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

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

Unit EPGV US1279

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

(TECAN).

SNP genotyping is widely used at many scale, both for research and breeding for diversity and genetic association studies, building genetic maps, confirm varieties identification, detect QTL and genomic selection. However, species diversity in term of genomic features (such as genome size, ploidy, heterozygosity) challenges the emergence of new and useful approaches. Chip arrays are considered as a “gold standard” but show limits in flexibility and associated costs. In addition, they return information on target SNP only, masking potential useful diversity around the target. We recently implemented a targeted Genotyping-By-Sequencing (tGBS) approach, namely the Allegro Targeted Genotyping v2

EPGV tGBS projects.

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

[1]. https://download.maizegdb.org/B73_RefGen_v2/
[2]. Hosmani et al. 2019. https://solgenomics.net/organism/Solanum_lycopersicum/genome/
[3]. 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.

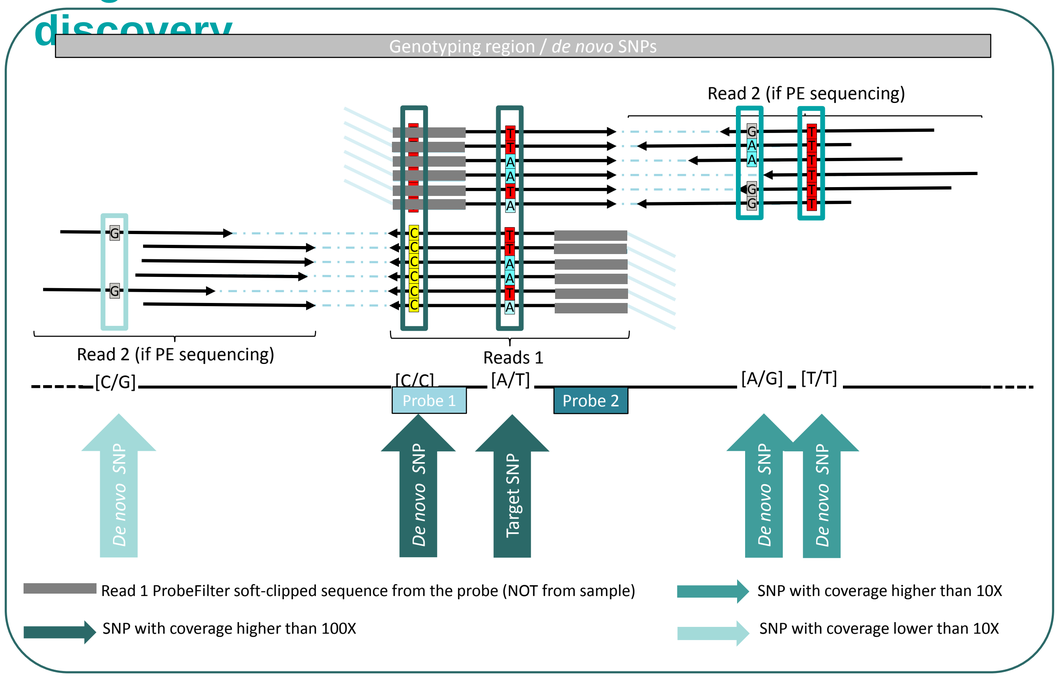
DESIGN
- Three parts construction of targeted SNP panel between geneticians, TECAN and EPGV.

EXPERIMENT
- TECAN protocols with EPGV adaptations (optimized normalisation and PCR amplification).

