



HAL
open science

A chromosome-level assembly of lavender ‘Maillette’ using long-reads and a newly generated linkage map

Aurélie Canaguier, Damien Hinsinger, Aurélie Bérard, Isabelle Le Clainche,
Berline Fopa Fomeju, Patricia Faivre Rampant

► To cite this version:

Aurélie Canaguier, Damien Hinsinger, Aurélie Bérard, Isabelle Le Clainche, Berline Fopa Fomeju, et al.. A chromosome-level assembly of lavender ‘Maillette’ using long-reads and a newly generated linkage map. PAG XXI. Plant and Animal Genome Conference, Jan 2024, San Diego (California), United States. 10.1371/journal.pone.0243853 . hal-04670399

HAL Id: hal-04670399

<https://hal.inrae.fr/hal-04670399v1>

Submitted on 12 Aug 2024

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.



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

INRAE

EPGV
Etude du Polymorphisme
des Génomes Végétaux

iteipmai



A chromosome-level assembly of lavender 'Maillette' using long-reads and a newly generated linkage map

Hinsinger Damien D.¹, Canaguier Aurélie¹,
Berard Aurélie¹, Le Clainche Isabelle¹, Fopa
Fomeju Berliné², Faivre-Rampant Patricia¹

¹ INRAE, Evry, France

² ITEIPMAI, Chemille en Anjou, France

Metrics of Lavender assemblies

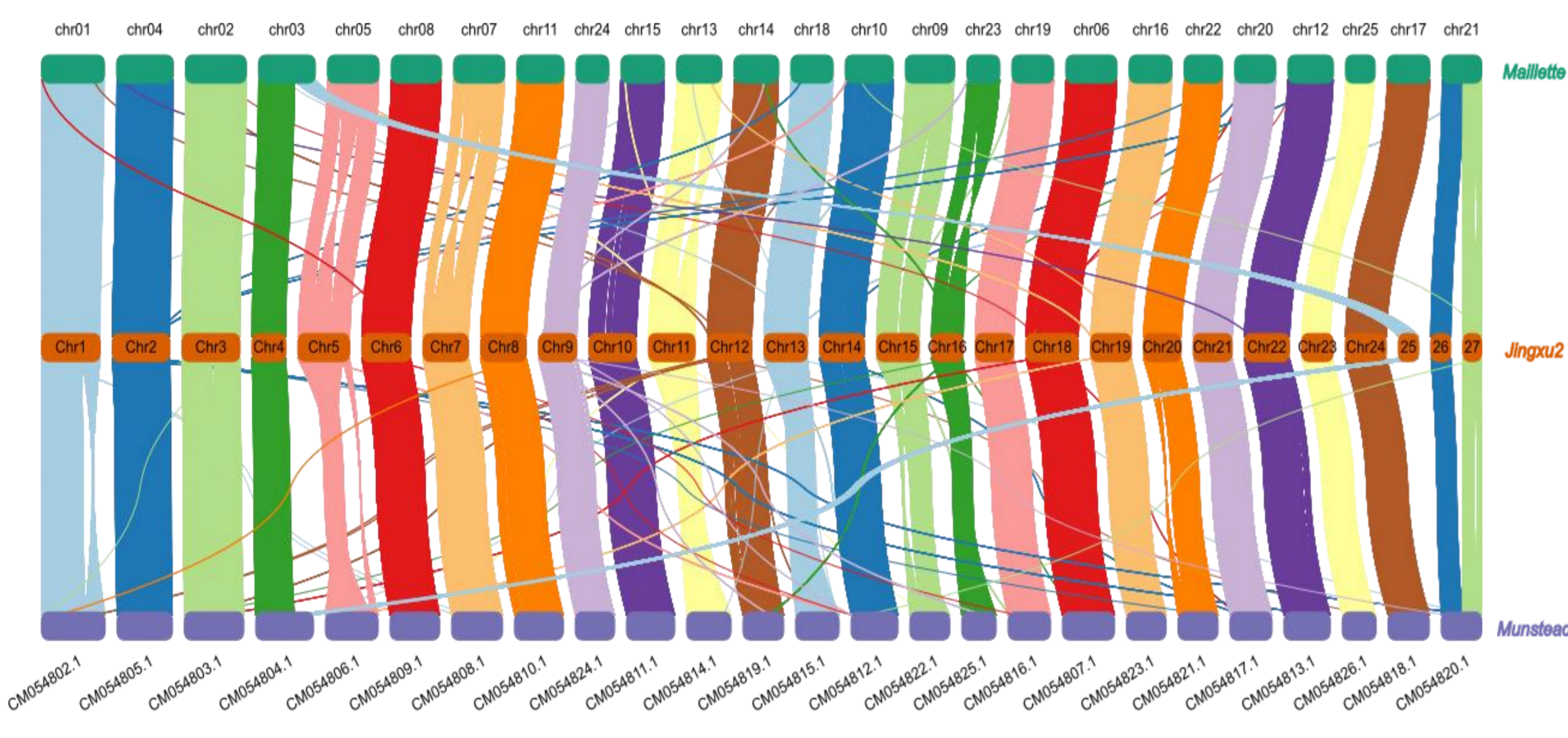
	Maillette (this study)	Maillette (Malli et al. 2019)	JingXu2 (Li et al. 2021)
Chromosome number	25	25	27
Sequencing technologies	Pacbio HiFi	Illumina + transcriptome	PacBio CLC + 10X + Hi-C
Estimate of genome size (Mb)	721.73	~850	1016.25
Total length of contigs (Mb)	891.5	688.04	911.14
Total number of contigs	579		1383
N50 of contigs (Mb)	20.35		1.22
Largest contig (bp)	35,738,470		9,968,313
Total length of scaffolds (Mb)	891.6	869.79	914.49
Total number of scaffolds	550	84,291	306
N50 of scaffolds (Mb)	33.57	0.097	36.20
Largest scaffold (Mb)	46.4	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	64,342	62,141	65,905

