



HAL
open science

Long Reads for Long-Term Goals: A Genome Resource for Lavander Genetic Improvement

Berline Fopa Fomeju, Aurélie Berard, Isabelle Le Clainche, Aurélie Canaguier, Elodie Belmonte, Veronique Gautier, Charles Poncet, Marie-Christine Le Paslier, Guillaume Frémondrière, Damien Hinsinger, et al.

► To cite this version:

Berline Fopa Fomeju, Aurélie Berard, Isabelle Le Clainche, Aurélie Canaguier, Elodie Belmonte, et al.. Long Reads for Long-Term Goals: A Genome Resource for Lavander Genetic Improvement. PAG XXX Plant and Animal Genome Conference, Jan 2023, San Diego (California), United States. hal-03937877

HAL Id: hal-03937877

<https://hal.inrae.fr/hal-03937877v1>

Submitted on 13 Jan 2023

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

INRAE

EPGV
Etude du Polymorphisme
des Génomes Végétaux

iteipmai

GENTYANE
PLATE-FORME de
GÉNOTYPAGE et SÉQUENÇAGE en AUVERGNE

Long Reads for Long-Term Goals : A Genome Resource for Lavender Genetic Improvement

Berline Fopa Fomeju¹, Aurélie Berard², Isabelle Le Clainche², Aurélie Canaguier², Elodie Belmonte³, Véronique Gautier³, Charles Poncet³, Marie-Christine Le Paslier², Guillaume Frémondrière⁴, **Damien D. Hinsinger²** and Patricia Faivre-Rampant²

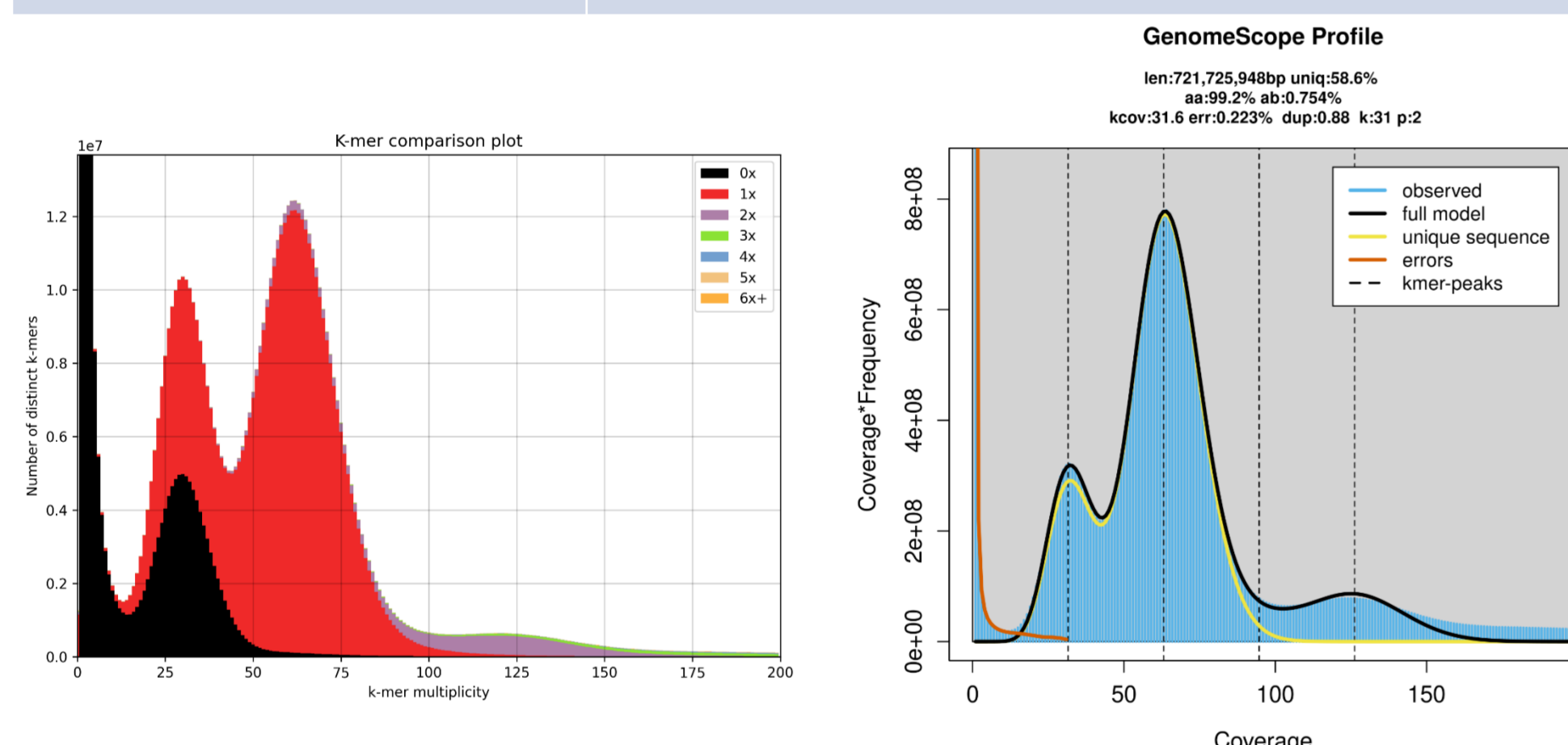
- ¹ ITEIPMAI, Chemille en Anjou, France
² INRAE, Evry, France
³ INRAE, Clermont-Ferrand, France
⁴ ITEIPMAI, Montboucher sur Jabron, France

