Genome-wide identification and expression profiling of *Alba* gene family members in response to abiotic stress in tomato (*Solanum lycopersicum* L.)

Antt Htet Wai
Yangon University of Education https://orcid.org/0000-0001-6822-368X
Lae-Hyeon Cho
Pusan National University
Muhammad Waseem
South China Agricultural University
Do-jin Lee
Sunchon National University
Je-Min Lee
Kyungpook National University
Chang-Kil Kim
Kyungpook National University
Mi-Young Chung (queen@scnu.ac.kr)
Sunchon National University

Research Article

Keywords: Alba genes, Solanum lycopersicum, abiotic stress, genome-wide analysis, expression profiling, subcellular localization, 3D structure

Posted Date: July 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-701131/v1

Version of Record: A version of this preprint was published at BMC Plant Biology on November 12th, 2021. See the published version at https://doi.org/10.1186/s12870-021-03310-0.

Abstract

Alba (Acetylation lowers binding affinity) proteins are an ancient family of nucleic acid-binding proteins that function in gene regulation, RNA metabolism, mRNA translatability, developmental processes, and stress adaptation. Here, we undertook the first comprehensive genome-wide characterization of the *Alba* gene family in tomato (*Solanum lycopersicum* L.). We identified eight tomato *Alba* genes, which were classified into two groups: genes containing a single Alba domain and genes with a generic Alba domain and RGG/RG repeat motifs. Cis-regulatory elements and target sites for miRNAs, which function in plant development and stress responses, were prevalent in *SlAlba* genes. To explore the structure–function relationships of tomato Alba proteins, we predicted their 3D structures, highlighting their likely interactions with several putative ligands. Confocal microscopy revealed that SlAlba–GFP fusion proteins were localized to the nucleus and cytoplasm, consistent with putative roles in various signaling cascades. Expression profiling revealed the differential expression patterns of most *SlAlba* genes across diverse organs. *SlAlba1* and *SlAlba2* were predominantly expressed in flowers, whereas *SlAlba5* expression peaked in 1 cm-diameter fruits. The *SlAlba* genes were differentially expressed (up- or downregulated) in response to different abiotic stresses. Furthermore, all but one of these genes were induced by abscisic acid treatment, pointing to their possible regulatory roles in stress tolerance via an abscisic acid-dependent pathway. Our characterization of *SlAlba* genes should facilitate the discovery of additional genes associated with organ and fruit development as well as abiotic stress adaptation in tomato.

1. Introduction

Alba (Acetylation lowers binding affinity) superfamily proteins belong to an ancient group of nucleic acid-binding proteins that originated prior to the divergence of archaea and eukaryotes and are widely distributed in nearly all kingdoms of life (Aravind et al., 2003; Goyal et al., 2016). These small, basic proteins interact with DNA and RNA as homodimers or heterodimers (Crnigoj et al., 2013; Jelinska et al., 2005) and contain a conserved nucleic acid-binding Alba domain with an IF3-C fold (Aravind et al., 2003). Alba proteins also show the distinct characteristic of being regulated via acetylation by PAT (protein acetyl transferase) and deacetylation by Sir2 (a NAD^+-dependent histone deacetylase, HDAC), indicating that they function as transcriptional regulators similar to histones (Bell et al. 2002; Zhao et al., 2003). Although Alba proteins were initially classified as chromosomal proteins and were considered to function in the maintenance of chromatin architecture and transcriptional repression, many studies have documented their functional diversity, including their roles in genome packaging and organization, transcriptional and translational regulation, RNA metabolism, and development and differentiation processes (Goyal et al., 2016).

Alba proteins exist in two forms: relatively small proteins with a single generic Alba domain; and larger proteins with an Alba domain plus RGG/RG repeat motifs or additional domain/s (Goyal et al., 2016; Náprstková et al., 2021). All archaeal Alba homologs, which belong to the Archaea Alba protein family,

are relatively small proteins with only a single Alba domain. Eukaryotic Alba proteins diverged into two paralogous lineages, the Rpp20-like family (with single Alba domains) and the Rpp25-like family (mostly larger proteins with RGG/RG repeat motifs or additional domains) (Aravind et al., 2003). The RGG/RG repeat motifs are thought to have roles in DNA damage signaling, SnRNP biogenesis, the regulation of apoptosis, transcription, pre-mRNA splicing, and translation, many of which are at least partly controlled by the arginine methylation of RGG/RG repeats (Thandapani et al., 2013).

Alba genes encoding proteins with RGG/RG repeat motifs have been identified in many species (Dupé et al., 2014; Gissot et al., 2013; Náprstková et al., 2021). Various types of domains with diverse biological functions have also been identified in combination with the Alba domain in numerous Alba family proteins in many domains of life, such as the F-box domain in the fungus *Taphrina*; the CLIP1 zinc finger domain in the nematode *Pristionchus*; the ATP synthase subunit H domain in the protozoan parasite *Theileria*; the sugar transporter domain in the Florida lancelet; the DEAD/DEAH box helicase domain in plants; and so forth (Goyal et al., 2016). The addition of RGG/RG repeat motifs and new domains to genes with a single Alba domain appears to have contributed to speciation and the functional diversification of Alba family proteins to meet the demands of increasing cellular variety and complexity.

Diverse abiotic stresses, including drought, salinity, heat, and cold stress, can hamper plant growth and development, resulting in significantly reduced crop productivity (Wai et al., 2020). Various environmental stresses trigger changes in *Alba* behavior. For example, LiAlba1 and LiAlba3, from the protozoan parasite *Leishmania infantum*, repress translation by interacting with RNA-binding proteins and ribosomal subunits. These proteins also function as translation factors and translocate from the cytoplasm into the nucleolus and flagellum in response to heat stress (Dupé et al, 2015). Similarly, all four Alba proteins of *Trypanosoma brucei* (TbAlba1, TbAlba2, TbAlba3, and TbAlba4) are RNA-binding proteins that localize to the cytoplasm as parts of stress granules (SG) upon exposure to nutrient stress (Mani et al., 2011). Stress-induced differential expression has also been observed for *Alba* genes from rice treated with different abiotic stresses and phytohormones (Verma et al., 2018). Mild heat stress (37 ◦C) altered the expression of most Arabidopsis *Alba* genes in inflorescences (Náprstková et al., 2021). In cotton (*Gossypium hirsutum*), *GhALBA4* and *GhALBA5* were significantly induced by water deficit and salinity treatment, and plants in which the expression of these genes was repressed by virus-induced gene silencing (VIGS) were highly sensitive to dehydration as well as salt stress, pointing to their putative roles in abiotic stress tolerance (Magwanga et al., 2019).

Genome-wide identification of *Alba* genes has been performed in several plant species, but no comprehensive study of the evolutionary relationships or characteristics of the *Alba* gene family in solanaceous crops has been reported. Tomato (*Solanum lycopersicum* L.) is an economically important fruit whose yields are severely affected by unfavorable environmental conditions (Wai et al., 2020). The molecular mechanisms regulating fruit development and ripening have been extensively studied in this model fruit crop with the aim of increasing fruit yield under diverse environmental stresses. Hence, in the current study, we performed a comprehensive genome-wide analysis and expression profiling of *SlAlba*

genes in response to various abiotic stresses. Our findings provide a basis for further functional characterization of *Alba* genes involved in the development and stress tolerance of tomato.

2. Materials And Methods

2.1. In silico identification and sequence analysis of tomato *Alba* genes

We identified tomato Alba gene family members from the Phytozome database (http://www.phytozome.net) using the keyword “Alba” and performed BLAST searches of the Sol genomics database (http://www. solgenomics.net/tools/blast/) using Arabidopsis *thaliana* Alba protein sequences obtained from TAIR (https://www.Arabidopsis.org/) as queries. To validate the presence of the Alba domain, the resulting eight non-redundant Alba protein sequences were subjected to searches using the NCBI CDD search (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) and the SMART web tool (http://smart.emblheidelberg.de/). The exon/intron structures were visualized using the Gene Structure Display Server-2.0 web server (http://gsds.cbi.pku.edu.cn/index.php) by loading both *SlAlba* genomic and coding sequences. The protein length (number of amino acids), molecular weight, GRAVY values (grand average of hydropathicity index), and isoelectric points of the SlAlba proteins were determined using ProtParam (http://cn.expasy.org/tools/protparam.html). The Open Reading Frame finder tool (https://www.ncbi.nlm.nih.gov/orffinder/) was used to analyze the open reading frames of the *SlAlba* genes. Clustal Omega (Clustal Omega < Multiple Sequence Alignment < EMBL-EBI) and ESPript web server (https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) were used for multiple-protein sequence alignment. The web tool “Immunomedicine Group” (http://imed.med.ucm.es/Tools/sias.html) was employed to analyze the sequence homology of the eight Alba proteins. The MEME suite motif search tool (http://memesuite.org/) was used to examine the conserved motifs in protein sequences with the following parameters: maximum number of motifs 15 and motif length between six and 50 amino acids. WoLF-PSORT (https://wolfpsort.hgc.jp/) was used to predict the subcellular localizations of the SlAlba proteins.

2.2. Phylogenetic analysis

The full-length Alba protein sequences retrieved from the Phytozome and NCBI databases were aligned using Clustal Omega, followed by phylogenetic analysis via the neighbor-joining (NJ) method with 1000 bootstrap replications in MEGA 6.0 (Tamura et al., 2013). The names of the genes used to build the phylogenetic tree, together with their accession numbers, are described in Table S1.

2.3. Prediction of miRNA target sites, cis regulatory elements, and chromosomal locations

The putative miRNA targets were analyzed using the psRNATarget web tool (http://plantgrn.noble.org/psRNATarget/analysis). The PlantCare database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was employed to investigate the promoter region of each gene 1500 bp upstream of the initiation codon [ATG] to predict putative cis-acting elements in these promoters. The *SlAlba* genes were mapped to chromosomes using the online tool

MapGene2Chrom web v2 (http://mg2c.iask.in/mg2c_v2.0/) after determining their chromosomal positions from the Sol genomic database.

2.4. Gene Duplication and Microsynteny Analysis

Gene duplications across tomato *Alba* genes were predicted with the TBtools software (Chen et al., 2020). The synonymous (Ks) and non-synonymous (Ka) nucleotide substitution rates of the *SlAlba* genes were computed following the method of Nei and Gojobori (1986). The Ka/Ks ratio was analyzed to determine the mode of selection (Nekrutenko et al., 2002). The formula T = Ks/2r MYA (millions of years ago) was used to estimate the divergence time (T) for the duplicated gene pairs. Ks is the synonymous substitution rate per site, and r is the constant for dicot plants of 1.5×10^{-8} substitutions per site per year (Koch et al., 2000). Using a reciprocal BLAST search against the entire genomes of tomato, Arabidopsis, and rice, the microsyntenic relationship of *Alba* genes across these species was analyzed. The TBtools software was used to visualize the duplicated gene pairs (Chen et al., 2020).

2.5. Comparative modeling of tomato Alba proteins

The 3D structures of the SlAlba proteins were predicted with the I-TASSER server using the protein sequences of SlAlba1-8 as input (Yang et al., 2015). 3D models were generated by multiple threading of the alignments with LOMETS and performing iterative TASSER assembly simulations. The best modeled structures with the maximum scores were selected and the template analogs were also identified. The resulting molecular models were refined with ModRefiner (Xu et al., 2011). The functions of the modeled proteins were predicted using the I-TASSER server based on global and local similarity to template proteins in PDB with known structures and functions.

