

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	2890055
ToLID	ddAndSaus6
Species	Androsace saussurei
Class	Magnoliopsida
Order	Ericales

Genome Traits	Expected	Observed
Haploid size (bp)	469,524,587	788,683,157
Haploid Number	10 (source: ancestor)	40
Ploidy	2 (source: ancestor)	4
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
- . Observed Ploidy is different from Expected

Curator notes

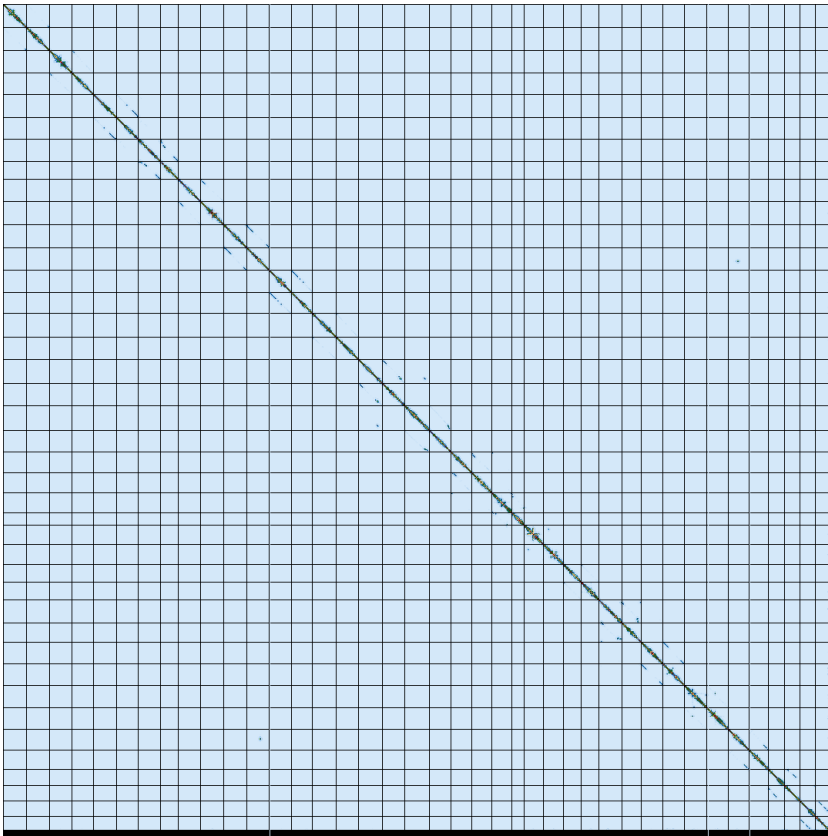
. Interventions/Gb: 123
. Contamination notes: "Few contaminant sequences, mostly viral sequences "
. Other observations: "Here we retained the Nextdenovo assembly. Note that Hifiasm generated an assembly of 782Mbp with a high N50 (21Mb) and a low N90 (<50Kb). Several ploidies were observed in the Androsace 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). 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. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	792,599,937	788,683,157
GC %	32.99	32.99
Gaps/Gbp	301.54	360.09
Total gap bp	23,900	37,100
Scaffolds	197	121
Scaffold N50	23,288,581	20,531,085
Scaffold L50	11	19
Scaffold L90	28	35
Contigs	436	405
Contig N50	5,110,328	5,004,711
Contig L50	50	51
Contig L90	167	168
QV	51.7172	51.7153
Kmer compl.	94.9672	94.8263
BUSCO sing.	7.2%	
BUSCO dupl.	89.2%	
BUSCO frag.	1.0%	
BUSCO miss.	2.6%	

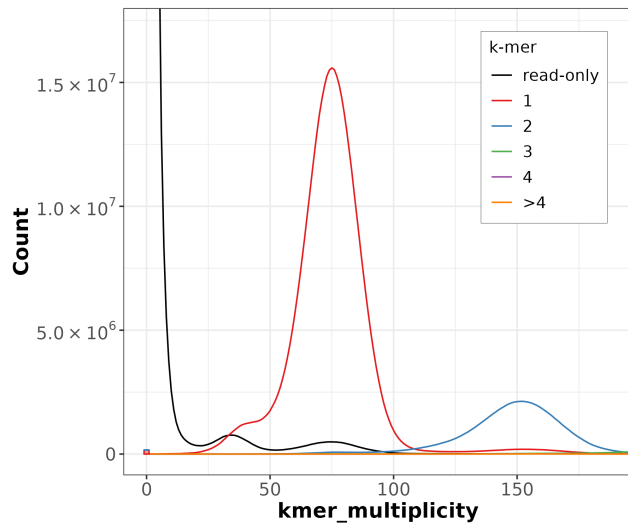
BUSCO 5.4.3 Lineage: eudicots_odb10 (genomes:31, BUSCOs:2326)

