ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	160988
ToLID	**drZelAbel2**
Species	Zelkova abelicea
Class	Magnoliopsida
Order	Rosales

Genome Traits	Expected	Observed
Haploid size (bp)	1,591,124,144	1,582,971,690
Haploid Number	14 (source: direct)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q67

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

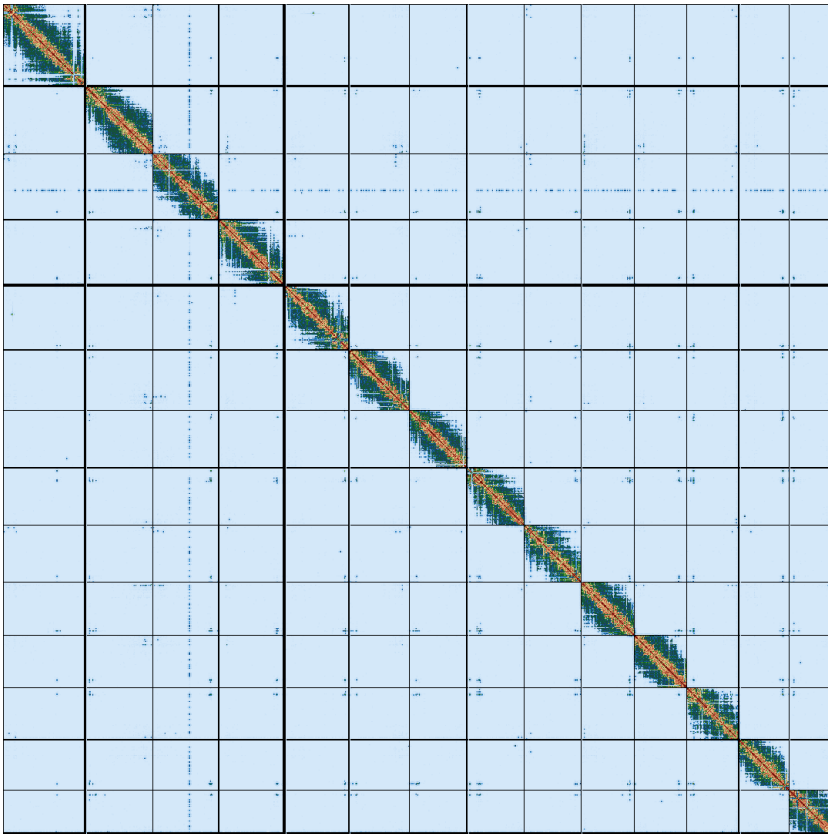. Kmer completeness value is less than 90 for collapsed

Curator notes

. Interventions/Gb: 140
. Contamination notes: ""
. Other observations: "The assembly of ZELKOVA ABELICEA (drZelAbel2) is based on 27X PacBio data and 90X Omni-C Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS. In total, 2 contigs were identified as contaminants (bacterial), totaling 23 Kb (with the largest being 15 Kb). We did not use purge_dups to remove haplotype duplications, as it did not work properly on this genome. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 47 haplotypic regions and 22 contaminant sequences were removed, totaling 29 Mb and 4 Mb, respectively (with the largest being 1.6 Mb and 1 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,153,367,633	1,582,971,690
GC %	34.52	34.6
Gaps/Gbp	11.27	38.54
Total gap bp	1,300	9,100
Scaffolds	25	165
Scaffold N50	101,492,064	109,271,146
Scaffold L50	5	7
Scaffold L90	10	13
Contigs	38	226
Contig N50	89,234,789	59,119,457
Contig L50	6	10
Contig L90	13	34
QV	67.1341	67.1577
Kmer compl.	56.0285	73.6274
BUSCO sing.	72.4%	97.0%
BUSCO dupl.	1.5%	1.5%
BUSCO frag.	1.5%	0.4%
BUSCO miss.	24.6%	1.1%

BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: embryophyta_odb10 (genomes:50, BUSCOs:1614)
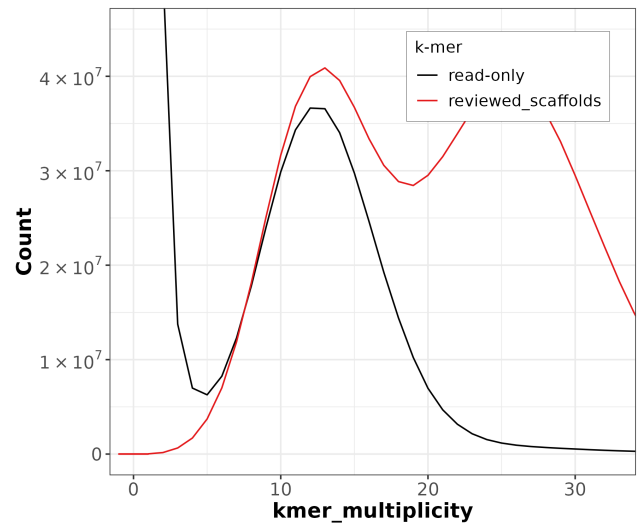
HiC contact map of curated assembly

collapsed [LINK]
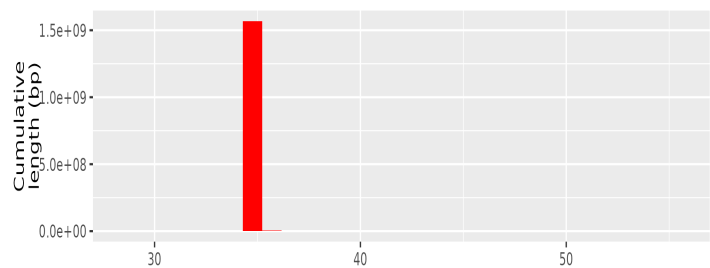
K-mer spectra of curated assembly

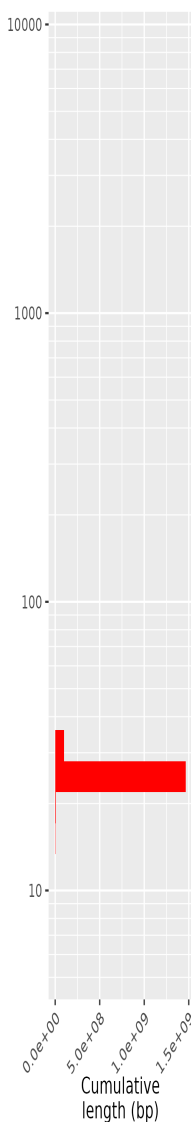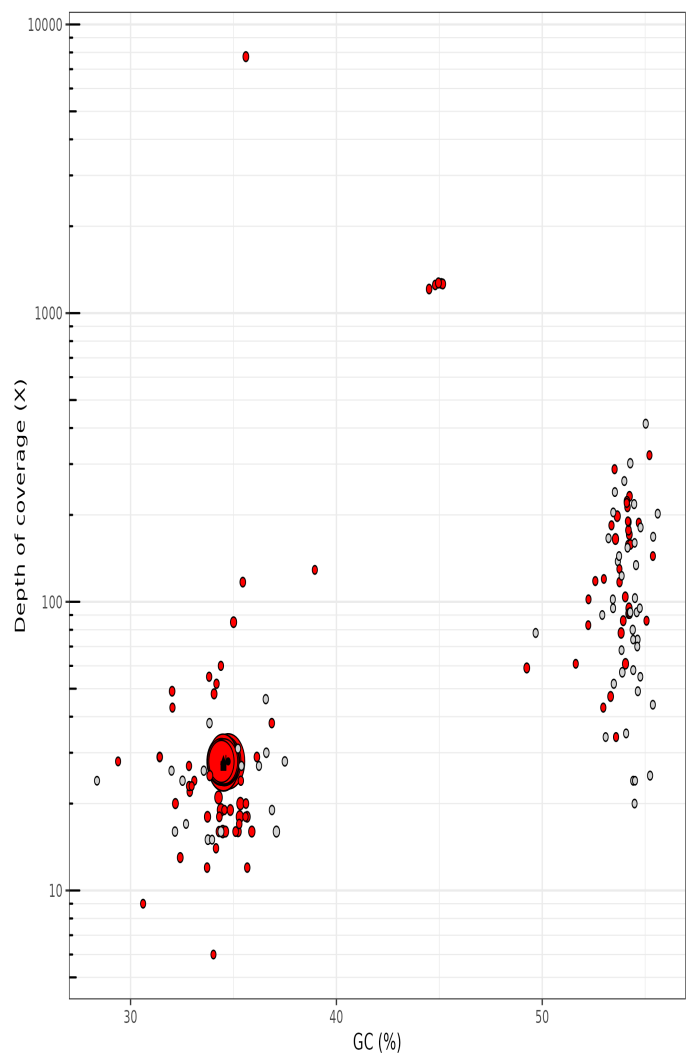
Distribution of k-mer counts per copy
numbers found in asm

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	27	74

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: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-05-15 10:07:12 CEST