LONG DNA TECHNOLOGIES EVALUATION FOR PLANT STRUCTURAL VARIATIONS DETECTION : NANOPORE ONT SEQUENCING vs BIONANO GENOMICS OPTICAL MAPPING

Aurélie Canaguier[1], Erwan Denis[2], Romane Guilbaud[1], Ghislaine Magdelenat[3], Caroline Belser[3], Benjamin Istace[3], Corinne Cruaud[3], Arnaud Lemainque[3], Patrick Wincker[3], Marie-Christine Le Paslier[1], Patricia Faivre-Rampant[1] and Valérie Barbe[2]

[1] Université Paris-Saclay, INRAE, Etude du Polymorphisme des Génomes Végétaux (EPGV), 91000, Evry, France, [1] IR INRAE Génomique ,
[2] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France , [3] Genoscope, Institut François Jacob, CEA, Université Paris-Saclay, 91057 Evry, France.

INTRODUCTION

Structural Variations (SVs) identification remains in infancy in plants mainly due to the limited useful genomic resources. Long-read sequencing obtained with the MinION instrument (Oxford Nanopore Technologies (ONT)) and DLE optical maps produced by the Saphyr device (Bionano Genomics), were evaluated on their ability to identify SVs, taking as study model two *Arabidopsis thaliana* ecotypes, Columbia-0 (Col-0) and Landsberg erecta-1 (Ler-1).

RESULTS

1- Assembly & Alignment : metrics and visualization

Technology	ONT		Bionano	
Ecotype	**Col-0**	**Ler-1**	**Col-0**	**Ler-1**
Cumulative size (Mb) sequences or molecules	13 036	8 174	577 460	610 944
N50 size (Kb)	12	14	153	135
Depth haploide genome of 130Mb	X100	X63	X4 440	X4 700
Cumulated size (Mb) Trimmed seq / sampled molecules	9 817	6 141	90 000	75 000
N50 size (Kb)	12.7	16.5	259	300
Depth haploide genome of 130Mb	X75	X47	X700	X580
Number of contigs or cmaps	**79**	**101**	**18**	**14**
N50 size in Mb (number)	12.5 (5)	10.7 (5)	15.5	15.4
Assembly size (Mb)	**117**	**117**	**133**	**131**