2.6. Subcellular localization

The tomato *Alba* cDNA was amplified using gene-specific primers (Table S2) and cloned into the pGA3452 vector harboring the maize *Ubi1* promoter, the synthetic GFP coding region, and the nos terminator (Kim et al., 2009). NLS-mRFP under the control of the 35S promoter was employed as a nuclear marker. The constructs were cotransformed via electroporation into protoplasts prepared from rice Oc cells, and the resulting transformants were incubated overnight at 28°C in the dark for 12 to 16 h. Fluorescent signals were visualized under a filter-equipped microscope (BX61; Olympus, Tokyo, Japan) using bright field illumination, the GFP channel, and the red fluorescent protein channel.

2.7. Preparation of plant materials and stress treatments

Tomato plants (*Solanum lycopersicum* cv. Ailsa Craig) were grown in soil in a growth room under controlled conditions at a temperature of 25 °C during the day and 20 °C at night and a 16 h light 8 h dark photoperiod. For tissue-specific expression analysis, fresh roots, stems, and leaves were collected from 28-day-old plants. The remaining plants were transferred to a greenhouse maintained at 25/20 °C day/night temperatures until the reproductive stage to collect flower and fruit samples. Three different

types of flower samples were harvested: floral buds, full-blooming flowers, and senescent flowers. Fruit samples were collected at six different stages: (i) young fruits approximately 1 cm in diameter at 2 weeks after pollination (1 cm fruits) (ii) immature fruits at 5 weeks after pollination (IM fruits), (iii) mature green fruits at 7 weeks after pollination (MG fruits), (iv) breaker fruits, when the color of mature fruits changes from green to faint yellow-orange (B fruits), (v) breaker after 3 days (B3 fruits), and (vi) breaker after 7 days (B7 fruits) (Wai et al., 2021).

To study expression patterns of *SlAlba* genes under different abiotic stress conditions (abscisic acid [ABA], heat, cold, salt [NaCl], and drought), leaf samples from 28-day-old seedlings of uniform growth and development were collected at 0, 1, 3, 6, 12, and 24 h after the start of the stress treatments (Wai et al., 2021). ABA treatment was applied by spraying leaves with 100 µM ABA. To impose heat and cold stress, the plants were incubated in a growth cabinet at 40 °C and 4 °C, respectively. Salt stress was imposed by submerging roots in a 200 mM NaCl solution. For drought treatment, whole plants were gently pulled out from the soil, carefully cleaned with fresh water, and incubated on dry paper towels at 25 ± 1 C without additional water supply (Khatun et al., 2016,2017). Plants grown in soil under normal conditions (25 °C) were used as the 0 h controls for all stress treatments. The samples were collected from three biological replicates, immediately frozen in liquid nitrogen, and stored at −80 °C for RNA isolation.

2.8. RNA extraction and qPCR expression analysis

Total RNA was isolated from the samples using an RNeasy Mini kit (Qiagen, Hilden, Germany) and purified with an RNase-free DNase I kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's protocols. RNA concentrations were determined with a NanoDrop® 1000 spectrophotometer (Wilmington, DE, USA), and 1 µg total RNA was used to synthesize cDNA using a Superscript® III First-Strand cDNA synthesis kit (Invitrogen, Carlsbad, CA, USA) as per the manufacturer's instructions. Primer3 software (http://frodo.wi.mit.edu/primer3/input.htm) was used to design the gene-specific primers for all tomato *Alba* genes (Table S3). Melting curve analysis was performed to verify the specificity of the amplicon for each primer pair (Verma et al., 2018), and the expression of 18S rRNA (F: AAAAGGTCGACGCGGGCT, R: CGACAGAAGGGACGAGAC) was used as an internal control for normalization (Balestrini et al., 2007). qRT-PCR was performed in a 10 µL reaction mixture consisting of 1 µL (50 ng) cDNA, 2 µL forward and reverse primers (5 pmol concentration), 5 µL of iTaq SYBR Green (Qiagen, Hilden, Germany), and 2 µL double distilled water. A Light cycler® 96SW 1.1 (Roche, Germany) was used to amplify and record the Cq value of each sample with the following conditions: pre-denaturation at 95 °C for 5 min followed by 40 cycles at 94 °C for 10 s, annealing at 58 °C for 10 s, and extension at 72 °C for 15 s. The $2^{-\Delta\Delta Ct}$ method was used to determine the relative transcript levels of each gene against each treatment (Schmittgen and Livak, 2008). Statistical analysis of the relative expression data was conducted with SigmaPlot 12.1. (SYSTAT and MYSTAT Products, United States, and Canada) using two-tailed Student's t-tests.

3. Results

3.1 Identification and sequence analysis of Alba family proteins in tomato

We identified eight non-redundant putative *Alba* genes in tomato, which were named *SlAlba1* to *SlAlba8* in accordance with their positions on the chromosomes. The open reading frames of the *SlAlba* genes showed considerable variation in length, ranging from 384 bp to 1188 bp for *SlAlba5* and *SlAlba7*, respectively, with a mean of 701 bp. The predicted sizes of the eight SlAlba proteins varied from 127 (*SlAlba5*) to 395 (*SlAlba7*) amino acids (aa), with a mean of 236 aa. The computed molecular weights of these proteins ranged from 14.24 to 43.29 with iso-electric points (pI) of 5.61 to 10.22. Based on their predicted pI values, the SlAlbas include both acidic and basic proteins. The grand average of hydropathicity (GRAVY) values of these proteins ranged from -1.126 to -0.312, indicating that they all are hydrophilic proteins (Table 1).

3.2 Phylogenetic and domain analysis of tomato Alba proteins

Phylogenetic analysis clearly placed Alba family members from multiple diverse species into three major families: one archaea-specific family and two eukaryote-specific families, the RPP20-like and RPP25-like families (Fig. 1). The archaea-specific family solely comprises archaeal Alba proteins, whereas the eukaryotic homologs of Alba are organized into two distinct families. We constructed an evolutionary tree, which revealed a major cluster containing proteins from dicots and monocots together with lower plants. The tomato Alba proteins were distributed in both eukaryote-specific families, with the largest number of SlAlbas being RPP25-like family proteins with RGG/RG repeat boxes. The remaining SlAlbas lacked RGG/RG motifs and were grouped in the RPP20-like family; no SlAlba protein was found in the archaea-specific family.

Phylogenetic tree analysis indicated that the Alba members from tomato were selectively paired with those from the closely related solanaceous crop potato in both eukaryotic families.

Interestingly, the RPP20-like Alba family members from the basal angiosperm *Amborella* showed a close evolutionary relationship with the corresponding proteins from the single-celled green alga *Chlamydomonas reinhardtii* and dicotyledonous grapevine, whereas the RPP25-like proteins of *Amborella* were grouped in the same clades as those of monocotyledonous crop species. Alba homologs from the moss *Physcomitrella* (PpAlb2, PpAlb3, and PpAlba4) were clustered together and showed closer relationships with those of solanaceous crops in the RPP25-like family, whereas the RPP20-like protein (PpAlb1) was clustered together with those of cereal crops. In the archaeal Alba protein group, SsoAlba1 and SsoAlba2 from *Sulfolobus solfataricus* were paired with their related orthologs SshAlba1 and SshAlba2, respectively, from *Sulfolobus shibitae.* However, *Archaeoglobus fulgidus* (AfAlba1) and *Aeropyrum pernix* (ApAlba1) were grouped in the same clades with their paralogs derived from linage-specific evolution (AfAlba2 and ApAlba2). These results point to the diversification of Alba proteins

across various kingdoms of life as well as the sequence conservation in related groupings during the evolutionary process.

Of the eight Alba proteins identified in tomato, five (SlAlba2, SlAlba3, SlAlba6, SlAlba7, and SlAlba8) displayed a long-form structure comprising an N-terminal Alba domain followed by a C-terminal region with several arginine-glycine (RGG) repeats, which function as RNA recognition motifs to interact particularly with guanine (G)-quadruplexes in RNA (Ozdilek et al., 2017). However, these motifs were absent from the three remaining, short-form SlAlba proteins (SlAlba1, SlAlba4, and SlAlba5) (Fig. 2). The clustering of the long-form SlAlba proteins and short-form SlAlba proteins in different eukaryotic families highlights the importance of RGG repeats in the structural and functional diversification of tomato Alba proteins during the course of evolution. All Alba proteins from tomato and potato belonging to the RPP20-like family share a conserved NRIQVS-hexapeptide motif at the start of their Alba domains, whereas most of their homologs from the RPP25-like family contain a distinct conserved NEIRIT-hexapeptide motif (Fig. 3). These findings indicate that these motifs were conserved in tomato and potato and are related to the structural and functional similarity of Alba family members from solanaceous crops. Multiple sequence alignment of SlAlba proteins with a well-characterized Alba homolog from the human malaria parasite *Plasmodium falciparum* highlighted the evolutionary conservation of RPP-20-like family Alba proteins from a unicellular parasite and a flowering plant, particularly the presence of similar amino acid residues in the hexapeptide motif at the start of the Alba domain and in the TKKP-tetrapeptide motif of the first loop (L1), which are related to the DNA binding affinity of the Alba domain (Figure 3).

3.3 Analysis of the exon-intron structures and conserved motifs of *SlAlba* genes

Analysis of the exon-intron structures of the *SlAlba* genes showed that the numbers of exons (4 or 5) in tomato *Alba* genes in the RPP20-like family were almost identical. By contrast, members of the RPP25-like family contained 6 to 9 exons, with an average of 7 (Figure S1). *SlAlba7* had the most exons (9), while *SlAlba1* had the fewest (4). However, the genes in the same phylogenetic group shared similar exon-intron compositions in terms of intron number and exon size.

An in-depth motif analysis revealed that Alba proteins from the RPP20-like family had fewer motifs (3 or 4) than those from the RPP25-like family (7 to 14 with a mean of 9) (Fig. 4). Motifs 1 and 2 in the Alba domain were conserved in all Alba proteins, while motifs 7, 8, and 13, which contain the RGG/RG repeat motif, were prevalent only in large Alba proteins belonging to the RPP25-like family. All RPP20-like family members shared similar motif structures and compositions, containing motif 3, which is characteristic of the members of the RPP20-like family, in addition to motifs 1 and 2. These findings highlight the importance of these motifs in the evolutionary conservation of Alba proteins from this family in both monocot and dicot plants. Motifs 1, 2, 4, 5, and 6 were highly conserved in all members of the RPP25-like family, pointing to their role in the structural similarity among Alba members belonging to this group in dicots and monocots. Several conserved motifs were unique to certain RPP25-like family members, such as motif 9 for SlAlba7, motif 11 for SlAlba1 and OSAlba7, motif 12 for SlAlba6 and SlAlba7, motif 14 for

SlAlba8 and OsAlba7, and motif 15 for OsAlba5 and OsAlba7. Interestingly, motif 10 was absent from Arabidopsis but present in many rice and tomato family members.

3.4. Chromosomal distribution, gene duplication, and microsynteny analysis

The eight tomato *Alba* genes were distributed on four of the twelve tomato chromosomes (1, 4, 6, and 9), with all genes residing close to the distal ends of the chromosomes. The *SlAlba* genes appeared to be distributed unevenly among the chromosomes, with a single gene on chromosome 1, two genes each on chromosomes 4 and 9, and three genes on chromosome 6 (Figure S2).

Of the eight *SlAlba* genes, only *SlAlba6* and *SlAlba7* were predicted to be a segmentally duplicated gene pair; these genes are located on chromosomes 6 and 9, respectively (Figure S3). No tandemly duplicated genes were predicted in the *SlAlba* gene family, since no genes resided within a 100-kb distance on the same chromosome. SlAlba proteins from the same phylogenetic clades exhibited higher sequence identity compared to those from different clades (Table S4). The Ka/Ks ratio of the duplicated gene pair (*SlAlba6*/*SlAlba7*) was > 1, indicating that these genes had undergone positive selection during the process of evolution. The predicted divergence time of the paralogous gene pair (*SlAlba6*/*SlAlba7*) indicates that the gene duplication event took place 64.09 million years ago (MYA) (Table 2). We constructed a comparative microsynteny map to analyze orthologous *Alba* gene pairs from tomato, rice, and Arabidopsis in order to investigate the evolutionary history and relationships across their genomes. This analysis predicted three orthologous gene pairs between tomato and Arabidopsis, but only one orthologous gene pair between tomato and rice, as well as between Arabidopsis and rice (Fig. 5).