Metrics of Lavender assemblies

	Maillette (this study)	Maillette (Malli et al. 2019)	JingXu2 (Li et al. 2021)
Chromosome number	25	25	27
Sequencing technologies	Pacbio HiFi	Illumina + transcriptome	PacBio CLC + 10X + Hi-C
Estimate of genome size (Mb)	721.73	~850	1016.25
Total length of contigs (Mb)	891.5	688.04	911.14
Total number of contigs	579	1383	1383
N50 of contigs (Mb)	20.35	1.22	1.22
Largest contig (bp)	35,738,470		9,968,313
Total length of scaffolds (Mb)		869.79	914.49
Total number of scaffolds		84,291	306
N50 of scaffolds (Mb)		0.097	36.20
Largest scaffold (Mb)		0.94	46.64
GC content (%)	39.18	38.1	38.58
Complete BUSCOs (%)	97.1	89.7	91.4
Repeat density (%)	68.23	42.8	58.28
Number of protein-coding genes	ongoing	62,141	65,905

Assembly

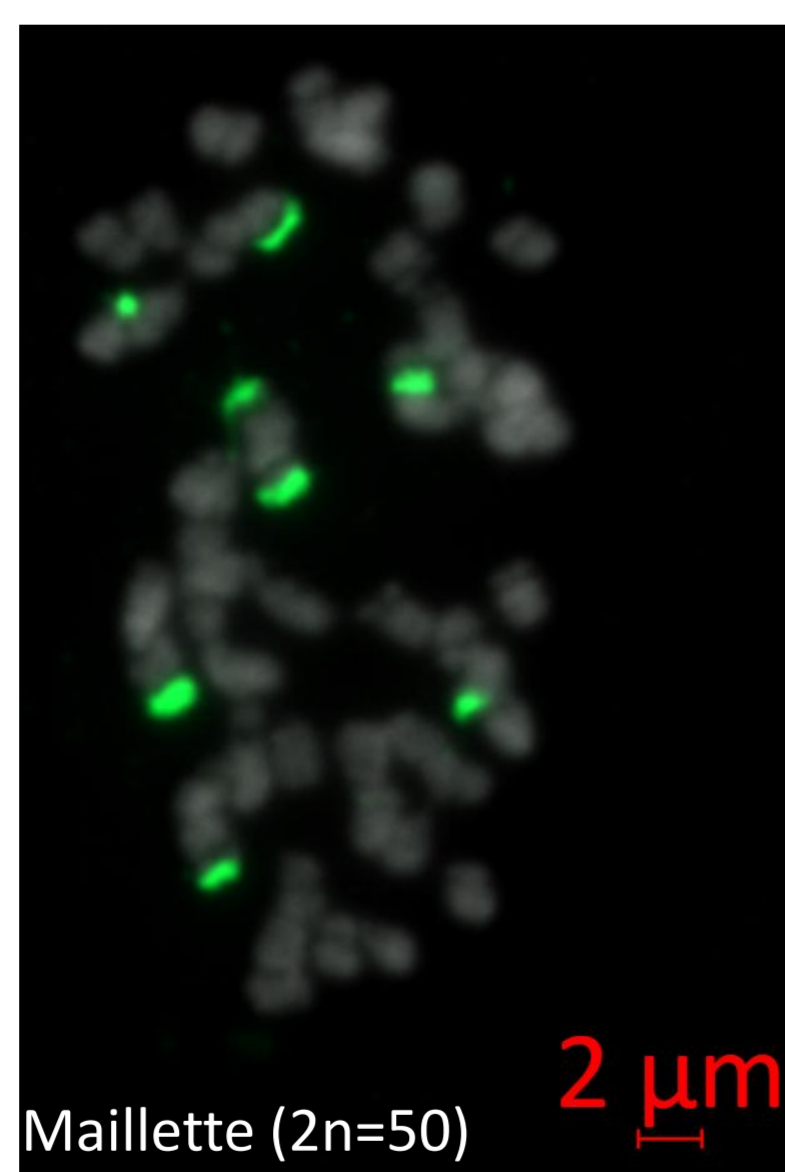
BUSCO (eudicot_odb10, n:2326)	C:97.1% [S:24.7%, D:72.4%], F:0.6%, M:2.3%
Mercury completeness	83.0479
Mercury QV (error rate)	41.7636 (6.66247e-05)



