



# In Silico Structural and Functional Mapping of *Arabidopsis thaliana* Chromosome 5



**Abstract** : We analyzed a 1,404 bp nucleotide sequence and after analyzing through BLASTn we found out that this sequence has a 100 percent similarity with *Arabidopsis thaliana* genome assembly chromosome:5, that has 29-megabase segment..Multi-level analyses reveal a conserved, **chloroplast-targeted light-harvesting chlorophyll a/b binding protein**, highlighting the power of computational biology in linking sequence data to photosynthetic function.

AATGGGCAACCAGAAAAGAAATATGGATAAAAAAGAGATGGAATTAGAGATACTTCAATCCAAAGCTGTTTCCGTATTGGCTAATCCCAACATCTCAC  
TCTCTATCTATATCCCATCTCTGATTTCTCCACCTCTCTCTCATCCCAAAATACTAACATAAAAGTCAAAGTCCCTGAGACCAATCTTCACCAAC  
AGCAAAAACAAGAAAGAACAACTAAGCAAAAAGAAAGCTCAAGCCGAGAGAGACAATGGCATCAACATTACAGAGCTCAAGCAGTGTCTTTA  
CCCCAACCAATTCCTTGGCCAGACTAAGGCTCAAGCTTTAACCCCTTCGTGATGTTGTCTCTCGGATCTCCCAAGTACACTATGGTCTGTTTCTT  
TTTTGGTCTTAATCTGGTCATATTAAGGTGTTTTCATCTTCTTGTAGTTTGTGTCTGACTTTCAGTAAATGCTATGTGAAGGGAATGATCTT  
TGGTATGGACCTGACAGAGTGAAGTACTTAGGACCTTTTCGGTTCAAACTCCGTCTTACCTCACCGGAGAATTCCTGGCGATTATGGTTGGGACCC  
GCCGTTTATCCGAGACCTGAAGCCTTTGCCAAAACAGAGCTCTTGAGGTAGTCTCTAAATCTCACTCACTCAATGTAAAAACAGAGTAATAAC  
TAATCATCTTTGCAGGTGATCCATGGGAGATGGGCAATGTTGGGAGCTTTGGTTGCATAACCCCTGAAGTCTCTCAAAGTGGGTCCGTGTGGACTT  
CAAGAACCAGTCTGGTTCAAAGCCGTTTCAAAATCTTCCGGAAGGCGTTTGGACTACTTAGGCAACCCCAACCTAGTCCATGCTCAGAGCATTTT  
AGCCGTCTTGGCTTCCAAGTTCATCTCATGGGTTTGGTTGAAGGTTTCCGCATCAACGCTTTAGTGGTGTGGCGAAGGCAACGACTTGTACCCGG  
TGGCAATACTTTGACCGGTTGGGTCTCGCTGATGATCCAGTTACTTTGCTGAGCTTAAAGTGAAGAGATCAAGAACGGAAGATTGGCTATGTTCTC  
TATGTTTGGCTTCTTGTTCAGGCAATGTTACTGGAAAAGGTCCTTTGGAGAATCTCCTTGACCATCTTGACAACCTCTTGTCTCAACAATGCA  
TTTCGCAACTAAGTTTGACCTGGAGCTTAAATTTCAAGTCTTGTAGTGAATCTTGAATGATGAAACATAACCTTTCTTGTCTCTCAAAATGACA  
AATCTTCTATTGCTATGCACAGAAGAGTCTCAAAATACATGTCTCACAGCTAAATCTACAATGCTCTTAATCTAACATTTTGGATCGTGATCT  
AATCAAGCTAAGTGG

DNA SEQUENCE ANALYSIS→RNA ANALYSIS→PROTEIN ANALYSIS→SUBCELLULAR LOCALIZATION→PHYLOGENETIC ANALYSIS→PROTEIN-PROTEIN INTERACTION

