

From the seed to the genome: Providing full solution to deliver high quality genome assembly

Nathalie Rodde, Caroline Callot, Stephane Cauet, Charlotte Cravero, Nicolas Théron, William Marande, Nadine Gautier, Nadège Arnal, Elisa Prat, Anthony Théron, Isabelle Dufau, David Pujol, Margaux-Alison Fustier, Arnaud Bellec and Sonia Vautrin

Relying on a **15 years expertise in plant genomics** and especially in **high quality DNA production** from plants, the CNRGV offers a full solution to deliver high quality genomes assemblies. We are in charge of projects **from seeds** or plants and manage all the steps from the HMW DNA production to the data assembly. We combine complementary technologies such as **PacBio HiFi reads** and **Bionano optical maps**. We have developed **versatile pipelines** to reach **chromosome level assemblies** that give the possibility to integrate data from a reference genome or genetic maps. We provide additional analysis such as annotation and structural variation detection.



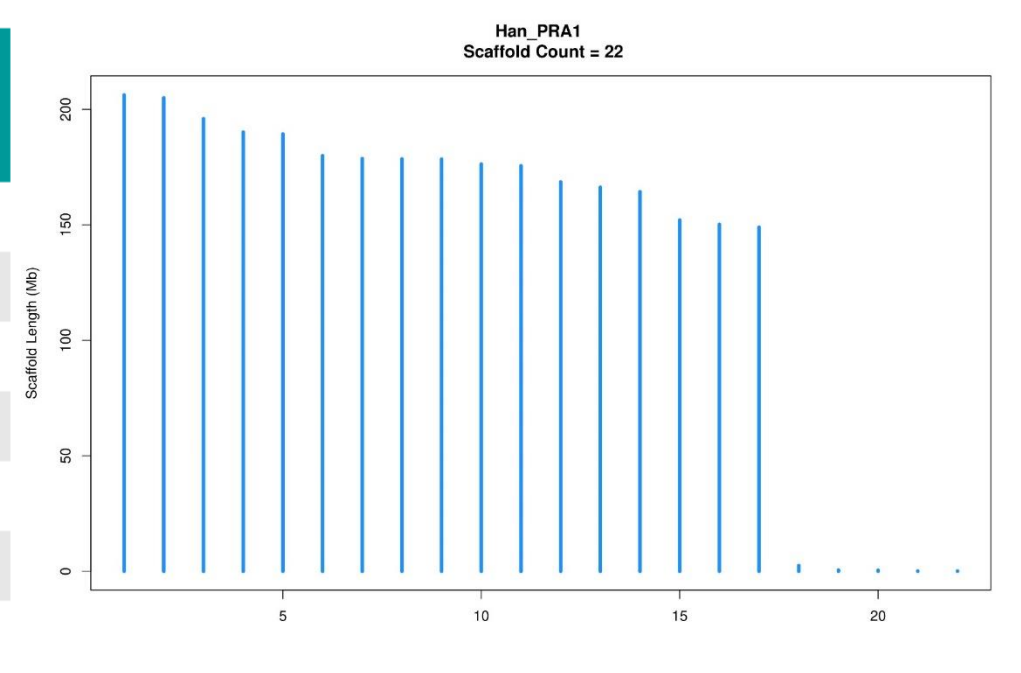
✓ Several protocols adapted to each plant species
 ✓ Fragments size : From 50 to 200kb



✓ Several protocols adapted to each plant species
 ✓ Fragments size : From 100 to 1 500kb

Combining PacBio and Bionano technologies to reach the chromosome level: example of sunflower

	NGS	Optical Map	Hybrid Scaffold	Hybrid Scaffold + NGS not scaffolded
Count	552	43	22	532
N50 length (Mb)	140,549	155,989	178,511	178,511
Max length (Mb)	199,767	214,125	206,253	206,253
Total length (Mb)	3042,466	3127,026	3008,080	3045,767
% of NGS in the hybrid scaffold		98,76		
% of "N" bases in the hybrid scaffold		0,01		
Gaps number in the hybrid scaffold		92		



Helianthus annuus cultivar. Estimated genome size : 3,2Gb - Estimated number of chromosome : 17

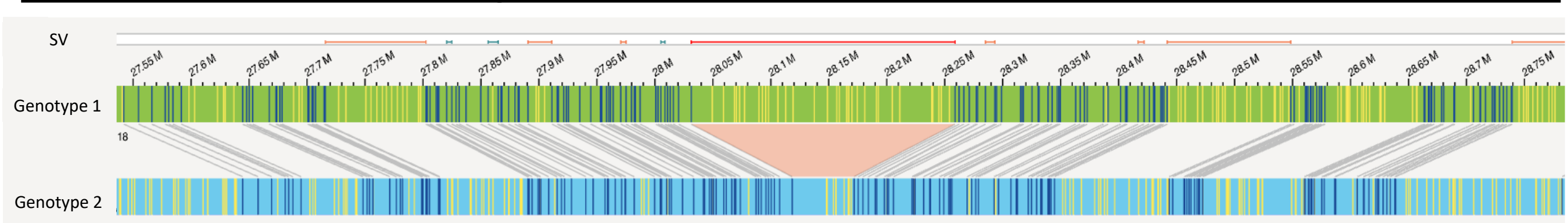
Working on various species

Species	Assemblies PacBio HiFi				Hybrid Scaffolds (HiFi + optical map)		
	Estimated length	Contig nb	Total length (Mb)	N50 (Mb)	Scaffold nb	Total length (Mb)	N50 (Mb)
Eucalyptus*	550 Mb	659	962	8,52	72	1010	30
Barley	5 Gb	4 015	4 351	32,64	111	4 190	105
Rapeseed	1,1 Gb	2 152	1 068	8,4	42	1 005	56
Sunflower	3,2 Gb	592	3 042	34,8	21	3 003	177
Wheat	15 Gb	5055	14657	30	324	14 459	204
Einkorn wheat	5,6 Gb	1142	5149	54	24	5103	522
Rose*	500 Mb	1105	1024	9,5	61	967	26

* Heterozygous species

We have developed several analysis pipelines dedicated to specific complex genomes. E.G. : For the rose sequencing project, we have worked on a haplotype separation method based on mapping and scaffolding with genetic map validation.

Genome data comparison and structural variation detection



Example of SV detection (deletion) : Comparing genome data (assemblies or optical map) allow structural variation detection (indel, inversion) from kb to Mb scale.

Perspective ? Capture method with Xdrop technology (Samplix®)



Targeted sequencing of regions with little prior knowledge on large panels

Prerequisite :

- Ultra specific PCR markers every 100kb max
- Design on conserved sequences among genotypes to allow diversity analysis

On-going development :

- First capture projects on sunflower and olive tree
- Bioinformatics expertise : adaptation of existing pipeline for chimera identification and elimination



24 Chemin de Borde Rouge
 Auzeville - CS 52627
 31326 Castanet tolosan cedex
 FRANCE
 Tél : 05.61.28.52.53
<https://cnrgv.toulouse.inrae.fr>



<https://twitter.com/CNRGV>