The lavender genome is composed of 68.23% of repeated sequences (mainly Copia – 36.13%).

Other Results - Cytology

Using chromosome counting and rDNA staining (realised at the Molecular Cytogenetics platform – INRAE Le Rheu), we found 25 chromosomes, as well as 8 45S rDNA clusters, in 'Maillette' root tip cells.

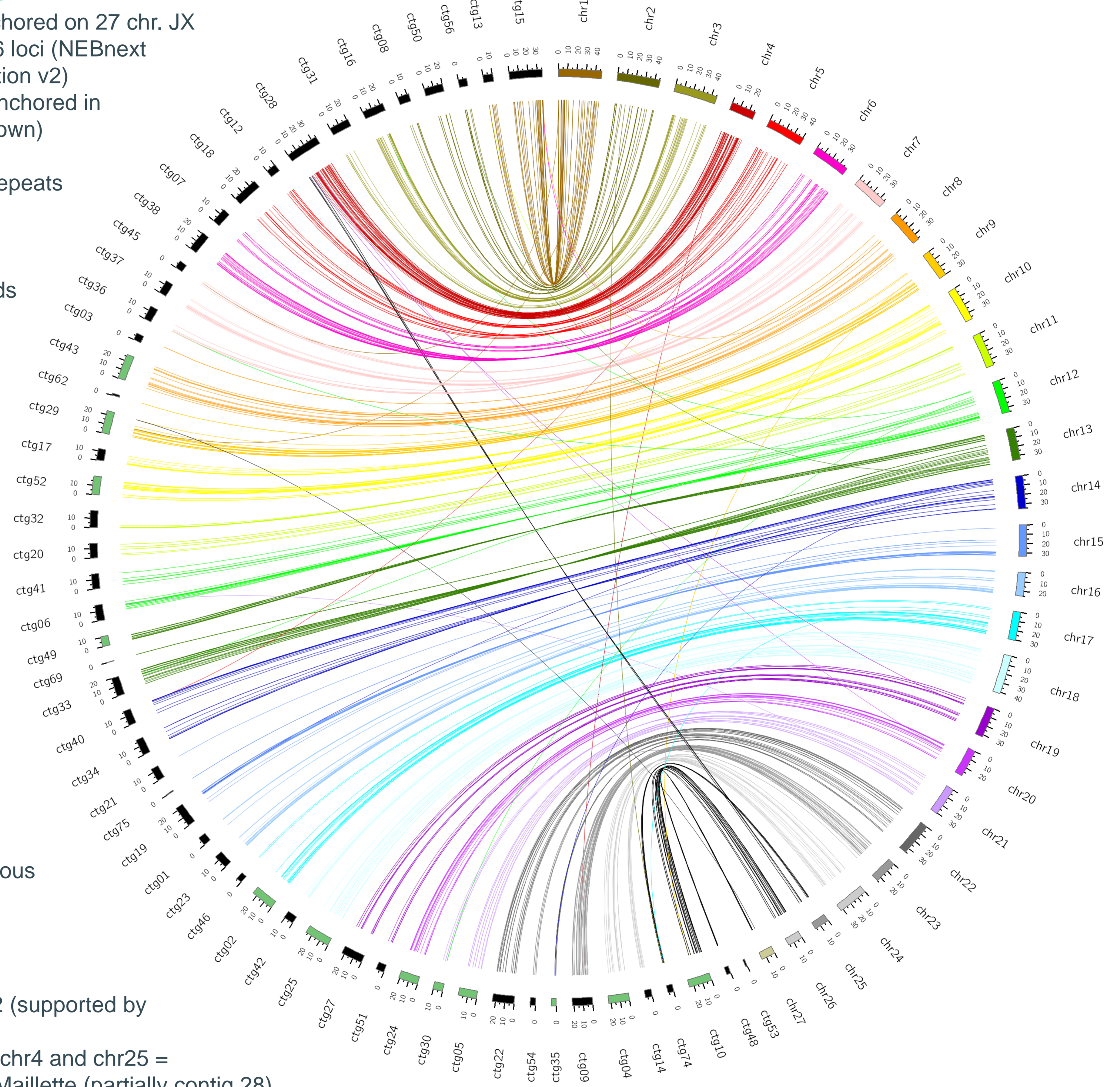


Materials & methods

Young leaves from an individual of the cultivar 'Maillette' were collected and immediately flash frozen. The protocol 'Plant Leaves' ([dx.doi.org/10.17504/protocols.io.bp21694yzlqg/v1](https://doi.org/10.17504/protocols.io.bp21694yzlqg/v1)) was slightly modified and the obtained DNA medusa was used for long reads sequencing. A **Pacbio HiFi** library was built following the manufacturer's instructions, then sequenced on a Sequel II at Gentyane (INRAE, Clermont-Ferrand) and HiFi reads generated. Assembly was performed with HiFiAsm using default parameters. **Assembly quality** was assessed using BUSCO, Mercury and KAT. Repeated Sequences were identified with RepeatModeler/RepeatMasker. **Synteny** with the genome sequence of the cultivar JingXu2 was assessed by mapping context sequences from a panel designed at EPGV (INRAE, Evry) during the GenoLavande project on sequences from both assemblies, then assessing the markers showing a '1 to 1' relationship.

Synteny with JingXu2 (JX)

- 54 Maillette contigs anchored on 27 chr. JX
