

From seed to genome: Delivering a Comprehensive Solution for High-Quality Genome Assembly

Anthony Theron, Caroline Callot, Stephane Cauet, Charlotte Cravero, William Marande, Nathalie Rodde, Robin Schwartz, Nadine Gautier, Nadège Arnal, Isabelle Dufau, David Pujol, Sonia Vautrin, Arnaud Bellec

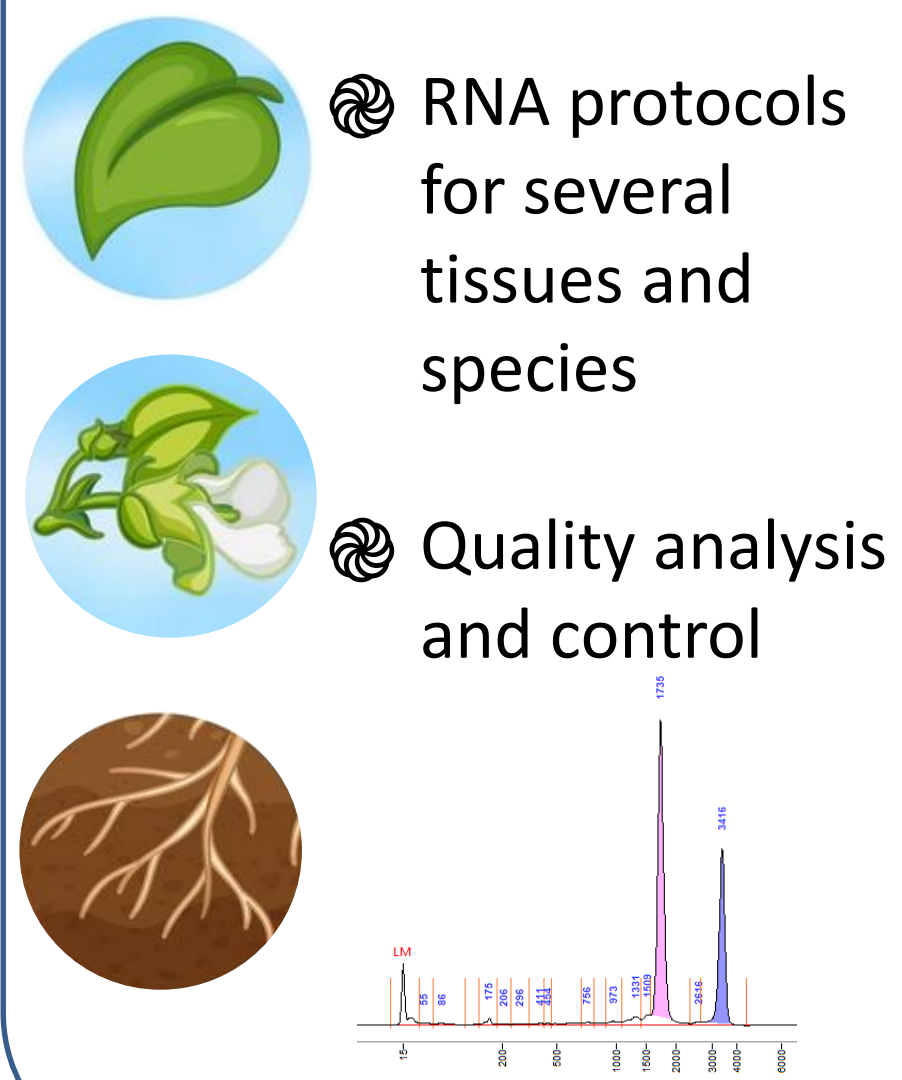


The french genomic center (CNRGV), specializes in genomics services, supporting both research teams and private enterprises in the characterization of plant diversity and the connection of genotype to phenotype through genome analysis. Depending on the specific objectives, we assist our partners in either comprehensive genome studies or towards approaches targeted on regions of interest. With over 15 years of specialized experience in plant genomics, with a particular emphasis on the high-quality extraction of DNA from plants, CNRGV provides a comprehensive solution for delivering top-notch genome assemblies. Our responsibility extends to managing projects from the initial stages, starting with seeds or plants, and overseeing every step of the process, from the production of High Molecular Weight DNA to the final data assembly. Our approach is characterized by the integration of complementary technologies, including PacBio HiFi reads with the brand new REVI0 system, Bionano optical map, and Hi-C data. Our adaptable pipelines allows us to deliver a full annotated chromosome-level assembly.