## DNA Sequence Analysis (Gene-Level)

❖ NCBI BLASTn

Arabidopsis thaliana genome assembly, chromosome: 5

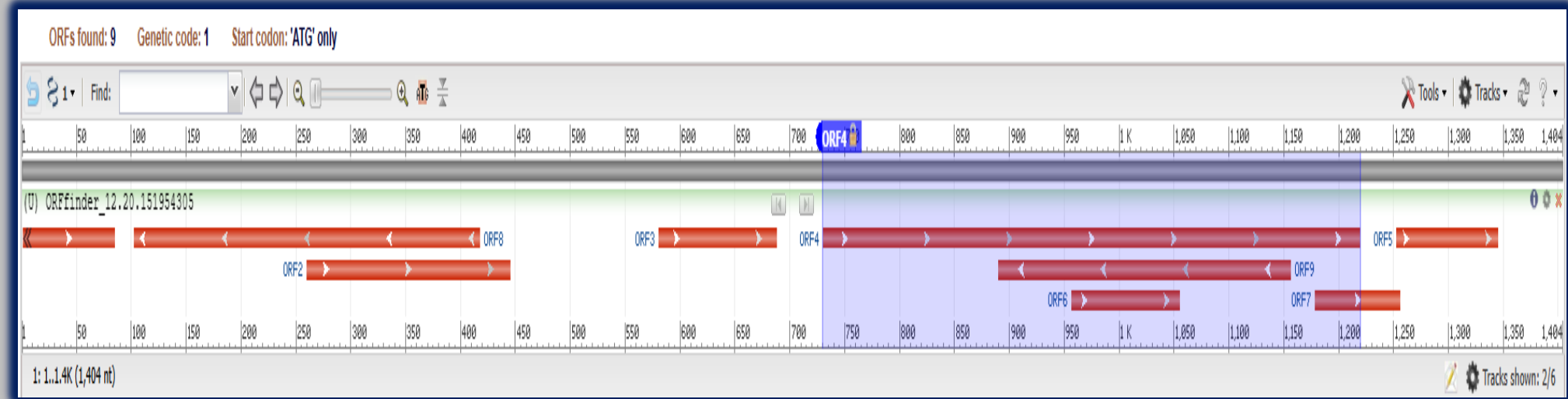
Sequence ID: **OZ364833.1** Length: 2902600 Number of Matches: 1

Range 1: 24096845 to 24098248 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
2593 bits(1404)	0.0	1404/1404(100%)	0/1404(0%)	
Query 1	AATGGGCAACCAGAAAAGAAATATGGATAAAAAAGAGATGGAATTAGAGATACTTCAAA	60		
Sbjct 24096845	AATGGGCAACCAGAAAAGAAATATGGATAAAAAAGAGATGGAATTAGAGATACTTCAAA	24096904		
Query 61	CAAAAGCTGTTCTGATGGGTAATCCCAAGATCTCACTCTCTATATATATATATATATAT	120		
Sbjct 24096905	CAAAAGCTGTTCTGATGGGTAATCCCAAGATCTCACTCTCTATATATATATATATATAT	24096964		
Query 121	TGATTTCCCAAGCTCTCTCTCACTCCCAAAATATCAACATAAAAGTCAAGTCCCTGAG	180		
Sbjct 24096965	TGATTTCCCAAGCTCTCTCTCACTCCCAAAATATCAACATAAAAGTCAAGTCCCTGAG	24097024		
Query 181	ACCAATATCTTTCACCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCA	240		
Sbjct 24097025	ACCAATATCTTTCACCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCA	24097084		
Query 241	GCTCAAGCTGAGAGAGCAAGATGCAATCAACATCAGCAAGCTCAAGACAGTGTCTTACCC	300		
Sbjct 24097085	GCTCAAGCTGAGAGAGCAAGATGCAATCAACATCAGCAAGCTCAAGACAGTGTCTTACCC	24097144		

❖ ORF FINDER



❖ Expasy  
Expasy Translate

BLAST analysis reveals 100% identity with [*Arabidopsis thaliana*]

5' Frame 1  
MLGAPGCTPEVLYQWVDEPEPWFYAGSGLIFSGSLDVLGNLVHAQSILVGFQVILNGLVGRINGLGVGENDLUPGGQVDFPLGLADPVTFAELKVEIKINGRLAMESE  
GFFVQLVIGRQPLDNLCHLDNFWANMAFATKPAQA-

