



**HAL**  
open science

## **Feedback on tGBS with Allegro : Does it hit the target ?**

Aurélie Berard, Aurélie Canaguier, Romane Guilbaud, Isabelle Le Clainche,  
Marie-Christine Le Paslier, Benjamin Loire, Damien Hinsinger, Patricia Faivre  
Rampant

### ► **To cite this version:**

Aurélie Berard, Aurélie Canaguier, Romane Guilbaud, Isabelle Le Clainche, Marie-Christine Le Paslier, et al..  
Feedback on tGBS with Allegro : Does it hit the target ?. PAG XXX. Plant and Animal Genome, Jan 2023, San  
Diego (California), United States. , <10.15454/4NYHD6>. <hal-03937923>

**HAL Id: hal-03937923**

**<https://hal.inrae.fr/hal-03937923v1>**

Submitted on 13 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No  
Derivative Works - International License

# Feedback on tGBS with Allegro : Does it hit the target ?

## Unit EPGV US1279

Aurélien Berard, Aurélien Canaguier, Romane Guilbaud, Isabelle Le Clainche, Marie-Christine Le Paslier, Benjamin Loire, Damien Hinsinger, Patricia Faivre Rampant (Direction)

## EPGV tGBS projects.

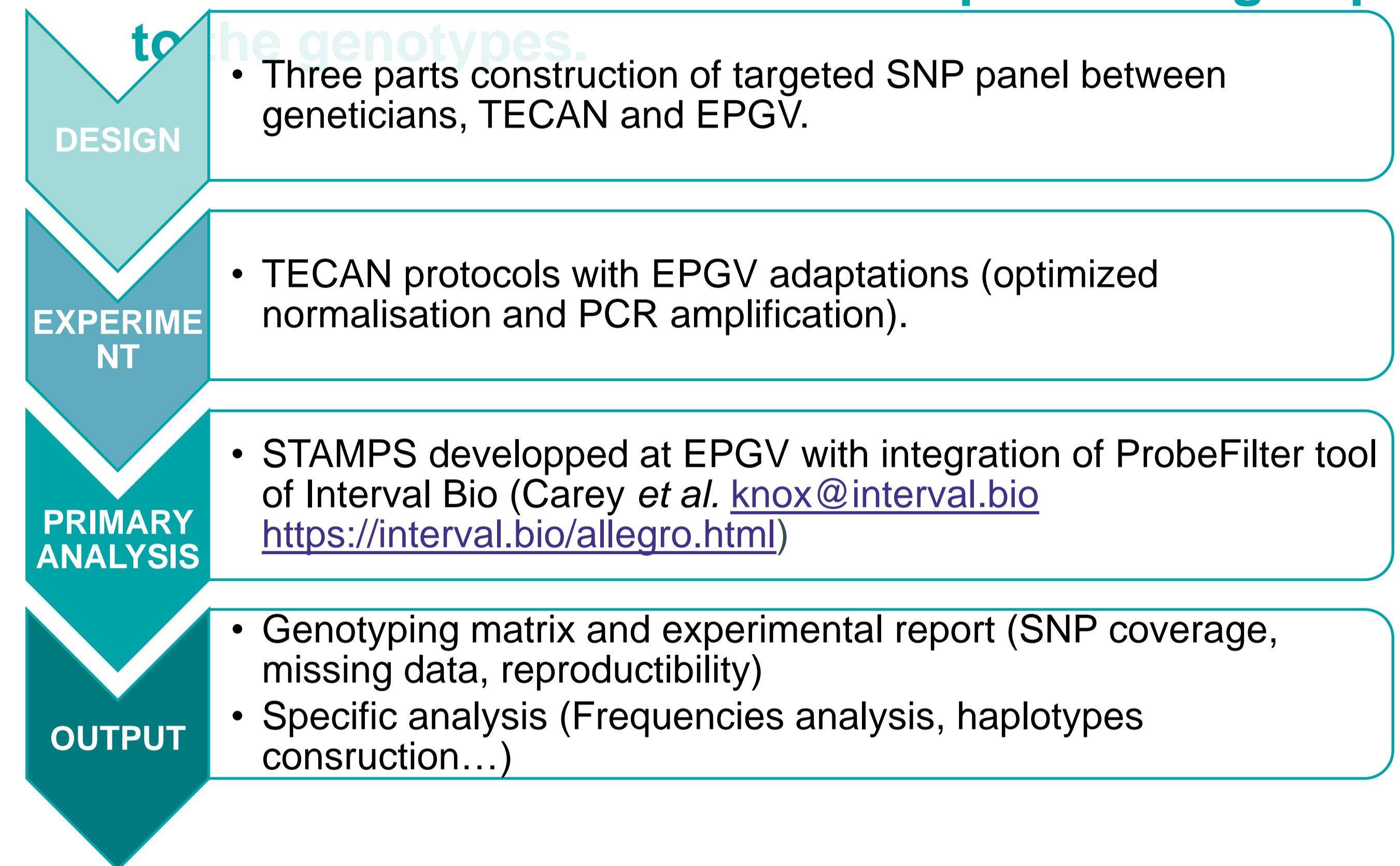
Organisms	Reference	Number of SNP Target	Number of samples	Studied samples
<i>Zea mais</i>	Genome v2.0 (2n=20, 2300 Mb) <sup>1</sup>	2 592	192	Lines, hybrids, pools
<i>Lens sp.</i>	Targeted regions (2n=14, ~25 Mb)	25 000	336	Varieties, wild accessions
<i>Solanum lycopersicum</i>	Genome v4.0 (2n=24, 950 Mb) <sup>2</sup>	19 998	384	Varieties
<i>Plasmopara viticola</i>	Genome (2n=34, 93 Mb) <sup>3</sup>	5 000	2x192	F1 populations

