving accurate genotyping of high-density SNP loci.
5. It is suggested that the identification threshold of SNP-based GS (Genetic Similarity) value for EDVs in Chinese varieties, based on high-density SNP data, should be above 92.0%.
6. Conventional systematic breeding can also develop Essential Derived Varieties (EDVs) of their parents.

13

THANK YOU FOR YOUR ATTENTION!

14

[End of Annex and of document]