Fasciclin-like arabinogalactan gene family in *Nicotiana benthamiana*: genome-wide identification, classification and expression in response to pathogens

Xinyang Wu
College of Agriculture and Biotechnology, Zhejiang University https://orcid.org/0000-0003-4877-7081
Yuchao Lai
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University
Lanqing Lv
State Key Laboratoty for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University
Mengfei Ji
State Key Laboratoty for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University
Kelei Han
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Insititute of Plant Virology, Ningbo University
Dankan Yan
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Insititute of Plant Virology, Ningbo University
Yuwen Lu
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University
Jiejun Peng
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Insititute of Plant Virology, Ningbo University
Shaofei Rao
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University
Fei Yan
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University
Hongying Zheng (zhenghongying@nbu.edu.cn)

Jianping Chen
State Key Laboratoty for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University

---

Research article

Keywords: N. benthamiana, Fasciclin-like arabinogalactan proteins, Gene expression, Abiotic stress, Turnip mosaic virus

Posted Date: June 3rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-18406/v2

Version of Record: A version of this preprint was published at BMC Plant Biology on July 1st, 2020. See the published version at https://doi.org/10.1186/s12870-020-02501-5.

Abstract

Background: *Nicotiana benthamiana* is widely used as a model plant to study plant-pathogen interactions. Fasciclin-like arabinogalactan proteins (FLAs), a subclass of arabinogalactan proteins (AGPs), participate in mediating plant growth, development and response to abiotic stress. However, the members of FLAs in *N. benthamiana* and their response to plant pathogens are unknown.

Results: 38 *NbFLAs* were identified from a genome-wide study. *NbFLAs* could be divided into four subclasses, and their gene structure and motif composition were conserved in each subclass. *NbFLAs* may be regulated by cis-acting elements such as STRE and MBS, and may be the targets of transcription factors like C2H2. Quantitative real time polymerase chain reaction (RT-qPCR) results showed that selected *NbFLAs* were differentially expressed in different tissues. All of the selected *NbFLAs* were significantly downregulated following infection by turnip mosaic virus (TuMV) and most of them also by *Pseudomonas syringae pv tomato* strain DC3000 (*Pst* DC3000), suggesting possible roles in response to pathogenic infection.

Conclusions: This study systematically identified *FLAs* in *N. benthamiana*, and indicates their potential roles in response to biotic stress. The identification of *NbFLAs* will facilitate further studies of their role in plant immunity in *N. benthamiana.*

Background

The plant cell wall is a dynamic and complex organelle, which is mainly composed of cellulose, hemicellulose, pectins, glycans and proteins. It is not only involved in mechanical protection and structural support, but also in signal transduction, intercellular communication and immunity[1-3].

Hydroxyproline-rich glycoproteins (HRGPs) are typical cell-wall proteins that participate in plant growth, development and immunity[4, 5]. HRGPs have a few repetitive glycosylation motifs containing hydroxyproline (Hyp) residues that are glycosylation sites. Based on the different levels of O-glycosylation, the HRGP superfamily can be classified into three subfamilies: the hyperglycosylated arabinogalactan proteins (AGPs), the minimally glycosylated Pro-rich proteins (PRPs) and the moderately glycosylated extensins (EXTs)[5]. AGPs are abundant in plants, and can themselves be subdivided into six main subclasses: the classical AGPs, AG peptides, Lys-rich AGPs, FLAs, non-classical AGPs and chimeric AGPs[6]. FLAs generally have one or two fasciclin domains, and have been discovered in fruit flies, mammals, sea urchins, plants, yeast and bacteria. Besides fasciclin domains, FLAs often contain an N-terminal signal peptide as well as a C-terminal glycosylphosphatidylinositol (GPI) anchor signal peptide. The GPI and fasciclin domains are functionally important and are believed to mediate cell adhesion[7, 8].

