

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

| | |
|---------|---------------------|
| TxID | 2890055 |
| ToLID | ddAndSaus20 |
| Species | Androsace saussurei |
| Class | Magnoliopsida |
| Order | Ericales |

| Genome Traits | Expected | Observed |
|-------------------|-----------------------|-------------|
| Haploid size (bp) | 469,524,587 | 791,205,177 |
| 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

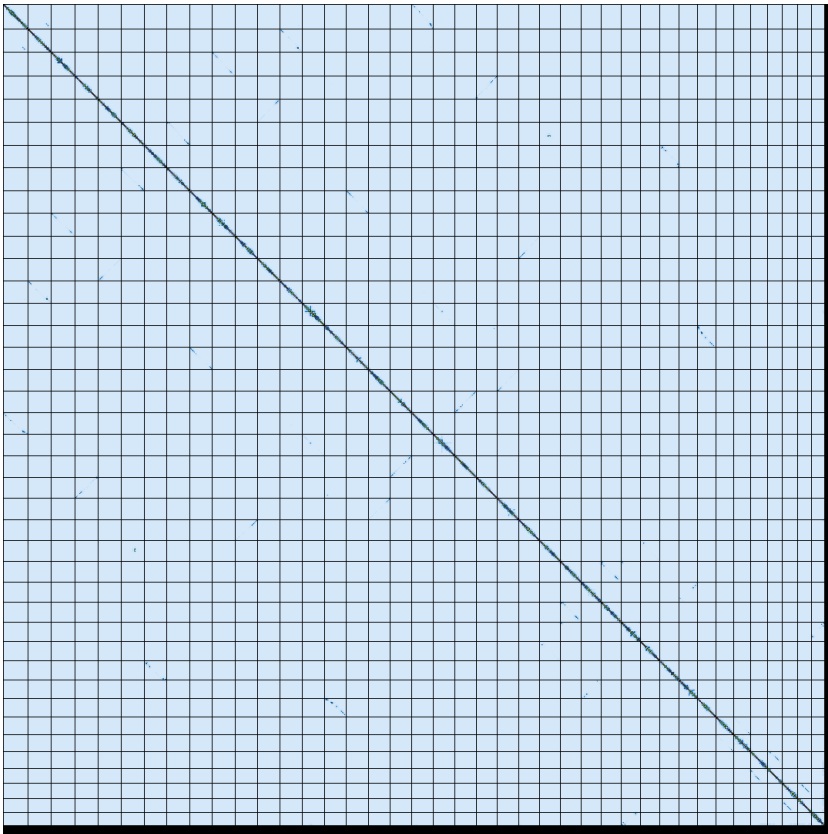
- . Interventions/Gb: 60
- . Contamination notes: "Few contaminant sequences, mostly viral sequences"
- . Other observations: "Several ploidies were observed in the Androsace genus, with 10 chromosomes. Here, we reconstructed 40 chromosomes (clearly 4 sets of 10 chromosomes based on the Hi-C map), suggesting an octoploid genome for Androsace saussurei. This is an interesting feature that adds a new element to the findings in the article by Weiss (2021), "Chromosome number and ploidy level of Androsace maxima (Primulaceae) in Austria" (https://www.zobodat.at/pdf/NEIL_1_0177-0180.pdf), which describes only diploid, tetraploid, and hexaploid genomes. 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 | 791,205,177 |
| GC % | 32.99 | 32.99 |
| Gaps/Gbp | 301.54 | 322.29 |
| Total gap bp | 23,900 | 29,300 |
| Scaffolds | 197 | 182 |
| Scaffold N50 | 23,288,581 | 20,531,085 |
| Scaffold L50 | 11 | 19 |
| Scaffold L90 | 28 | 36 |
| Contigs | 436 | 437 |
| Contig N50 | 5,110,328 | 5,037,000 |
| Contig L50 | 50 | 50 |
| Contig L90 | 167 | 169 |
| QV | 51.7172 | 51.7239 |
| Kmer compl. | 94.9672 | 94.9019 |
| BUSCO sing. | 7.2% | 7.0% |
| BUSCO dupl. | 89.2% | 89.4% |
| BUSCO frag. | 1.0% | 1.0% |
| BUSCO miss. | 2.6% | 2.6% |