HiC contact map of curated assembly

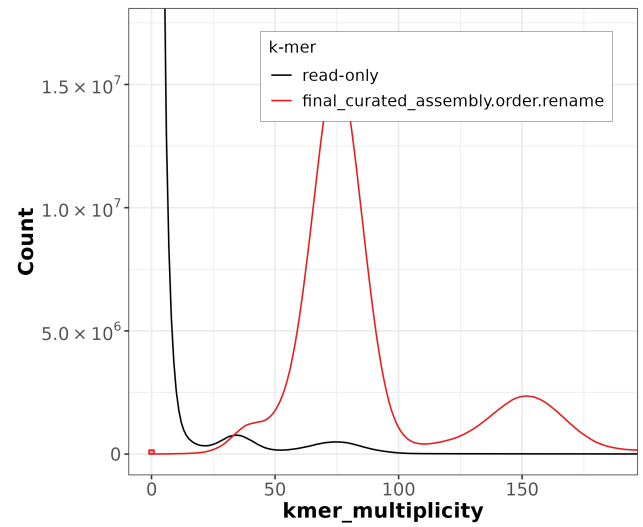


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

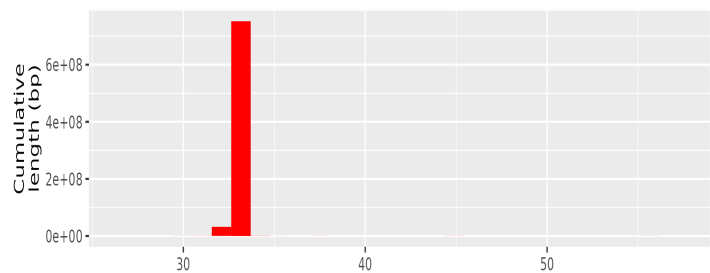


Distribution of k-mer counts per copy numbers found in asm

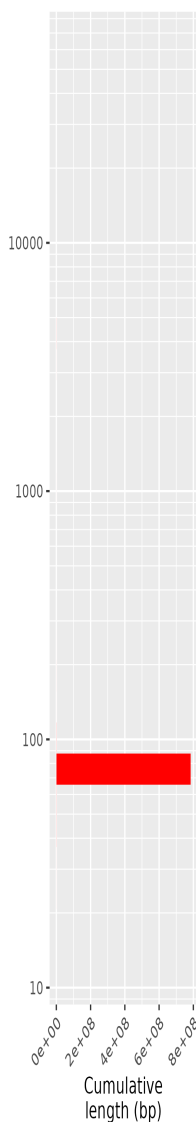
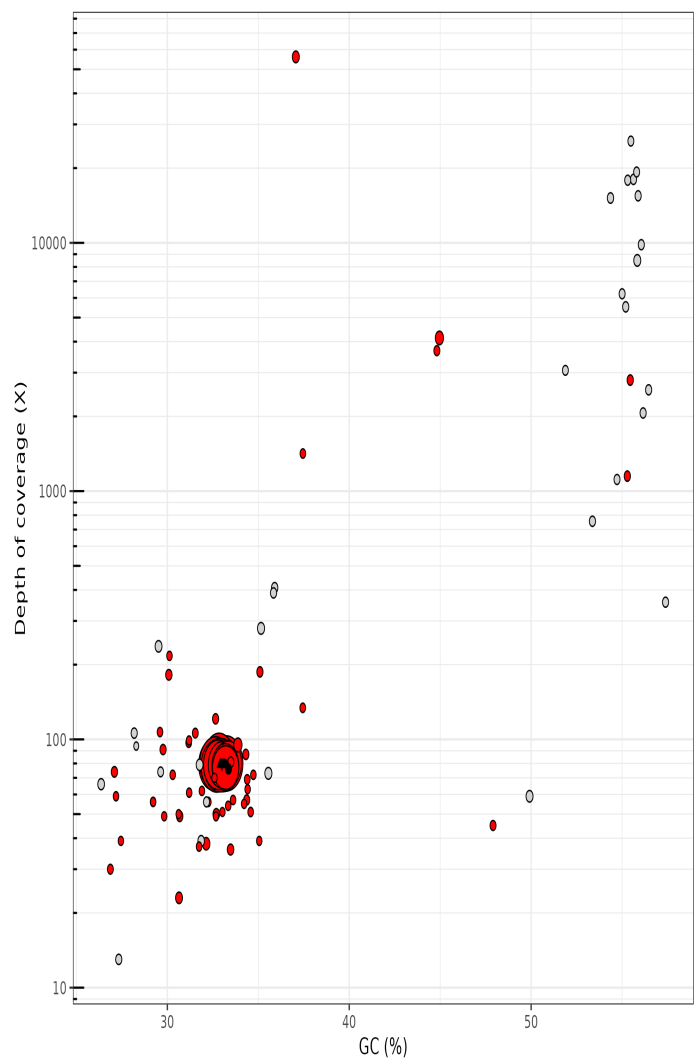


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



superkingdom

- Eukaryota
- N/A
- Viruses

Longest sequences (bp)

- ddAndSaus6_19 - 23550734 (Eukaryota)
- ▲ ddAndSaus6_1 - 22551102 (Eukaryota)
- ddAndSaus6_17 - 22515911 (Eukaryota)
- + ddAndSaus6_15 - 22255123 (Eukaryota)
- ddAndSaus6_5 - 22167175 (Eukaryota)

Length (bp)

- 5.0e+06
- 1.0e+07
- 1.5e+07
- 2.0e+07

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

- **Nextdenovo**
 - |_ *ver*: 2.5.1
 - |_ *key param*: NA

Curation pipeline

- **Nextdenovo**
 - |_ *ver*: 2.5.1
 - |_ *key param*: NA
- **yahs**
 - |_ *ver*: 1.2
 - |_ *key param*: NA

Submitter: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2024-08-02 14:59:55 CEST