Assembly Quality

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%).

Synteny with JingXu2 and Munstead



- Chromosomes 26 -27 in Jingxu2 correspond to the chromosome 21 in Maillette
- Chromosome 25 in JingXu2 corresponds to a part of the chromosome 3 in Maillette
- Several large Structural Variations (SVs) among the three genomes, despite global synteny

Conclusions

- The linkage map encompassed the 25 chromosomes of Maillette previously determined by cytology, and allowed to scaffold 95.3% of the initial assembly in pseudomolecules
- Final assembly is more contiguous (N50 of contigs + 11 chromosomes with 2 telomeres)
- We identified two major rearrangement events explaining the two supernumerary chromosomes in the Chinese cultivar JingXu2
- Several large structural variations were identified, that need further evaluation

Lavender (*Lavandula angustifolia* L.) is a patrimonial and economically important species in Southern part of Europe. However, increasing pathogenic pressure related to climate change challenges cultivated varieties yields and quality of essential oils. The cultivar 'Maillette', that is most-valued for the quality of its essential oil, is widely used in genetic improvement programs in France.

We previously generated a long-reads based genome assembly from an individual of the lavender 'Maillette' using PacBio HiFi data (579 contigs assembled, representing 891.5 Mb, N50=20.35Mb, QV score of 41.8 and 97.1% of complete BUSCO genes).

Herein we built a genetic map for 'Maillette', using a genotyping tool we developed. This linkage map (822 polymorphic SNP ~33 SNP per linkage group) encompassed the expected 25 linkage groups. This linkage map was then used to scaffold our assembly. Fifty-four contigs were grouped in 25 pseudomolecules (N50=34.78 Mb) containing 94.9% (846.2/891.5 Mb) of the assembly, plus 551 unplaced contigs (min=17.8 kb ; max= 3.8 Mb). Twenty-two of twenty-five chromosomes showed telomere repeats at either both ends (11 chromosomes) or one end (11 chromosomes). Benefiting from available genome sequences showing different chromosome numbers, we highlighted the good synteny between Maillette (25 chromosomes) and Munstead (25 chromosomes) and identified 2 large chromosomal fissions (in Maillette chr03 and chr21) differentiating Maillette from JingXu2 (27 chromosomes).

This chromosome-level assembly for an economically important variety of this emblematic species will allow a better estimates of the cultivar diversity and benefit to the improvement programs through reliable GWAS.