Table 1. ONT and Bionano metrics

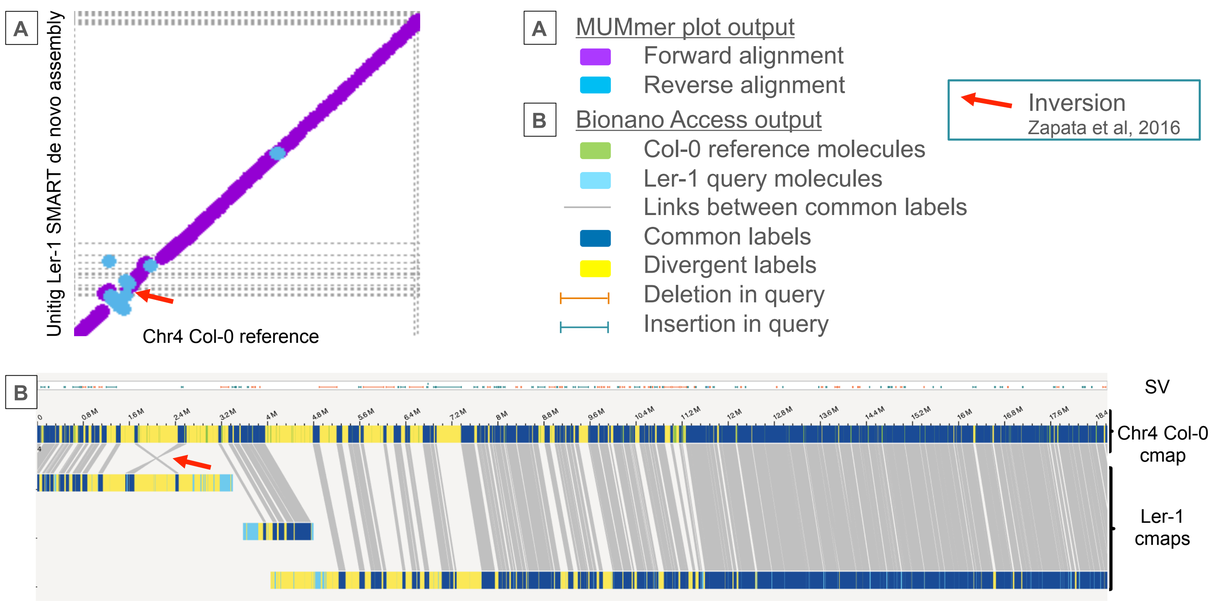


Figure 2. Alignments of Ler-1 assembly and cmaps against the **Chr4** Col-0 reference

2- SV description

Technology	Ecotype	Ref.	# SV	Cumulated Size (Mb)	Median Size (bp)	# INS	# DEL	# INV	# Others
ONT	Ler-1	Col-0	**1 187**	12.3	**3 455**	591	579	12	5
	Col-0	Ler-1	**1 047**	5.4	**3 180**	537	496	10	4
Bionano	Ler-1	Col-0	**589**	6.7	**4 332**	288	293	6	2
	Col-0	Ler-1	**520**	3	**3 741**	280	236	4	0

No duplication was reported in this study.

Table 2. Filtered SV and other rearrangements (jump & translocation) description

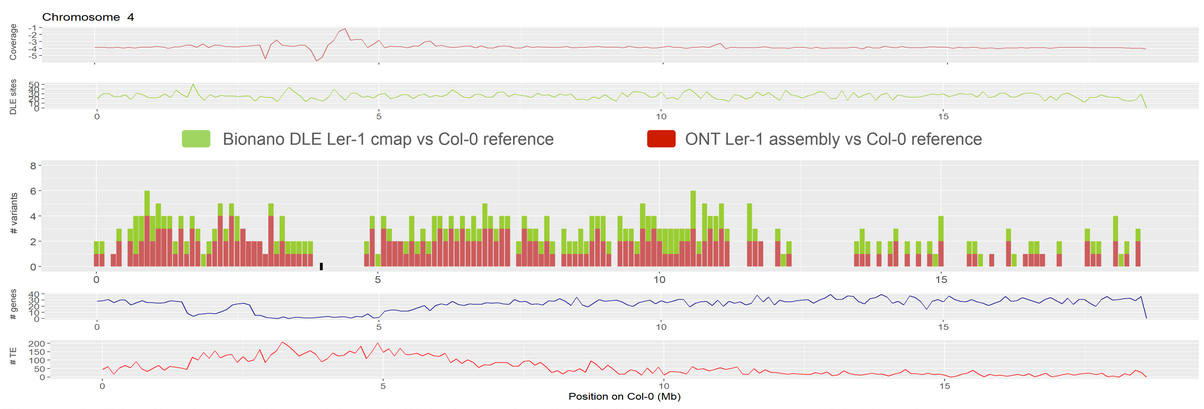
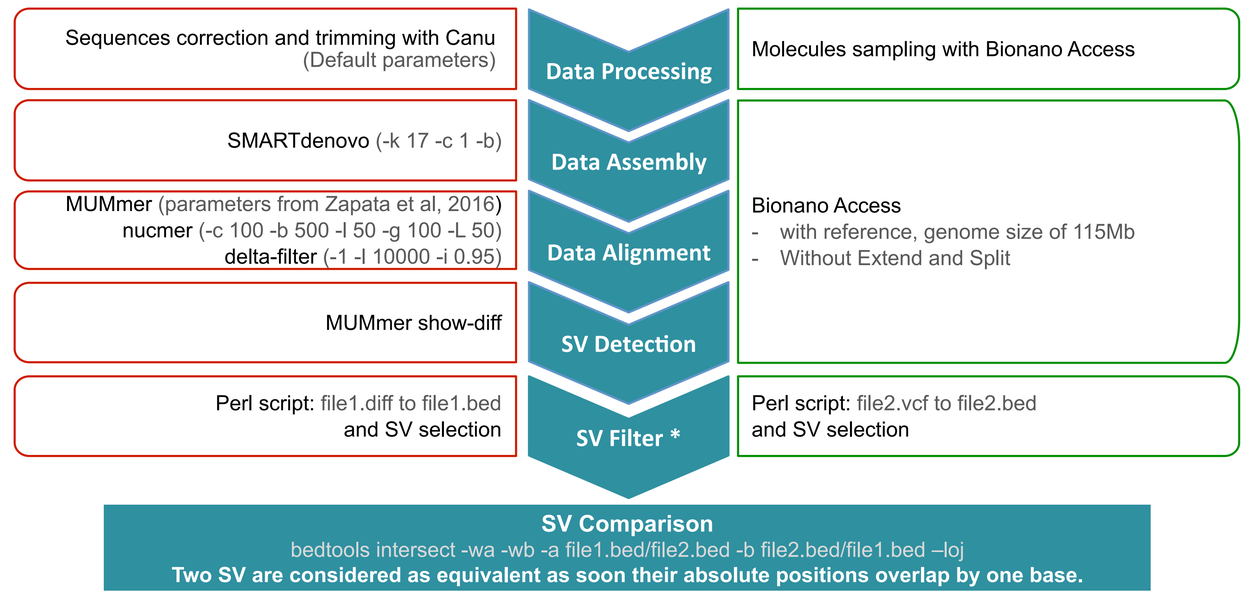


