UPOV
International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Testing Methods and Techniques **TWM/3/17**

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

Original: English
Date: April 9, 2025

PAD – AN ALGORITHM FOR PROGENY-ANCESTOR DETECTION BASED ON GENETIC PROFILES

Document prepared by an expert from the International Seed Federation (ISF)

Disclaimer: this document does not represent UPOV policies or guidance

The annex to this document contains a copy of a presentation "PAD – an algorithm for Progeny-Ancestor Detection based on genetic profiles", to be made by an expert from the International Seed Federation (ISF), at the third session of the TWM.

[Annex follows]

International Seed Federation
Seed is Life

PAD – an algorithm for Progeny-Ancestor Detection based on genetic profiles

Emerson Limberger and Achik Aznag
TWM/3
April 2025

1

Progeny-Ancestor Detection based on genetic profiles

Varieties are the most valuable result of plant breeders' work and characterizing them is a key component of PVP.

Traditional methods for characterization include:

- the collection of morphological characteristics that describe phenotypes
- the annotation of formulas that represent breeding schemes or variety components (parental lines)

The advent of marker technologies could be used to support both cases. Today markers are already implemented to optimize variety collections, to offer an alternative method for examining characteristics whereby the phenotypic expression remains conclusive and by a few authorities to confirm pedigrees in species where the parental lines are part of the DUS examination.

This work presents an algorithm for progeny-ancestor detection to perform marker analysis that could support different PVP related activities.

International Seed Federation
Seed is Life

2

Compatibility with different DNA assays

DNA assays are extremely relevant when considering the analysis to be applied.

The current version of the algorithm is intended to work with MNPs as described by Dr. Peng Hai (UPOV document **BMT/18/15**).

The analysis can also be performed with SNPs when those are structured in Genetic Tags under the following assumptions:

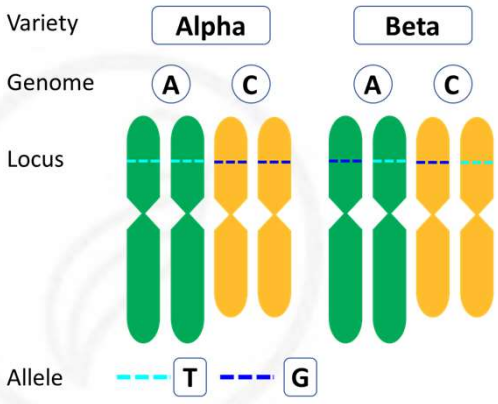
- The tags are built with linked markers respecting the physical or genetic order
- There is no recombination between the markers within a tag
- The number of alleles is complete

3

Species considered

The analysis can be performed in any species, but the algorithm considers allele inheritance from a single locus

- Level of ploidy: the preferred ploidy is diploid but there is no restriction of use. If the presence of heterologous loci affect the allele calling process, that issue will be reflected in the results. In those cases, a segregation test is recommended and only markers that segregate in 1:1 should be used.



4

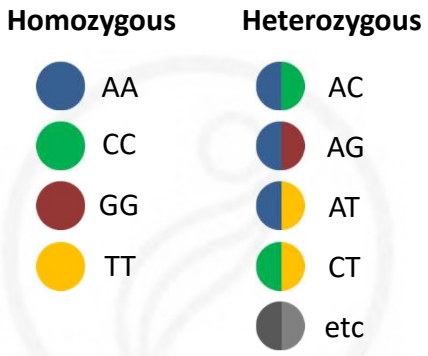
Requirements to run the test

The loci are represented at the SNP level, with their respective biallelic variation and possible heterozygote.

The loci can be an individual SNP or a combination of two or more.

The combination of SNPs generates a MNP or a Genetic Tag (depending on the genotyping assay) and computed as such. This methodology was adopted to enable analysis of datasets derived from any SNP chemistry, thus allowing a broader usage.

The nucleotides are represented by their first letters: Adenine=A, Cytosine=C, Guanine=G and Thymine=T, at the diploid level.



5

Requirements to run the test

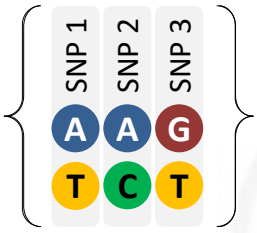
Number of expected alleles (e) based on the biallelic property of SNPs=2 and the length of the MNP (L), in this example=3.

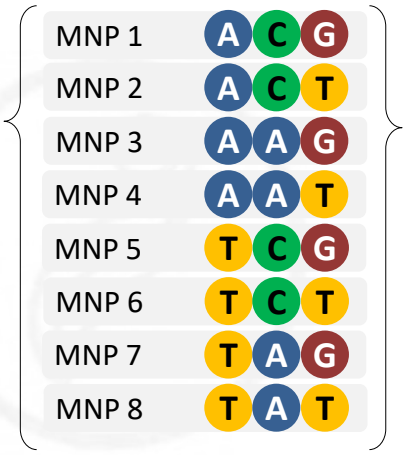
The permutation allowing repetitions

$$Alleles = e^L$$

e=2 and L=3

The number of expected alleles is 8. Loci lacking the complete expected alleles can still be used.