Materials & methods

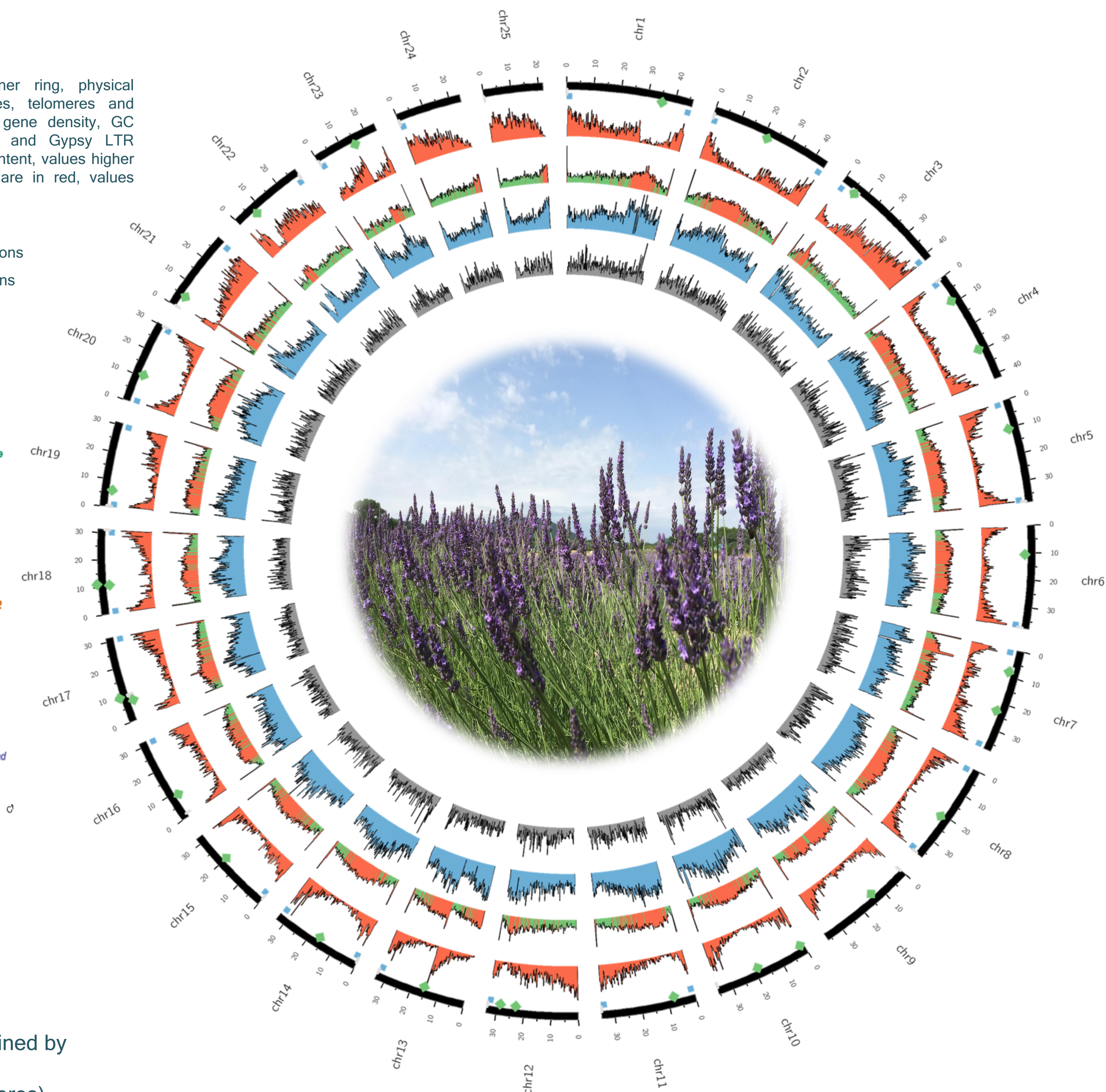
Contigs previously generated were scaffolded with AllMaps, using a linkage map constructed based on the genotyping of 215 offsprings from a self crossing of Maillette. This genotyping resulted in 822 markers properly segregating in the offsprings.

Gene coding annotation was performed with Helixer, and repeated elements were annotated using RepeatModeler/RepeatMasker. Canonical telomere repeats were searched using a custom-made script, while putative centromeres were identified with Quartet Centrominer. Synteny between Maillette and the lavender cultivars Munstead and JingXu2 was evaluated using NGenomeSyn.

Genomic landscape

From the outer to the inner ring, physical pseudomolecules, centromeres, telomeres and gaps locations (see below), gene density, GC content, Copia LTR density and Gypsy LTR density are shown. For GC content, values higher than the average GC value are in red, values below are in green.

- ◆ Remaining gaps
- Putative centromeric regions
- Identified telomeric regions



Perspectives

- Close remaining gaps with available Nanopore long reads (data not shown)
- Refine Gene and repeat annotations, generate functional annotations
- Compare with scaffolding from Hi-C linked reads
- Validate identified SVs



Center
Île-de-France - Versailles-Saclay

Aknowledgements



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>



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

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