Collaborators

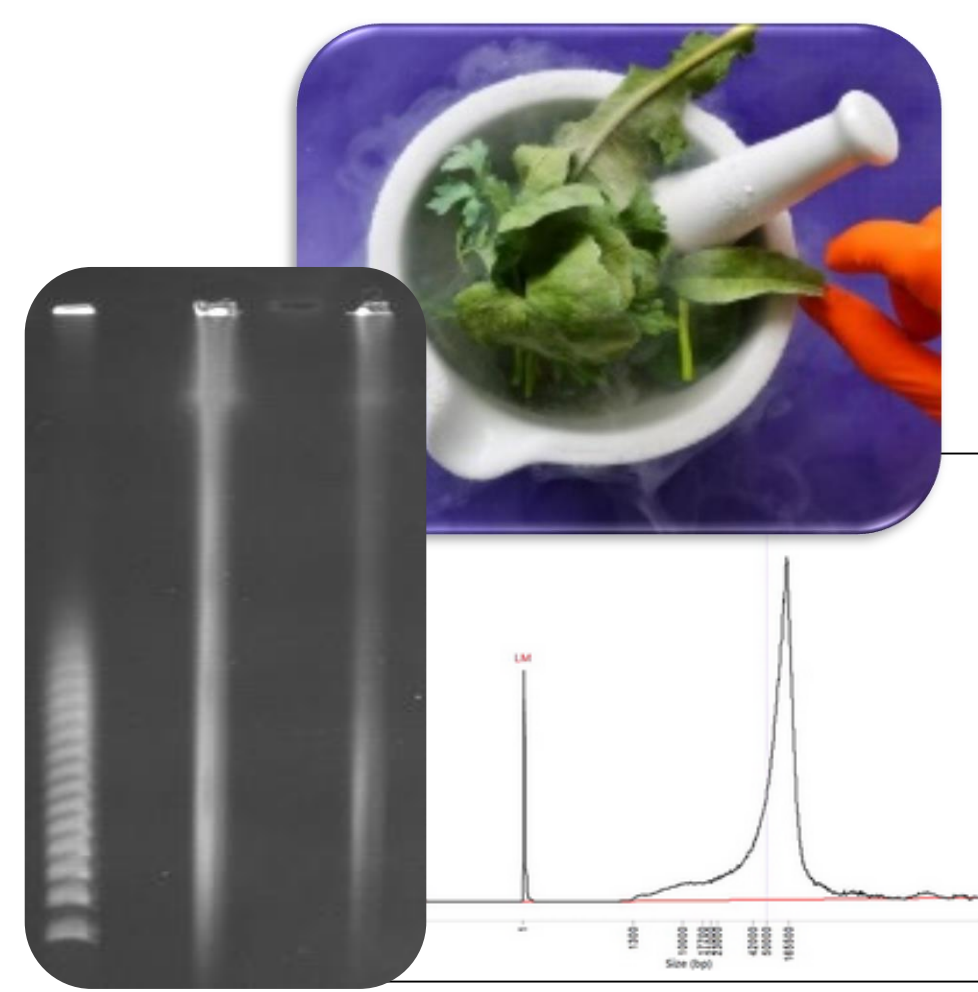


RNA extraction



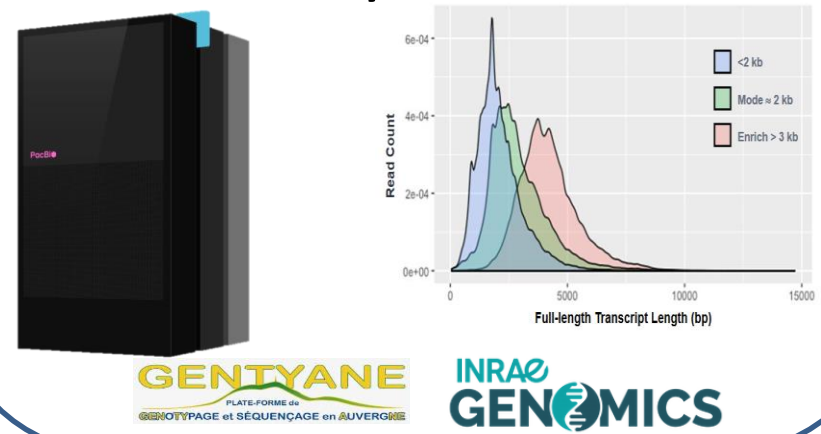
HMW & uHMW DNA extraction

- Robust and flexible DNA extraction protocols for several species
- Quality analysis and control
- DNA fragment size form 100kb to 2,5Mb