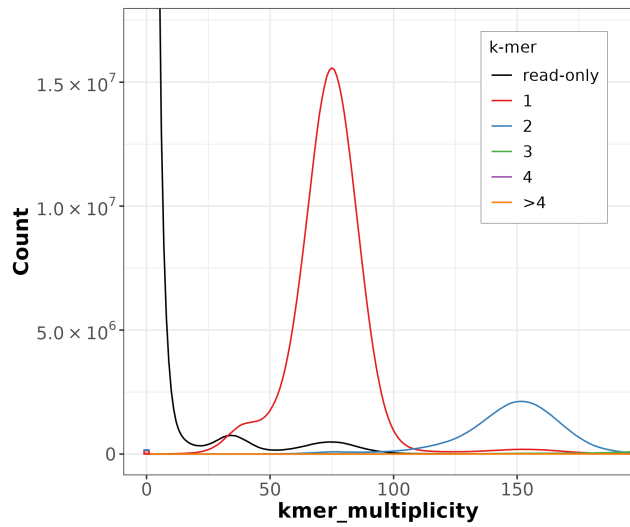
BUSCO 5.4.3 Lineage: eudicots_odb10 (genomes:31, BUSCOs:2326)

HiC contact map of curated assembly

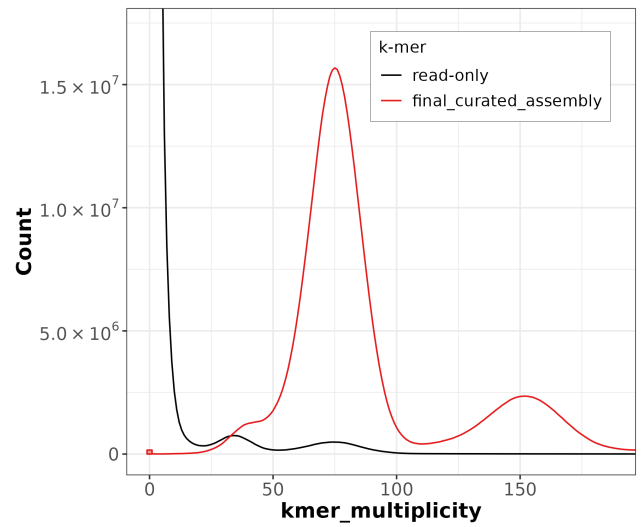


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

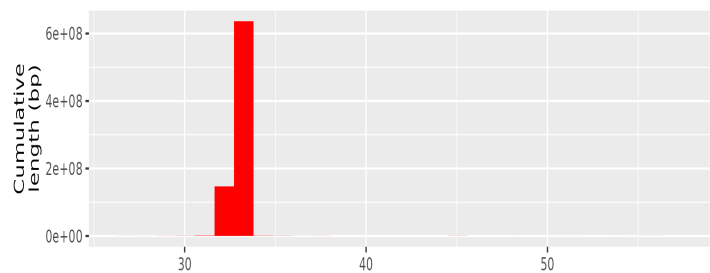


Distribution of k-mer counts per copy numbers found in asm

