


Technical Working Party on Testing Methods and Techniques**TWM/1/18****First Session****Virtual meeting, September 19 to 23, 2022****Original:** English**Date:** August 31, 2022

INTERNATIONAL HARMONISATION AND VALIDATION OF A SNP SET FOR THE MANAGEMENT OF TOMATO REFERENCE COLLECTION*Document prepared by an expert from the Netherlands**Disclaimer: this document does not represent UPOV policies or guidance*


The annex to this document contains a copy of a presentation on “International harmonisation and validation of a SNP set for the management of tomato reference collection”, prepared by an expert from the Netherlands, to be made at the first session of the TWM.

[Annex follows]



International harmonisation and validation of a SNP set for the management of tomato reference collection


UPOV-TWM/1 – September 19-23, 2022




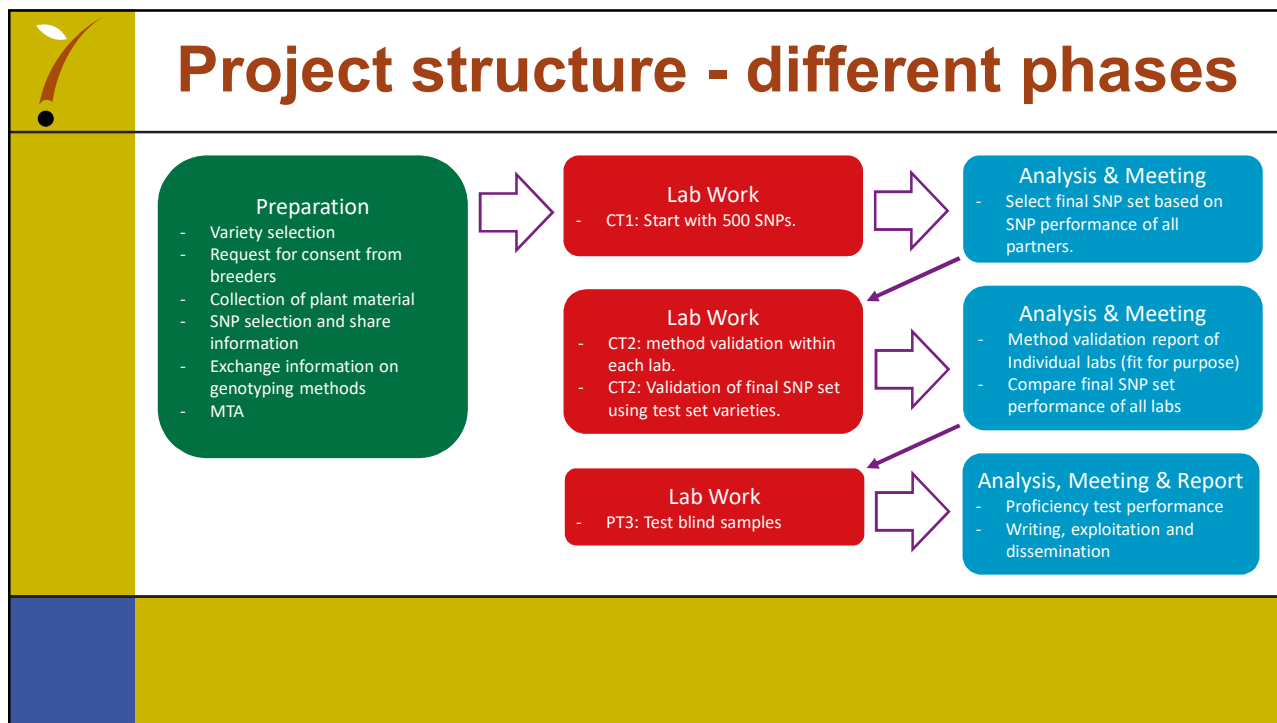
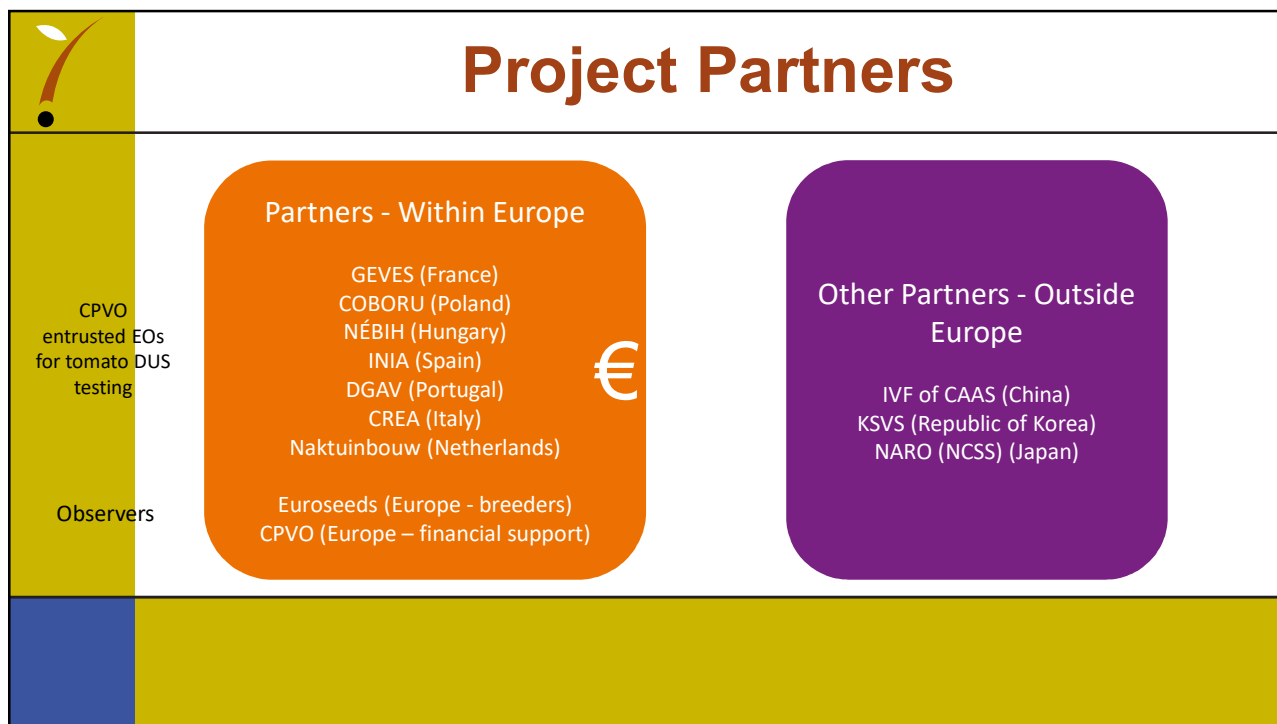
Objective and Scope

