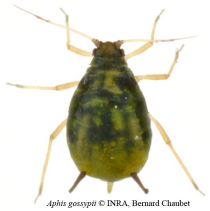
Hybrid de novo genome assembly using Oxford Nanopore Technology, 10X Genomics Linked-Reads sequencing, and Bionano optical map of a non-model species: the case of the melon-cotton aphid Aphis gossypii

Jacques Lagnel, Rafael Feriche-Linares, Pierre Bertrand, William Marande, Anne Loiseau, Amalia Sayeh, Maxime Manno, Nathalie Boissot

Hybrid *de novo* genome assembly using Oxford Nanopore Technology, 10X Genomics Linked-Reads sequencing, and Bionano optical map of a non-model species: the case of the melon-cotton aphid *Aphis gossypii*


Aphis gossypii © INRA, Bernard Chaubet

Jacques LAGNEL[1], Rafael Feriche-Linares[1], Pierre BERTRAND[1],
William MARANDE[2], Anne LOISEAU[3], Amalia SAYEH[4], Maxime MANNO[4] and Nathalie BOISSOT[1]

[1]INRAE, GAFL, 84140 Avignon, France; [2]INRAE, CNRGV 31326 Castanet Tolosan, France;
[3] INRAE, CBGP 34988 Montferrier sur Lez, France; [4] GeT-PlaGe, 31100 Toulouse, France

Context

Insect species challenge high quality genome production because of their small size, the combination of high polymorphism and heterozygosity, the presence of repeat regions, and the pooling of polymorphic individuals to form libraries. No highly resolved genome is available for aphids which are major pests of cultivated plants.

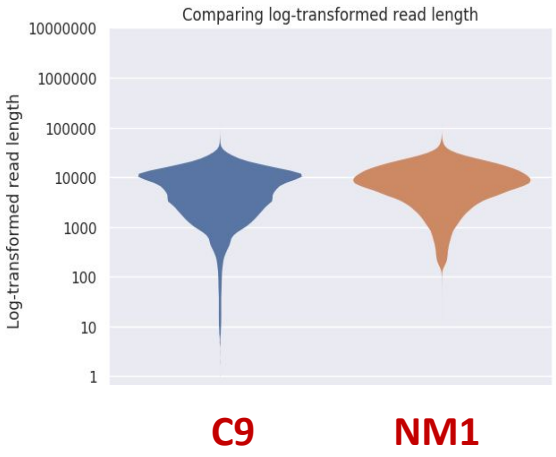
Aphis gossypii genome is expected with 339 Mbp in 4 pairs of chromosomes. To minimize heterozygosity, we focused on a lineage, **C9**, with a low estimated heterozygosity, and we pooled individuals deriving from clonal reproduction. We produced a high quality genome from **10X Genomics Linked-Reads (10X Genomics)**, **long reads (Oxford Nanopore Technology: ONT)**, and **Optical Map (Bionano)**, in order to compare advantages of combining different methods. We also produced a genome of a more heterozygous clone, **NM1**, but without an optical map support.

Raw Data

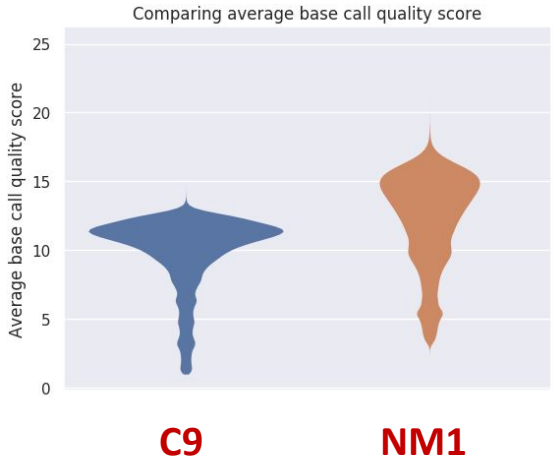
- 3 flow cells MinIon (ONT) for clones C9 et NM1.

Metrics ONT	C9	NM1
Mean read length	7.4 Kb	9.3 Kb
Mean read quality	9.9	12.3
Median read length	5.4 Kb	7.5 Kb
Median read quality	10.8	13.1
Number of reads	2,629,764	3,009,026
Read length N50	12.1 Kb	13.7 Kb
Total bases	19.33 Gb	28 Gb
Coverage	57X	82X