3.5 Analysis of putative Stress- and Hormone-Responsive Cis-Elements and miRNA Target sites in *SlAlba* Genes

To assess the transcriptional regulation of *SlAlba* genes under abiotic stress conditions, we analyzed the 1500-bp sequences in their promoter regions. We identified various numbers of phytohormone- and stress-responsive cis-regulatory elements in the promoters of *SlAlba* genes, including the following: TC-rich repeats (implicated in defense and stress responses); drought-responsive MYB-binding site (MBS); cold- and hypersalinity-responsive low temperature-responsive elements (LTRs); the CGTCA-motif (associated with the jasmonic acid [JA] response); ABRE elements (related to the ABA response); gibberellic acid (GA) response-related TATC-elements and GARG-motifs; TCA-elements (involved in the SA response); WUN motifs (with roles in wounding responses); and TGA elements and AuxRR-core (involved in the auxin response; Figure S4 and Table S5). We also identified the target sites for several different types of miRNAs involved in various biological processes in the *SlAlba* genes (Table S6), many of which are related to development and stress responses, such as *sly-miR9477-3p*, *sly-miR5303*, *sly-miR9474-5p*, *sly-miR9471a-3p*, *sly-miR9473-3p*, *miR156*, *miR167*, *miR169*, *miR171*, *miR172*, *miR393*, *miR395*, *miR396*, *miR397*, *miR399*, and *miR408*, and *miR827* (López-Galiano et al., 2019; Liu 2018; Tripathi et al., 2018; Zhang et al., 2015).

3.6. Prediction of the three-dimensional structures and functions of SlAlba proteins

We predicted the 3D structures of SlAlba proteins using the optimal templates listed in Table S7 and other relevant information, such as the percentage of sequence identity, coverage, and Z-scores within the reliable range. An analysis of the resulting models by UFSC chimera 1.15 revealed varied numbers of α-helices, β-strands, and coils in the SlAlba proteins, as shown in Table S8 and Fig. 6 (Pettersen et al., 2004). TM-scores of < 0.17 and > 0.5 indicate models with random similarity and a high level of homology, respectively. The TM-scores and RMSD values of the models fell within the reliable range, implying that the models were accurate (Table 3). The C-scores of the predicted models ranged from -4.30 (SlAlba8) to -0.70 (SlAlba4), and other parameters, such as the number of decoys and cluster density, demonstrated that the results were within a credible range (Table 4).

Table 3. Secondary structure prediction for SlAlba proteins by I-TASSER.

No.	GI Number	Type	TM-Score	RMSD
1	1104569164	SlAlba1	0.36+-0.12	12.6+-4.3
2	723693946	SlAlba2	0.35+-0.12	13.5+-4.0
3	460385381	SlAlba3	0.37 ± 0.13	13.1 ± 4.1
4	460392354	SlAlba4	0.62 ± 0.14	5.9 ± 3.6
5	460392188	SlAlba5	0.61 ± 0.14	6.1 ± 3.8
6	460393594	SlAlba6	0.32 ± 0.10	14.8 ± 3.6
7	1104635736	SlAlba7	0.56 ± 0.15	9.5 ± 4.6
8	723732812	SlAlba8	0.26 ± 0.08	17.0 ± 2.8

GI number, GenInfo Identifier number; TM-score, Template modeling score; RMSD: root-mean-square deviation between residues structurally aligned by TM-align.

Table 4. Parameters for 3D structure modeling of SlAlba proteins.

No.	Type	GI Number	C-Score	No. of Decoys	Cluster Density
1	SlAlba1	1104569164	-3.16	1608	0.0197
2	SlAlba2	723693946	-3.30	1231	0.0141
3	SlAlba3	460385381	-3.06	615	0.0187
4	SlAlba4	460392354	-0.70	7038	0.1925
5	SlAlba5	460392188	-0.83	7083	0.1685
6	SlAlba6	460393594	-3.65	621	0.0108
7	SlAlba7	1104635736	-1.20	600	0.1370
8	SlAlba8	723732812	-4.30	418	0.0056

GI number, GenInfo Identifier number; C-Score, confident score.

We investigated the binding residues for the SlAlba proteins based on an alignment between the template and the newly generated models. Five of the eight SlAlba proteins (SlAlba1, SlAlba2, SlAlba4, SlAlba5, and SlAlba8) were predicted to bind to both DNA and RNA in addition to other ligands, such as peptides, arginine, and manganese. SlAlba3 and SlAlba7 displayed no affinity for nucleic acids, but they did have affinity for other ligands, such as β-carotene and magnesium, whereas SlAlba6 exhibited a specific affinity for RNA (Figure S5). Gene ontology (GO) analysis suggested that the tomato Alba family members play different roles in diverse biological processes, including chromosome condensation, cellular response to stress, and carbohydrate metabolic process, a putative role in cytolysis and protein complex oligomerization, and rRNA methylation (Table S9).

3.7. Expression analysis of *SlAlba* genes in different organs

Expression profiling revealed that most *SlAlba* genes showed differential expression patterns across the various organs investigated. Four out of the eight *Alba* genes were predominantly expressed in stem tissues, while two other *Alba* genes were preferentially expressed in flowers and the remaining two genes showed maximum expression in 1 cm fruits and leaves, respectively.

SlAlba1 was highly expressed in flowers at various stages of development, such as floral buds (550-fold relative to roots), full blooming flowers (800-fold relative to roots), and senescent flowers (250-fold relative to roots), whereas it displayed minimal expression in roots and 1 cm fruits and no expression in other organs. By contrast, homologs of *SlAlba1* in the same phylogenetic group, *SlAlba4* and *SlAlba5*, were broadly expressed across all organs examined. *SlAlba4* showed the highest expression in stems (3-fold vs. the control [leaves]), but its transcript levels were significantly lower (2-fold to 4-fold) in fruits at all stages of development except for 1 cm fruits compared to the control. *SlAlba5* was 3- to 8.5-fold more highly expressed in fruits at all six developmental stages except immature fruits compared to the control.

The expression of *SlAlba5* peaked (8.5-fold vs. the control) in 1 cm fruits, followed by stems (>5-fold relative to the control).

SlAlba2, belonging to the RPP25-like family, is another flower-specific gene. This gene showed maximum expression in flower buds (>96-fold higher than the control), followed by full blooming flowers (>4-fold compared with the control). The expression of *SlAlba2* was 2.5-fold higher in breaker fruits than the control but significantly downregulated in 1 cm fruits, immature fruits, and B7 fruits during fruit development (-1.7 to 4-fold vs. the control). Intriguingly, *SlAlba3* was the only gene whose transcripts were most abundant in leaves, followed by stems, and its expression was significantly reduced (by 5.5- to 14.8-fold relative to the control) throughout all stages of fruit development. *SlAlba6*, *SlAlba7*, and *SlAlba8*, homologous genes belonging to the same phylogenetic clade, exhibited a 2-fold upregulation in stem tissues. Compared to the control, their transcript levels were significantly reduced in fruits during most stages of development (Fig. 7).

3.8. Expression analysis of tomato *Alba* family genes under different abiotic stresses and phytohormone treatments

SlAlba1 transcripts were undetectable in all leaf samples from control, abiotic stress-, or ABA-treated plants, whereas many of the other *Alba* genes displayed diverse expression patterns over time following exposure to these treatments (Fig 8a-e).

Heat stress triggered significant changes in the transcript profiles of the *SlAlba* genes at various time points (Fig. 8a). Compared to the control (0 h), *SlAlba4* and *SlAlba5* expression significantly increased (1.6-to 2-fold) at 24 hours after heat treatment. *SlAlba5* expression was repressed (1.7-to 2-fold) during the early phases of heat stress but gradually recovered after 9 hours and was upregulated (2-fold) at 24 hours after heat stress. By contrast, the majority of *Alba* genes (*SlAlba2*, *SlAlba3*, *SlAlba6*, *SlAlba7*, and *SlAlba8*) showed similar responses to heat treatment, with decreased expression levels (1.3-fold to 6-fold vs. the control) following heat exposure.

Most of the *Alba* genes were responsive to salt treatment but *SlAlba4* and *SlAlba7* showed no significant change in expression under salt treatment compared to the control (Fig. 8b). *SlAlba6* was induced by salt stress; its transcript levels did not significantly change during the early phases of salt treatment but gradually increased and peaked (2-fold vs. the control) at 24 hours of exposure to salt stress. Several *Alba* genes, *SlAlba2*, *SlAlba3*, *SlAlba5*, and *SlAlba8*, displayed reduced transcript levels (1.5-to 4-fold lower than the control) at one or two time points during salt-stress treatment.

Seven *Alba* genes were differentially regulated in response to low temperature stress (Fig. 8c). *SlAlba8* was significantly upregulated (1.7- to 2-fold vs. the control) by cold treatment, while multiple *Alba* genes (*SlAlba3*, *SlAlba4*, *SlAlba5*, *SlAlba6*, and *SlAlba7*) were downregulated (1.6- to 4-fold relative to the control) in response to cold stress.

All the tomato *Alba* genes showed similar expression profiles under drought conditions, and they were significantly deregulated (1.5- to 80-fold vs. the control) at multiple time points during the drought stress period (Fig. 8d). The *SlAlba* genes also showed similar expression patterns in response to the stress hormone ABA. Most of the genes were considerably upregulated (1.8- to 5-fold higher than the control) under ABA treatment; only one gene, *SlAlba3*, was unresponsive to ABA treatment (Fig. 8e).

3.9. Subcellular location analysis of tomato Alba proteins

In silico subcellular location analysis suggested that the SlAlba proteins are localized to different parts of the cell, such as the cytoplasm, chloroplasts, and nucleus (Table S10). To verify their predicted localizations, three *Alba* genes (*SlAlba4*, *SlAlba5*, and *SlAlba6*) were used to generate fusion proteins with GFP. These proteins were expressed in rice protoplasts and their subcellular locations examined. SlAlba4 was predominantly located in the cytosol, but could also be observed in the nucleus. SlAlba5 was primarily located in the cytosol, with strong signals detected in a large portion of the cytoplasm. SlAlba6 was detected in the nucleus as well as the cytoplasm (Fig. 9).

4. Discussion

In this study, we identified eight *Alba* genes from tomato via genome-wide analysis. The number of *Alba* gene family members varies across diverse species of the plant kingdom (Fig. 1). *Alba* genes exist as a multi-gene family in tomato with the average number compared with other plant species, implying that they have biological importance in this plant species.

In agreement with the previous reports (Goyal et al., 2016; Náprstková et al., 2021), in-depth analysis of the phylogenetic relationships among Alba proteins revealed two types. Some of the proteins are relatively large, with an RGG/RG repeat motif or other domain in addition to the generic single Alba domain; most of these proteins belong to the RPP25-like family. The other Alba proteins are relatively small, with only the generic single Alba domain; most of these proteins belong to the RPP20-like family in plants or the archaeal Alba family (Fig. 1). These findings suggest that the emergence of additional motifs and domains in Alba family proteins may have played a significant role in the expansion and diversification of *Alba* gene family in plant species during the course of evolution.

