

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2890055
ToLID	<b>ddAndSaus6</b>
Species	<i>Androsace saussurei</i>
Class	Magnoliopsida
Order	Ericales

Genome Traits	Expected	Observed
Haploid size (bp)	469,524,587	788,387,943
Haploid Number	10 (source: ancestor)	40
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q51

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

### Curator notes

. Interventions/Gb: 123  
. Contamination notes: "A few contaminant sequences, mostly fungal and viral "  
. Other observations: "Genome assembly of *Androsace saussurei*. This assembly is provided by the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). Several ploidies were observed in the *Androsace* genus, with 10 chromosomes (Weiss (2021), "Chromosome number and ploidy level of *Androsace maxima* (Primulaceae) in Austria", [https://www.zobodat.at/pdf/NEIL\\_1\\_0177-0180.pdf](https://www.zobodat.at/pdf/NEIL_1_0177-0180.pdf)). Here, we reconstructed 40 chromosomes (clearly 4 sets of 10 chromosomes based on the Hi-C map), suggesting a tetraploid or an octoploid genome for *Androsace saussurei*. If the genome is tetraploid, it is unusual as the assembler reconstructed complete maternal and paternal alleles, indicating high heterozygosity. In contrast, if the genome is octoploid, it indicates near homozygosity. We can clearly see regions with higher coverage that may indicate gene conversions between the homologous chromosomes, which would require additional techniques for complete assembly. Tetraploidy was also

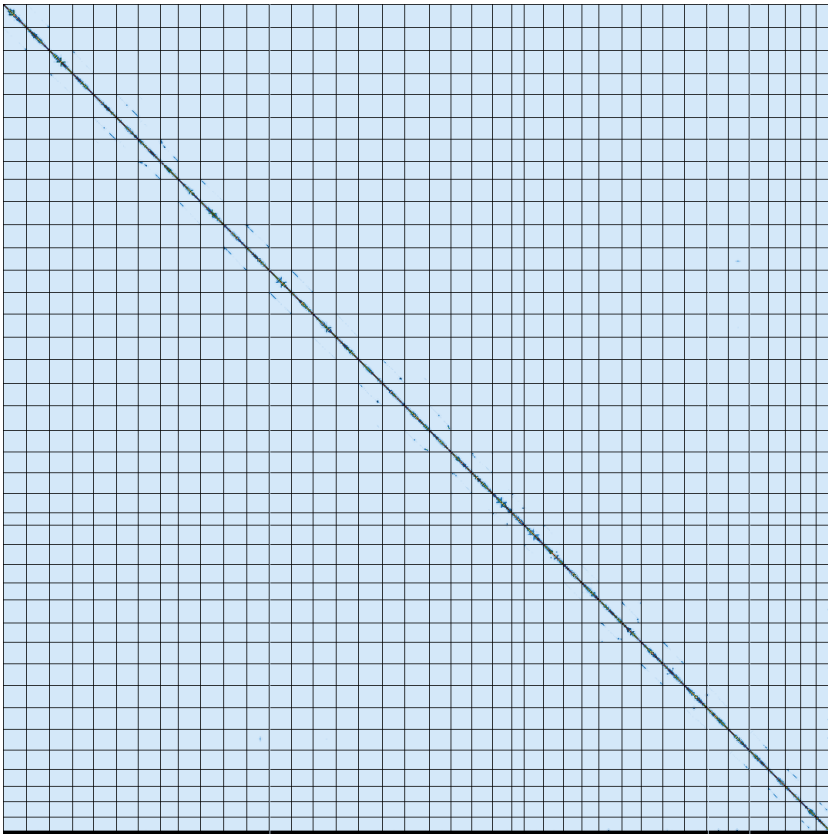
suggested by Smudgeplot, and SNPs showed few variations, supporting the tetraploid hypothesis. The twenty most complete chromosomes will be submitted as the reference haplotype, and the other twenty as the alternate haplotype. "

# Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	792,599,937	788,387,943
GC %	32.99	32.99
Gaps/Gbp	301.54	360.23
Total gap bp	23,900	37,100
Scaffolds	197	117
Scaffold N50	23,288,581	20,531,085
Scaffold L50	11	19
Scaffold L90	28	35
Contigs	436	401
Contig N50	5,110,328	5,004,711
Contig L50	50	51
Contig L90	167	167
QV	51.7172	51.7137
Kmer compl.	94.9672	94.8087
BUSCO sing.	7.2%	7.1%
BUSCO dupl.	89.2%	89.4%
BUSCO frag.	1.0%	1.0%
BUSCO miss.	2.6%	2.5%

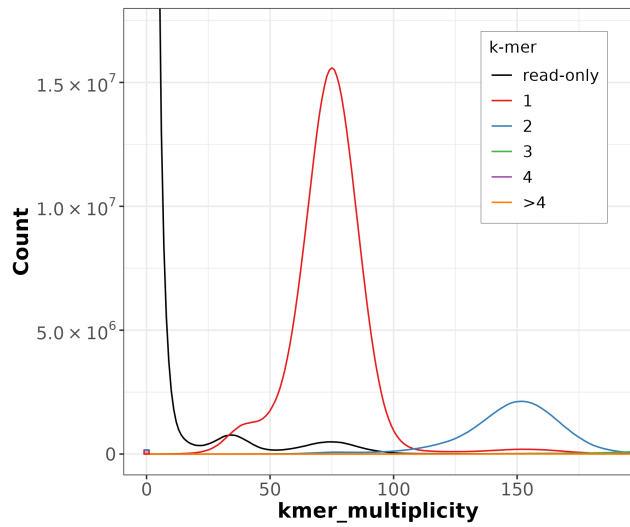
BUSCO: 5.4.3 (euk\_genome\_met, metaeuk) / Lineage: eudicots\_odb10 (genomes:31, BUSCOs:2326)