1. Length threshold (>100 AA).
2. Presence of canonical start/stop codons.
3. Conservation of high-identity domains.

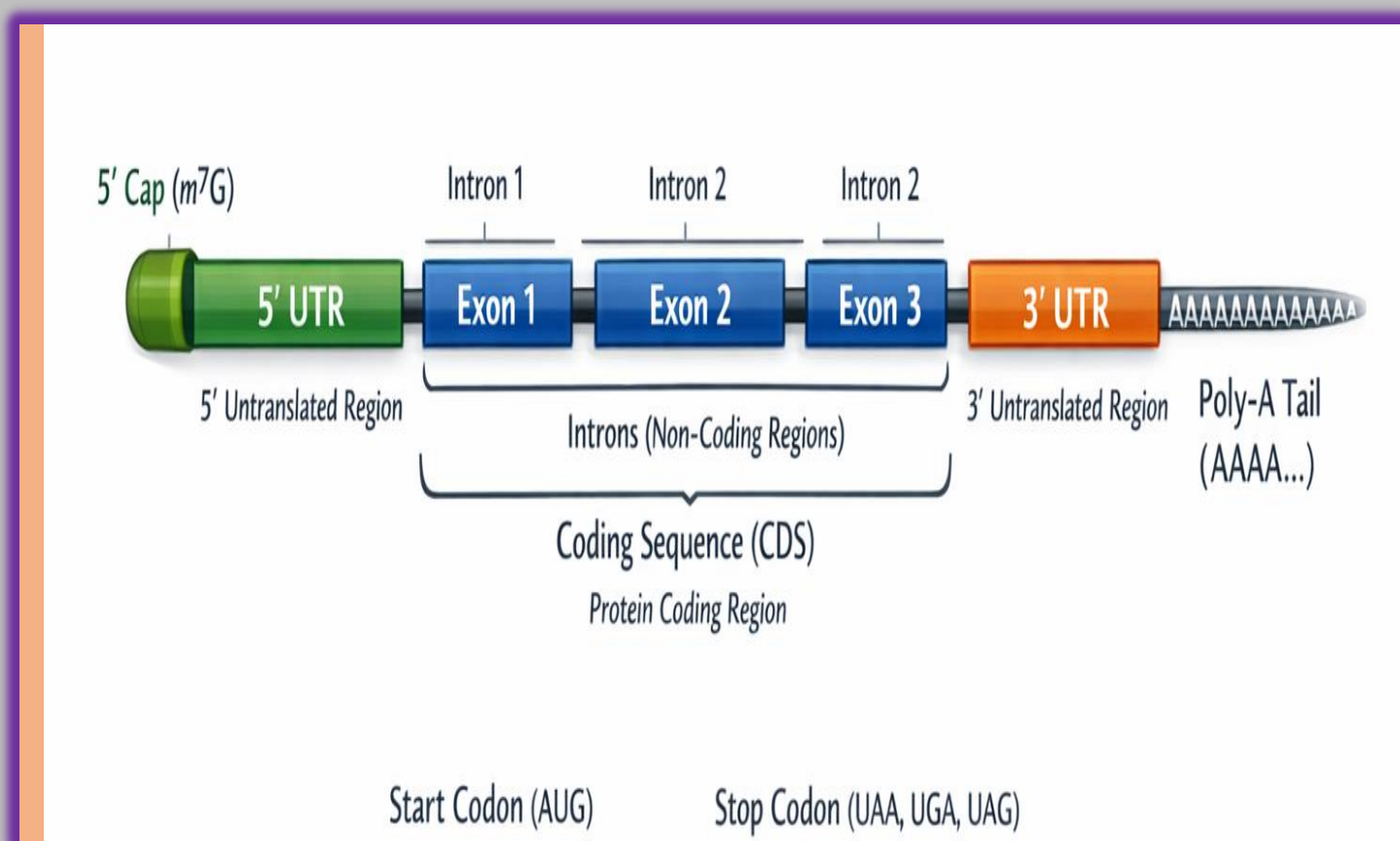
**GC CONTENT: 42.5**  
**REPEATS: TOTAL 6 REPEATS**

Top Three Hits

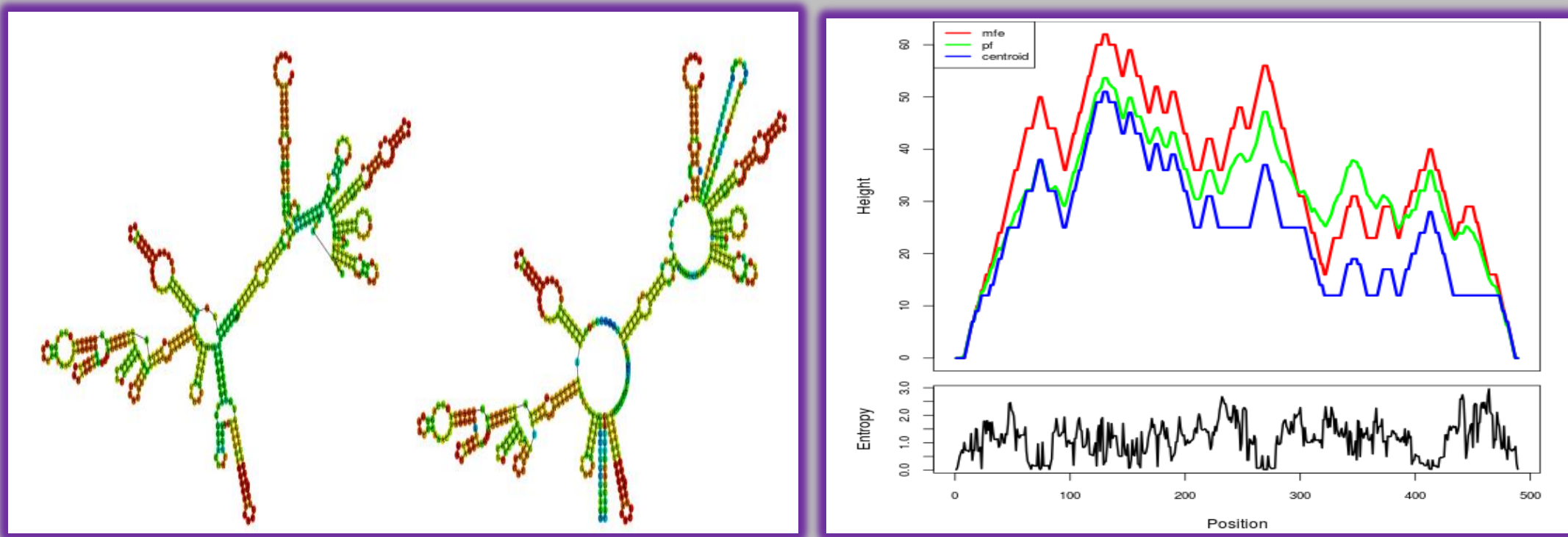
Description	Scientific Name	Common Name	Taxid	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Arabidopsis thaliana genome assembly, chromosome: 5	Arabidopsis thaliana	thale cress	3702	2593	2593	100%	0.0	100.00%	2902600	OZ364833.1
Arabidopsis thaliana genome assembly, chromosome: 5	Arabidopsis thaliana	thale cress	3702	2593	2593	100%	0.0	100.00%	2750405	OZ364823.1
Arabidopsis thaliana genome assembly, chromosome: 5	Arabidopsis thaliana	thale cress	3702	2593	2593	100%	0.0	100.00%	2882578	OZ364818.1