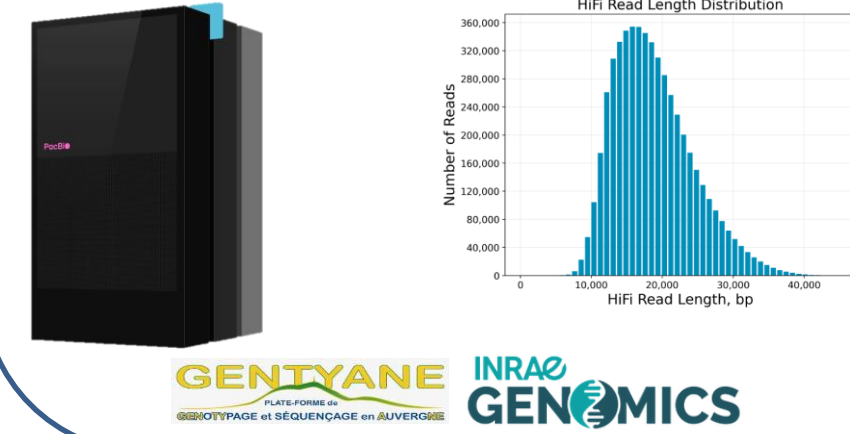
Iso-Seq & Kinnex

- Yield : Up to 40M reads/SMRT cell on REVI0 system
- Full-length cDNAs → No need of transcript assembly
- Characterize alternative splicing event → Isoform discovery



HiFi Library Prep

- Yield : Up to 110 Gb CCS 3 passes/SMRT cell on REVI0 system
- Long read lengths up to 25 kb
- Methylation included (5mC)
- High read accuracy >99.9%

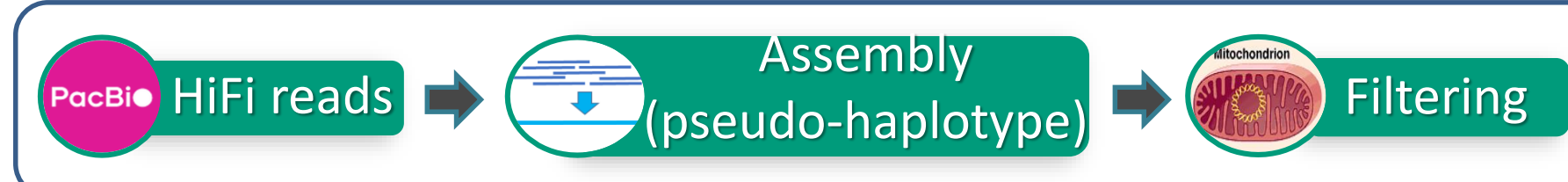


Optical mapping

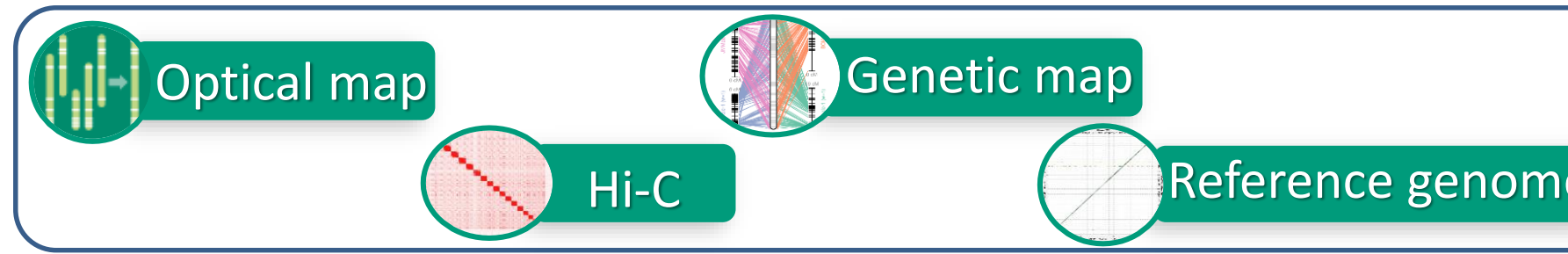
- Yield : Up to 2,5Tb /Flowcell on Saphyr system
- 3 Flowcells /chip
- N50 up to 300 kb



