



Sandrine ARRIBAT<sup>1</sup>, William MARANDE<sup>1</sup>, Pierre FRASSE<sup>2</sup>, Pauline HEUILLARD<sup>3</sup>, Olivier BOUCHEZ<sup>3</sup>, Mondher BOUZAYEN<sup>2</sup>, Mohamed ZOUINE<sup>2</sup> and Hélène BERGES<sup>1</sup>

1. INRA/CNRGV - Chemin de Borde Rouge - 31326 Castanet-Tolosan - France. [helene.berges@inra.fr](mailto:helene.berges@inra.fr); 2. Laboratoire Génomique et Biotechnologie du Fruit - GBF, UMR990, INRA/INP-ENSAT - Avenue de l'Agrobiopole - BP 32607 - 31326 Castanet-Tolosan Cedex; 3. INRA - GetPlage - Chemin de Borde Rouge - 31326 Castanet-Tolosan

The French Plant Genomic Resources Center (CNRGV) is dedicated to the analysis of plant genome complexity. Indeed, among living organisms, plants display a high level of genome complexity due to their large size, variations in polyploidy levels and high percentage of repetitive elements. In a context of climate change, population growth and limited energy resources, increasing plant genomes knowledge is essential for a better understanding of mechanisms driving plant adaptation and evolution.

With the long reads sequencing technology, such as Pacific Bioscience that produces reads of several kilobases, the genome quality assembly was largely improved with a considerable contig number reduction and a better resolution of repeated sequences localization. More recently, a new technology of linked reads, named 10X Genomics, was used to create an additional scaffolding that gives a real input to the quality assembly. However, it remains challenging to obtain high quality assemblies at the genome scale.

At the CNRGV, in complement to different NGS technologies, we use whole genome optical maps to improve the genome assembly quality from several plant species. The Optical Mapping technology (ie. Irys system from Bionano Genomics) is based on direct visualization of high molecular weight DNA (from 150kb to 2Mb) labeled on specific sequences patterns. This hybrid scaffolding strategy (NGS and optical maps) could allow the production of genome resolved at the chromosome level. We will present a comparison of results obtained for the tomato genome with a combination of optical maps and either PacBio or 10X genomics sequencing technologies (the results of the combination of the 3 technologies are presented in another poster).