In agreement with the findings of Verma et al. (2018), our phylogenetic analysis clearly grouped the plant *Alba* genes into two distinct subclusters: *Alba* genes from monocots (rice, sorghum, and maize) and those from dicots (tomato, potato, Arabidopsis, chickpea, and grapevine). In addition, the only *Alba* gene (*CreAlba1*) identified in the genome of the single-celled green alga *Chlamydomonas reinhardtii* was clustered together with those of land plant species in the Rpp20-like phylogenetic group, indicating that common *Alba* genes are shared by chlorophytes and streptophytes, which diverged over one billion years ago (Merchant SS 2007). Additionally, *Alba* genes from the moss *Physcomitrella patens*, the basal angiosperm *Amborella*, monocots (sorghum, rice, and maize), and dicots (tomato, potato, grapevine, Arabidopsis, and chickpea) were present within both eukaryotic-specific phylogenetic groups (Fig. 1), suggesting that these two families evolved before bryophytes and angiosperms diverged approximately

450 million years ago (Rensing et al., 2008). Overall, these findings enabled us to retrace ancient evolutionary transitions in this gene family.

From an evolutionary standpoint, gene duplication events increase the number of genes in a particular gene family, which can help plants adapt to adverse environmental stresses [Cannon et al., 2004; Li et al., 2014]. One duplicated gene pair, *SlAlba6*/ *SlAlba7*, was predicted in the tomato genome (Figure S3), indicating that the eight *SlAlba* genes appear to have been derived from an original set of seven ancestral genes. Microsynteny analysis revealed three segmentally duplicated gene pairs between tomato and Arabidopsis, but only a single duplicated gene pair between tomato and rice as well as between Arabidopsis and rice. These results are consistent with the closer evolutionary relationship between tomato and the dicotyledonous model plant Arabidopsis than between tomato and the monocotyledonous model plant rice (Fig. 5).

Analysis of genetic structural diversity is indispensable for the evolutionary analysis of a multi-gene family. Detailed analyses of the conserved motifs and exon-intron structures of the *Alba* genes revealed that exons and introns as well as conserved motifs were organized in a similar pattern among pylogenetically closely related *Alba* genes but in a different manner among those from the different clusters. These findings point to the functional redundancy across phylogenetically closely related Alba family genes and present a likely rationale for the functional divergence among the divergent *Alba* genes during the evolutionary process (Figure S1, Fig. 4).

Predicting the three-dimensional structure of a protein provides valuable information about its possible molecular functions and ligand-binding sites. A previous study uncovered the likely binding affinity of rice Alba proteins to several molecules including DNA and RNA. In the current study, 3D-modeling of SlAlba proteins also predicted possible binding interactions with DNA, RNA, and peptide molecules (Figure S5), supporting their putative functions in transcriptional and translational regulation (Table S9).

The proper transport of a protein to its specific subcellular locations is critical for its optimal activity. A previous subcellular localization experiment revealed that OsAlba1 localizes to the nucleus and cytoplasm (Verma et al., 2014). In support of this finding, tomato *Alba* fusion proteins also showed GFP signals in the nucleus and cytoplasm (Fig. 9). These findings suggest that the diverse Alba proteins are involved in a variety of cellular signaling processes in the cytoplasm and nucleus.

Analyzing the expression patterns of a gene during growth and development and upon exposure to stress stimuli may help determine its functions. In support of previous findings (Verma et al., 2018), majority of *SlAlba* genes except *SlAlba1* exhibited differential expression profiles across the organs examined, suggesting these genes play distinct regulatory roles in growth and development (Fig. 7). Several of the tomato *Alba* genes (*SlAlba3*, *SlAlba4*, *SlAlba6*, *SlAlba7*, and *SlAlba8*) were predominantly expressed in vegetative organs, but others, including *SlAlba1*, *SlAlba2*, and *SlAlba5*, showed higher transcript levels in reproductive organs such as flowers and fruits. These results suggest that these genes play preferential roles in these organs and developmental phases in tomato.

The initiation and development of a floral bud, a process modulated by multiple floral genes and environmental factors, plays a vital role in fruit set and crop yield (Koutinas et al., 2010). Intriguingly, *SlAlba1* was expressed at strikingly higher levels in flower buds, fully bloomed flowers, and senescent flowers compared to root tissues and 1 cm fruits, and no expression was detected in other organs. *SlAlba2* was predominantly expressed in flower buds, with its expression level many times higher than that in any other organs (Fig. 7). These findings suggest that *SlAlba2* regulates floral bud formation in tomato and that *SlAlba1* regulates flower development at all stages via interactions with other regulatory genes.

The roles of *Alba* genes in fruit development have not previously been studied in any vegetable crop. Tomato is a model organism for the study of climacteric fruits. Therefore, the molecular pathways controlling tomato fruit enlargement (which includes a cell division stage and a cell elongation stage) as well as fruit ripening have been extensively explored (Giovannoni et al., 2007; Lemaire-Chamley et al., 2005). Interestingly, *SlAlba5* is the only gene that showed higher expression in tomato fruits at all stages of development (except the immature stage), with its peak expression in 1 cm fruits. By contrast, the remaining seven *SlAlba* genes showed lower expression levels in fruits at all developmental stages compared to the control (leaves; Fig. 7). Our findings suggest that *SlAlba5* plays a regulatory role in fruit enlargement and ripening, particularly during the initial stages of fruit development.

Most tomato *Alba* genes that were highly expressed in vegetative organs had their highest expression levels in stems; only one gene had its highest expression level in leaves, and no gene has its highest expression level in roots. In addition to providing mechanical support to the aerial portions of the plant, the stem facilitates the long-distance translocation of water and nutrients to sustain plant growth under both normal and stressful conditions. The transcript levels of *SlAlba3*, *SlAlba4*, *SlAlba6*, *SlAlba7*, and *SlAlba8* were markedly higher in stems than in any other organ examined, suggesting that they likely function in stem growth, the long-range movement of water and nutrients, and stress tolerance. The higher expression level of *SlAlba3* in leaves suggests that it might play a role in leaf development and signaling cascades in leaves (Fig. 7). Taken together, these findings point to the diverse functions of *SlAlba* family genes in plant development.

Plant responses and adaptation to environmental stresses often involve differential gene expression, which is regulated by a dynamic network of numerous transcription factors and various stress tolerance genes in an ABA-dependent or -independent manner (Banerjee and Roychoudhury, 2017; Qin et al., 2011). Alba proteins are associated with stress tolerance due to their involvement in genome packaging and organization, transcriptional and translational regulation, post-translational regulation, and RNA metabolism, in addition to responses to different environmental stresses (Goyal et al., 2012, 2016; Verma et al., 2018). The possible roles of *Alba* family genes in plant stress responses were supported by previous studies reporting the stress-induced expression of *Alba* genes in several plant species, such as cotton, Arabidopsis, and rice (Magwanga et al., 2019; Náprstková et al., 2021; Verma et al., 2018). The role of OsAlba1 as a DNA binding protein involved in oxidative stress tolerance was revealed by complementation analysis in yeast. Moreover, the susceptibility of *ghAlba4*- and *ghAlba5*-silenced cotton

plants to drought and salinity conditions highlighted the possible involvement of *Alba* genes in plant stress tolerance (Magwanga et al., 2019; Verma et al., 2014).

In agreement with previous reports, we observed that tomato *Alba* genes were differentially expressed in response to various abiotic stresses (Fig. 8a-e). *SlAlba4* and *SlAlba5* expression was significantly induced by heat treatment (Fig. 8a), which is in agreement with the finding that many *Alba* family genes in Arabidopsis and rice were markedly upregulated in response to mild heat stress (37 C) and moderate heat stress (42 C), respectively (Náprstková et al., 2021; Verma et al., 2018). Here we showed that *SlAlba6* was sharply upregulated in plants under saline conditions (Fig. 8b), suggesting its possible role in salt tolerance in tomato. This result is in agreement with the recent study finding that *Alba* genes were expressed at higher levels in cotton upon exposure to salt stress and that *ghAlba4* and *ghAlba5* cotton plants showed increased sensitivity to salt treatment compared to wild type and control plants (Magwanga et al., 2019). Tomato *Alba* genes were downregulated under drought stress (Fig. 8d), while several *Alba* genes in rice and cotton were upregulated by dehydration treatment, pointing to the functional divergence of *Alba* family genes in different plant species (Magwanga et al., 2019, Verma et al., 2018). The expression level of *SlAlba8* was considerably elevated following cold stress (Fig. 8c), which is consistent with the finding that a few rice *Alba* genes were upregulated in response to low temperature stress (Verma et al., 2018). The phytohormone ABA is well known for its role in regulating plant acclimation to adverse environmental stresses, including heat, cold, salt, and drought (Luo et al., 2017; Suzuki et al., 2016). All tomato *Alba* genes except *SlAlba3* were markedly upregulated under ABA treatment (Fig. 8e), which agrees with the finding that multiple *Alba* genes in rice were induced by ABA treatment (Verma et al., 2018). Therefore, *SlAlba* genes might function in abiotic stress tolerance via their direct roles in ABA signaling.

The regulation of tomato *Alba* genes under stress conditions was also corroborated by the prevalence of multiple cis-elements related to stress tolerance and hormonal responses in their promoter regions. Such elements might facilitate the regulation of these genes under different abiotic stress conditions (Figure S4, Table S5).

Post-transcriptional regulation of numerous miRNA families plays a vital role in various biological processes, including development and stress tolerance (Filipowicz et al., 2008). Several target sites for miRNAs related to stress tolerance and developmental processes were predicted in *SlAlba* genes (Table S6). This result is consistent with the finding that miRNA target sites are present in the *Alba* genes of various plant species including rice, Arabidopsis, maize, and sorghum (Verma et al., 2018).

5. Conclusions

In the current study, we identified eight *SlAlba* genes in tomato. These genes were clustered into two phylogenetic groups based on their domain and motif architectures. The likely involvement of these *SlAlba* genes in diverse signaling pathways, developmental cascades, and plant stress responses is underscored by the presence of development- and stress response-associated cis-regulatory elements

and miRNA target sites in their sequences, their dual localization to the cytoplasm and nucleus, and their putative binding to several ligands including DNA, RNA, and peptides. Expression profiling revealed that most *SlAlba* genes are differentially expressed across various organs and in response to different stimuli. The expression pattern of *SlAlba5* suggests that it functions in fruit enlargement and ripening, especially during the cell division phase. Several *SlAlba* genes were significantly induced under different abiotic stresses, such as *SlAlba4* and *SlAlba5* by heat, *SlAlba6* by salt, and *SlAlba8* by cold stress, pointing to their possible roles in tolerance to these stresses. Our findings lay the foundation for further exploring the *SlAlba* gene family and provide new perspectives for the genetic improvement of tomato.

Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

D.-J.L., J.-M.L., C.-K.K. and M.-Y.C. conceived the study and design the work. M.-Y.C., D.-J.L., J.-M.L. and C.-K.K. supervised and monitored the experimental study; A.H.W. and M.W. carried out in silico analysis. A.H.W. grew plants, isolated RNA, conducted expression analysis, analyzed the data and write the original draft following the guidance of M.Y.C.L.H.C performed subcellular localization experiment. All authors made contributions to the manuscript.

Funding

This research was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean government (MSIT) (no. NRF-2018R1A2B6001781).