Assembly



Scaffolding & validation



Annotation



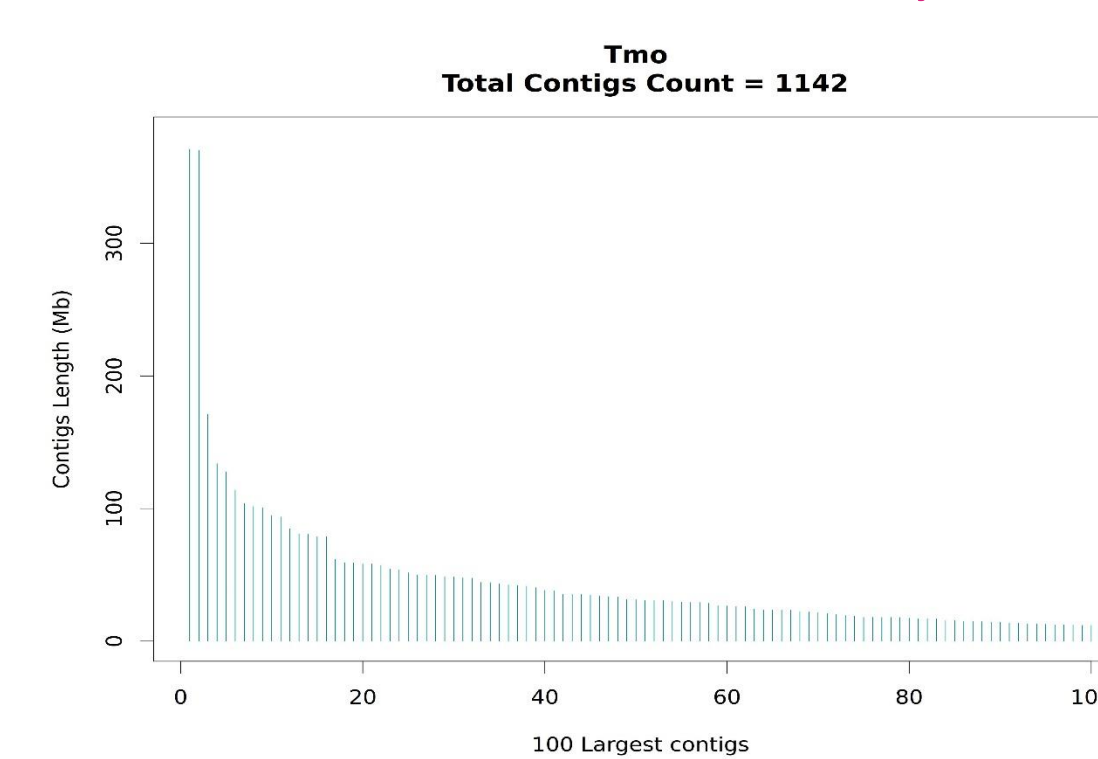
Evaluation / QC

- K-mer analysis : genome size estimation, QV, completeness, heterozygosity, duplication
- Contigs filtering : organelle, low quality
- Assembly metric, Busco score and reporting



Chromosome size assembly combining Hifi reads and optical maps: The example of Einkorn Wheat (*Triticum Monococcum*)

