The complete genome sequence of the parasitic weed *Orobanche cumana* (sunflower broomrape)

Nicolas Pouilly[1], Jérôme Gouzy[1], Marie-Claude Boniface[1], Olivier Bouchez[2], Sébastien Carrère[1], Olivier Catrice[1], Stéphane Cauet[3], Clotilde Claudel[4], Ludovic Cottret[1], Sébastien Faure[4], Álvaro Calderón González[5], Xavier Grand[1], Luyang Hu[1], Céline Jéziorski[3], Marc-Marie Lechat[6], Ludovic Legrand[1], Johann Louarn[1], William Marande[3], Nicolas Ribière[4], Erika Sallet[1], Philippe Simier[6], Leonardo Velasco[5], Cécile Donnadieu[2], Christophe Jestin[7], Philippe Delavault[6], Hélène Bergès[3], Marie Coque[4], Begoña Pérez-Vich[5] and Stéphane Munos[1] *

(1) Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France, (2) INRA GeT Genomics Facility, Castanet-Tolosan, France, (3) INRA - CNRGV, Castanet-Tolosan, France, (4) BIOGEMMA, Mondonville, France,

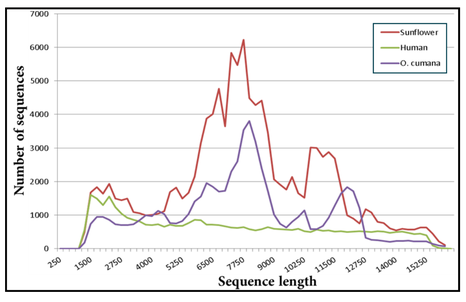
(5) Instituto de Agricultura Sostenible - CSIC, Córdoba, Spain, (6) Laboratoire de Biologie et Pathologie Végétales, SFR 4207 QUASAV, Université de Nantes, Nantes, France, (7) Terres Inovia, Thiverval-Grignon, France

* Corresponding author: stephane.munos@inra.fr

Abstract *Orobanche cumana* Wallr. is an obligatory and non-photosynthetic root parasitic plant of the sunflower crop, causing important yield losses on infested fields. Located in Earthen Europe, Spain and Asia, this parasitic weed can rapidly spread to new areas and his emergence has been observed in France since 2007 (Jestin *et al.*, 2014). In sunflower, breeding for resistance was mainly based on single major resistance genes. New more virulent races of *O. cumana* appeared, leading to a breakdown of resistance genes. A better understanding of the mechanisms involved in the interaction between sunflower and *O. cumana* may improve sustainability of the resistance by using resistance genes acting at different steps of the life cycle of the parasitic plant.

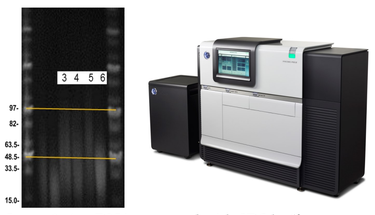
Combining long read sequencing, optical mapping, SNP-based genetic mapping and RNA-seq expression analysis, we have produced a first version of the 1.42 Gb genome sequence of *O. cumana* (2n=38) (Schneeweiss *et al.*, 2004; Weiss-Schneeweiss *et al.*, 2006). Our de novo strategy resulted in an assembly of 1.40 Gb, constituted by 622 scaffolds with a N50 of 5.9 Mb, provided to the public research community through a Web Genome Browser. We aim to obtain the sequences of the pseudomolecules through an improved genetic map thanks to polymorphism located on the scaffolds, identified by whole genome re-sequencing of the parental lines of a F2 segregating population. The genome sequence of *O. cumana* will contribute to the characterization of its physiology and development and in the understanding of the host-parasite interactions. This release should allow identifying avirulence genes, as putative interactor with sunflower proteins, and considering the identification of new resistance genes in sunflower.