<sup>1</sup>. [https://download.maizegdb.org/B73\\_RefGen\\_v2/](https://download.maizegdb.org/B73_RefGen_v2/)

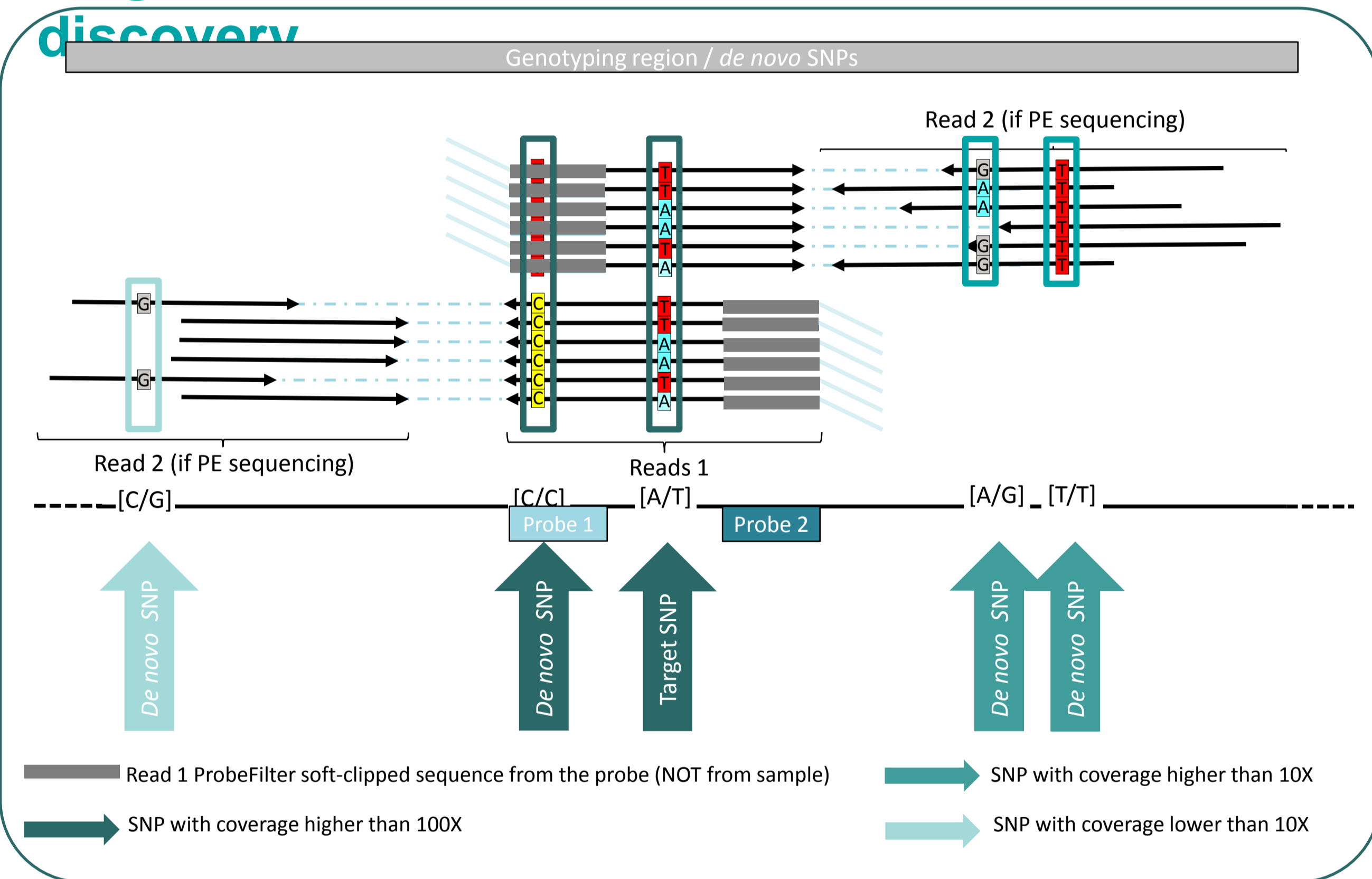
<sup>2</sup>. Hosmani et al. 2019. [https://solgenomics.net/organism/Solanum\\_lycopersicum/genome/](https://solgenomics.net/organism/Solanum_lycopersicum/genome/)