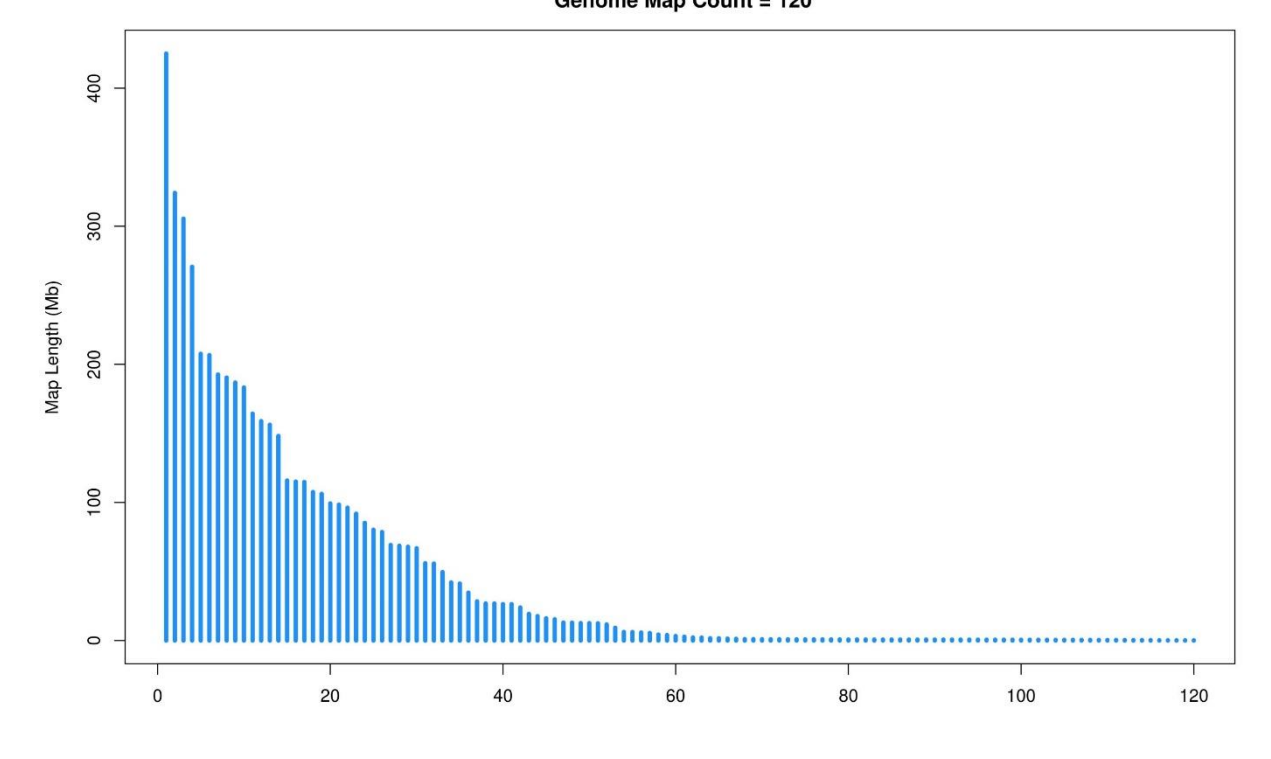
Pacbio Hifi Sequencing



Sequence assembly (hifiasm v15)