So far, the FLA family members have been identified in several plant species. 21 *FLA*s have been identified in *Arabidopsis thaliana*[8], 27 in rice (*Oryza sativa*)[9, 10], 34 in wheat (*Triticum aestivum*)[10], 35 in poplar (*Populus trichocarpa*)[11], 19 in cotton (*Gossypium hirsutum*)[12], 33 in Chinese cabbage

(*Brassica rapa*)[13], 18 in *Eucalyptus grandis*[14] and 23 in textile hemp (*Cannabis sativa*)[15]. FLAs are cell wall structural glycoproteins that mediate cellulose deposition and cell wall development. They are believed to participate in fiber development, elongation and stem dynamics, affecting the quality of fiber and wood in cotton and woody plants like poplar and eucalyptus[16] and are abundant in the xylem[17]. Knock down of *PtFLA6* resulted in a decrease of stem hardness and xylem cellulose lignin, and down-regulation of genes involved in cell wall synthesis[18]. Overexpression of *GhGalT1* promoted cotton fiber development by controlling the glycosylation of FLAs[19] and in plants where *GhAGP4* was knocked down, fiber initiation and elongation were strongly inhibited and there was suppression of the cytoskeleton network and of cellulose deposition in fiber cells[20]. During cell wall regeneration from cotton protoplasts, there is up regulation of proline-rich protein (PRPL), glycine-rich protein (GRP), and extensin (EPR1) but also of FLA2, which may mediate the construction and modification of the cell wall[21]. In addition, *AtFLA11*, *AtFLA12, EgrFLA2* and *EgrFLA3* have similar functions[14, 22]. *FLA*s can also regulate pollen development. In Arabidopsis and maize, *AtFLA9* and *ZmFLA7* showed negative correlation with abortion, and reductions in the expression of *FLA*s increased the abortion of fertilized ovaries[23]. *AtFLA3*-silenced Arabidopsis had abnormal pollen grains, also suggesting a function in pollen formation[24]. *FLA*s have also been implicated in cell-to-cell communication[13], shoot development[25, 26], seed mucilage adherence[27], glycan stabilization[28] and in response to stresses from salt[29-31], cold[32] and hydrogen peroxide[33].

Although *FLA*s have multiple roles in plant growth and development, very little is known about any involvement they may have in response to pathogens. *N. benthamiana* is a model plant for studying plant immunity, but the structure, function and expression of its *FLA* gene family members is unknown. In this study, we have identified and characterized the members of the *FLA* gene family in *N. benthamiana* and also reported their subcellular localization, expression patterns, and their response to viral and bacterial pathogens.

Results

Identification of Members of the NbFLA Family

Based on previous studies[8], FLAs have an AGP-like glycosylated region, a fasciclin domain and an N-terminal signal peptide. We followed these criteria to identify putative FLAs in *N. benthamiana*. The sequences of the 21 identified AtFLAs were downloaded[8] and the *N. benthamiana* genome was downloaded from the Sol Genomics Network (https://solgenomics.net/)[34]. A total of 38 NbFLAs were identified by two round BLASTP and signal peptide prediction (Table 1 and Additional file 1: Table S1). Most of these (66%) have lengths of 200-300aa, while the largest (NbFLA10) has 495aa and the smallest (NbFLA26) has only 182aa. The predicted isoelectric points range from 4.29 to 9.77, and the molecular weights (MWs) derived only from the amino acid sequences (not including glycans) are in the range 19.68-52.32kDa. The protein properties of the NbFLAs are similar to those of other plant species[8, 11].

Phylogenetic Analysis and Multiple Sequence Alignment of NbFLAs

To better reveal their evolutionary relationships and to help the classification of NbFLAs, the sequences of all 21 AtFLAs and 38 NbFLAs were used to construct a phylogenetic tree (Fig.1). Because of the low sequence similarity between some FLAs, phylogenetic analysis alone could be misleading and therefore pair-wise sequence similarity, presence and number of fasciclin domains and GPI were also used to create a classification, as previously described[8]. Most NbFLAs were sufficiently classified by phylogenetic analysis, but for a few (NbFLA8/15 and NbFLA10/14) their protein properties including the presence and number of fasciclin domains and GPI had also to be taken into account.

