

# Genome Annotation Report

Report generated: 2025-12-16

The species name TBD genome assembly accession TBD was annotated by TBD (contact ORCID/Bioproject TBD). This annotation includes 264 transcribed mRNAs from 264 genes. The average transcript length is 1758.89 bp, with an average of 1 coding transcripts per gene and 5 exons per transcript. The annotation file is available at [file path TBD](#).

## Summary

<b>Number of genes:</b>	264
<b>Number of CDSs:</b>	264
<b>Number of transcripts:</b>	264
<b>Mean transcript length:</b>	1758.89 bp
<b>Mean transcripts per gene:</b>	1
<b>Mean exons per transcript:</b>	5
<b>BUSCO summary:</b>	C:1.1%[S:1.0%,D:0.1%],F:0.2%,M:98.7%,n:1614
<b>BUSCO lineage dataset:</b>	embryophyta_odb10
<b>BUSCO mode:</b>	proteins
<b>OMArk completeness:</b>	S:1.10%,D:0.09%[U:0.08%,E:0.01%],M:98.81%
<b>OMArk lineage:</b>	Brassicaceae
<b>OMArk consistency:</b>	A:97.57%[P:0.00%,F:2.08%],I:0.35%[P:0.00%,F:0.00%],C:0.00%[P:0.00%,F:0.00%],U:2.08%

## Metadata

<b>Project ID:</b>	Bioproject TBD
<b>Study name:</b>	Test annotations
<b>Project description:</b>	Annotation benchmarking project
<b>Contact name:</b>	TBD

<b>Contact ID:</b>	contact ORCID/Bioproject TBD
<b>Contact email:</b>	email address TBD
<b>Taxon ID:</b>	TBD
<b>Scientific name:</b>	species name TBD
<b>Assembly accession:</b>	assembly accession TBD
<b>Assembly seqcol digest:</b>	TBD
<b>Assembly aliases:</b>	test-assembly.fasta
<b>Evidence ID:</b>	TBD
<b>Evidence type:</b>	TBD
<b>Evidence version or date of retrieval:</b>	TBD
<b>Evidence source:</b>	TBD
<b>Annotation tools:</b>	TBD
<b>Annotation tool versions:</b>	TBD
<b>Annotation protocol:</b>	TBD
<b>Annotation file label:</b>	test-annotation-file.gff
<b>Annotation file URL:</b>	file path TBD
<b>Annotation file type:</b>	GFF
<b>Annotation file checksum:</b>	TBD

## AGAT Full Stats

### Counts

<b>AGAT stats calculated from:</b>	annotated transcripts (without isoforms)
<b>Genes:</b>	264
<b>Transcripts:</b>	264
<b>Exons:</b>	1312

<b>CDS:</b>	264
<b>Introns:</b>	1048
<b>Single exon genes:</b>	74
<b>Single exon transcripts:</b>	74

### Mean Ratios

<b>Transcripts per gene:</b>	1
<b>Exons per transcript:</b>	5
<b>Exons per CDS:</b>	5
<b>CDS per transcript:</b>	1
<b>Introns per transcript:</b>	4

### Mean Lengths (bp)

<b>Gene:</b>	1758.89
<b>Transcript:</b>	1758.89
<b>Exon:</b>	248.53
<b>CDS:</b>	1235.14
<b>Intron:</b>	131.94

### Median Lengths (bp)

<b>Gene:</b>	1587.5
<b>Transcript:</b>	1587.5
<b>Exon:</b>	134.5
<b>CDS:</b>	134.5
<b>Intron:</b>	92

## Total Lengths (bp)

<b>Genes:</b>	464346
<b>Transcripts:</b>	464346
<b>Exons:</b>	326076
<b>CDS:</b>	326076
<b>Introns:</b>	138270

## Longest Features (bp)

<b>Gene:</b>	8111
<b>Transcript:</b>	8111
<b>Exon:</b>	2748
<b>CDS:</b>	4575
<b>Intron:</b>	2081

## Shortest Features (bp)

<b>Gene:</b>	201
<b>Transcript:</b>	201

## BUSCO Full Stats

<b>Version:</b>	5.8.3
<b>Lineage dataset:</b>	embryophyta_odb10
<b>Mode:</b>	proteins
<b>One line summary:</b>	C:1.1%[S:1.0%,D:0.1%],F:0.2%,M:98.7%,n:1614
<b>Complete BUSCOs:</b>	1.1%

<b>Single-copy:</b>	1%
<b>Duplicated:</b>	0.1%
<b>Fragmented:</b>	0.2%
<b>Missing:</b>	98.7%
<b>Total markers:</b>	1614

## OMArk Full Stats

### Assessment information

<b>OMAmer version:</b>	2.1.0
<b>OMAmer database version:</b>	2.0.3

### Completeness Assessment

<b>Clade:</b>	Brassicaceae
<b>Total HOGs:</b>	17996
<b>Single-copy:</b>	1.1%
<b>Total duplicated:</b>	0.09%
Duplicated (expected):	0.01%
Duplicated (unexpected):	0.08%
<b>Missing:</b>	98.81%

### Consistency Assessment

<b>Total proteins:</b>	288
<b>Consistent:</b>	97.57%
Partial consistent:	0%

Fragmented consistent: 2.08%

**Inconsistent:** 0.35%

Partial inconsistent: 0%

Fragmented inconsistent: 0%

**Contaminants:** 0%

Partial contaminants: 0%

Fragmented contaminants: 0%

**Unknown:** 2.08%

## Detected Species

Clade	Taxon ID	Protein count	Proteome %
-------	----------	---------------	------------

Arabidopsis thaliana	3702	282	97.92
----------------------	------	-----	-------