BSA-Seq : an efficient method to decipher a complex trait on Poplar, a highly heterozygous diploid genome

Aurélie Canaguier[1], Véronique Jorge[2], Vanina Guérin[2], Odile Rogier[2], Vincent Segura[2], Aurélie Chauveau[1], Elodie Marquand[1], Aurélie Bérard[1], Marie-Christine Le Paslier[1], Catherine Bastien[2] and Patricia Faivre Rampant[1]

1 - INRA, US 1279 Etude du Polymorphisme des Génomes Végétaux, F-91000 Evry, France *2 – INRA, UR 0588 AGPF, Centre INRA Val de Loire, Orléans, France*

INTRODUCTION

The efficiency of the Bulk Segregant Analysis (BSA) has clearly been demonstrated to detect genomic regions and genes involved in various traits**.** It allows large experiments reducing the cost and time and preserving the power of full individual's population analysis. Over the past few years the combination of BSA and Next Generation Sequence (NGS) data (BSA-Seq) has given a new accuracy and robustness to the discovery of genes and genomics regions underlying traits of interest, mainly on crop and model species (1).

In our study, we applied the BSA-Seq on Poplar, a heterozygous and diploid genome, to decipher the genetic determinism of leaf rust resistance. As a proof of concept, we focused on R_{us}, a major gene previously fine-mapped on Chromosome 19 and controlling the uridenia size during the rust-Poplar interaction (2,3).

MATERIAL AND METHODS

MATERIAL

- Phenotyping for traits associated to the resistance to *Melampsora larici-populina* (*Mlp*) leaf rust of parents and **1414 progenies** from an interspecific cross :

***Populus deltoides* clone 73 028-62 (*Pd*) x *Populus trichocarpa* clone 101-74 (*Pt*)**

- Independently DNA extractions with Qiagen Kit **and genotyping** (2,3).

METHOD – BULK CONSTITUTION

The selection of 62 progenies, based on genotyping of markers physically linked to R_{us} and the phenotypic information, was realized as described in **Figure 1**. Then, the corresponding DNA were pooled equimolarly into 4 bulks.

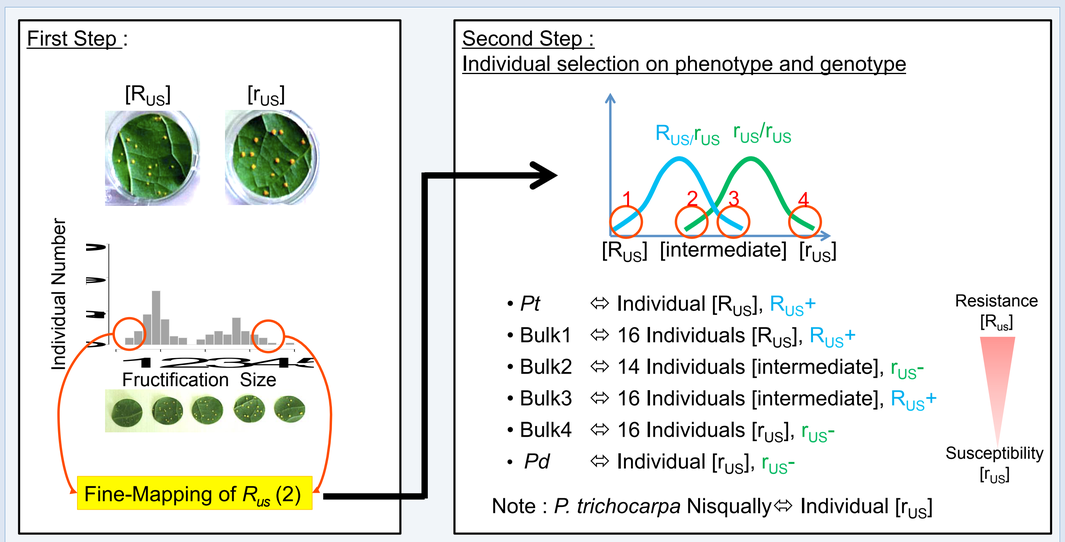

Figure 1. Population phenotyping and genotyping to select progenies for extreme and intermediate Bulks.

METHOD – BULK COMPARISON

The **RUS** markers are expected to be polymorphic between the 2 parents and to cosegregate with [R_{US}] and R_{US}+. More precisely, a variant was considered as **RUS** whenever it fulfilled the following two conditions: (i) its alleles differed between the 2 parents; and (ii) its *P. trichocarpa* allele was present in bulks 1 and 3 and absent in bulks 2 and 4.

Positions which segregates in conformity with the resistance leaf rust and not with R_{US}+.are called « other » (**OTH**).