The 38 NbFLAs we identified could be divided into the same four subclasses previously reported for the AtFLAs[8], named I to IV (Fig. 1). NbFLA2/8/12/15/22/25/26/27/29/32/33/36 belong to subclass I, and have a single fasciclin domain and GPI anchored signal (except NbFLA36), as do the related AtFLAs and PtrFLAs[8, 11]. NbFLA6/9/16/17 belong to subclass II. Subclass II is the smallest group and members contain two fasciclin domains but have no C-terminal GPI anchor site. Members of subclass III (NbFLA3/4/5/7/10/14/18/19/23/24/34/38) have either one or two fasciclin domains, and most (77%) have a C-terminal GPI anchor site. The remaining NbFLAs (NbFLA1/11/13/20/21/28/30/31/35/37) constitute subclass IV, which contains NbFLAs that are quite distantly related to the other NbFLAs and which have no consistent pattern in the number of fasciclin domains or the presence of a GPI signal.

We also constructed separate phylogenetic trees for each subclass of NbFLAs, including the sequences from the other 8 plant species in which FLAs have been identified (Arabidopsis, rice, wheat, poplar, cotton, Chinese cabbage, *Eucalyptus grandis* and textile hemp) (Additional file 2: Fig. S1). In general, FLAs have a relatively high homology among closely related species, like AtFLAs/BrFLAs and OsFLAs/TaFLAs. FLAs from the same species often exist in pairs, like NbFLA26/29 and TaFLA19/27, suggesting that they may be paralogous genes. Subclasses I and III are the two largest groups and the clustering patterns are complicated. FLAs from the same species do not generally group together, and there are some closely-related pairs from different species suggesting that they are orthologous genes (e.g. NbFLA12/BrFLA22 and TaFLA2/OsFLA2). In subclasses II and IV, most FLAs from the same species group together (e.g. NbFLA6/9/16/17 and TaFLA6/7/8/29). Subclass II has fewest members and most of them are not GPI anchored, but the OsFLAs are a significant exception.

Previously reported fasciclin domains contain about 110-150 amino acid residues and have two highly conserved regions (H1 and H2) and a [Phe/Tyr]-His ([Y/F] H) motif[12]. An alignment of the amino acid sequences of the fasciclin domains of the NbFLAs constructed using MUSCLE and some manual analysis showed a similar pattern (Fig. 2). The Thr residue in the H1 region is highly conserved and is followed by other conserved residues such as Val/ Ile (one position after Thr) and Asn/Asp (six positions after Thr). These residues may play a role in maintaining the structure of the fasciclin domain and/or cell adhesion[12]. As reported for other fasciclin domains[11, 31, 35], small hydrophobic amino acids such as Leu, Val and Ile are abundant in the H2 region. In the [Y/F] H motif, His and Pro residues are also relatively conserved.

Analysis of the Structural and Conserved Motifs of *NbFLAs*

Further analysis of gene structure and motifs of the NbFLAs is shown in Fig. 3. The phylogenetic tree confirmed that NbFLAs could be grouped into four subclasses (Fig. 3a). Analysis of the genomic DNA sequences showed that *NbFLAs* usually had 0, 1 or 2 introns (Fig. 3b). All of the members in subclass II have one or two introns while most members of subclasses I and III have none (Fig. 3b). The most closely related members of each subclass, usually have a similar exon/intron structure, with little difference in the length of introns and exons. However, a few *NbFLA* gene pairs showed different intron/exon arrangements. For example, *NbFLA1* and *NbFLA31* have high sequence similarity, but *NbFLA1* has no introns while *NbFLA31* has one.