Supplementary Materials

The following are attached supplementary related to this article, Figure S1. Schematic representation of the exon-intron distribution of *SlAlba* gene family. Figure S2. Chromosome distribution of tomato *Alba* genes. Figure S3. Gene duplication investigation of *Alba* genes in the tomato genome. Figure S4. Putative cis-acting elements in the upstream of *SlAlba* genes. Figure S5. The predicted binding of putative patterner ligands to SlAlba proteins. Figure S6. Overview of conserved motifs of Alba proteins from tomato, *Arabidopsis* and rice determined using MEME web tool. Table S1. List of the Alba amino acid sequences used for phylogenetic investigation. Table S2. The primer sequences used for subcellular localization analysis. Table S3. The primer sequences of *SlAlba* genes used for qRT-PCR analysis. Table S4. Sequence identity among 8 tomato Alba proteins. Table S5. List of *cis* elements in the promoter regions tomato *Alba* genes. Table S6. Prediction of miRNA target sequences in tomato Alba genes. Table S7. Templates used for 3D structure modeling of SlAlba proteins. Table S8. Secondary structural components in SlAlba proteins. Table S10. Subcellular localization of SlAlba proteins predicted by in silico analysis.

References

Aravind, L., Iyer, L.M., Anantharaman, V., 2003. The two faces of Alba: The evolutionary connection between proteins participating in chromatin structure and RNA metabolism. Genome Biol. 4, R64.

Balestrini, R., Gómez-Ariza, J., Lanfranco, L., Bonfante, P., 2007. Laser microdissection reveals that transcripts for five plant and one fungal phosphate transporter genes are contemporaneously present in arbusculated cells. Mol. Plant Microbe Interact. 20, 1055–1062.

Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P., White, M. F., (2002). The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. Science. 296, 148-151.

Banerjee, A., Roychoudhury, A., 2017. Abscisic-acid-dependent basic leucine zipper (bZIP) transcription factors in plant abiotic stress. Protoplasma. 254, 3–16.

Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D., May, G., 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol. 4, 10.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., Xia, R., 2020. TBtools: An integrative toolkit developed for interactive analyses of big biological data. Mol. Plant. 13, 1194–1202.

Črnigoj, M., Podlesek, Z., Zorko, M., Jerala, R., Anderluh, G., Ulrih, N. P., 2013. Interactions of archaeal chromatin proteins Alba1 and Alba2 with nucleic acids, PLoS One. 8, e58237.

Dupé, A., Dumas, C., Papadopoulou, B., 2014. An Alba-domain protein contributes to the stage-regulated stability of amastin transcripts in Leishmania. Mol. Microbiol. 91, 548–561.

Dupé, A., Dumas, C., Papadopoulou, B., 2015. Differential Subcellular Localization of Leishmania Alba-Domain Proteins throughout the Parasite Development. PLoS One. 10, e0137243.

Filipowicz, W., Bhattacharyya, S. N., Sonenberg, N., 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. rev. genet. 9, 102-114.

Giovannoni, J.J.,2007. Fruit ripening mutants yield insights into ripening control. Curr. Opin. Plant Biol. 10, 283–289.

Gissot, M., Walker, R., Delhaye, S., Alayi, T.D., Huot, L., Hot, D., Callebaut, I., Schaeffer-Reiss, C., Van Dorsselaer, A., Tomavo, S., 2013. Toxoplasma gondii Alba proteins are involved in translational control of gene expression. J. Mol. Biol. 425, 1287–1301.

Goyal, M., Alam, A., Iqbal, M.S., Dey, S., Bindu, S., Pal, C., Banerjee, A., Chakrabarti, S. and Bandyopadhyay, U., 2012. Identification and molecular characterization of an Alba-family protein from human malaria parasite *Plasmodium falciparum*, Nucleic Acids Res. 40, 1174–1190.

Goyal, M., Banerjee, C., Nag, S., Bandyopadhyay, U.,2016. The Alba protein family: Structure and function. Biochim Biophys Acta. 1864, 570-83.

Khatun, K., Robin, A.H.K., Park, J.I., Ahmed, N.U., Kim, C.K., Lim, K.B., Kim, M.B., Lee, D.J., Nou, I.S. and Chung, M.Y., 2016. Genome-wide identification, characterization and expression profiling of LIM family genes in *Solanum lycopersicum* L. Plant Physiol. Biochem. 108, 177–190.

Khatun, K., Robin, A.H.K., Park, J.I., Ahmed, N.U., Kim, C.K., Lim, K.B., Kim, M.B., Lee, D.J., Nou, I.S. and Chung, M.Y., 2017. Molecular characterization and expression profiling of tomato GRF transcription factor family genes in response to abiotic stresses and phytohormones. Int. J. Mol. Sci.18, 1056.

Kim, S.R., Lee, D.Y., Yang, J.I., Moon, S., An, G., 2009. Cloning vectors for rice. J. Plant Biol. 52, 73–78.

Koch, M.A., Haubold, B., Mitchell-Olds, T., 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, Arabis, and related genera (Brassicaceae). Mol. Biol. Evol.17, 483–98.

Koutinas, N., Pepelyankov, G., Lichev, V., 2010. Flower induction and flower bud development in apple and sweet cherry. Biotechnol. Biotechnol. Equip. 24, 1549–1558.

Jelinska, C., Conroy, M.J., Craven, C.J., Hounslow, A.M., Bullough, P.A., Waltho, J.P., Taylor, G.L., White, M.F., 2005. Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. Structure. 13, 963-971.

Lemaire-Chamley, M., Petit, J., Garcia, V., Just, D., Baldet, P., Germain, V., Rothan, C., 2005. Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. Plant Physiol. 139, 750–769.

Li, Z., Jiang, H., Zhou, L., Deng, L., Lin, Y., Peng, X., Yan, H., Cheng, B., 2014. Molecular evolution of the HD-ZIP I gene family in legume genomes. Gene. 533, 218-228.

Liu, M., Yu, H., Zhao, G., Huang, Q., Lu, Y. Ouyang, B., 2018. Identification of drought-responsive microRNAs in tomato using high-throughput sequencing. Funct. Integr. Genomics 18, 67–78.

Luo, D. L., Ba, L. J., Shan, W., Kuang, J. F., Lu, W. J., Chen, J. Y., 2017. Involvement of WRKY transcription factors in abscisic-acid-induced cold tolerance of banana fruit. J. agric. food chem. 65, 3627-3635.

López-Galiano, M. J., Sentandreu, V., Martínez-Ramírez, A. C., Rausell, C., Real, M. D., Camañes, G., Ruiz-Rivero, O., Crespo-Salvador, O., García-Robles, I., 2019. Identification of Stress Associated microRNAs in *Solanum lycopersicum* by High-Throughput Sequencing. Genes. 10, 475.

Magwanga, R.O., Kirungu, J.N., Lu, P., Cai, X., Xu, Y., Wang, X., Zhou, Z., Hou, Y., Agong, S.G., Wang, K. and Liu, F., 2019. Knockdown of ghAlba_4 and ghAlba_5 Proteins in Cotton Inhibits Root Growth and Increases Sensitivity to Drought and Salt Stresses. Front. Plant Sci. 10, 1292.

Mani, J., Güttinger, A., Schimanski, B., Heller, M., Acosta-Serrano, A., Pescher, P., Späth, G., Roditi, I., 2011. Alba-domain proteins of Trypanosoma brucei are cytoplasmic RNA-binding proteins that interact with the translation machinery. PLoS One. 6, e22463.

Náprstková, A., Malínská, K., Záveská Drábková, L., Billey, E., Náprstková, D., Sýkorová, E., Bousquet-Antonelli, C. and Honys, D., 2021. Characterization of ALBA Family Expression and Localization in *Arabidopsis thaliana* Generative Organs. Int J Mol Sci. 22, 1652.

Nekrutenko, A., Makova, K.D., Li, W.H., 2002. The KA/KS ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. Genome Res. 12, 198-202.

Ozdilek, B.A., Thompson, V.F., Ahmed, N.S., White, C.I., Batey, R.T., Schwartz, J.C., 2017. Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. Nucleic Acids Res. 45, 7984-7996.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera-A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605-1612.

Qin, F., Shinozaki, K., Yamaguchi-Shinozaki, K., 2011. Achievements and challenges in understanding plant abiotic stress responses and tolerance. Plant Cell Physiol. 52, 1569-1582.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y. and Tanahashi, T., 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science. 319, 64-69.

Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. 3, 1101–1108.

Suzuki, N., Bassil, E., Hamilton, J.S., Inupakutika, M.A., Zandalinas, S.I., Tripathy, D., Luo, Y., Dion, E., Fukui, G., Kumazaki, A. and Nakano, R., 2016. ABA is required for plant acclimation to a combination of salt and heat stress. PLoS One. 11, e0147625.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Thandapani, P., O'Connor, T.R., Bailey, T.L., Richard, S., 2013. Defining the RGG/RG motif. Mol. Cell. 50, 613–623.

Tripathi, A., Goswami, K., Tiwari, M., Mukherjee, S.K., Sanan-Mishra, N., 2018. Identification and comparative analysis of microRNAs from tomato varieties showing contrasting response to ToLCV infections. Physiol Mol Biol Plants.24,185-202.

Verma, J. K., Gayali, S., Dass, S., Kumar, A., Parveen, S., Chakraborty, S., Chakraborty, N., 2014. OsAlba1, a dehydration-responsive nuclear protein of rice (*Oryza sativa* L. ssp. indica), participates in stress adaptation. Phytochemistry 100, 16–25.

Verma, J.K., Wardhan, V., Singh, D., Chakraborty, S., Chakraborty, N., 2018. Genome-Wide Identification of the Alba Gene Family in Plants and Stress-Responsive Expression of the Rice Alba Genes. Genes. 9, 183.

Xu, D., Zhang, Y., 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J.101, 2525-2534.

Wai, A.H., Naing, A.H., Lee, D.J., Kim, C.K., Chung, M.Y., 2020. Molecular genetic approaches for enhancing stress tolerance and fruit quality of tomato. Plant Biotechnol. Rep. 14, 515–537.

Wai, A.H., Waseem, M., Khan, A.B.M., Nath, U.K., Lee, D.J., Kim, S.T., Kim, C.K. Chung, M.Y., 2021. Genome-Wide Identification and Expression Profiling of the *PDI* Gene Family Reveals Their Probable Involvement in Abiotic Stress Tolerance in Tomato (*Solanum lycopersicum* L.). Genes. 12, 23.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: Protein structure and function prediction. Nat. Methods. 12, 7-8.

Zhang, B., 2015. MicroRNA: A new target for improving plant tolerance to abiotic stress. J. Exp. Bot. 66, 1749-1761.

Zhao, K., Chai, X., Marmorstein, R., 2003. Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding. J. Biol. Chem. 278, 26071-26077.