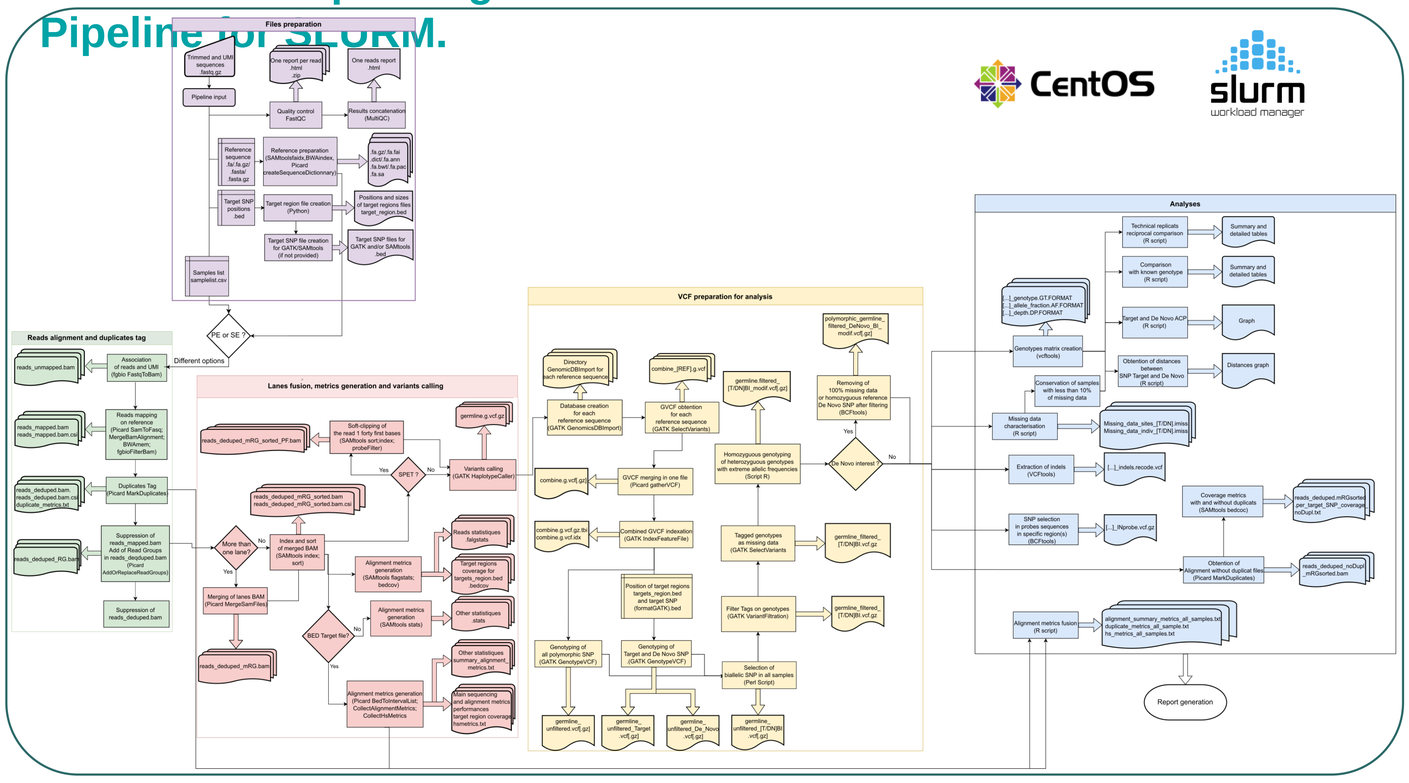
PRIMARY ANALYSIS
- STAMPS developped at EPGV with integration of ProbeFilter tool of Interval Bio (Carey *et al.* knox@interval.bio https://interval.bio/allegro.html)

OUTPUT
- Genotyping matrix and experimental report (SNP coverage, missing data, reproductibility)
- Specific analysis (Frequencies analysis, haplotypes consruction…)

Target and de novo SNP discovery



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



Conclusions

SPET is a powerfull 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 geneticians, TECAN and EPGV leads to specific probes design illustrated by the hight 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 inferred 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. Additionnally, the tGBS workflow can be adapted to genotype structural variations.

Result

Organism	SE/PE	Design Sucess Rate	On Target Bases Percent	Median,Min, Max coverage of Target SNP	Percent of genotype Target SNP after filtering[a]	Target SNP Repeatability Percent[b]	SPET tGBS vs Array Validation	De novo SNP Number	X-times Target SNP Number
Z. Mais	PE	99.4% 2 571/2 587)	89	238X,10X,>450X	99.9	96.5	98.8[c]	60 269	~23x
L. species	PE	100% (25 000/25 000)	98.7	75X,31X,167X	95.9	99.0	-	188 159	~7.5x
S. lycopersicum	SE	99.9% (19 998 /20 000)	82.2	77X,10X,139X	94.1	99.9	-	23 522	~2x
P. viticola	PE	100% (5 000/5 000)	98.3	226X,10X,>450X	99.9	98.7	-	122 647	~24.5x

[a]. Filter : DP>10X, GQ>20 ; heterozygouous allelic frequency (0.2<AF<0.8).
[b]. 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).
[c]. 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

INRAE UMR GQE – F - Le Moulon
INRAE – UMR Agroecologie – F - Dijon
INRAE UR GAFL – F - Avignon
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

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