- 839 markers from 3,766 loci (NEBnext direct Genotyping Solution v2)
- + 36 Maillette contigs anchored in 80 JX scaffolds (not shown)
- Contigs with telomere repeats (one end)
 - ✓ 12 anchored on JX chr (green contigs in view)
 - ✓ 24 anchored on scaffolds
 - ✓ 6 without anchorage



Conclusions

- Assembly more contiguous (N50 of contigs)
- Better completion
- 839 markers used to check collinearity
 - ✓ Synteny with JingXu2 (supported by several markers)
 - ✓ Hypothesis: JingXu2 chr4 and chr25 = one chromosome in Maillette (partially contig 28)

Perspectives

- Expand contigs with available Nanopore long reads (data not shown)
- Scaffolding to pseudomolecules with genetic map (currently under construction)
- Gene and repeat annotations
- Compare with scaffolding from Hi-C linked reads

Center
Île-de-France – Versailles-Saclay

Aknowledgements

Molecular Cytogenetics platform – INRAE Le Rheu

Genolavande



INRAE - Unité EPGV US1279
CEA - Institut de biologie François Jacob
Site d'Evry - Bat G1
2 rue Gaston Crémieux
91057 - Evry Cedex
France
support-epgv@inrae.fr

Associated publication :



Genespace lavande : Fopa Fomeju B, Brunel D, Bérard A, Rivoal J-B, Gallois P, Le Paslier M-C, et al 2020

Quick and efficient approach to develop genomic resources in orphan species: Application in *Lavandula angustifolia*.

PLoS ONE 15(12): e0243853. <https://doi.org/10.1371/journal.pone.0243853>

<https://www6.versailles-grignon.inrae.fr/epgv>