Assembly of the genome



IN-23

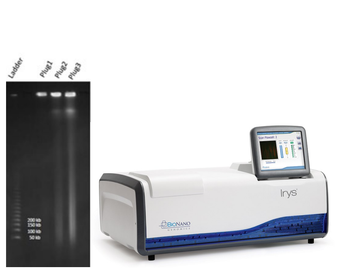
Long Read Sequencing



Lanes 3, 4 : gDNA extracted with SDS buffer (Mayjonade *et al.*, 2016); lane 5 : gDNA extracted with CTAB/PEG buffer (Stadermann *et al.*, 2015) ; lane 6: sheared gDNA (40kb, Megaruptor®Diagenode)

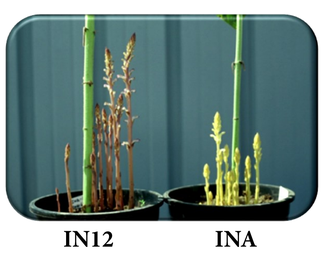
88X coverage
126 SMRT Cells

Optical Mapping



370X coverage
N50=143kb

From contigs to chromosomes

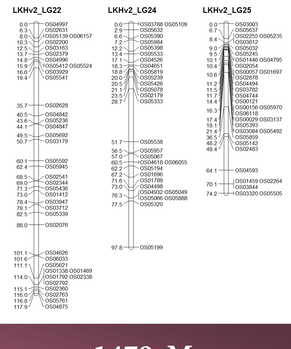


IN12 INA

***O. cumana* segregating population**

IN12 X INA
⇩
F1
⇩
91F2

Genetic Mapping



1479cM,
28 linkage groups

Annotation of the genome

Transcriptomic

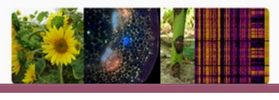
Steps	NUM	MAX (Mp)	N50 (Mp)	MEAN	Total (Gb)
Raw data (subreads)	13.2M	85.05			149.9
Corrected reads (CANU)	7.04M	55.53	13.98	10651bp	75.06
Genome assembly (CANU)	905	16.88	3.57	1.53Mb	1.388
Remove spurious + Sequence based scaffolding + polishing (QUIVER)	793	16.98	4.21	1.74Mb	1.380
+ Optical map scaffolding (Bionano hybrid scaffolding)	**622**	**22**	**5.92**	**2.25Mb**	**1.40 (*1.5% of N*)**

- **256 contigs representing 90% of the genome assembly**
- **95 contigs (593Mb) anchored thanks to the genetic map**
- **Improvement of the genetic map :**
  - **Re-Sequencing for IN12 and INA parental lines**
  - **13511SNPs identified on 145 unmapped contigs**
  - **Genotyping of 278 SNPs located on the 145 contigs (*in progress*)**

EuGene Plant pipeline annotation	
Total number of genes	55726
Number of protein coding genes	46447
Mean gene length (bp)	3569
Number of non protein coding genes	9279

Genome Browser

INRA Sunflower Bioinformatics Resources



https://www.heliagene.org

Sequencing High Molecular Weight genomic DNA (HMW gDNA) was obtained from IN-23, a highly homozygous broomrape race F line. HMW gDNA was isolated according to the protocol developped by Mayjonade *et al.* 2016. Libraries were produced from 100µg of HMW gDNA then sequenced with P6-C4 chemistry and 360min movie times (PacBio RSII, Pacific Biosystems).

Scaffolding An optical mapping (Irys System, BioNano Genomics) was established by a BspQQ1 nicking enzyme digestion of gDNA (estimation of 11,5 labels per 100kb).

Bioinformatic Correction of raw sequences and assembling were realized with CANU software. Consensus contigs sequences were polished with QUIVER.

Genetic Mapping A set of 1536 SNPs maximizing the diversity among 12 populations of O. cumana was defined by an exome capture (Biogemma). A segregating population (n=91F2) from parental lines IN12 and INA was genotyped with these 1536 SNPs and an additional set of 168SSR (Pineda-Martos *et al.* 2014). The genetic map including 509 SNPs + 18 SSR was building using CarthaGène software (INRA) with a high stringency.