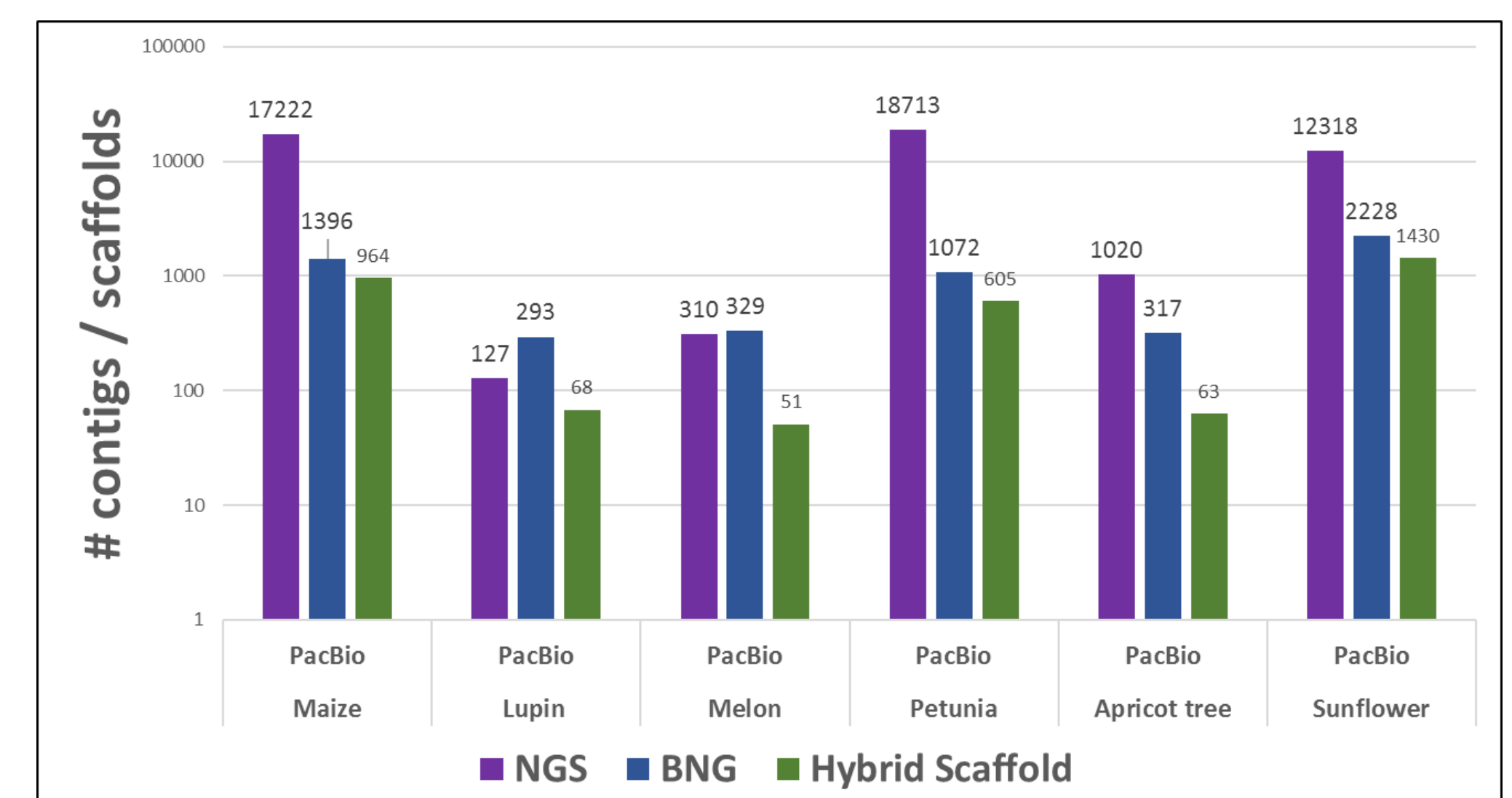
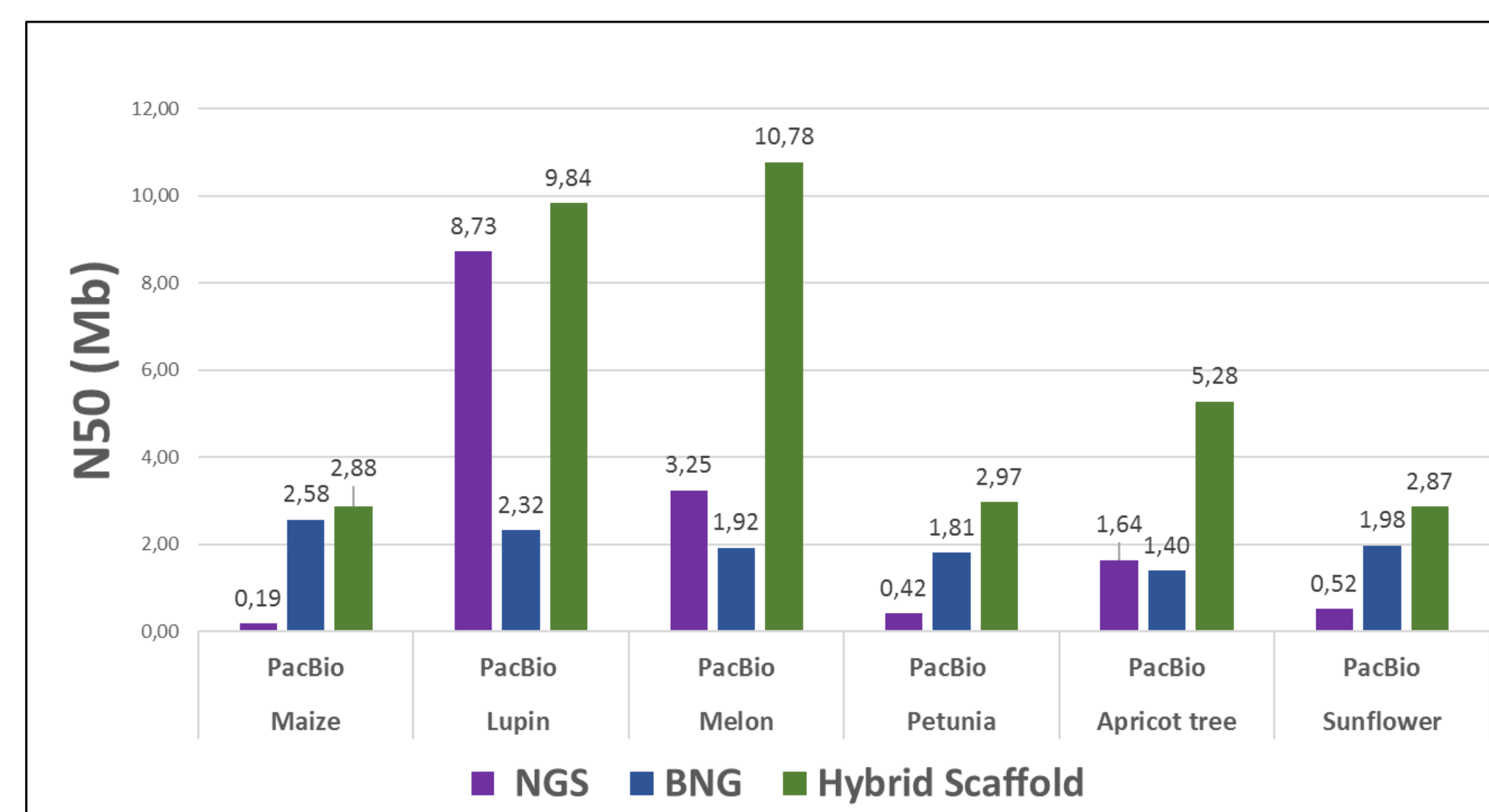