RESULT 1

ANALYSIS WORKFLOW

STEP	TOOLS	DESCRIPTION	RESULTS
Sequencing	• TruSeq DNA Illumina • GA or HiSeq2000	• Paired-end reads (2x76bp and 2x101bp) • Sizing 450-650bp	Raw reads Pt : 131 823 814 Pd : 314 757 684 B1 : 242 963 534 B2 : 216 556 986 B3 : 174 374 986 B4 : 197 757 684
Trimming	• Trimmomatic/0.32 (5)	• Min Length = 36pb • No N ; PHRED score > 20 on 4 bases (sliding windows)	Trimmed reads Pt : 72 634 382 Pd : 284 660 324 B1 : 206 186 462 B2 : 186 221 932 B3 : 152 344 640 B4 : 178 663 234
Mapping	• BWA/0.7.12 (6)	• Ref : *P. trichocarpa* Nisqually v3.0 softmasked (4) (Default parameters)	Mapped reads Pt : 70 713 777 Pd : 269 209 371 B1 : 200 949 595 B2 : 181 161 530 B3 : 148 398 543 B4 : 173 148 769
Merging and filtration	• SAMtools/1.3.1 (7)	• Reads with unique mapping in pair with quality score ≥ 30	Filtered mapped reads Pt : 51 903 451 Pd : 173 489 464 B1 : 129 698 163 B2 : 121 026 257 B3 : 99 723 832 B4 : 114 707 114
Variant detection	• FreeBayes/0.9.21 (8)	• Multi samples detection on Chromosomes and scaffolds ≥ 50Kb • Multi samples Depth ≥ 30	Variant frequency : Chr. : 32,7 variant/kb Sc. >50kb : ~25 variant/kb
Variant validation	• R/3.3.1	• Conformity of the Mendellian segregation in the 4 Bulks (Pearson's Chi2) • Parental depth > 10 • Information on the 4 bulks	Filtered variant frequency : Chr. : 5 variant/kb Sc. >50kb : ~0,7 variant/kb
Segregation identification	• In-house Perl scripts	• Identification in each Bulk, of the specific allele(s) of *P. trichocarpa* and/or *P. deltoides*. • Characterization of the segregation by bulk comparison	On chromosomes : OTH 0,02 to 0,07% ~32% ; RUS 0,01 to 0,43% ~26% On scaffolds :
Interval identification	• In-house Perl scripts	• At least 2 consecutive variants with the same segregation.	Intervals **OTH** : 74 int. Intervals **RUS** : 91 int.
Interval selection	• Manual selection	• Interval size > 20nt and/or • More than 1 interval in 500Kb	Int. **OTH** on Chr01, 02, 03, 04, 05, 06, 08, 09, 10, 11, 12, 15, 16, 18 and sc.20 : 47 int. 23 regions Int. **RUS** on Chr19, sc.20, sc.25, sc.232 : 32 int. 4 regions
Interval validation	• Primer3/BLAST (9,10)	• Primer3 on 32 Int. sequences (4 Regions) • Validation by BLAST of primers and amplicon on BAC sequences.	Design of 128 primer pairs : 128 NewM 29 validated primer pairs on 29 int. : 29 NewM

B1 : Bulk1 **B3** : Bulk3 **B2** : Bulk2 **B4** : Bulk4 ***Pd*** *: Populus deltoides* clone 73 028-62 ***Pt*** *: Populus trichocarpa* clone 101-74

NewM : New marker **Int.** : Interval **Chr.** : chromosome **sc.** : scaffold

Figure 2. Analysis workflow and general results.

RESULT 2

DESIGN OF NEW MARKERS ON A SPECIFIC REGION

Legend

- Chromosome 19 or scaffolds of *P. trichocarpa* Nisqually v3.0
- Interval on *P. trichocarpa* Nisqually v3.0
- 1 partial new marker : one primer and partial amplicon
- 1, 2 or 3 new marker(s) : 2 primers and amplicon
- *Pt* BACs related to R_{us} allele
- *Pt* BACs related to r_{us} allele
- **74H06**, **97C12** **25F19**... BACs anchoring by new markers on one scaffold and on the Chromosome 19
- Previous genetic and physical markers linked to R_{us} (2).