## RNA Analysis (mRNA Features)

❖ UTR SCAN



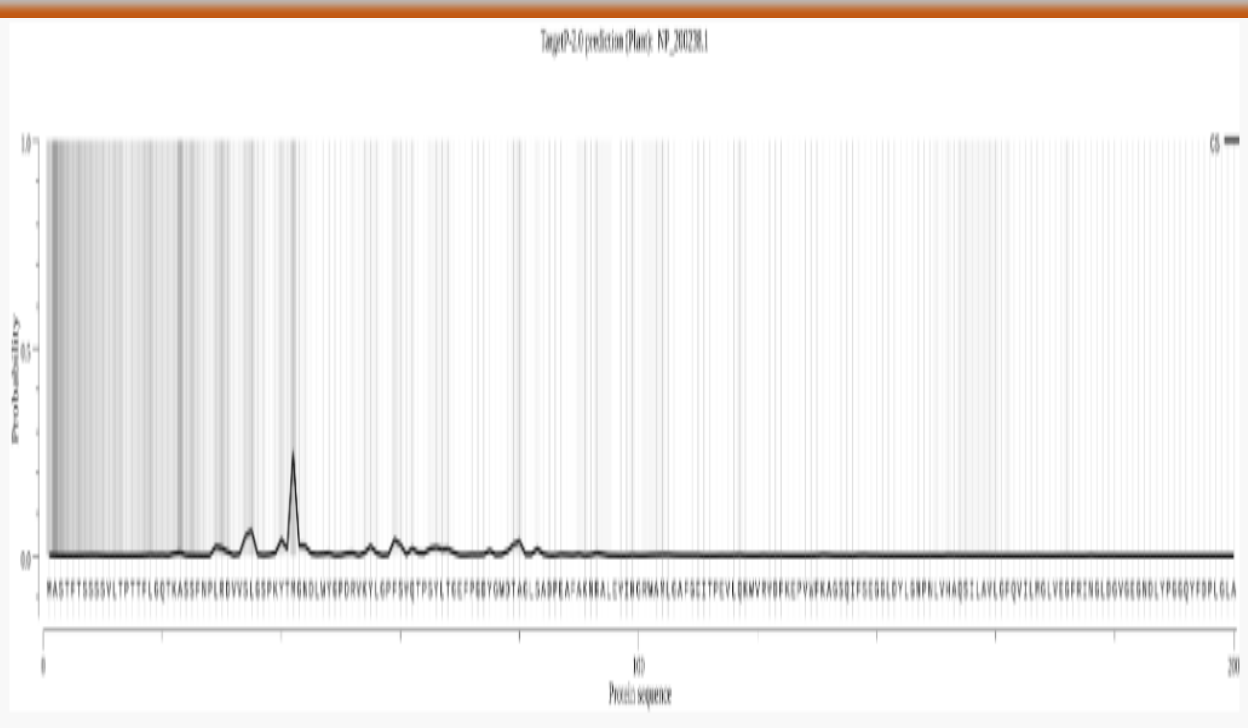
❖ RNA FOLD



-145.90 KJ/mol  
indicate more stable and spontaneous fold

## Subcellular Localization Prediction

❖ TARGETp



❖ CELLO

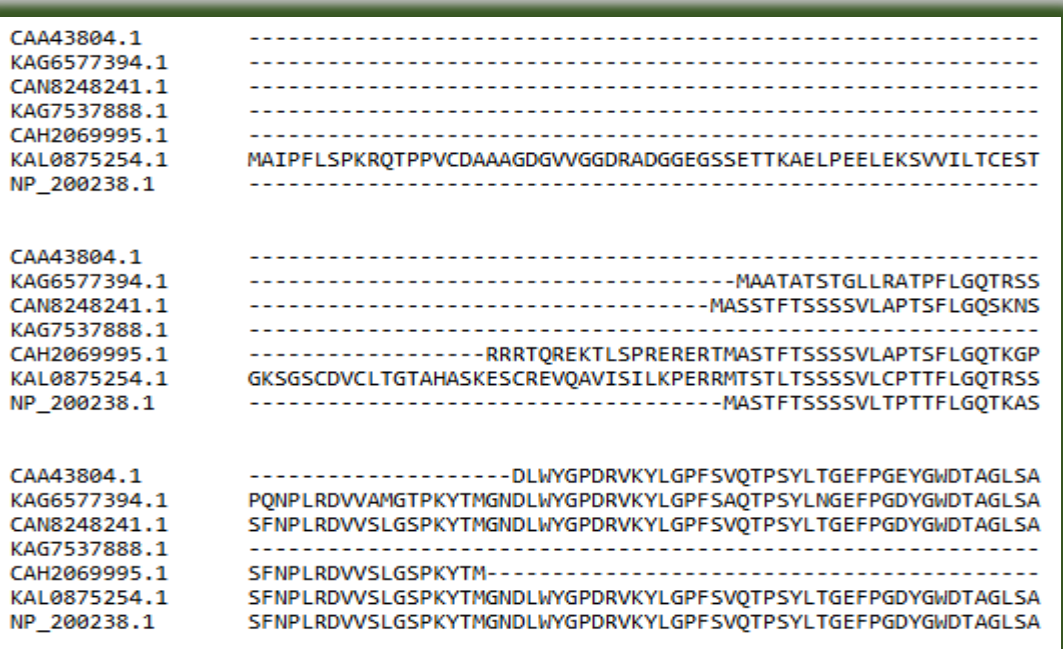
CELLO RESULTS		
SeqID: NP_200238.1 light-harvesting chlorophyll B-binding protein 3 [Arabidopsis thaliana]		
Analysis Report:	LOCALIZATION	RELIABILITY
SVM	Chloroplast	0.361
Amino Acid Comp.	Chloroplast	0.739
N-peptide Comp.	Chloroplast	0.949
Partitioned seq. Comp.	Chloroplast	0.826
Physico-chemical Comp.	Chloroplast	0.894
Neighborhood seq. Comp.	Chloroplast	0.894
CELLO Prediction:	Chloroplast	3.768 *
	Vacuole	0.390
	Cytoplasmic	0.173
	ER	0.157
	PlasmaMembrane	0.126
	Lysosomal	0.122
	Mitochondrial	0.095
	Extracellular	0.083
	Peroxisomal	0.052
	Nuclear	0.019
	Golgi	0.013
	Cytoskeletal	0.005

Predicted Location and Justification

It is a protein of chloroplast thylakoid membrane. Predicted as a chloroplast targetted protein via an N-terminal transit peptide, LHCb3 localizes to thylakoid membrane where it integrates into the LHCII heterotrimeric complex to facilitate light absorption and energy transfer for Photosystem II.