<sup>3</sup>. Dussert et al 2018. <https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.15454/4NYHD6>

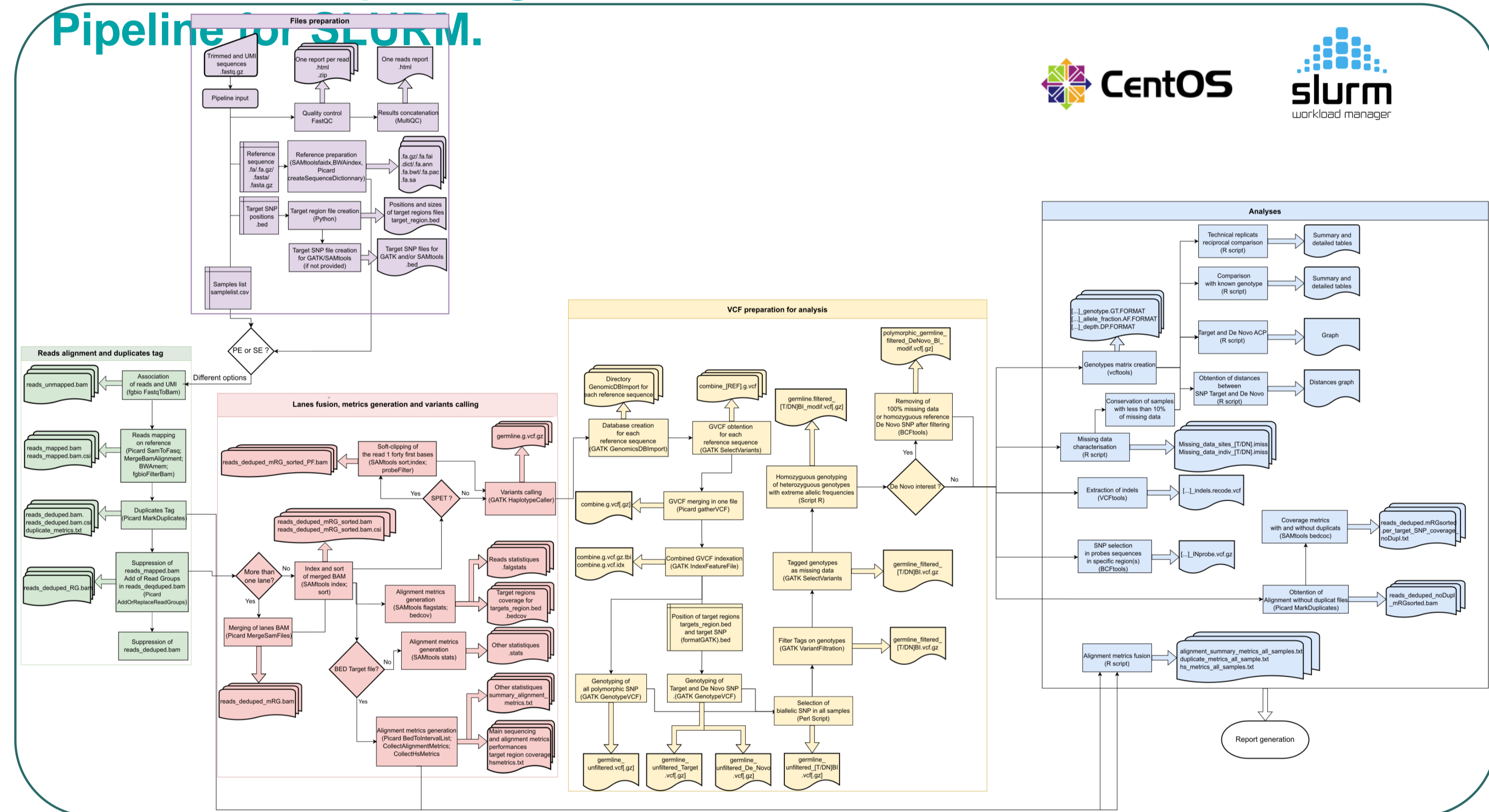
## tGBS Workflow : from the SNP panel design up to the genotypes