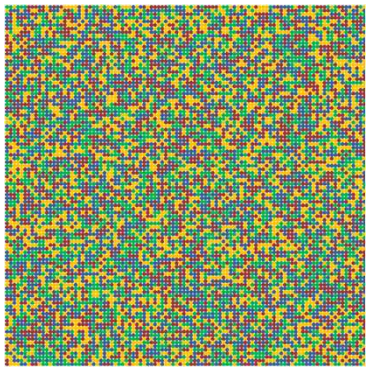




6

Requirements to run the test

Reference collection of varieties with genotypes



Target genotype

Result

The line or lines matching the criteria, that is, having a compatible genetic profile, will be listed as the potential parents.

7

Reference set

The reference set contains marker names in the first column, the linkage block ID in the second column, followed by inbred lines in subsequent columns.

Marker	LB	Variety 1	Variety 2	Variety 3	Variety 4	Variety 5	Variety 6	Variety 7	Variety 8
SNP 1	1	AA	AA	AA	AA	CC	CC	CC	CC
SNP 2	1	CC	CC	GG	GG	CC	CC	GG	GG
SNP 3	1	GG	TT	GG	TT	GG	TT	GG	TT
SNP 4	2	TT	GG	TT	GG	TT	GG	TT	GG
SNP 5	2	TT	TT	TT	TT	AA	AA	AA	AA
SNP 6	2	GG	GG	AA	AA	GG	GG	AA	AA
SNP 7	3	CC	CC	AA	AA	CC	CC	AA	AA
SNP 8	3	TT	CC	TT	CC	TT	CC	TT	CC
SNP 9	3	AA	AA	AA	AA	CC	CC	CC	CC

8

Reference set

The reference set contains marker names in the first column, the linkage block ID in the second column, followed by inbred lines in subsequent columns.

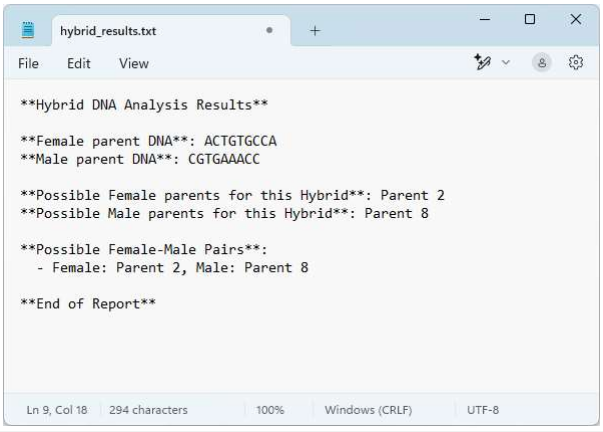
MNPs within the same sequence are identified by the same ID

Marker	LB	Variety 1	Variety 2	Variety 3	Variety 4	Variety 5	Variety 6	Variety 7	Variety 8
SNP 1	1	AA	AA	AA	AA	CC	CC	CC	CC
SNP 2	1	CC	CC	GG	GG	CC	CC	GG	GG
SNP 3	1	GG	TT	GG	TT	GG	TT	GG	TT
SNP 4	2	TT	GG	TT	GG	TT	GG	TT	GG
SNP 5	2	TT	TT	TT	TT	AA	AA	AA	AA
SNP 6	2	GG	GG	AA	AA	GG	GG	AA	AA
SNP 7	3	CC	CC	AA	AA	CC	CC	AA	AA
SNP 8	3	TT	CC	TT	CC	TT	CC	TT	CC
SNP 9	3	AA	AA	AA	AA	CC	CC	CC	CC

Test set

Users should prepare the reference set with the marker names in the first column, the linkage block in the second column, followed by inbred lines in subsequent columns.

Marker	Hybrid
SNP 1	AC
SNP 2	CG
SNP 3	TT
SNP 4	GG
SNP 5	TA
SNP 6	GA
SNP 7	CA
SNP 8	CC
SNP 9	AC



Next steps

The best performance for this analysis is achieved with MNPs as described in the BMT/18/15.

It can be used with genetic tags within the same recombination blocks but needs further testing.

Eventually the algorithm could be adapted to support other chemistries, including older techniques like SSR.

A test version will be made available for those who are interested in using the algorithm.

Please contact: emerson.limberger@corteva.com for a test version

11

Thank you for your attention!

Seed is Life

12

[End of Annex and of document]