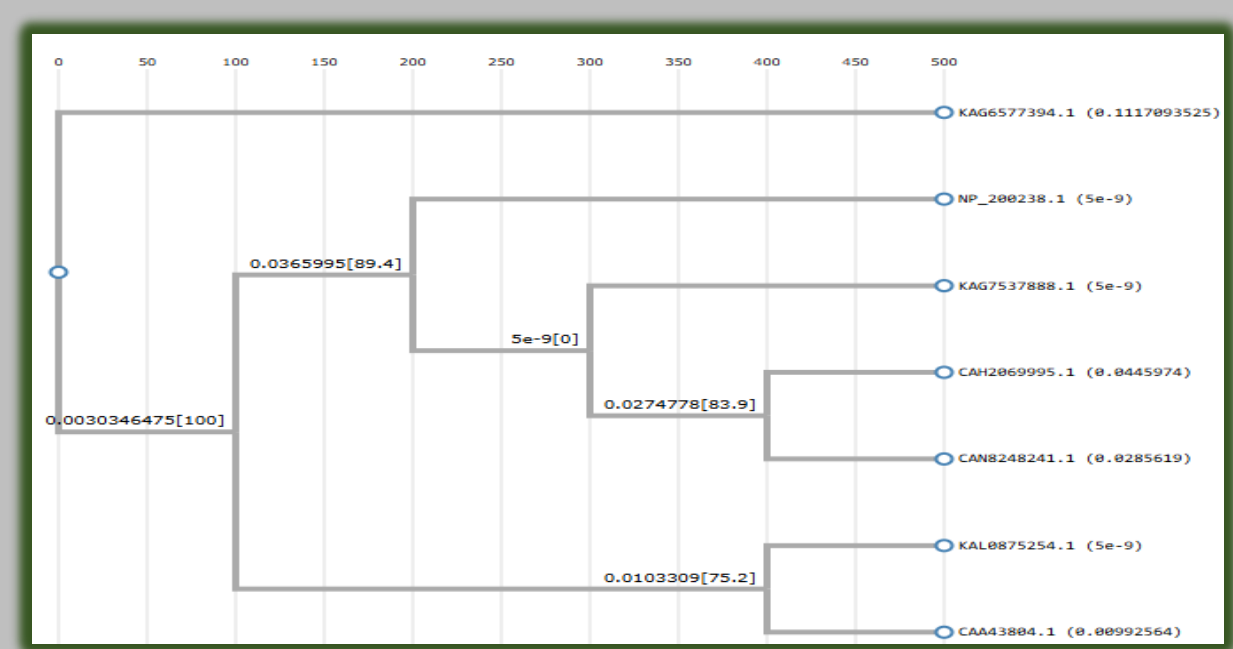
## Phylogenetic Analysis

MSA



❖ CLUSTAL OMEGA

PHYLOGENETIC TREE



This phylogenetic tree shows the evolutionary relationships among homologous protein sequences, where closer branch clustering indicates higher sequence similarity and shared ancestry. Branch lengths reflect evolutionary distance, and bootstrap values indicate the confidence of each node.

