ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	3708
ToLID	**ddBraNapu2**
Species	Brassica napus
Class	Magnoliopsida
Order	Brassicales

Genome Traits	Expected	Observed
Haploid size (bp)	519,373,373	1,041,442,488
Haploid Number	18 (source: direct)	19
Ploidy	4 (source: direct)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q73

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

. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed

Curator notes

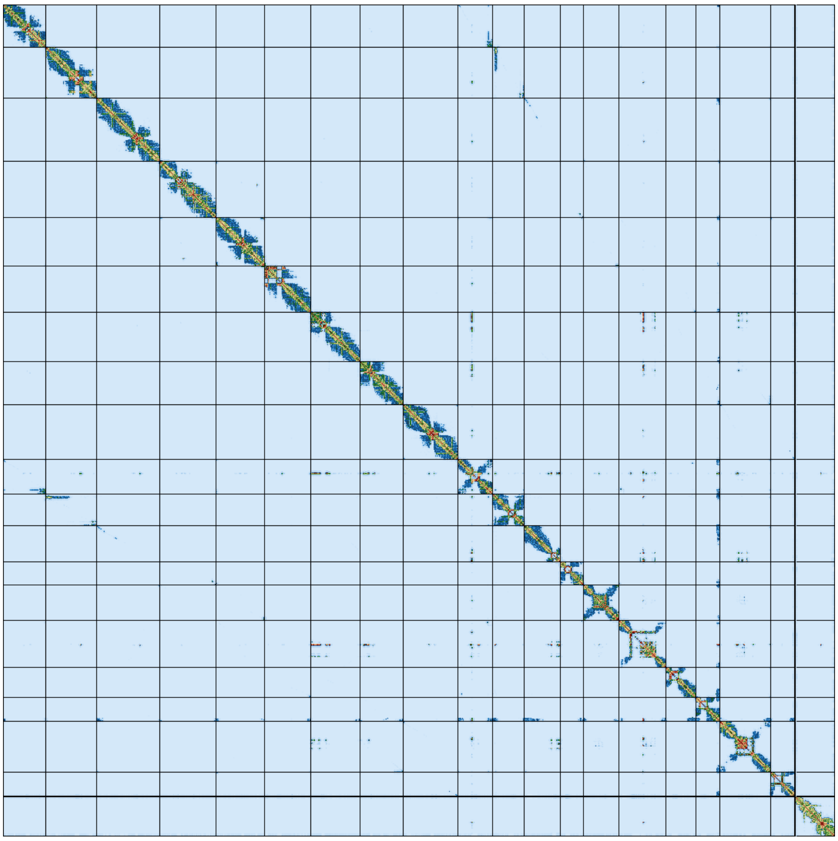
. Interventions/Gb: 38
. Contamination notes: ""
. Other observations: "The assembly of Brassica napus (ddBraNapu2) is based on 57X PacBio data and 116X OmniC Hi-C data. The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 43 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.188 Mb (with the largest being 0.222 Mb). Additionally, 1894 regions totaling 76.888 Mb (with the largest being 1.821 Mb) were identified as haplotypic duplications and removed. The mitochondrial and the chloroplastic genomes were assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. Chromosome-scale scaffolds confirmed by Hi-C data were named as in the reference Brassica napus Darmor-bzh genome assembly (GCA_905183035.1). "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,041,979,797	1,041,442,488
GC %	37.13	37.13
Gaps/Gbp	73.9	78.74
Total gap bp	7,700	10,300
Scaffolds	61	33
Scaffold N50	57,297,403	59,266,812
Scaffold L50	8	8
Scaffold L90	16	17
Contigs	138	115
Contig N50	19,179,000	19,179,000
Contig L50	21	21
Contig L90	53	54
QV	72.3483	73.4375
Kmer compl.	97.9986	97.998
BUSCO sing.	4.9%	2.4%
BUSCO dupl.	93.7%	97.0%
BUSCO frag.	0.1%	0.0%
BUSCO miss.	1.3%	0.6%

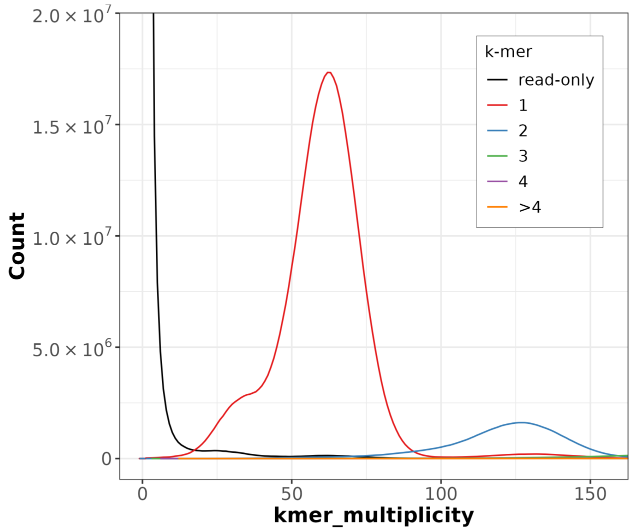
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: brassicales_odb12 (genomes:6, BUSCOs:4311)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: brassicales_odb12 (genomes:6, BUSCOs:4311)