An online MEME analysis was done to identify additional motifs among the 38 NbFLAs. Twenty conserved motifs were predicted (Fig. 3c and Additional file 3: Table S2) and each NbFLA contained between five and ten of these. Some motifs were common to most members, while the others were unique to one or few subclasses. For example, most NbFLAs (84%) contained motif 17. Motifs 10 and 11 were present only in subclass III and motifs 9, 16, 18 and 19 were found only in subclass II. Motif 7 was unique to subclasses II and IV, and most members of subclasses I and III contained both motifs 3 and 8 except NbFLA4/5/7/26/38. Subclass IV was clearly less closely related to the other subclasses, and motifs 12, 13 and 15 were unique to this subclass.

Prediction of cis-acting Elements and Transcription Factors among the *NbFLAs*

The cis-acting elements in the promoter regions of the *NbFLAs* were analyzed and the results are shown in Fig. 4 and Additional file 4: Table S3. There were seven types of cis-acting elements: environmental stress-related elements, hormone responsive elements, development related elements, light responsive elements, promoter related element, site-binding related elements and other elements. 105 cis-acting elements were predicted and showed great diversity (Fig. 4a). The most abundant elements were light-responsive elements, including G-box, GT1-motif and GATA-motif. 15 hormone responsive elements were identified and these are mainly involved in response to abscisic acid (ABA) or methyl jasmonate (MeJA) (Fig. 4b). Among the predicted environmental stress-related elements, STRE, MBS and ARE were the most abundant (Fig. 4c). Several abundant predicted cis-acting elements are known to mediate plant immunity. For example, *VdMYB1* binds to the MBS in the *VdSTS2* gene promoter, thus activating *VdSTS2* transcription and positively regulating defense responses[36]. *Machi3-1* and *TaRIM1* also bind MBS cis-acting elements to increase host resistance[37, 38].

By binding to transcription factors (TFs), cis-acting elements regulate the precise initiation and efficiency of gene transcription. We then therefore predicted potential TFs which may regulate the transcription of *NbFLAs* (Fig. 5 and Additional file 5: Table S4). The *NbFLAs* had an average of five TFs, but it appears that *NbFLA4* and *NbFLA27* may be regulated by more TFs, including specific TFs like RAV and CPP, while *NbFLA8/15/38* may each be regulated by only two TFs. In total, 25 TFs were predicted of which C2H2, BBR-BPC, Dof, Myb and MIKC were the most abundant. Previous studies have demonstrated the role of TFs in regulating plant immunity. NbCZF1, a novel C2H2-Type zinc finger protein, is a regulator of plant

defense[39] and VvDOF3 enhances powdery mildew resistance in *Vitis vinifera*[40]. In addition, AtMyb15 and MdMyb30 also participate in enhancing disease resistance[41, 42].

Subcellular Localization Analysis of NbFLAs

Bioinformatics analysis based on the NbFLA amino acid sequences suggested that all of them could locate to membranes, and only NbFLA4 was predicted to locate in both the nucleus and membranes (Table 1). To validate these predictions, we selected one NbFLA in each subclass (NbFLA4/6/31/32) to analyze their localization by laser confocal microscopy. AtP1P2A-GFP was used as membrane marker[43]. The results showed that while NbFLA6 and NbFLA32 were only located in membranes, NbFLA4 was present both in membranes and the nucleus, consistent with the predictions (Fig. 6).

A GPI anchored signal is vital for membrane localization and is predicted in about two thirds of AtFLAs and PtrFLAs and in 20 of 38 (53%) of NbFLAs (Table 1). Among the four selected NbFLAs, only NbFLA31 was not GPI anchored. Correspondingly, although a plasmolysis experiment confirmed the membrane localization of NbFLA31, a diffused red fluorescence could also be observed in the cytoplasm (Fig. 6 and Additional file 6: Fig. S2).