Figures

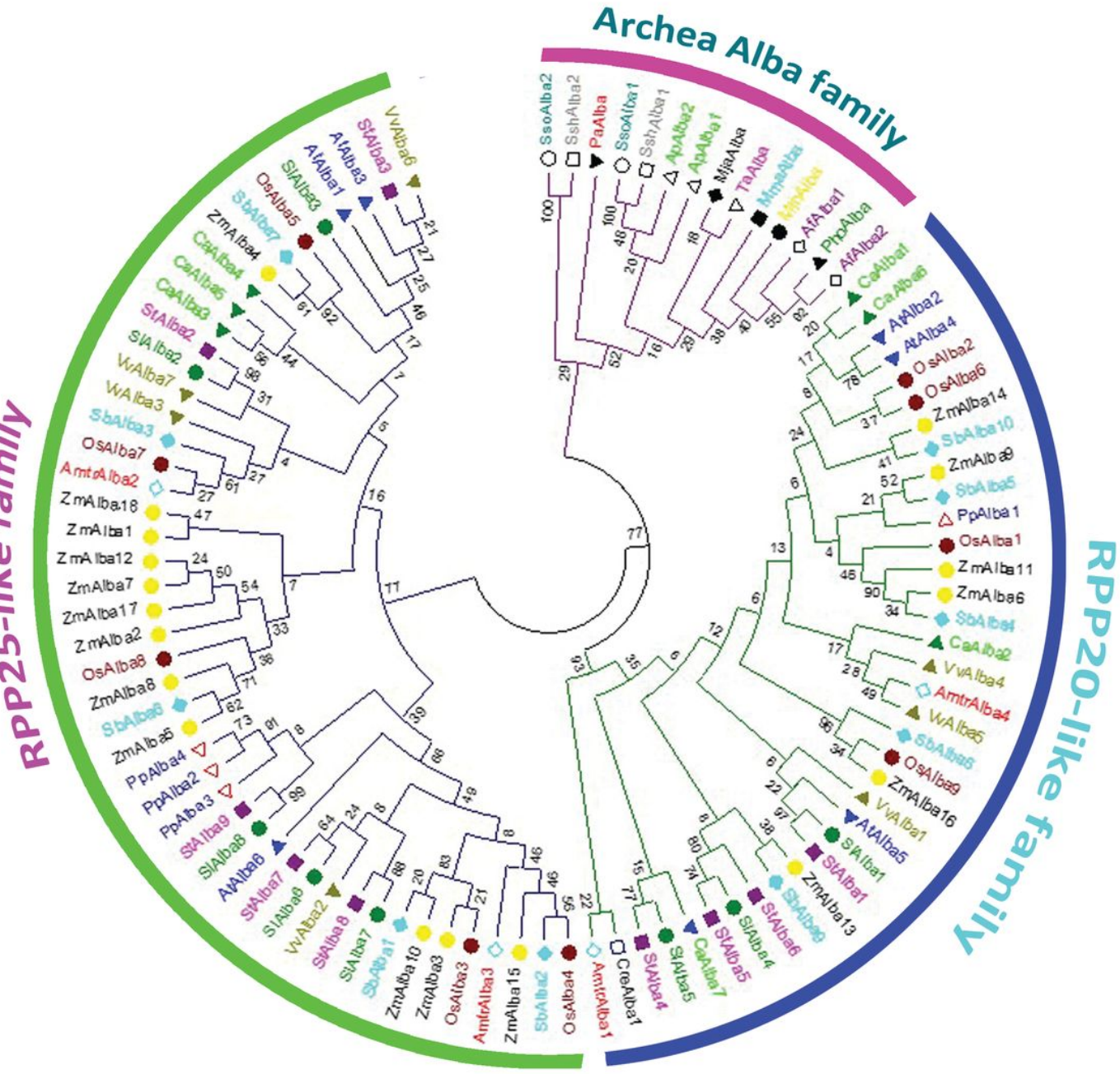

Figure 1

Phylogenetic analysis of Alba proteins in tomato and different organisms. The phylogenetic tree was constructed by the neighbor-joining method in MEGA6 software with 1000 bootstrap replicates using full-length Alba protein sequences. The Alba family proteins were clustered into three distinct families: archaeal Alba family, RPP20-like family and RPP25-like family. A species acronym was added before each Alba protein name: Sl, Solanum lycopersicum; St, Solanum tuberosum; At, Arabidopsis thaliana; Vv, Vitis vinifera; Ca, Cicer arietinum; Os, Oryza sativa; Zm, Zea mays; Sb, Sorghum bicolor; Amtr, Amborella

trichopoda; Pp, Physcomitrella patens; Cre, Chlamydomonas reinhardtii; Sso, Sulfolobus solfataricus; Ssh, S. shibitae; Ap, Aeropyrum pernix; Pho, Pyrococcus horikoshii; Mth, Methanobacterium thermoautotrophicum; Mja, Methanococcus jannaschii; Mma, Methanococcus maripaludis; Af, Archaeoglobus fulgidus; Pa, Pyrobaculum aerophilum and Ta, Thermoplasma acidophilum.

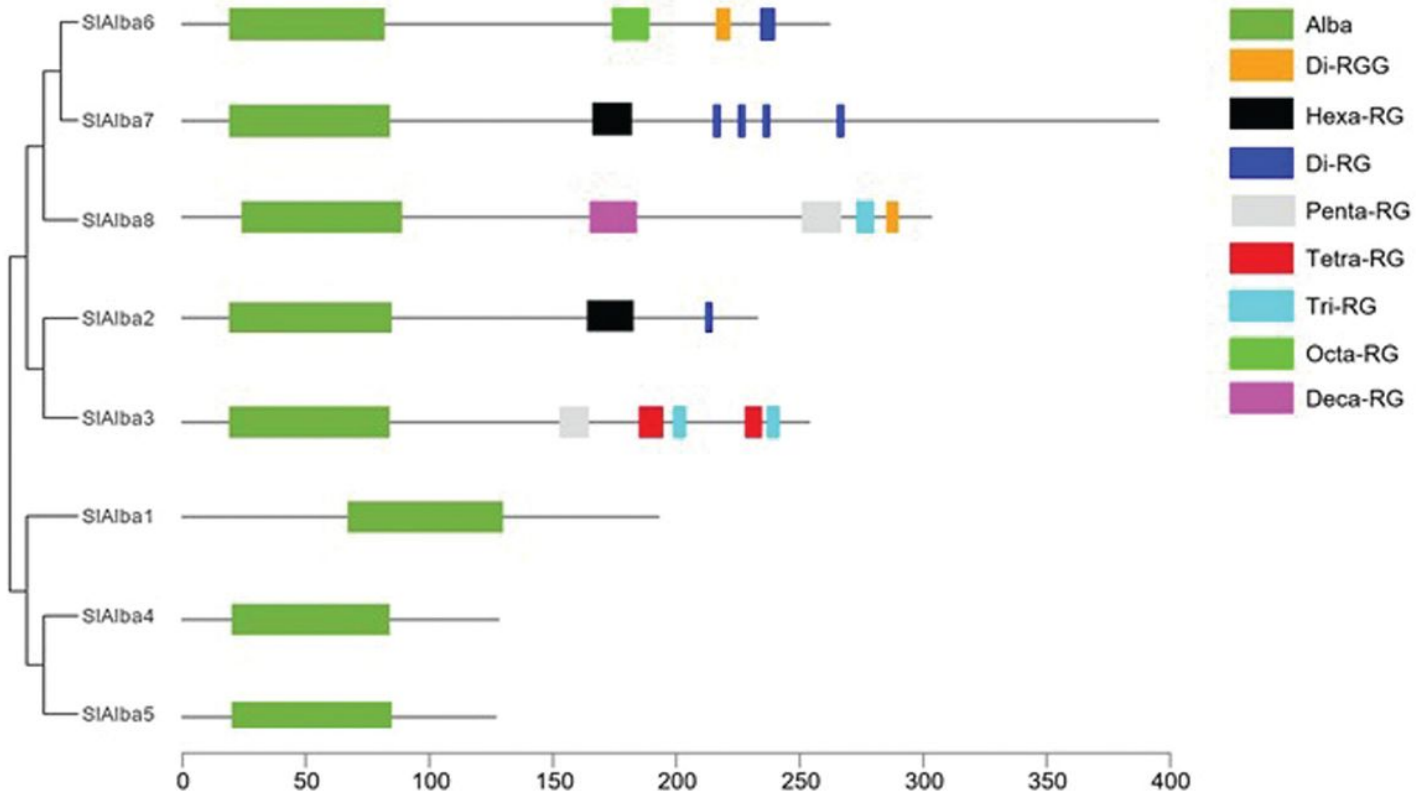

Figure 2

Schematic representation of the Alba domain and RGG/RG repeat motifs identified in SlAlba proteins. RGG/RG repeat motifs are defined in accordance with Thandapani et al. (2013) as follows: the di-RGG motif consists of two repeated RGG sequences spaced 0–4 residues apart, denoted as RGG(X0–4)RGG; the deca-, octa-, hexa-, penta-, tetra-, tri-, and di-RG motifs comprise ten, eight, six, five, four, three, and two repeated RG sequences, respectively, 0–4 residues apart.

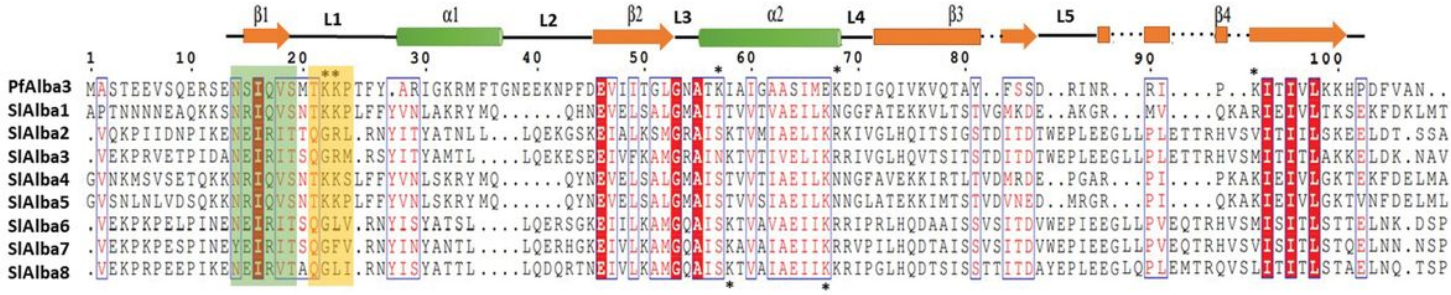

Figure 3

Multiple sequence alignment of the tomato Alba proteins. The protein sequences of the SlAlbas were aligned with that of the homolog from the human malaria parasite using Clustal Omega. The secondary structural elements defined by the ESPript 3.0 web server are also displayed above the alignment. The hexapeptide motifs at the beginning of the Alba domains and the tetrapeptide motifs in the first loop, which are critical for DNA binding, are marked by green and orange boxes, respectively. Asterisks indicate surface-exposed lysines engaged in DNA binding.