Figure 3. Distribution of filtered Ler-1 SV along the Col-0 Chr4 (❙centromer, 100-Kb windows)

MATERIAL AND METHODS

Arabidopsis thaliana Col-0 (accession 186AV) and Ler-1 (accession 213AV) seeds were obtained from the Versailles Arabidopsis Stock Center, INRA (France).

SV were obtained using Ler-1 and Col-0 assembly and cmap, respectively versus the Col-0 reference (Arabidopsis Genome Initiative, 2000) and the Ler-1 reference (Zapata et al 2016).



* filters : not in reference gap, not in organelles, types insertion (INS) deletion (DEL) inversion (INV), SV size > 1kb.

Figure 1. Analysis pipeline from **ONT** and **Bionano** data to SV comparison

3- SV ONT vs SV Bionano

Ecotype	Ref.	ONT	ONT – Bionano		Bionano	Bionano – ONT	
		specific	conflict by number/by type	concord.	specific	conflict by number/by type	concord.
Ler-1	Col-0	25.0 %	27.0 % / 0.4 %	**47.6 %**	1.0 %	14.5 / 1.2 %	**83.3 %**
Col-0	Ler-1	38.0 %	14.3 % / 0.4 %	**47.3%**	2.3 %	11.0 / 1.0 %	**85.7 %**

Table 3. Percent of filtered SV, discordant (specific, conflictual by number or by type) and concordant (concord.) between ONT and Bionano data analysis