ONT Read length



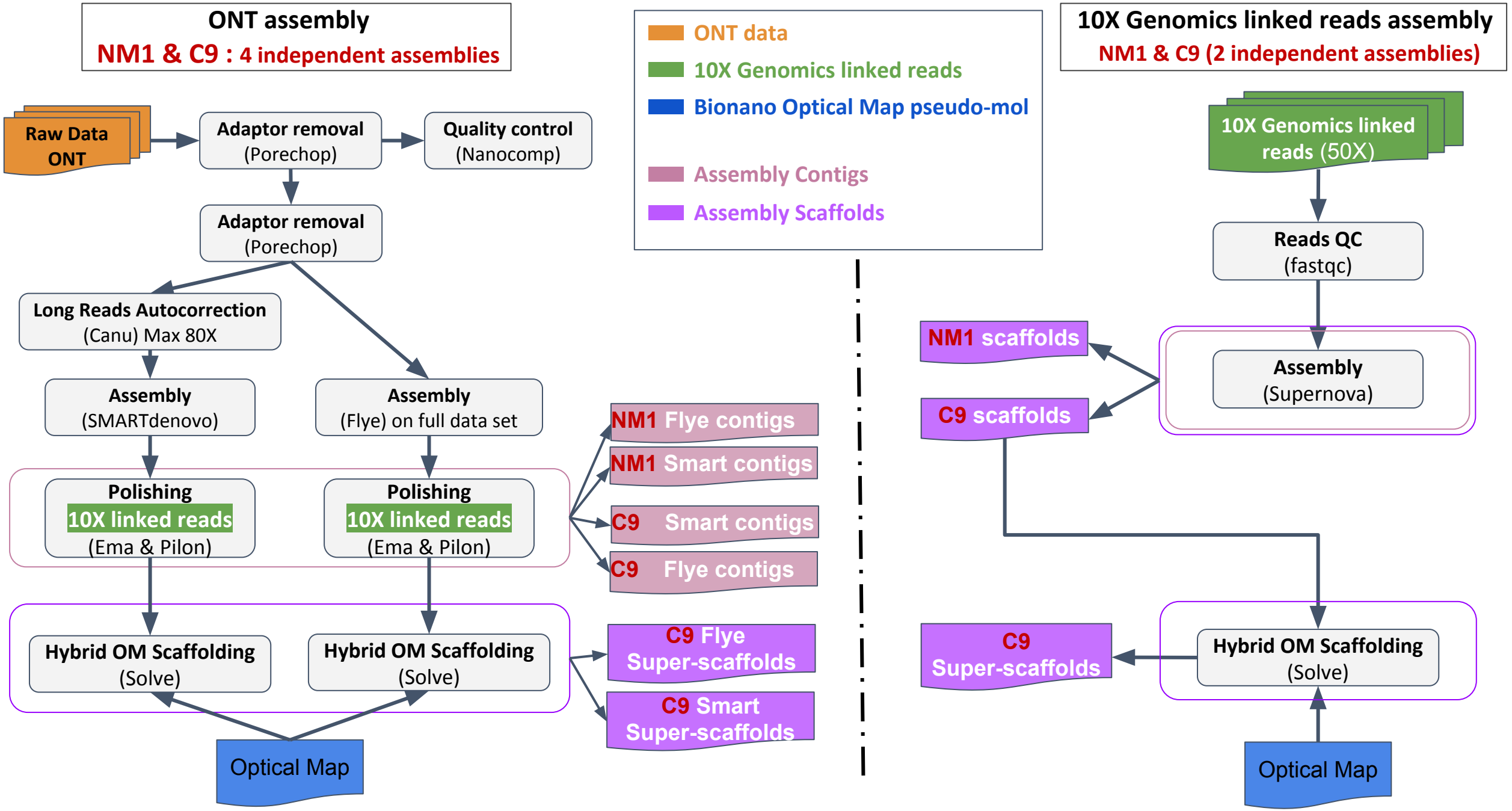
ONT Average base call quality score



- 10X Genomics linked reads sequencing for clones NM1 et C9. Give about 50X and total of 17 Gbp per clone

- Optical Mapping (OMAP) for clone C9 => Assembly (solve) 55 pseudo molecules, 424.9 Mbp et N50= 42.8 Mbp

Assembly process



Results : Assembly comparison

	ONT				ONT + OM		10X		10X+OM	Illumina
	Canu & SMARTdenovo		Flye		Solve		Supernova		Solve	Allpaths-lg
Metrics	C9	NM1	C9	NM1	C9 Canu & Smartdenovo	C9 Flye	C9	NM1	C9 Supernova	Draft Quan et al, 2019
Contigs	**1,022**	**848**	1,514	1,917						22,569
Scaffolds					23	27	13,211	15,149	25	4,718
Largest contigs (Mb)	11.68	14.67	11.82	6.85						
Largest scaffold (Mb)					91.70	90.0	4.16	14.20	88.57	5.5
N50 contigs (Mb)/L50	1.83/54	1.98/50	1.93/40	0.99/94						
N50 scaffold (Mb)/L50					50.2/3	49.3/3	0.78/120	2.61/34	49.22/3	0.44/195
Assembly size (Mb)	386	395.1	382.4	398	389	384.2	334	333.4	298	294
Busco3 insecta odb9 (%)										
Complete genes	93.8	94.0	93.2	94.1	93.8	92.9	92.5	93.5	90.0	91.0
Duplicated genes	2.4	4.1	2.8	3.3	2.4	2.3	3.1	2.7	2.2	4
Fragmented genes	1.7	1.3	2.1	1.6	1.7	2.1	2.2	2.1	1.7	3
Missing genes	4.5	4.7	4.7	4.3	4.5	5	5.3	4.4	8.3	6

Conclusion

- **The sequencing technology that gave the less congruent results with the optical map was the 10X Genomics link-reads.**

- **Sequencing using Oxford Nanopore technology was quality / cost effective and time effective for the assembly of *Aphis gossypii.* Both Flye and Canu/SMARTdenovo assemblies gave coherent contigs with optical map (Bionano), with an advantage for Canu/SMARTdenovo in terms of contiguity.**
  The combination of the contigs from ONT assemblies by the two methods with the Optical map built 27-23 super-scaffolds. Most Scaffolds produced, one hand by Flye, and on the other hand by Canu/smartdenovo, mainly matched by pairs. Furthermore, we didn't observe inversion or miss-assembly between pairs.

- **The longest scaffold, about 90 Mb, probably covered one full chromosome among the 4 present in *A. gossypii.* One scaffold was identified to be related to the endosymbiont bacteria *Buchnera*.**

- **The best assemblies (about 390 Mb) suggested that the *A. gossypii* genome size could be greater than expected (339 Mb).**

- **We are currently investigating the use of the 10X Genomics linked-reads and the optical map (Bionano) to produce a phased genome.**

Aphis gossypii © INRA, Bernard Chaubet