# HiC contact map of curated assembly

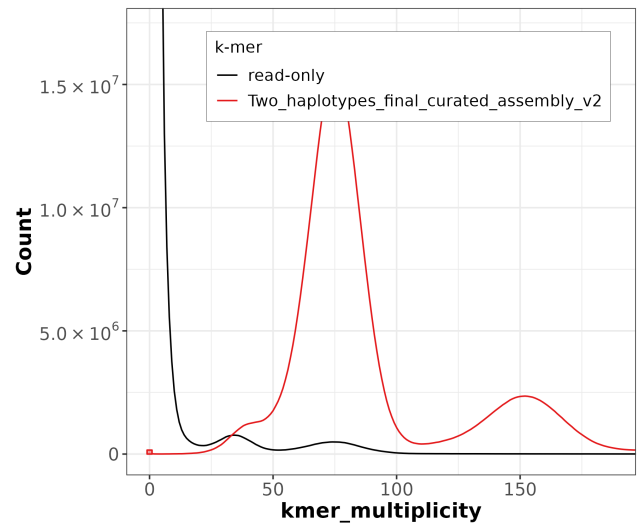


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

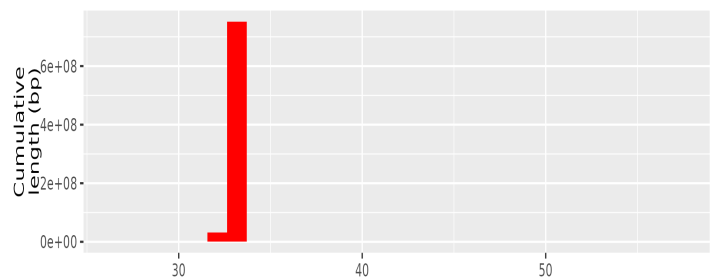


Distribution of k-mer counts per copy numbers found in asm

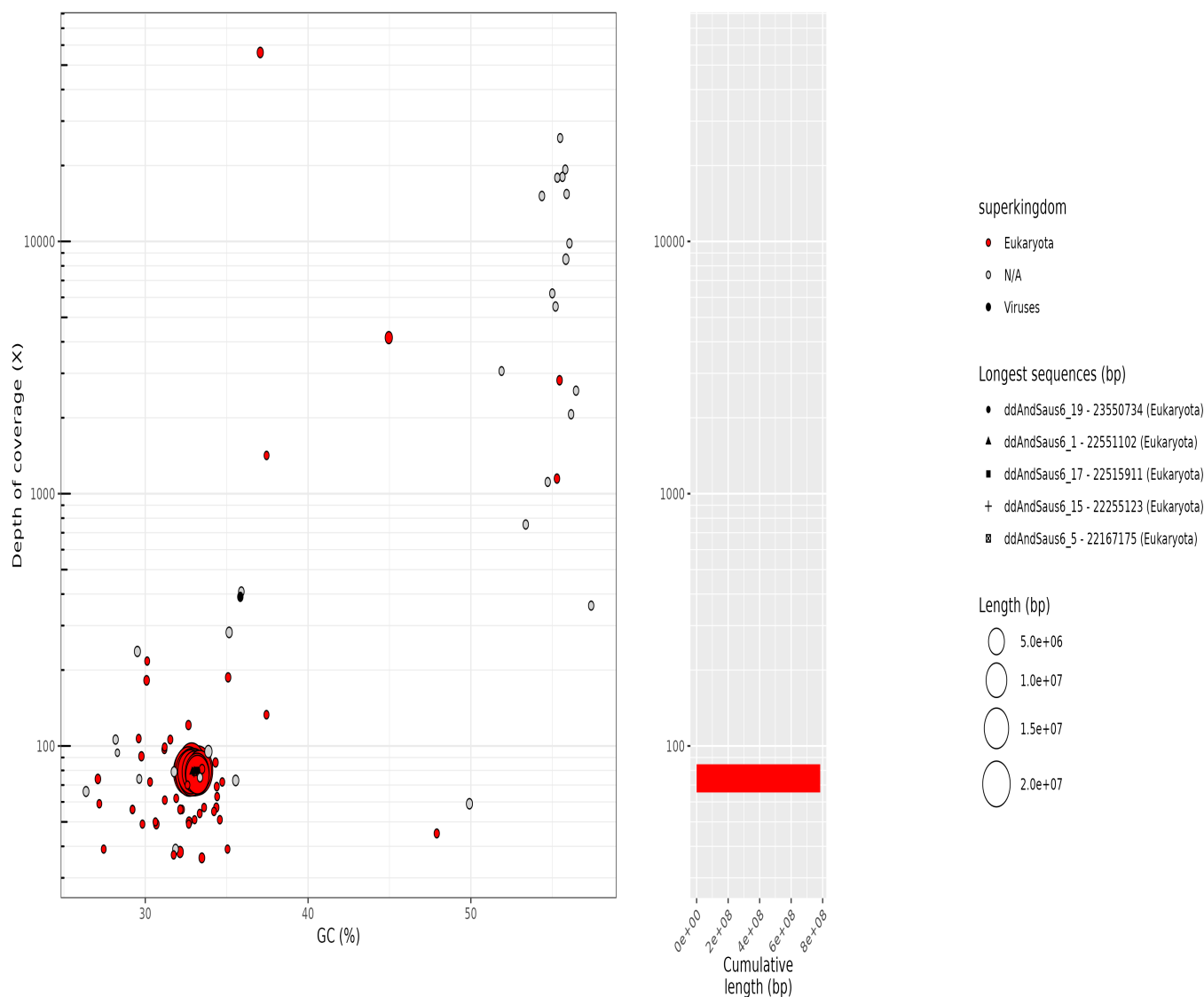


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



TAPAs summary Graph



**collapsed.** Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

# Data profile

Data	PACBIO Hifi	Omnic
Coverage	97	44

# Assembly pipeline

- **Hifiasm**
  - |\_ *ver*: 0.19.5-r593
  - |\_ *key param*: NA
- **purge\_dups**
  - |\_ *ver*: 1.2.5
  - |\_ *key param*: NA
- **YaHS**
  - |\_ *ver*: 1.2
  - |\_ *key param*: NA

# Curation pipeline

- **PretextMap**
  - |\_ *ver*: 0.1.9
  - |\_ *key param*: NA
- **PretextView**
  - |\_ *ver*: 0.2.5
  - |\_ *key param*: NA

Submitter: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2025-02-21 12:48:40 CET