Tomato-specific SNP set that is internationally accepted to be fit for purpose (*validated*)

The SNP genotypes of the selected tomato varieties are consistent regardless of *where* (different labs) or *when* (in time) or *how* (different genotyping technologies) the SNP genotypes are produced and analysed. (*harmonised*)

	<h2>Objective and Scope</h2>
	<p>Validated and harmonised SNP set and SNP genotypes are prerequisites for...</p> <ul style="list-style-type: none">- International DNA database for tomato- Use the database in DUS procedure for the management of reference collection (UPOV-models) <p>First step: this project</p>

	<h2>General Project information</h2>
	<p>Project started July 2019 (grant agreement between CPVO and Naktuinbouw)</p> <p>Budget €295.000; co-financed by CPVO for 90%</p> <p>Duration 30 months (December 2021) – extended with 20 months (August 2023)</p> <p>Delay:</p> <ul style="list-style-type: none">• Legal arrangements like Project Partner Agreement and Agreement on ownership and use of plantmaterial and DNA samples• Requesting consent of the titleholders• Covid 19



Variety Selection Criteria

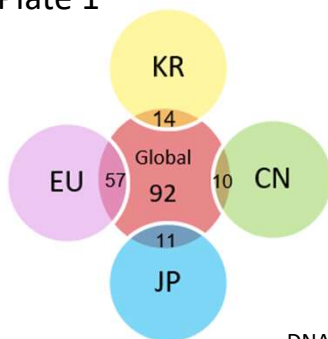


Criteria for 'training'-set:

- Representation of a broad genetic diversity (all types and all characteristics)
- varieties that are morphologically close but distinct, (variety pairs that might have caused some discussion in the DUS test and/or an extra year of testing was required to consider them distinct)
- different companies (different germplasms)
- No wild species

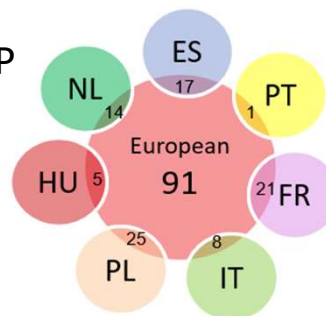
GLB set vs EU set

Plate 1



NO OVERLAP


Plate 2



DNA quality was very high
200 ul of 20 ug/ul

Global set of Varieties
Tested by all 7 lab partners


European set of Varieties
Tested only by 4 lab partners



Providing SNP information to partners

- We start with the best 500 SNPs from the Axiom Array screen (51.214 SNPs)
- Positions of these 500 SNPs on the reference genome were shared
- Flanking sequences 100 bp upstream and 100 bp downstream of the SNP position suitable for primer design were shared
- Inventory of genotyping methods and service providers

Coordinator provide 500 – 1000 SNPs and flanking regions

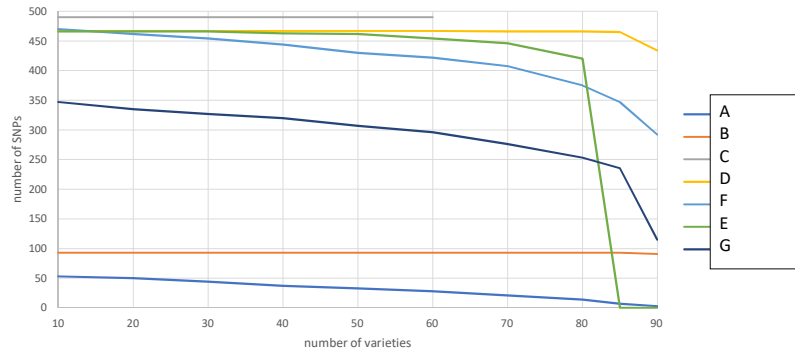


Data analysis on SNP 'Performance'

1. How many SNP assays are successful in producing genotypes for the varieties?
2. Are the genotypes for the SNPs produced consistent between the partners?
3. Can the varieties of the developmental sets be distinguished?

Successful SNPs in GLB set

Number of successful SNPs per number of varieties from the Global training-set for each partner



a SNP is 'successful' when a genotype was obtained for at least one sample (variety).

Conclusions

- The number of successful SNP assays is very variable between the 7 partners
- For most partners the number of successful SNPs drops for >80 varieties. So, in most of the SNP datasets we can observe missing data for 0-20 varieties
- Not the same varieties are missing in the several datasets. The missing varieties are randomly divided and different for each partner. From this observation we can conclude that DNA quality is not the reason for genotype failure of a particular variety.
- From these results we cannot draw a conclusion on which technology or genotyping method is preferable
- The number of successful SNPs for each of the EU partners is very consistent: the results on the GLB and EU sets of varieties are very consistent for each partner
- Whether a SNP is successful or not, is independent on the set of varieties

Data analysis on SNP 'Performance'

1. How many SNP assays are successful in producing genotypes for the varieties?
2. Are the genotypes for the SNPs produced consistent between the partners?
3. Can the varieties of the developmental sets be distinguished?

Consistency of genotypes

We want to select SNPs that produce consistent genotypes by all partners for each variety

	SL3.0ch01_346524_E	SL3.0ch01_346524_D	SL3.0ch01_346524_F	SL3.0ch01_346524_C	SL3.0ch01_507890_E	SL3.0ch01_507890_D	SL3.0ch01_507890_F	SL3.0ch01_507890_C
SL3.0ch01_346524_E	100	40,97	100	100	18,91	24,1	21,52	17,86
SL3.0ch01_346524_D	40,97	100	42,94	36,93	28,66	31,53	30,11	23,08
SL3.0ch01_346524_F	100	42,94	100	100	18,91	23,37	21,03	16,93
SL3.0ch01_346524_C	100	36,93	100	100	20	24,62	22,58	16,93
SL3.0ch01_507890_E	18,91	28,66	18,91	20	100	89,64	91,03	94,55
SL3.0ch01_507890_D	24,1	31,53	23,37	24,62	89,64	100	100	90,77
SL3.0ch01_507890_F	21,52	30,11	21,03	22,58	91,03	100	100	91,94
SL3.0ch01_507890_C	17,86	23,08	16,93	16,93	94,55	90,77	91,94	100