Tissue-Specific Expression of *NbFLAs*

To comprehensively understand the functions of *NbFLAs*, two or three *NbFLAs* from each subclass were randomly selected to analyze their expression in five different tissues (root, stem, young leaf, mature leaf and flower) by RT-qPCR (Fig. 7 and Additional file 7: Fig. S3). The expression level of all selected *NbFLAs* (except *NbFLA4*) was higher in young leaves than in mature ones. *NbFLA11/18/31/32/34* were highly expressed in young leaves, and *NbFLA4* were expressed highly in flowers. It was earlier reported that *PtFLA6* is specifically expressed in tension wood (TW) and that decreased transcripts of *PtFLA6* influenced stem dynamics[18]. In this study, *NbFLA2/6/15/17*, belonging to subclasses I and II, were highly expressed in stems, suggesting that they may play a role in stem dynamics.

Expression of *NbFLA*s Under Biotic Stress

To investigate whether *NbFLAs* participate in the response to pathogens, leaves of *N. benthamiana* were inoculated with turnip mosaic virus (TuMV), potato virus X (PVX), pepper mottle mosaic virus (PMMoV) and the bacterial pathogen *Pseudomonas syringae pv tomato* strain DC3000 (*Pst* DC3000). At 5 days post virus inoculation (dpi), or 2 days post *Pst* DC3000 infection, leaves were collected to study the expression pattern of 11 *NbFLA* genes by RT-qPCR (Fig. 8).

TuMV infection led to a huge reduction in expression of all the *NbFLAs* tested, especially *NbFLA15/18/32/34,* which all decreased by more than 99%. PVX or PMMoV infection usually induced a modest reduction in expression, although *NbFLA6* was slightly upregulated by PVX. The bacterial pathogen *Pst* DC3000 decreased expression of most *NbFLAs* by 73-99% but, in contrast, *NbFLA4* and *NbFLA7* were substantially upregulated. These results show that most *NbFLAs* are substantially affected by TuMV and *Pst* DC3000 and may therefore play roles in post-infection responses.

Discussion

FLA families have been identified and characterized in several plants including Arabidopsis[8], rice[9, 10], wheat[10], poplar[11], cotton[12], Chinese cabbage[13], *Eucalyptus grandis*[14] and textile hemp[15]. In this study, we identified 38 *FLA*s in *N. benthamiana* and found that their structural domains were conserved by studying phylogenetic trees, gene structure and conserved motifs (Fig. 3). In general, NbFLAs could be divided into four subclasses and NbFLAs in each subclass had similar gene structure, motifs and conserved domains. Consistent with the FLAs in Arabidopsis[8], subclass II contained fewest NbFLAs and NbFLAs in subclass IV were the most variable. The FLAs of other dicotyledonous plant species had similar properties in each subclass, but while dicot members of subclass II have no GPI, most OsFLAs and TaFLAs in the subclass are GPI anchored[10]. In addition, OsFLAs in subclass II have only one fasciclin domain, unlike the FLAs of the dicotyledonous species[10]. Thus a different classification of FLAs in monocotyledonous plants may be required.

25 of the 38 NbFLAs had a single fasciclin domain, 13 of them had two domains and 20 of the 38 were GPI anchored. A GPI-anchored signal together with a fasciclin domain are known to be important for cell adhesion, for membrane localization and for enabling more stable interactions between adhesion complexes. It has been suggested that plants may have FLAs with GPI-anchoring for maintaining the integrity of the plasma membrane and FLAs that are not GPI-anchored for mediating cell expansion[8].

Previous studies have shown different expression patterns of *FLAs* in the tissues of other plants. For example, *AtFLA11/12* were highly expressed in stems[22], as were *BrFLA6/9/22* (homologous to *AtFLA11*). Some *EgrFLAs* were also highly expressed in stems[14, 22] and ten *PopFLAs* were highly expressed in poplar tension wood[35]. *PtFLA6* and *ZeFLA11* were exclusively expressed in xylem tissues[18, 44]. These studies suggest that some *FLAs* play important roles in stem dynamics and cell wall elongation. In our study, *NbFLA2/6/15* were also expressed highly in stems whereas *NbFLA7/34* were highly expressed in roots, as were *PtrFLA12/21/22/24/27/28/30*[11], indicating that they may participate in root apical meristem development. Many *NbFLAs* were expressed highly in young leaves[11], as reported for *GhFLA5/8/9/12* and *Br4/5/10/21/27/33*[8, 12, 13], but no *PtrFLAs* tested had high expression in young leaves[11]. This may be because *N. benthamiana* more closely resembles cotton and Chinese cabbage in being a herbaceous annual.