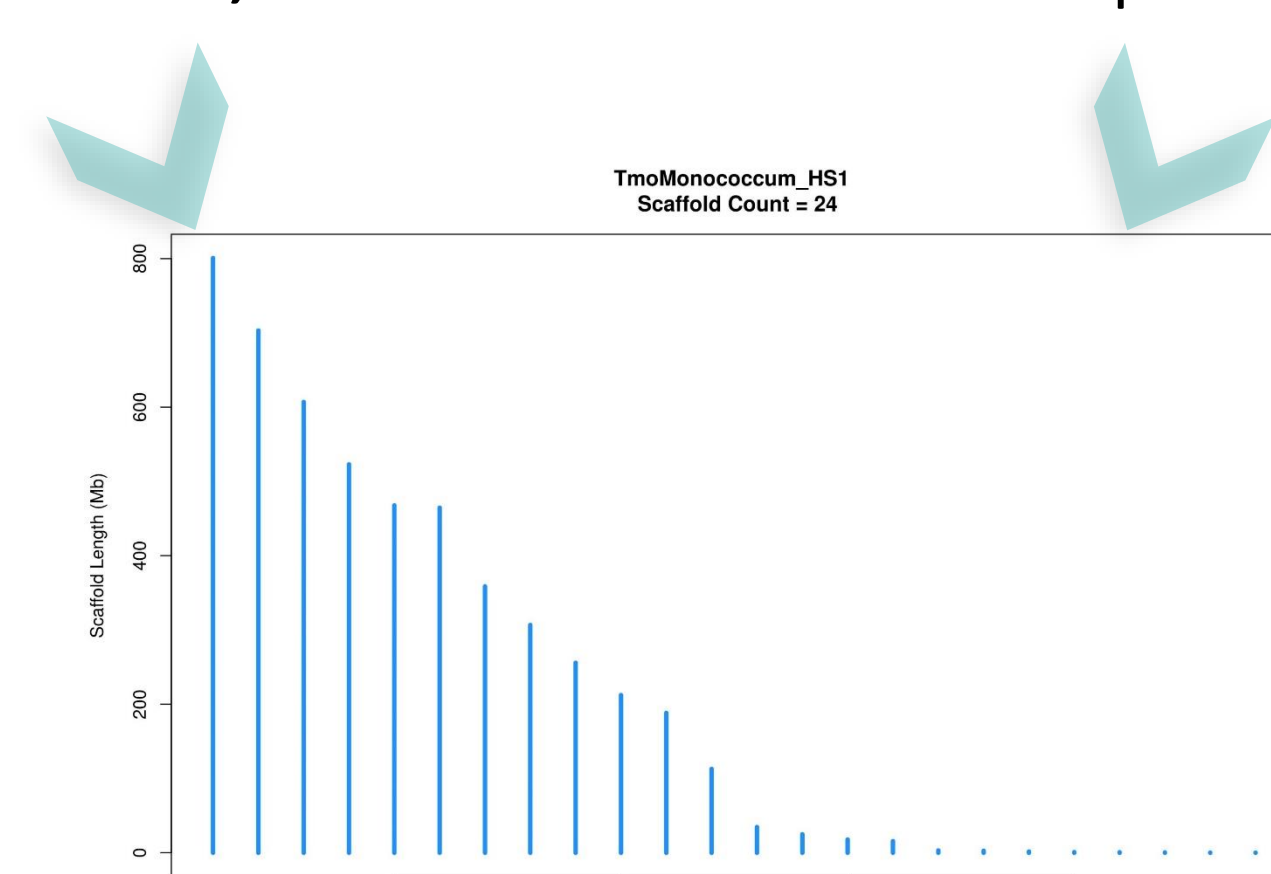
- Length: 5,15 Gb
- Contig nb: 1142
- Contig N50: 54,4 Mb

Bionano Optical mapping



Optical map assembly (Bionano access)

- Length: 5,25 Gb
- Map nb: 120
- Map N50: 164 Mb



Hybrid Scaffolding (Bionano access)

- Length: 5,1 Gb
- Scaffold nb: 24
- Scaffold N50: 523 Mb
- Scaffold nb >100Mb : 12

Genome assembly on several 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 |
| Rose* | 500 Mb | 1105 | 1024 | 9,5 | 61 | 967 | 26 |

* Heterozygous species

What's Next?

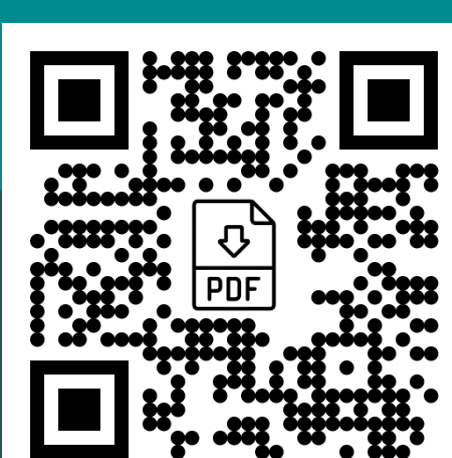


P2 solo from Oxford Nanopore Technologies

Ultra long reads to fill remaining gaps in complex genome (ex: wheat for repetitive elements)

Adaptive Sampling: Targeted sequencing method based on a software controlled enrichment. No need for upfront sample preparation.

*Eugene Pipeline
Carrere, Sébastien, & Gouzy, Jérôme. (2023).
<https://doi.org/10.5281/zenodo.7648710>



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>