Matrix comparing
(all SNPs for all partners) x (all SNPs for all partners)

dataset is not complete: missing SNP assays for partners – input is successful SNPs for each partner

Per SNP we compare the genotypes obtained by partner X to the genotypes obtained by partner Y

When the genotypes are consistent, the similarity is 100

a snapshot of the total similarity matrix for the pair-wise comparison of successful SNPs per partner for the Global Developmental set

We calculated the average similarity per SNP over the genotypes of all partners as an expression of consistency

Green SNP=70,14 Blue SNP=92,99

Consistency of genotypes

For 494 SNPs we obtained successful genotypes for at least 2 partners
The number of pair-wise combinations that is used to calculate the average similarity for all 494 SNPs is 4614.

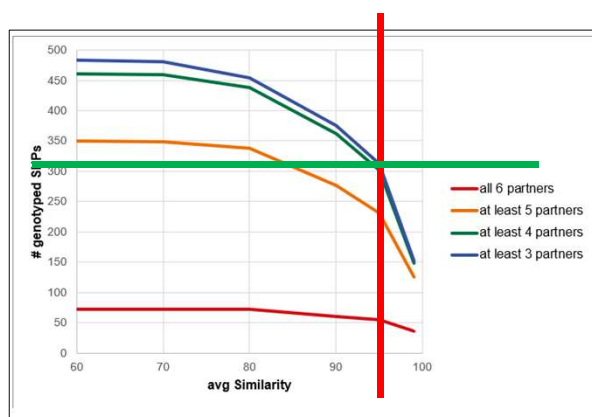
	#SNPs genotyped by N partners				
average similarity range	N = 6	N = 5	N = 4	N = 3	N = 2
>99	36	90	22	5	1
>95	55	176	71	11	3
>90	60	217	85 (Blue SNP)	13	3
>80	72	266	100	16	5
>70	72	277	110 (green SNP)	22	6
>60	72	278	111	23	6
<60	0	1	0	0	3

varieties with a successful genotype was not taken into consideration in this table. However, to determine SNP 'performance' also includes the successful genotyping on as many as possible varieties.

Consistency of genotypes

Average similarity =
expression for
consistency of the
genotypes.

Selection of high performance SNPs
- Average sim is most important!
- Be strict! >95
- Not much difference between 3 and 4 partners. Be strict! At least 4 partners



Consistency of genotypes

For 494 SNPs we obtained successful genotypes for at least 2 partners

The number of pair-wise combinations that is used to calculate the average similarity for all 494 SNPs is 4614.

average similarity range	#SNPs genotyped by N partners				
	N = 6	N = 5	N = 4	N = 3	N = 2
>99	25	20	23	5	1
>95	55	176	71	1	3
>90	68	227	103 (green SNP)	23	3
>80	72	266	100	16	5
>70	72	277	110 (green SNP)	22	6
>60	72	278	111	23	6
<60	0	1	0	0	3

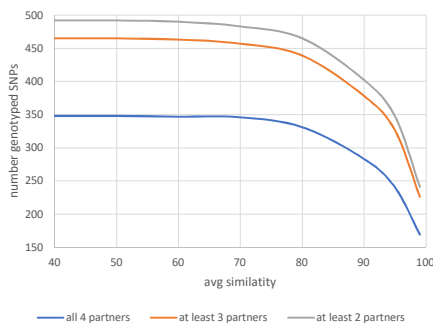
Best performing SNPs:
Very high average sim: >95
At least 4 partners: N≥4

#SNPs: 55+176+71=302

As the Global Set

Consistency of genotypes

average similarity range	#SNPs genotyped by N partners		
	N = 4	N = 3	N = 2
>99	169	57	15
>95	240	86	21
>90	283	95	24
>80	331	108	26
>70	346	111	26
>60	347	116	27
<60	1	1	0