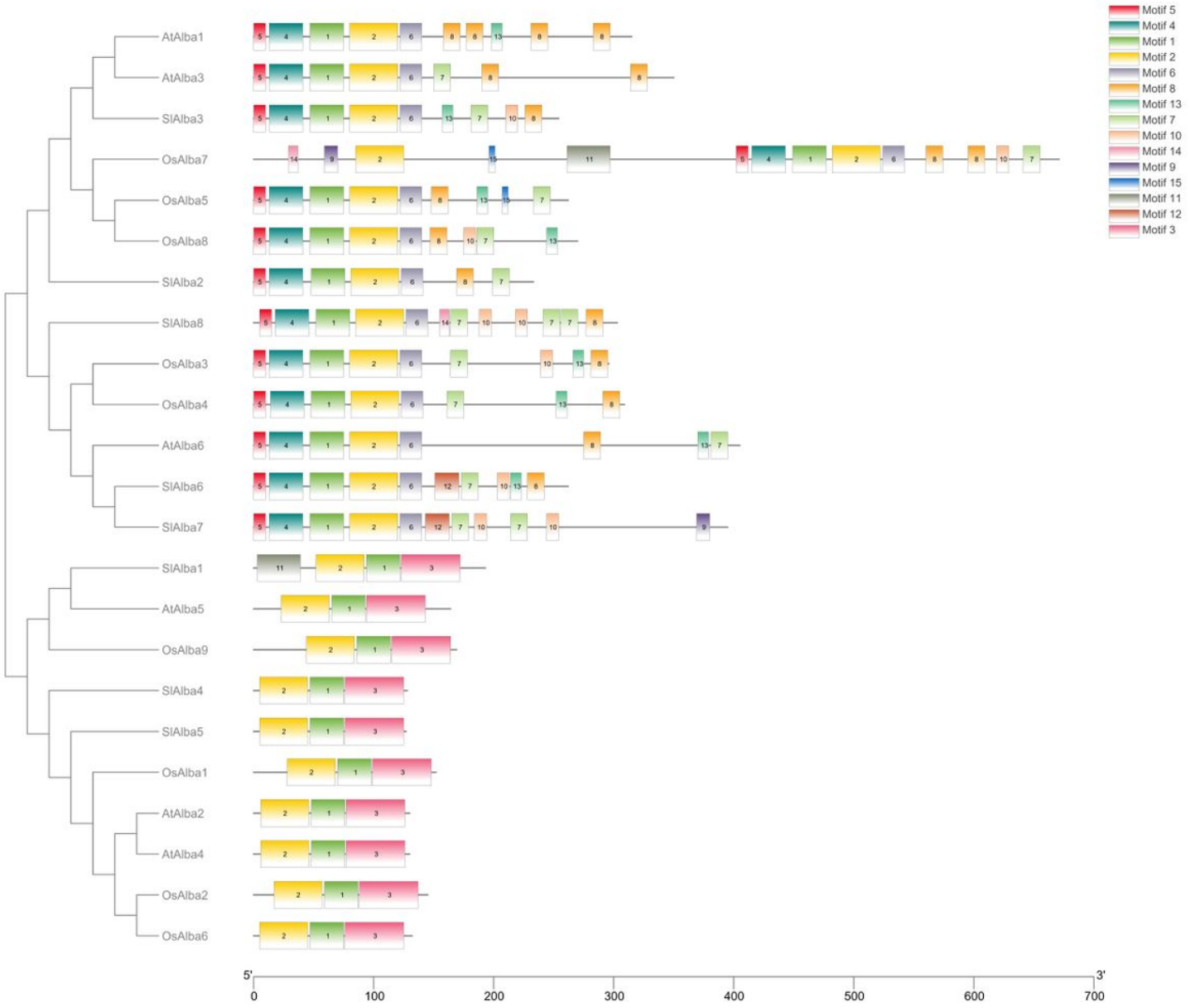

Figure 4

Schematic representation of 15 conserved motifs in Alba proteins from tomato, Arabidopsis, and rice as predicted by Multiple Em for Motif Elicitation (MEME) web server. The Alba domain contains two types of motifs viz, motif 1 and motif 2. Different motifs are denoted by different colored boxes with the motif names indicated in center.

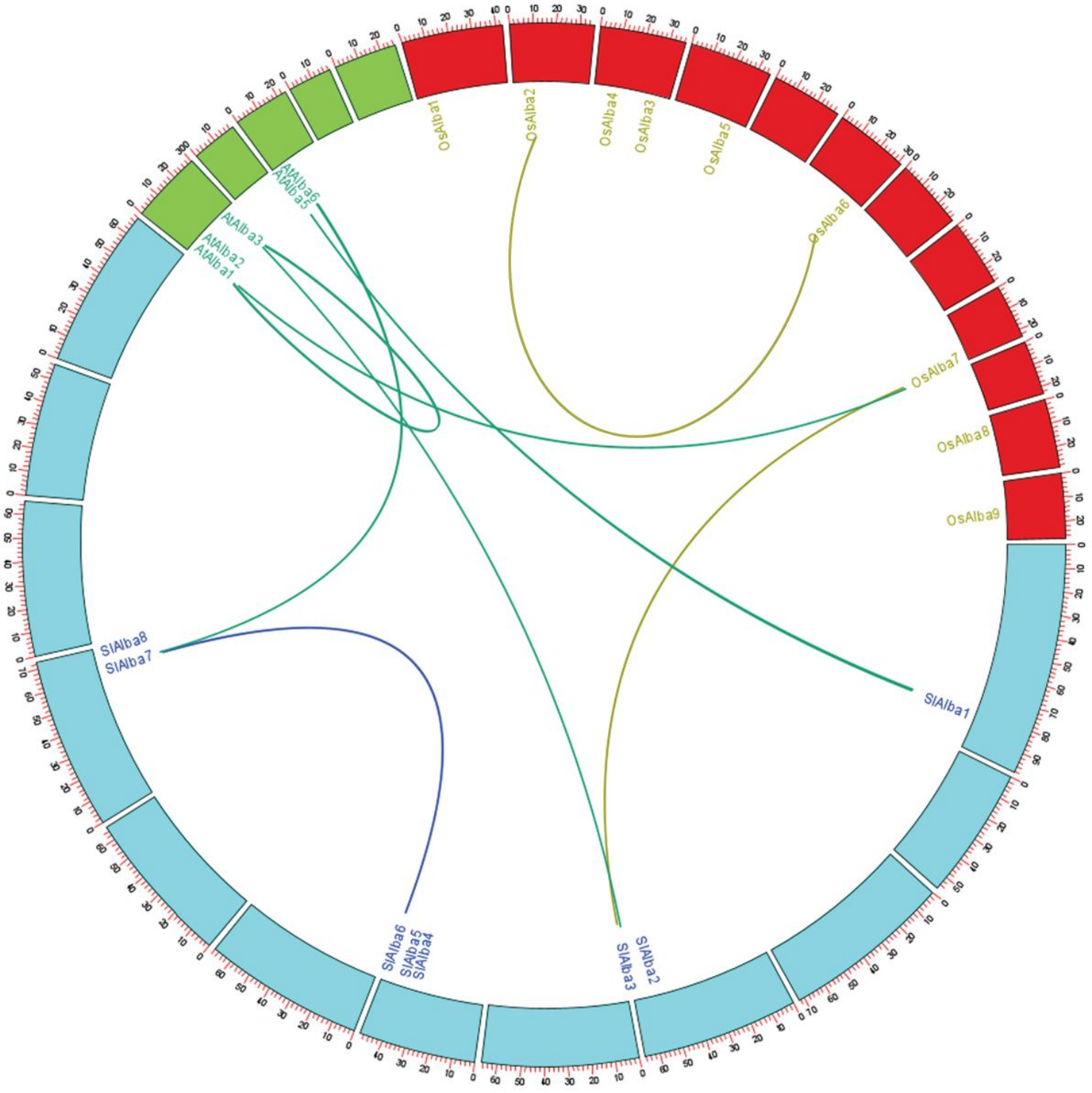

Figure 5

Microsyntenic relationship of Alba genes among tomato, Arabidopsis, and rice. The chromosomes of the three species are indicated by different colors: tomato, aqua; Arabidopsis, lime; and rice, red. All chromosomes are depicted to scale in megabase pairs (Mbp).

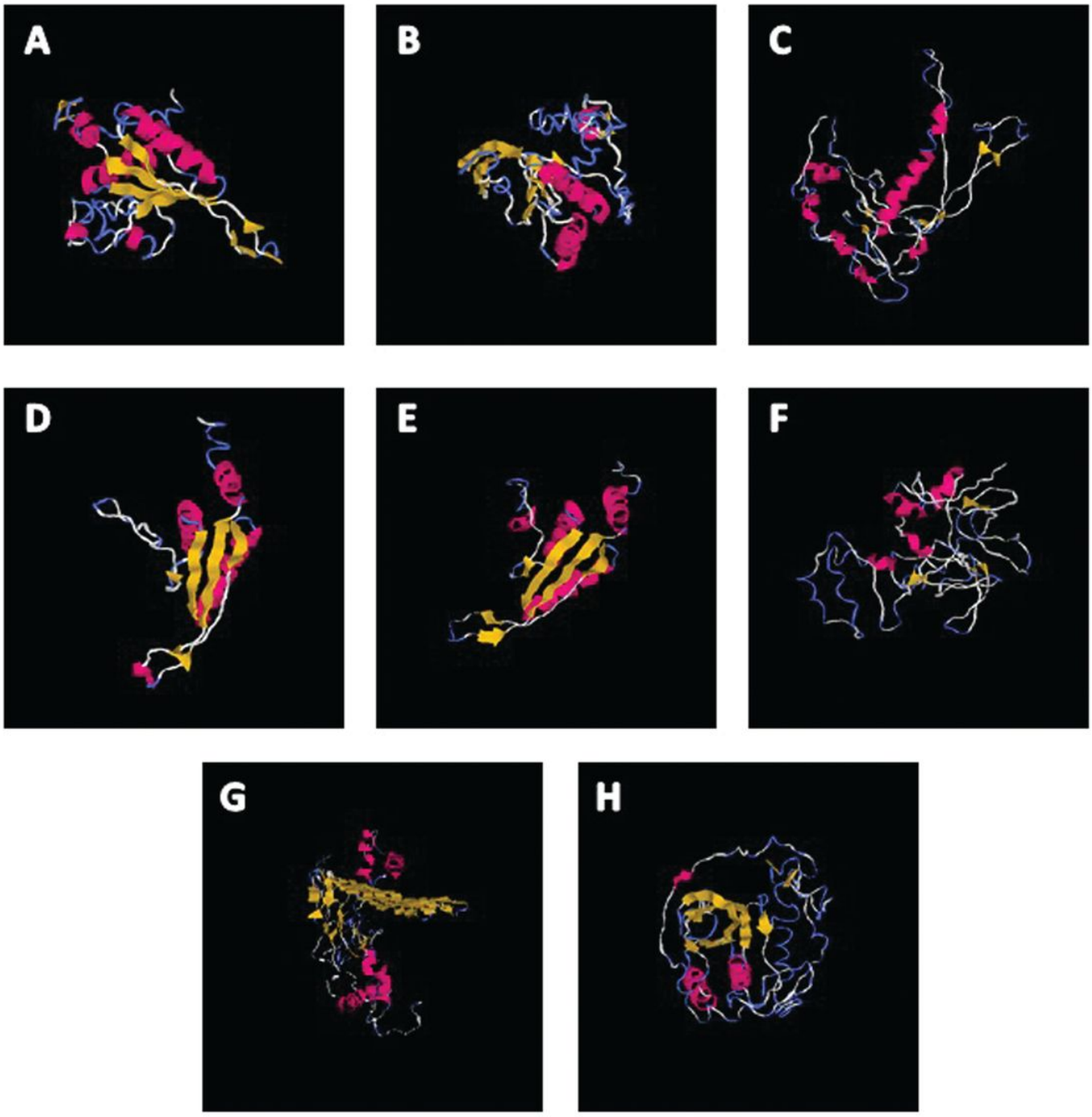

Figure 6

The 3-dimensional model structures of SlAlba proteins. The best model out of five models predicted by I-TASSER was selected. The secondary structural elements: α-helices (red), β-sheets (yellow), and coils (blue) are depicted for the generated 3D model structures of (A) SlAlba1; (B) SlAlba2; (C) SlAlba3; (D) SlAlba4; (E) SlAlba5; (F) SlAlba6; (G) SlAlba7; and (I) SlAlba8.