## Target and de novo SNP discovery



## STAMPS - Simple Targeted GBS Automated and Modular Pipeline for SLURM.



## Conclusions

SPET is a powerful flexible genotyping solution (from 1000 SNP up to 100 000, and 192 to 3 072 samples in one experiment) for organisms with reference genome or not.

The collaboration between geneticists, TECAN and EPGV leads to specific probes design illustrated by the high percent of bases alignment on the targeted regions. The *P. viticola* exception is potentially due to the lower specificity of SE mapping regarding to the PE mapping.

The targeted SNP coverage variability obtained, had no impact on the genotyping success rate. And a good sample repeatability was observed within the same experiment. According to our comparison results between genotypes from tGBS and available SNP array considered as golden standard, we can infer that tGBS could be an effective alternative to SNP array.

Moreover, the SNP *de novo* discovery allows haplotype reconstruction that could be useful to limit ascertainment bias. Additionally, the tGBS workflow can be adapted to genotype structural variations.

## Result

Organism	SE/PE	Design Success Rate	On Target Bases Percent	Median, Min, Max coverage of Target SNP	Percent of genotype Target SNP after filtering <sup>a</sup>	Target SNP Repeatability Percent <sup>b</sup>	SPET tGBS vs Array Validation	De novo SNP Number	X-times Target SNP Number
<i>Z. Mais</i>	PE	99.4% (2 571/2 587)	89	238X, 10X, >45 0X	99.9	96.5	98.8 <sup>c</sup>	60 269	~23x
<i>L. species</i>	PE	100% (25 000/25 000)	98.7	75X, 31X, 167 X	95.9	99.0	-	188 159	~7.5x
<i>S. lycopersicum</i>	SE	99.9% (19 998/20 000)	82.2	77X, 10X, 139 X	94.1	99.9	-	23 522	~2x
<i>P. viticola</i>	PE	100% (5 000/5 000)	98.8	226X, 10X, >45 0X	99.9	98.7	-	122 647	~24.5x

<sup>a</sup> Filter: DP > 10X, GQ > 20, heterozygous sites frequency (0.3 < AF < 0.8).  
<sup>b</sup> Filtered Target SNP genotypes comparison of technical repetitions (*Z. mais*: 4 rep. of B73; *L. sp.*: 6 rep. of Anicia; *S. lycopersicum*: 8 rep. of CR001; *P. viticola*: 4 replicats of PV1419 and PV412).  
<sup>c</sup> Comparison on 26 samples and 2 599 Target SNP with exclusion of missing data points. Results are validated in case of genotypes identity between both technologies.

Center Île-de-France - Versailles-Saclay

## Aknowlegdments

- Yamaizing INRAE UMR GQE – F - Le Moulon
- RésiLens INRAE – UMR Agroecologie – F - Dijon
- invite INRAE UR GAFL – F - Avignon
- GENOD EMIL INRAE UMR SAVE – F - Villenave d'Ornon



INRAE - Unit EPGV US1279 - Department BAP Research Infrastructure INRAE GENOMICS CEA - Institut de biologie François Jacob 2 rue Gaston Crémieux - 91057 Evry Cedex FRANCE

[support-epgv@inrae.fr](mailto:support-epgv@inrae.fr)

<https://www6.versailles-grignon.inrae.fr/epgv>