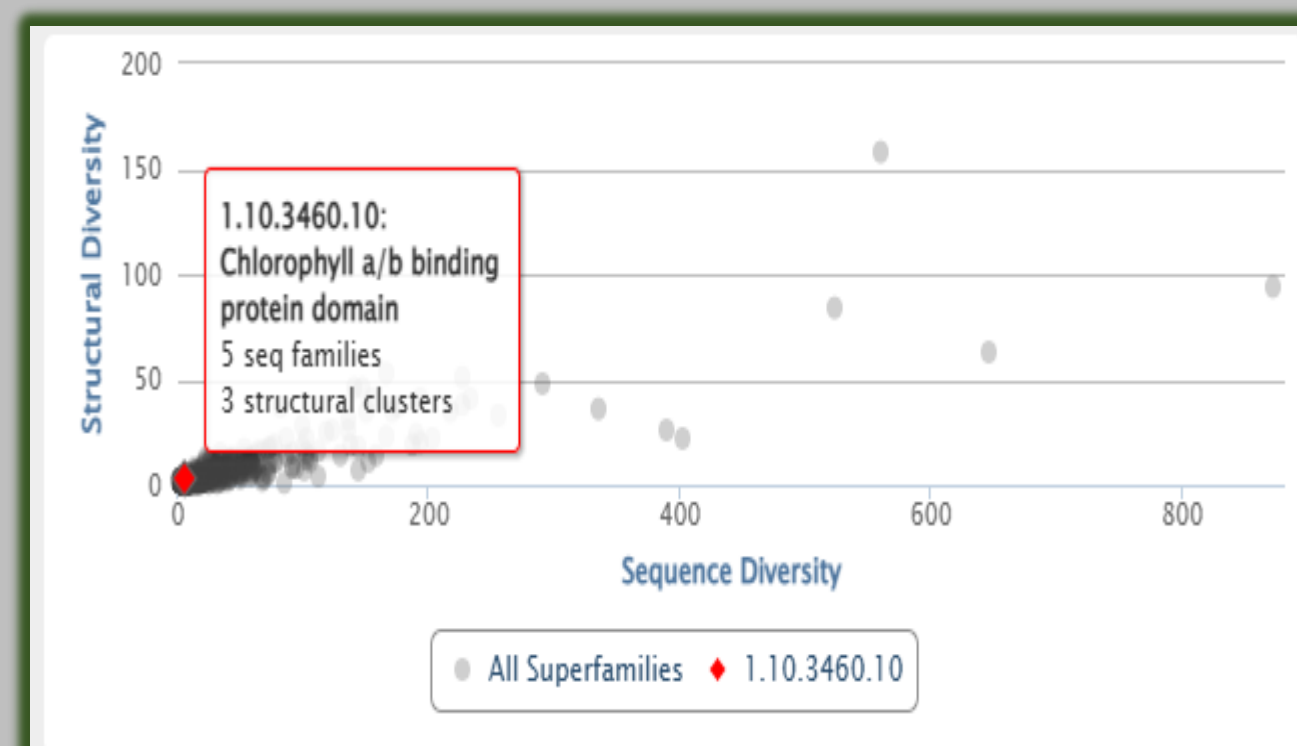
## Protein Analysis (Structure & Properties)