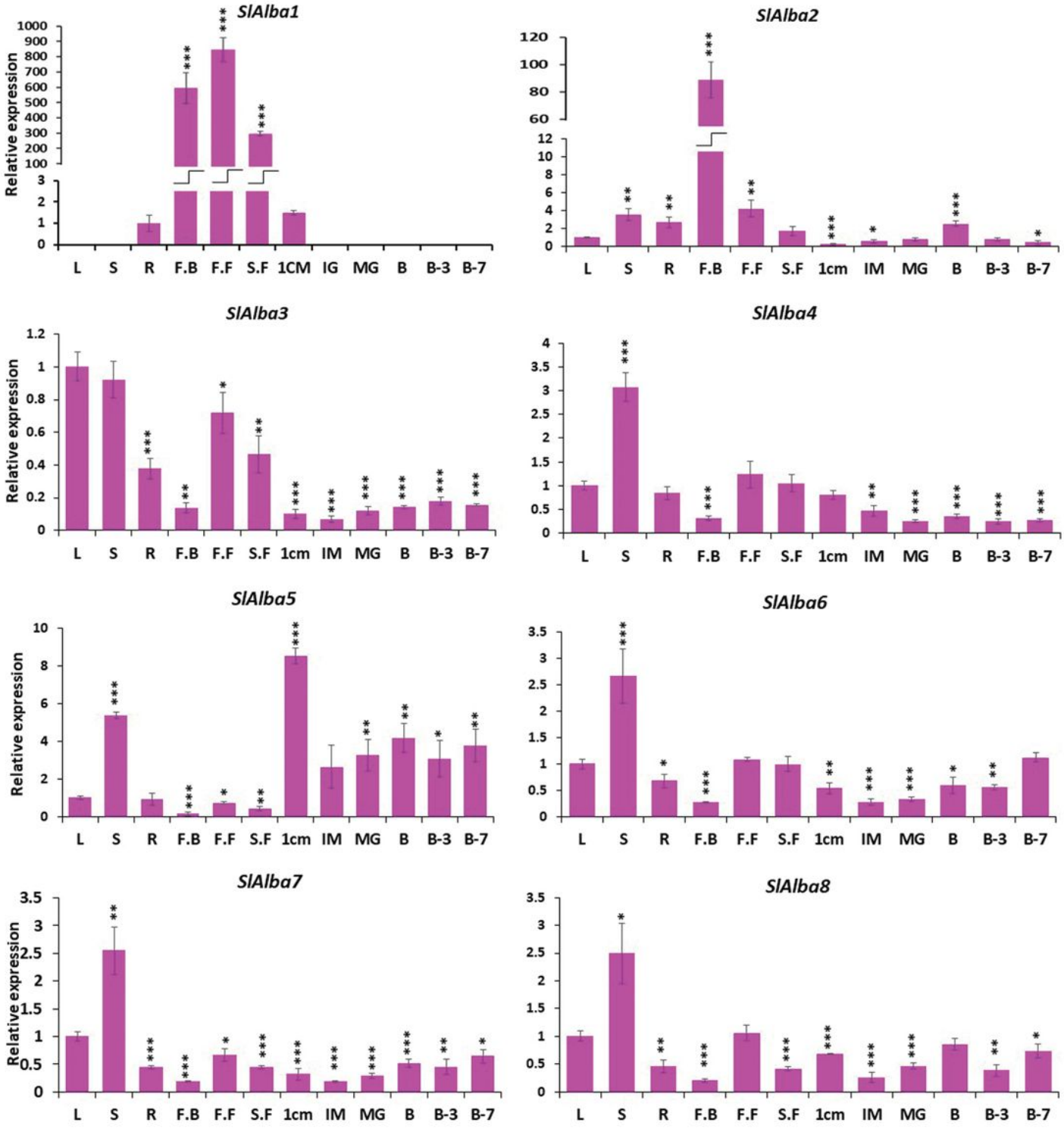

Figure 7

Quantitative reverse-transcription PCR (RT-qPCR) analysis of SlAlba gene expression in 12 organs: leaves, roots, stems, flower buds (FB), full blooming flowers (FF), senescent flowers (SF), 1 cm fruits, immature fruits (IM), mature green fruits (MG), breaker fruits (B), fruits at 3 days after the breaker stage (B3), and fruits at 7 days after the breaker stage (B7). Error bars represent the standard errors of the means of three

independent replicates. *, **, and *** indicate significant differences, as determined by t-test, at p-values ≤ 0.05, ≤ 0.01, and ≤ 0.001, respectively.

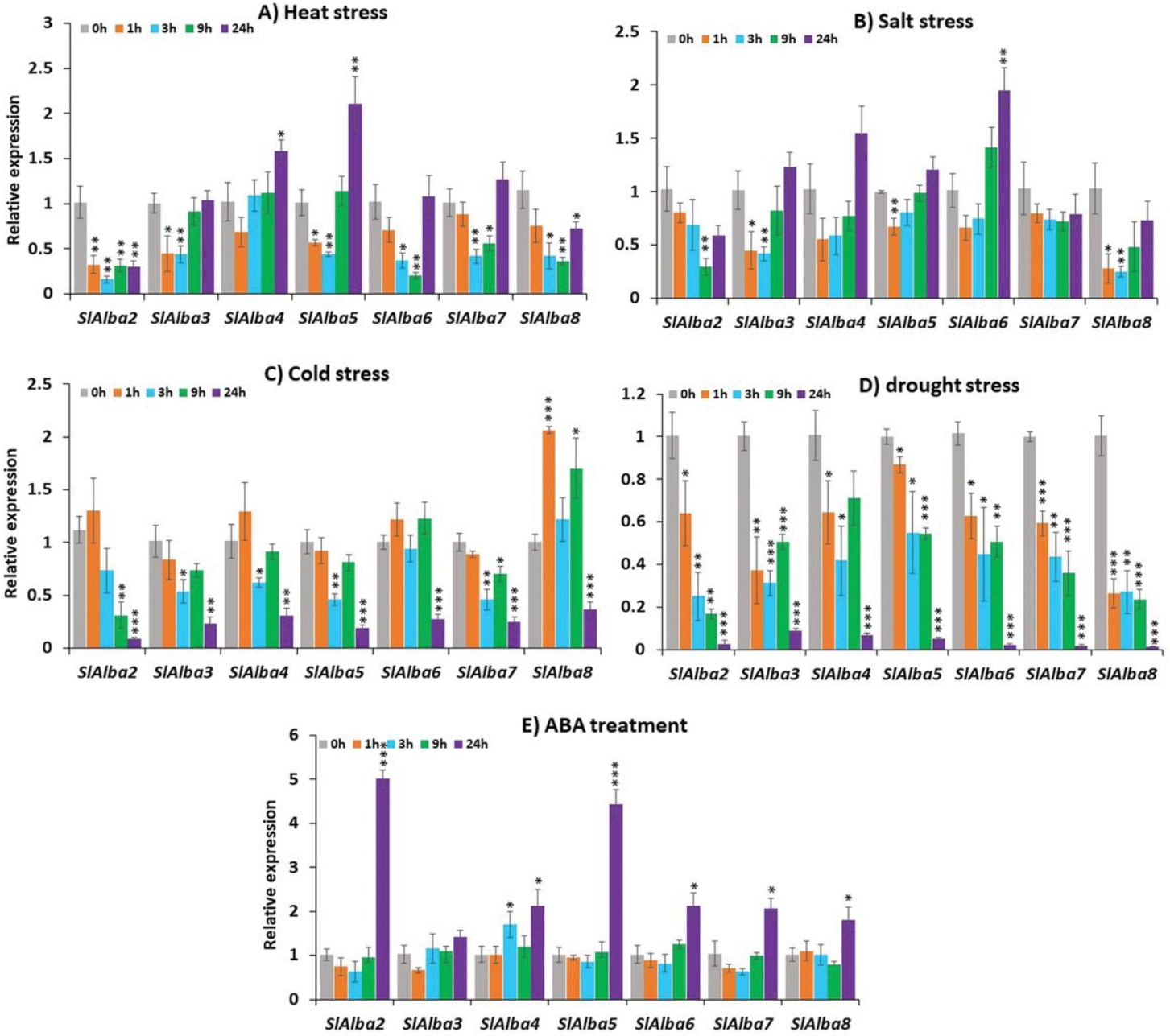

Figure 8

Relative expression levels of SlAlba genes in response to a) heat, b) salt (NaCl), c) cold, d) drought, and e) ABA treatment. Error bars denote standard errors of the means of three independent replicates. The asterisks indicate significant differences, as determined by t-test (* p-value ≤ 0.05, ** p-value ≤ 0.01, and *** p-value ≤ 0.001).

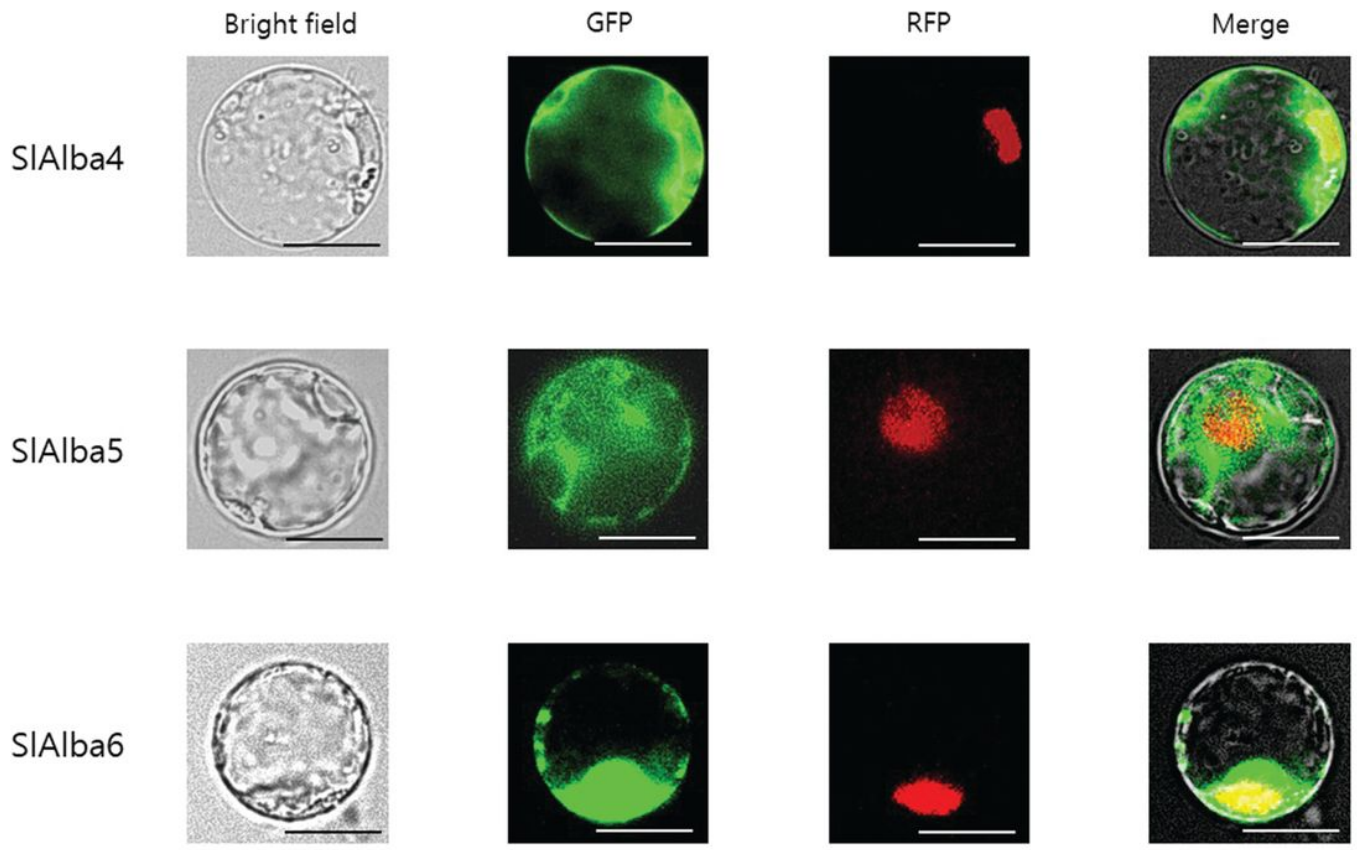

Figure 9

Subcellular localization of Alba proteins. Sub-cellular localization of SlAlba4, SlAlba5 and SlAlba6 were analyzed by transient expression using Oc cell protoplasts after 16 h incubation in dark at 28 C and observing with confocal microscope. NLS-mRFP construct was used as a nuclear localization marker. Scale bars = 10 µm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarymaterials.pdf
- TableS6.xlsx
- TableS10.xlsx