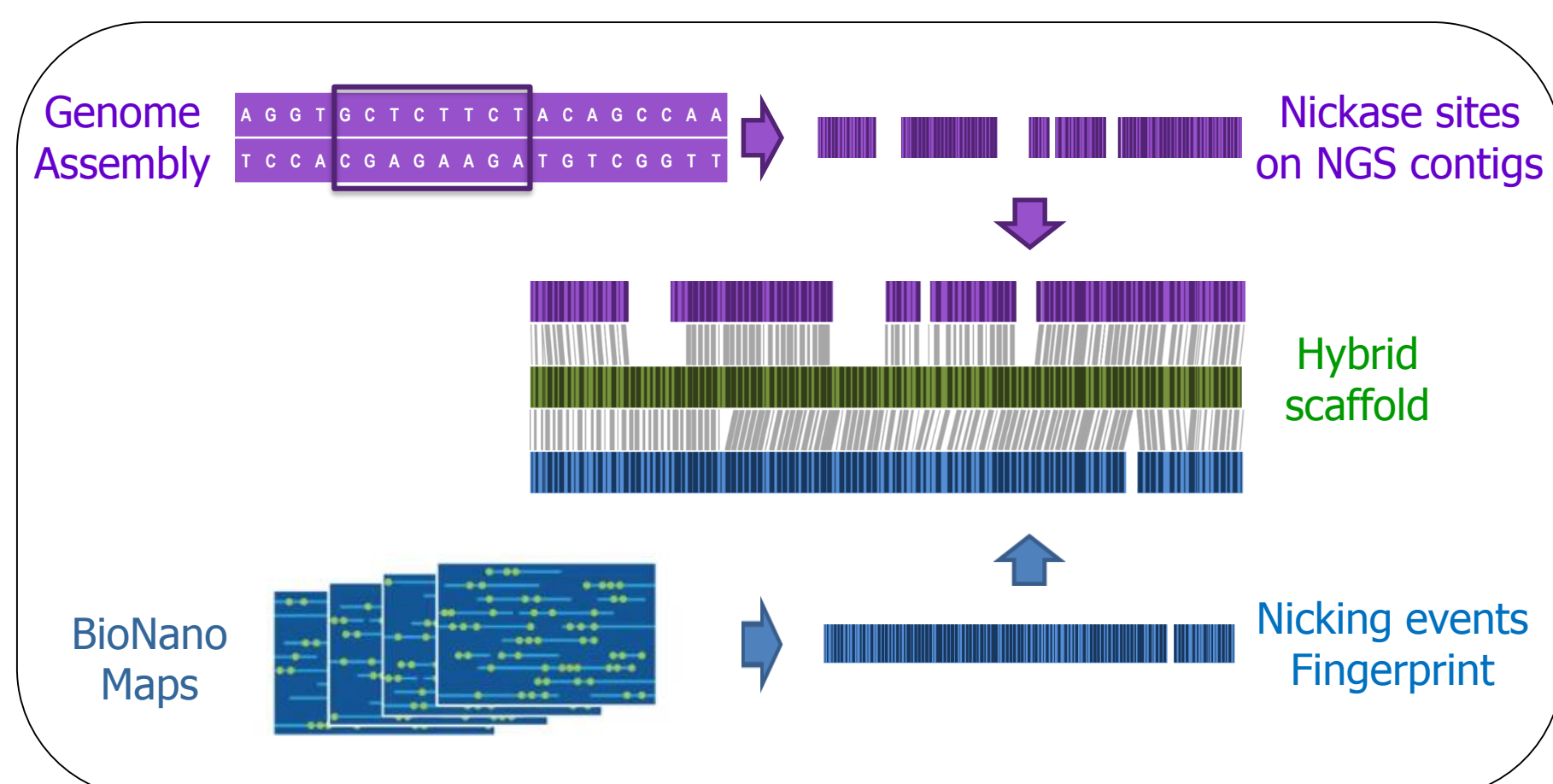
## Building Whole Genome Optical Map

