



**HAL**  
open science

## A phase-resolved, chromosome-scale de novo assembly of the diploid 'Regina' sweet cherry genome.

Loïck Le Dantec, Nabil Girollet, Jérôme Gouzy, Erika Sallet, Armel Donkpegan, Shawn Sullivan, Hayley Mangelson, Kyle Langford, Mathieu Fouché, Jose Quero-Garcia, et al.

### ► To cite this version:

Loïck Le Dantec, Nabil Girollet, Jérôme Gouzy, Erika Sallet, Armel Donkpegan, et al.. A phase-resolved, chromosome-scale de novo assembly of the diploid 'Regina' sweet cherry genome.. 10th International Rosaceae Genomics Conference (RGC 10), Dec 2020, Barcelone, Spain. hal-03318885

**HAL Id: hal-03318885**

**<https://hal.science/hal-03318885>**

Submitted on 11 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# A PHASE-RESOLVED, CHROMOSOME-SCALE DE NOVO ASSEMBLY OF THE DIPLOID

## 'REGINA' SWEET CHERRY GENOME

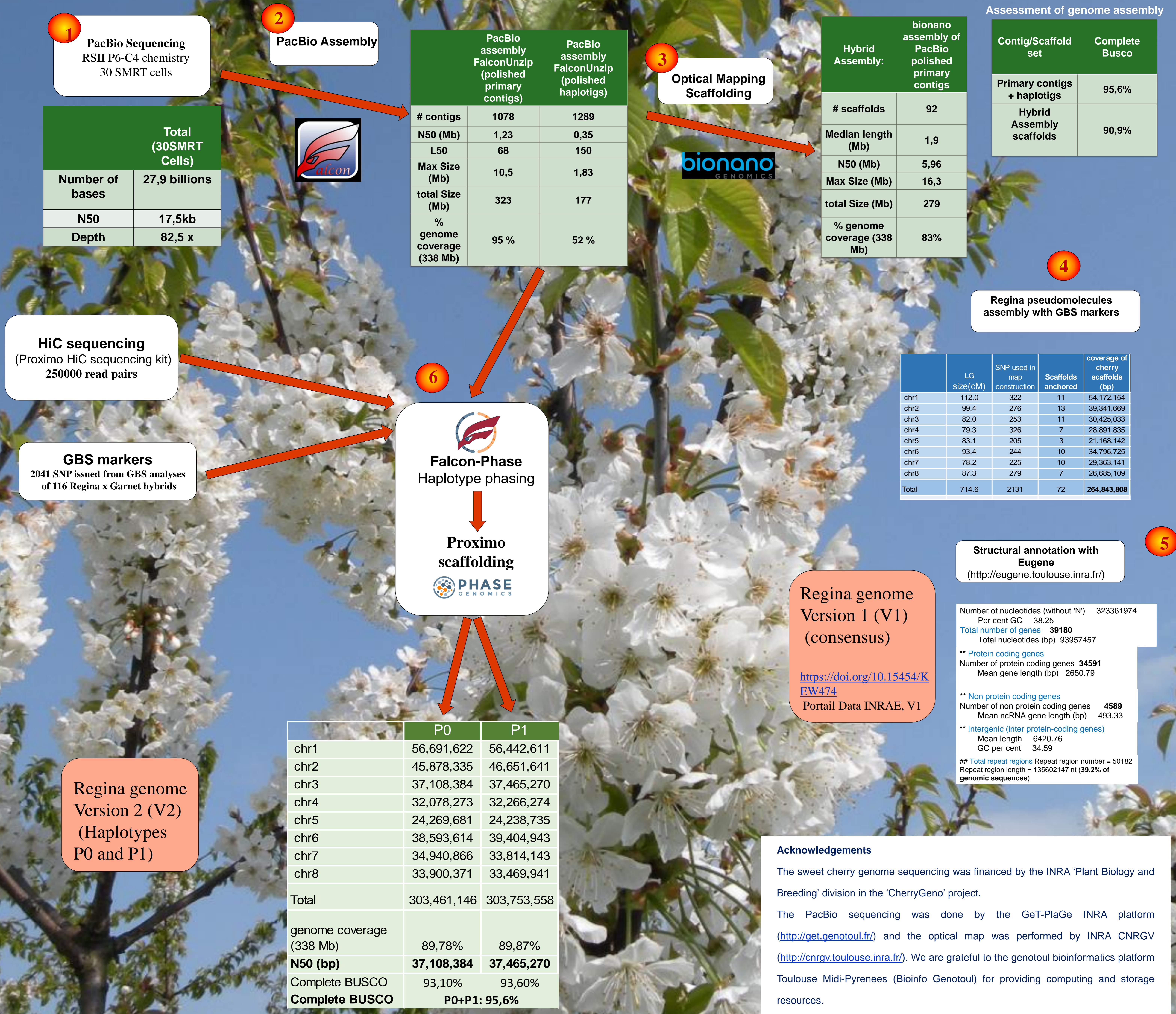
Loïck Le Dantec<sup>1\*</sup>, Nabil Girollet<sup>2</sup>, Jérôme Gouzy<sup>3</sup>, Erika Sallet<sup>3</sup>, Armel Donkpegan<sup>1</sup>, Shawn Sullivan<sup>4</sup>, Hayley Mangelson<sup>4</sup>, Kyle Langford<sup>4</sup>, Mathieu Fouché<sup>1</sup>, Jose Quero-Garcia<sup>1</sup>, and Elisabeth Dirlwanger<sup>1</sup>

<sup>1</sup>INRA-UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon, France, <sup>2</sup>UMR EGFV 1287 – Ecophysiologie et Génomique Fonctionnelle de la Vigne, Villenave d'Ornon, France, <sup>3</sup>Laboratoire des Interactions Plantes Micro-organismes (LIPM) -INRA/CNRS, Castanet-Tolosan, France, <sup>4</sup>Phase Genomics Inc, Seattle, WA, United States

\*Please send correspondence to loick.le-dantec@inrae.fr

### SUMMARY

Sweet cherry (*Prunus avium* L.) is a diploid species with an estimated genome size of 338 MB (Arumuganathan et al. 1991). It is mostly self-incompatible, and therefore has a heterozygous genetic background. We have sequenced the genome of the 'Regina' sweet cherry variety, which is a late blooming cultivar. Here we present whole-chromosomes assembly and phasing using a combination of long read sequencing, proximity ligation (HiC) data and GBS markers.



### CONCLUSION

Regina PacBio long reads were de novo assembled using FALCON UNZIP. A first assembly using optical mapping and high density genetic linkage maps with GBS data resulted in a genome of 264,8 Mb (78,4 % of estimated genome size) distributed on 8 linkage groups with a scaffold N50 ~6Mb (Regina V1). A new assembly (ReginaV2) combining Falcon Unzip assembly of PacBio long reads, proximity ligation (HiC) data and GBS markers has been done with Falcon Phased and the Proximo pipeline of Phase Genomics. It produced complete, fully phased, diploid chromosome scale scaffolds with a genome coverage of ~90%, a N50 of ~37 Mb and a complete BUSCO score of 95,6%. Structural and functional annotations of Regina V2 are in progress. Obtaining these phased chromosomes will allow very detailed studies of the allelic effects on the traits studied, in particular by making it possible to study the allelic expressions of candidate genes.