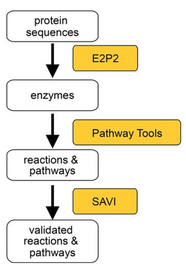
Re-Sequencing IN12 & INA lines were sequenced using 2x100nt HiSeq sequencing (Illumina). 40 912 polymorphic SNPs were identified between these two parental lines.

Transcriptomic 20 broomrape development stages from seeds to flowering, 3 replicates per stage, were used to construct 60 RNASeq libraries. cDNAs were sequenced using 2x100nt HiSeq sequencing (Illumina). Annotation was completed using of an automatic annotation EuGene Plant pipeline (Foissac *et al.* 2008).

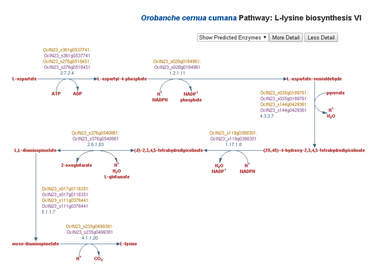
A first *de novo* genome assembly of a parasitic plant.

A first genetic map of *Orobanche cumana* Wallr.

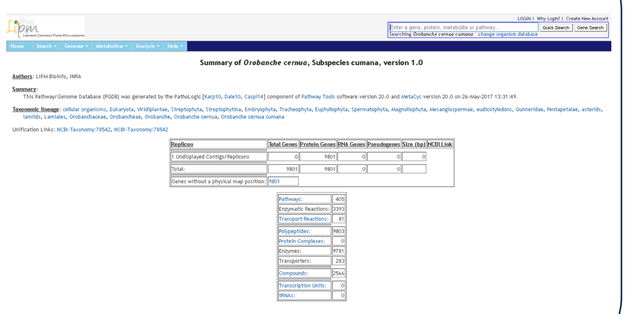
From this genomic resource, we start to produce genetic and biological information



Biochemical information



Link between pathways, reactions, enzymes and genes



Metabolic database

Cited litterature

- Foissac S., Gouzy J., Rombauts S., Mathé C., Amselem J., Sterck L., Van de Peer Y., Rouzé P. & Schiex T. (2008). Genome annotation in plants and fungi: EuGene as a model platform. Current Bioinformatics, 3(2), 87-97.
- Ještin C., Lecomte V., Duroueix F. (2014). Current situation of sunflower broomrape in France. Third International Symposium On Broomrape (Orobanche ŝpp.) In Sunflower, Cordoba, Spain, June 3rd to 6th, 2014.
- Koren S., Walenz B.P., Berlin K., Miller J.R., Bergman N.H. & Phillippy A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome research, 27(5), 722-736.
- Mayjonade B., Gouzy J., Donnadieu C., Pouilly N., Marande W., Callot C., Langlade N. & Muños S.(2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. BioTechniques, 61(4), 203-205.
- Pineda-Martos R., Velasco L., & Pérez-Vich B. (2014). Identification, characterisation and discriminatory power of microsatellite markers in the parasitic weed Orobanche cumana. Weed research, 54(2), 120-132.
- Schneeweiss G.M., Palomeque T., Colwell A.E., Weiss -Schneeweiss H. (2004). Chromosome numbers and karyotype evolution in holoparasitic Orobanche (Orobanchaceae) and related genera. American Journal of Botany, 91: 439-448.
- Weiss-Schneeweiss H., Greilhuber J., Schneeweiss G.M. 2006. Genome size evolution in holoparasitic Orobanche (Orobanchaceae) and related genera. *American Journal of Botany*, 93: 148-156.

INRA Centre Occitanie-Toulouse
Laboratoire des Interactions Plantes Micro-organismes
UMR CNRS-INRA 2594/441
24 Chemin de Borde Rouge - Auzeville CS 52627
31326 CASTANET TOLOSAN Cedex FRANCE
http://www6.toulouse.inra.fr/lipm_eng/Research/Sunflower-Genetics-and-Genomics

PAGXXVI-P1166