Best performing SNPs:
Very high average sim: >95
At least 3 partners: N≥3

#SNPs: 240+86=326

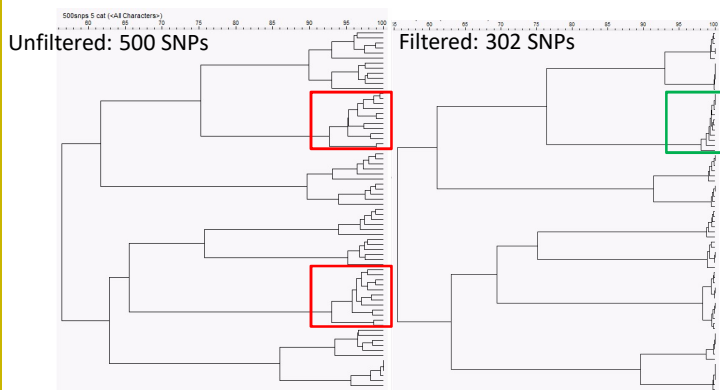
As the European set

Data analysis on SNP 'Performance'

1. How many SNP assays are successful in producing genotypes for the varieties?
2. Are the genotypes for the SNPs produced consistent between the partners?
3. Can the varieties of the developmental sets be distinguished?

Discriminative power - GLB

Are all varieties in GLB developmental set be distinguished with SNP set?



Before filtering:

- 88 varieties were distinct
- 2 pairs of varieties not distinct

After filtering:

- 1 pair was still not distinct

After filtering:

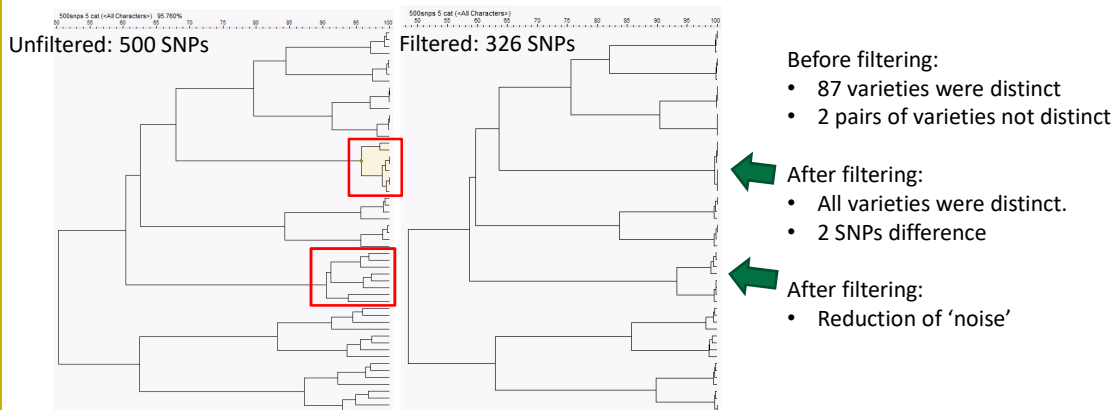
- Reduction of 'noise'

The effect of filtering the SNPs on 'performance'

different EOs – same company: explanation = same variety, different names?

Discriminative power - EU

Are all varieties in EU developmental set be distinguished with SNP set?



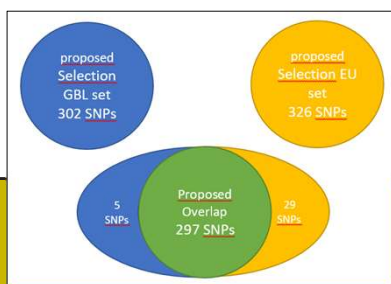
The effect of filtering the SNPs on 'performance'

Final selection of SNPs


Agreement to proceed with 1 International SNPs set.

Agreement to use average similarity >95 and $N \geq 4$ for GLB set

Agreement to use average similarity >95 and $N \geq 3$ for EU set



overlap between GLB and EU
297 SNPs

	<h2>Work in progress</h2>
	<p>Genotyping validation / test set varieties by all partners (1 plate; 90 samples)</p> <p>Method validation by each partners individually for every method - repeatability, reproducibility and robustness</p> <p>3rd Lab meeting</p> <p>Blind test</p>