❖ PROTPARAM  
Protein Properties

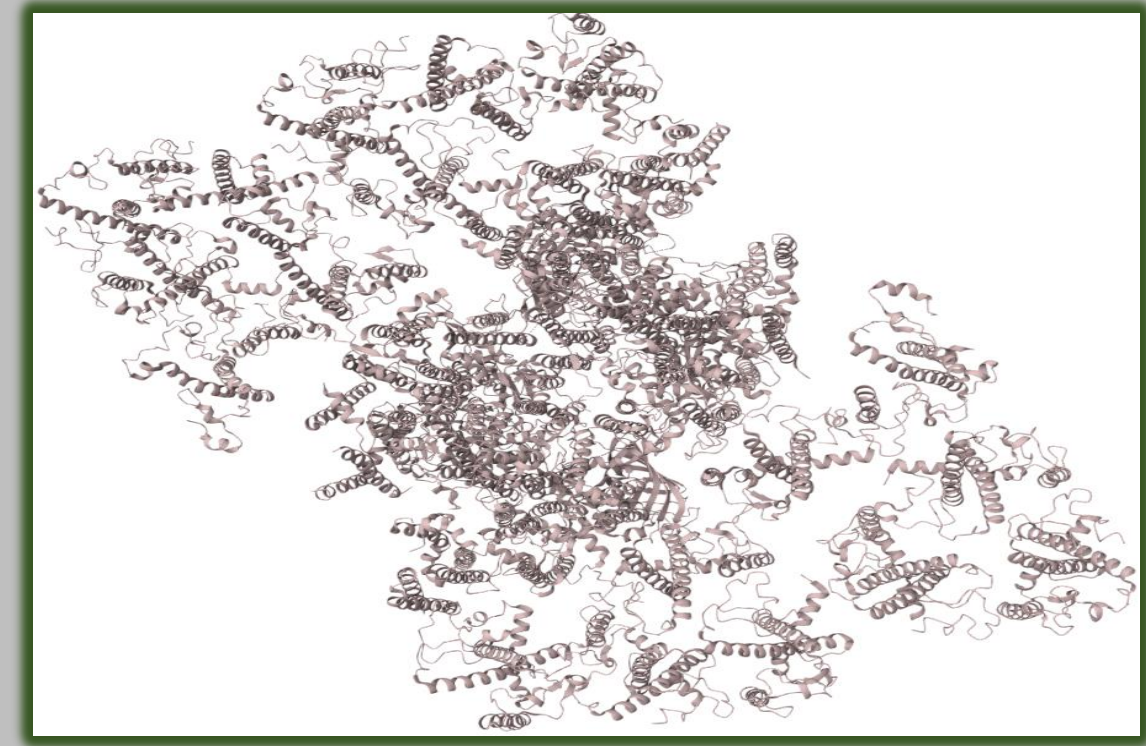
MW	28706.68
pI	4.96
AI	83.17
II	22.10

❖ Pfam

DOMAINS



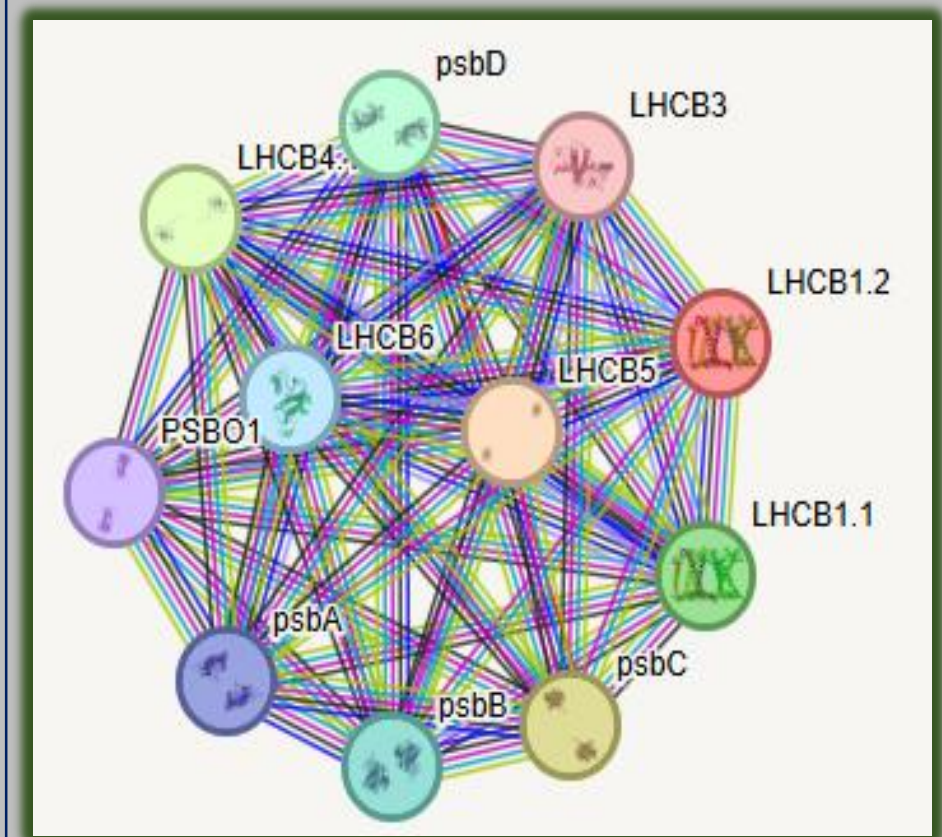
❖ Predict Protein  
Secondary Structure



Secondary structure revealed high density of hydrophobic alpha-helices consistent with its role as a trans membrane pigment-binding protein.

## Protein-Protein Interaction Analysis

❖ STRING



This interaction network highlights the tight functional coordination among light-harvesting chlorophyll a/b binding proteins and photosystem core subunits, which together optimize light capture and energy transfer during photosynthesis. The dense connectivity reflects their essential role in maintaining photosystem II stability and photosynthetic efficiency.

## Conclusion

This in silico study integrates DNA, RNA, protein, phylogenetic, and interaction analyses to comprehensively characterize a gene from *Arabidopsis thaliana* chromosome 5. The results confirm its identity as a chloroplast-localized light-harvesting chlorophyll a/b binding protein with a conserved structure and interaction network essential for efficient photosynthesis.

Arabidopsis Genome Initiative (2000); Altschul et al. (1990); Tamura et al. (2021).