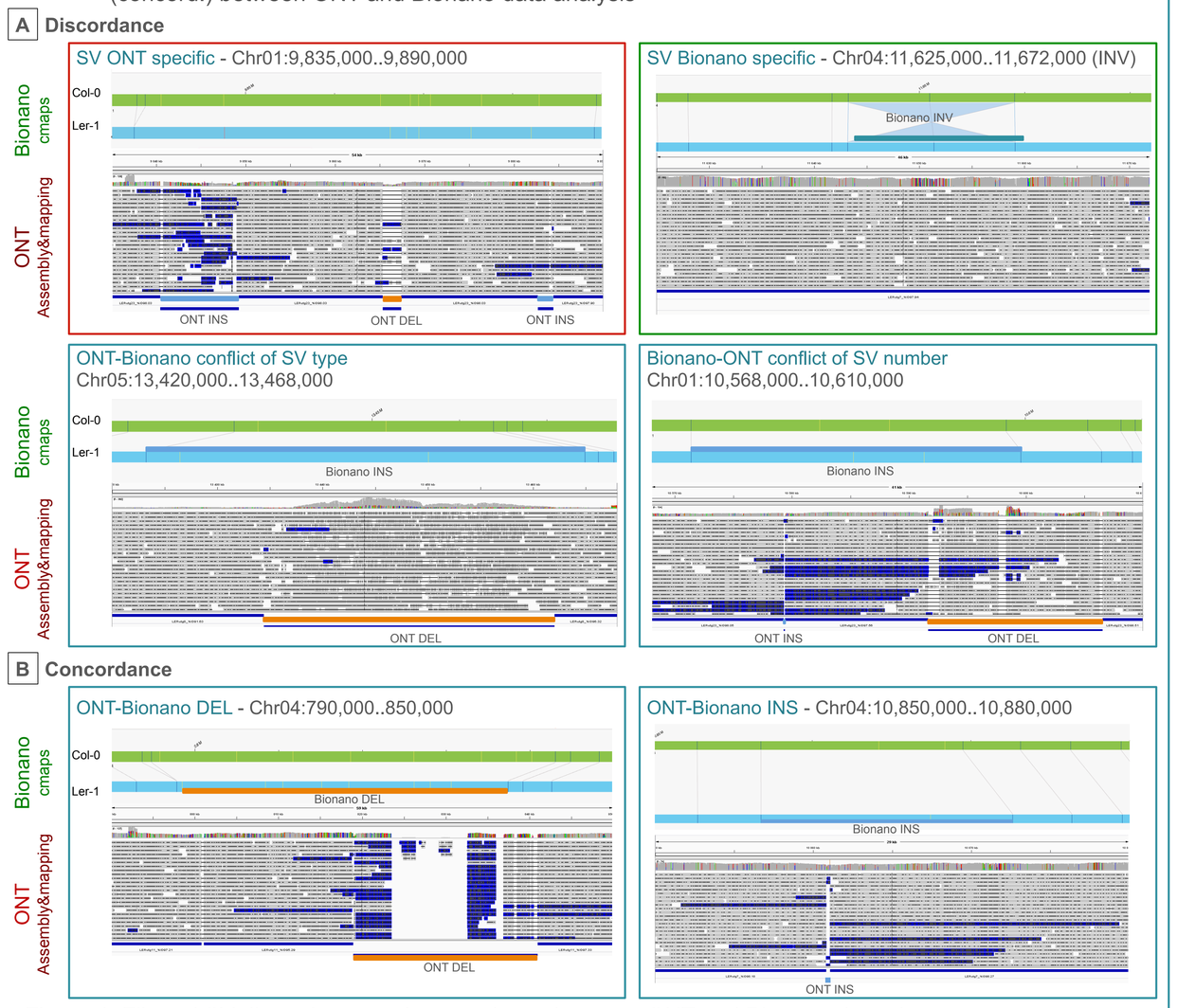


Figure 4. Filtered SV discordance and concordance between ONT and Bionano detection

CONCLUSION

99% of the SV detected were INS and DEL. Absence of duplication could be explained by absence or by elimination of alignments shorter than 10kb in ONT analysis. In Bionano analysis it may be due to the incapacity to catch such SV precision or too high divergence between the duplicated sequences.

ONT data analysis showed around two fold SV than Bionano, more over **the spectra of size is larger**. As expected, **Bionano SV are bigger and fewer**. In our study, **97% of SV detected with Bionano** data are detected with ONT, **more than 83 %** of them are **concordant**. **ONT** and **Bionano SV are respectively at 86% and 100% located around genes and/or TE**. Whereas Bionano did no SV detection in complex and too divergent regions, MUMmer can detect some. **ONT SV are more fragmented and this explains 10 to 14% of conflicts by number between both technologies**. Description and relevance have to be carefully check.

With our draft ONT assembly we recovered 80% of SV published in Zapata et al (2016). When we tested the detection with several sampling of corrected and trimmed sequences, we demonstrated that **the assembly obtained with X20 sequences depth, allowed catching more than 90% of the ONT SV.**

Where Bionano Access is an integrative way from molecules filtering to SV detection, applying quality control based on the coverage and the label density, the ONT SV detection is the combination result of alignment, filtering parameters and detector tool for the description. According to these results, taking into account cost and difficulties to develop standard protocol of HMW plant DNA extraction, ONT technology seems to be more affordable to detect SV to date.

Plant Biology and Breeding Division
F-78026 Versailles, France

http://www.bap.inra.fr/

Poster Online

EPGV group contact
Patricia.Faivre-Rampant@inra.fr