HiC contact map of curated assembly

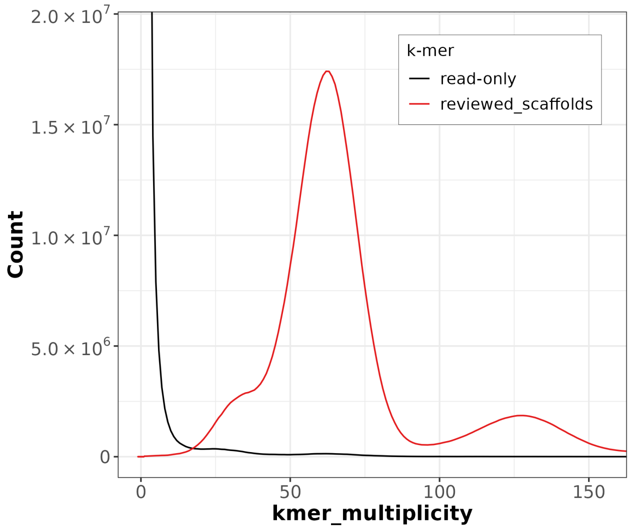


collapsed [LINK]

K-mer spectra of curated assembly

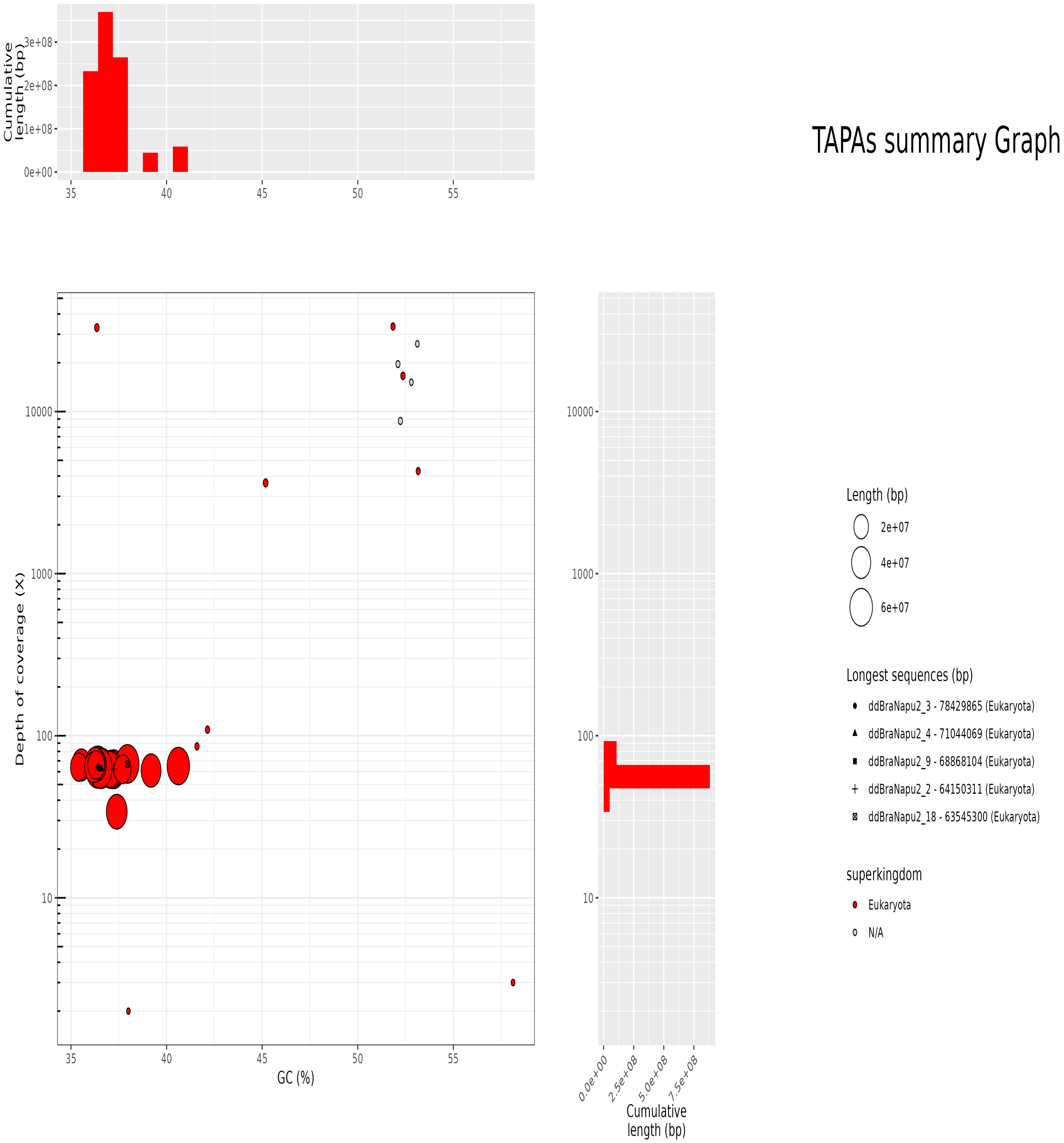


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	Long reads	Omnic
Coverage	132	116

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: 2025-10-10 09:50:28 CEST