### Irys System Workflow



## Improving genome assembly quality trough Hybrid Scaffolding

Hybrid scaffolding strategy → Genome assembly quality improvement examples



Maize: Amazing Project - Clémentine VITTE ([clementine.vitte@inra.fr](mailto:clementine.vitte@inra.fr)); Lupin: ERC LUPIN ROOTS - Benjamin PERET ([benjamin.peret@cnrs.fr](mailto:benjamin.peret@cnrs.fr)); Melon: Abdelhafid BENDAHMANE ([abdelhafid.bendahmane@inra.fr](mailto:abdelhafid.bendahmane@inra.fr)); Petunia: Michel MOSER ([michel.moser@ips.unibe.ch](mailto:michel.moser@ips.unibe.ch)); Apricot tree: Véronique DECROCOQ ([veronique.decrocoq@inra.fr](mailto:veronique.decrocoq@inra.fr)); Sunflower: SUNRISE - Stéphane MUNOS ([stephane.munos@inra.fr](mailto:stephane.munos@inra.fr))  
 NGS = Next Generation Sequencing; BNG = BioNano Genomics; Hybrid Scaffold = hybrid scaffolding combining sequencing data from Next Generation Sequencing and optical map data from Bionano Genomics

## The tomato genome : optical map with either PacBio or 10X Genomics sequencing technologies

Tomato genome : diploid (2n = 24) - 950 Mb

### Pacific Biosciences

	PacBio Assembly	Optical Map BspQ1	Hybrid Scaffold 1 step	Optical Map BssS1	Hybrid Scaffold 2 step
# contigs / Scaffolds	743	301	74	444	40
Min length (Mb)	0,006	0,220	0,300	0,163	1,005
Median length (Mb)	0,137	1,908	8,029	1,636	15,347
Mean length (Mb)	1,068	2,557	10,674	1,994	20,490
<b>N50 length (Mb)</b>	<b>3,401</b>	<b>3,710</b>	<b>17,674</b>	<b>2,672</b>	<b>33,881</b>
Max length (Mb)	13,628	13,002	40,899	12,226	59,827
Total length (Mb)	793,851	769,748	789,908	885,329	819,580

### 10X Genomics

	10X Assembly	Optical Map BspQ1	Hybrid Scaffold 1 step	Optical Map BssS1	Hybrid Scaffold 2 step
# contigs / Scaffolds	24579	301	116	444	80
Min length (Mb)	0,001	0,220	0,504	0,163	0,395
Median length (Mb)	0,002	1,908	5,095	1,636	7,496
Mean length (Mb)	0,032	2,557	6,646	1,994	10,542
<b>N50 length (Mb)</b>	<b>1,801</b>	<b>3,710</b>	<b>9,869</b>	<b>2,672</b>	<b>17,085</b>
Max length (Mb)	11,337	13,002	25,523	12,226	51,293
Total length (Mb)	795,187	769,748	770,962	885,329	843,341

Tomato: Mohamed ZOUINE ([mohamed.zouine@ensat.fr](mailto:mohamed.zouine@ensat.fr))

10 fold

9,5 fold

Thanks to high quality DNA molecules, optical mapping uses physical reality to link and correct NGS scaffolds. It is now possible to improve assembly quality at the genome level.