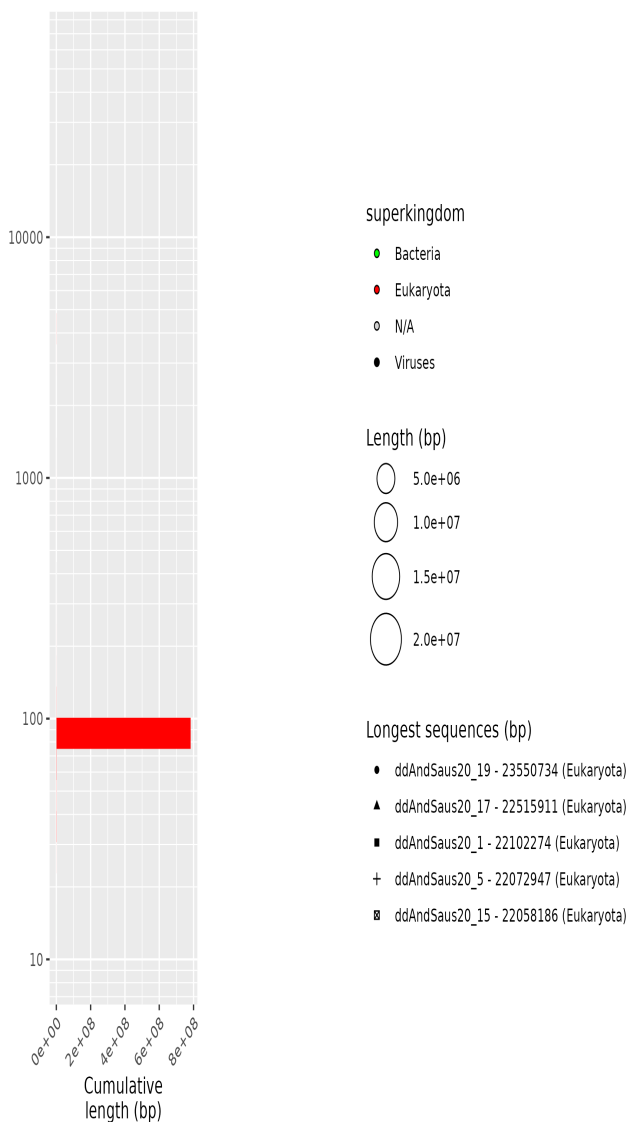
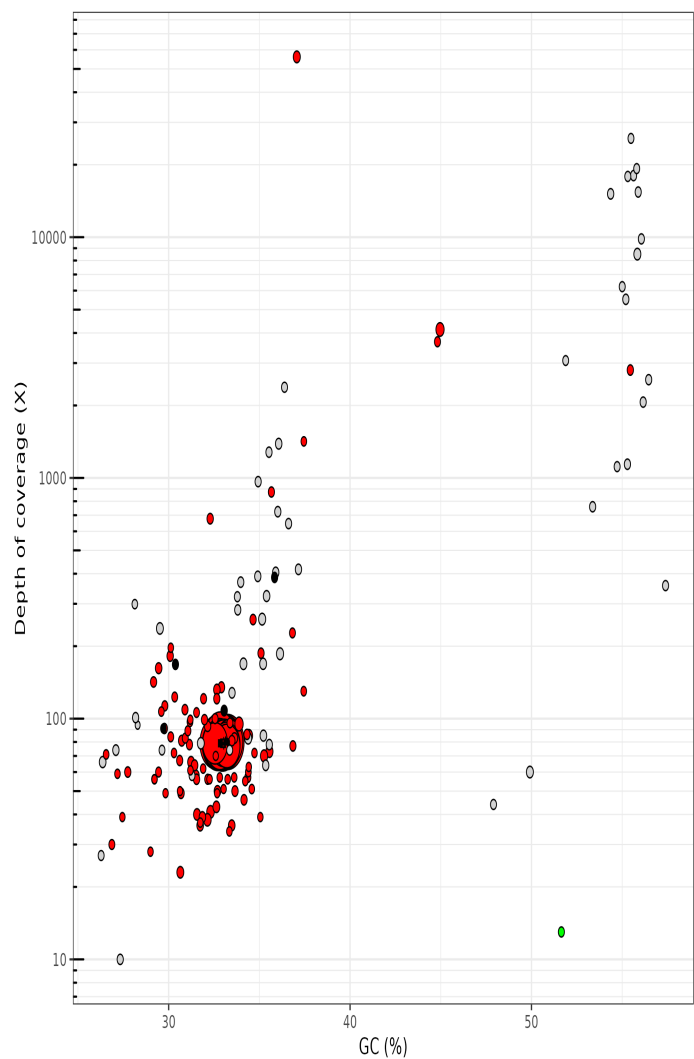


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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

Curation 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

Submitter: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2024-07-22 16:57:15 CEST