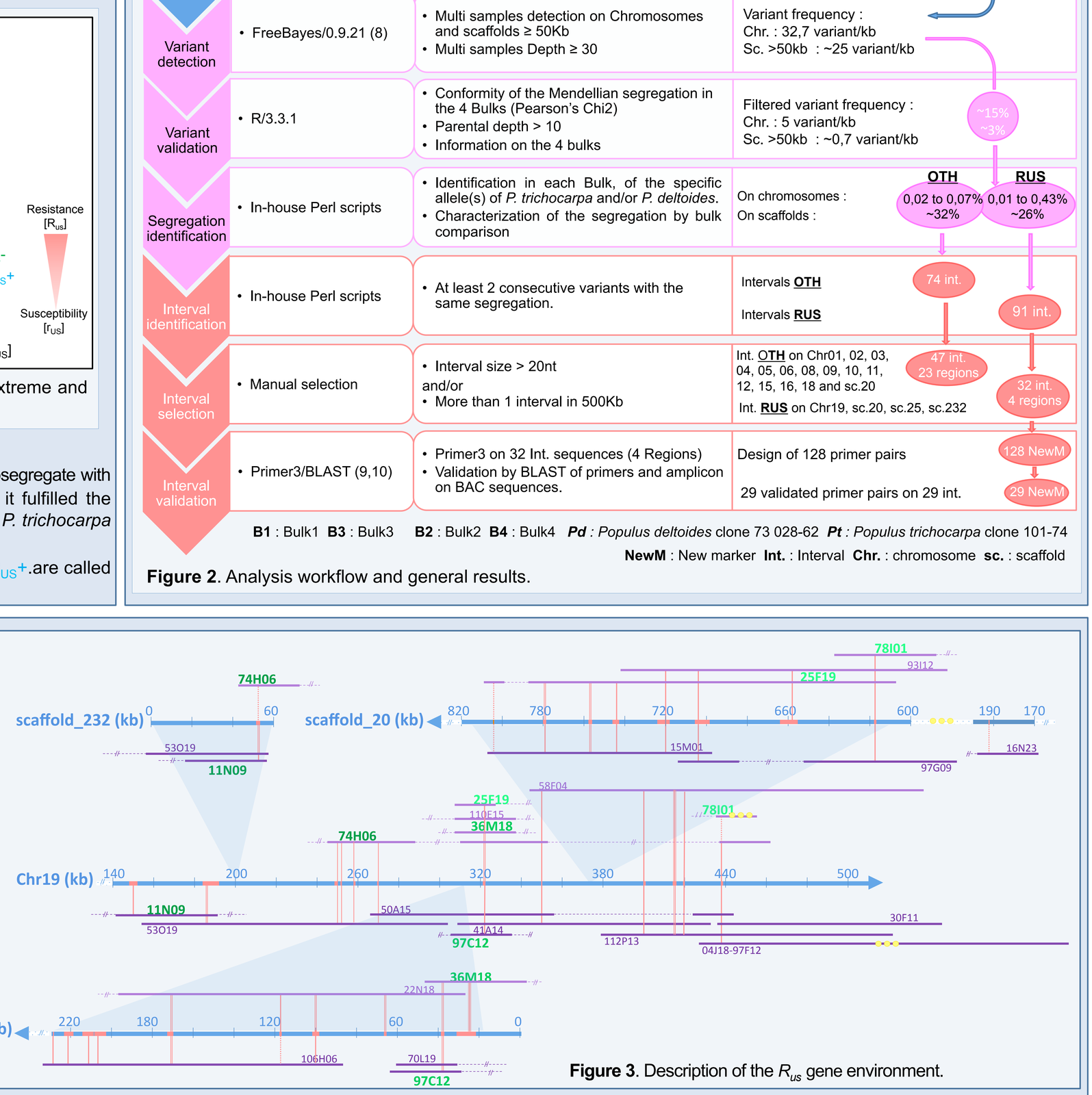

Figure 3. Description of the R_{us} gene environment.

CONCLUSION & PERSPECTIVES

BSA-seq method allows identification of *P. trichocarpa* and/or *P. deltoides* specific variants for complex trait in a diploid and heterozygous context and this, even if the mapping reference doesn't carry the searched region of interest.

Next steps are first to proceed the PCR experiments with the new markers on the parents and progenies to enrich the R_{us} fine-map ; second to characterize the 23 other regions in segregation with the resistance to leaf rust.

Moreover this pipeline, usable on any heterozygous species, releases to the scientific community a high-confidence set of variant positions based on the conformity of the allele frequencies within the bulks.

REFERENCES

- *(1) Cheng Z et al.:* **Genetic sample analysis in genetics and crop improvement**. *Plant Biotechnology Journal* (2016) **14**: 1941-1955.
- *(2) Bresson A et al.:* **Qualitative and quantitative resistances to leaf rust finely mapped within two nucleotide-binding site leucine-rich repeat (NBS-LRR)-rich genomic regions of chromosome 19 in poplar**. *New Phytologist* (2011) **192**: 151–163.
- *(3) Jorge V et al.:* **Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection**. *New Phytologist* (2005) **167**: 113-127.
- *(4) Tuskan G et al.*:**The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray)** *Science* (2006) **313**:1596-1604.
- *(5) Bolger A et al.*: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* (2014) **30**: 2114–2120.
- *(6) Li H and Durbin R*: **Fast and accurate short read alignment with Burrows–Wheeler transform**. *Bioinformatics* (2009) **25**: 1754–1760.
- *(7) Li H et al.*: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* (2009) **25**: 2078–2079.
- *(8) Garrison E and Marth G.*: **Haplotype-based variant detection from short-read sequencing**. *arXiv preprint arXiv:1207.3907 [q-bio.GN]*2012.
- (9) *Untergasser A et al.*: **Primer3—new capabilities and interfaces.** *Nucleic Acids Res* (2012) **40**(15): e115.
- *(10) Altschul S et al.*: **Basic local alignment search tool.** *Journal of Molecular Biology* (1990) **215**(3): 403-410.

THANKS TO CEA-CNG : A. Boland-Augé group for performing DNA samples QC, M.T Bihoreau and D. Lechner for providing INRA-EPGV group with access to their Illumina Sequencing Platform.

INRA SCIENCE & IMPACT

Plant Biology and Breeding Division
F-78026 Versailles, France

Poster Online

EPGV group contact
aurelie.canaguier@inra.fr