Some biotic and abiotic stresses lead to significant changes in the transcription of *FLAs*. For example, Under H_2O_2 stress, the expression levels of wheat FLA proteins were increased, which may contribute to H_2O_2 tolerance[33]. Similarly, *AtFLA3* was expressed more highly under cold stress[32]. Under salt stress, *OsFLA10/18* expression was reduced[9] while *PtrFLA2/12/20/21/24/30* were upregulated[11]. In addition, T*aFLA3/4/9* were downregulated after heat, ABA or NaCl treatment [10]. *OsFLA24* and *AtFLA1/2/8* were also significantly reduced following ABA treatment[8, 9]. Many of the frequently predicted TFs in the *NbFLAs*, including C2H2, Dof and Myb, have been reported to play a role in the ABA pathway[45-48] and therefore, as in other species, *NbFLAs* may be regulated by the ABA pathway. While the function of *FLAs* in the signaling pathway during abiotic stresses has been investigated, little is

known about their potential role in response to pathogens. *AtFLA1/2/8* were decreased by pathogen challenge, oxidative stress and in ascorbate-deficient vtc mutants[49]. The fungus *Ophiostoma novo-ulmi* reduced the expression of *FLAs* in English elm ramets[50]. Our results show that almost all *NbFLAs* were specifically downregulated by TuMV and *Pst* DC3000 infection and this suggests that *NbFLAs* may have specific roles in pathogen infection.

Because of their role in cell adhesion and their membrane localization, AGPs (including FLAs) may interact with receptor-like kinases as wall-associated kinases and thus be involved in signal transduction[51]. For example, *AtFLA4* (*SOS5*) mediated root growth and seed adhesion through cell wall receptor-like kinase (*FEI1/2*)[27], and modulated ABA signaling to regulate cell wall biosynthesis and root growth[25, 27]. The known functions of GPI and the fasciclin domain suggest that NbFLAs might be involved in host-pathogen interactions. Thus, a further role of *NbFLAs* in plant resistance is worth exploring.

Conclusion

In this study, 38 *NbFLAs* were identified and could be divided into four subclasses. In general, the closest members of NbFLAs from the same subclass have similar structure and conserved motifs. The expression patterns of selected *NbFLA*s in different tissues were diverse and selected *NbFLAs* were downregulated following infection by TuMV or *Pst* DC3000. Our results will help to lay the foundation for understanding of the structure and characteristics of the FLA family and for exploring the relationship between *FLAs* and immunity in *N. benthamiana.*

Methods

Identification of the NbFLAs family

The sequences of the 21 identified AtFLAs were downloaded and the *N. benthamiana* genome was downloaded from the Sol Genomics Network (https://solgenomics.net/) [34]. NbFLAs were identified by two rounds of BLASTP. Firstly, all AtFLAs were used to search possible NbFLAs using TBtools[52]. Then NCBI Batch CD-Search [53, 54]was used to confirm whether candidate NbFLAs contained a fasciclin domain including FAS1 (smart00554), Fasciclin superfamily (cl02663) or Fasciclin (pfam02469). Next, we predicted the N-terminal signal peptide by SignalP5.0[55], the C-terminal GPI anchor addition signal by big-PI Plant Predictor[56], and the glycosylation site by NetGlycate 1.0[57]. Finally, using criteria previously established, sequences that contained an AGP-like glycosylated region, fasciclin domains and an N-terminal signal peptide were considered as NbFLAs[11]. The CDS length, pI and molecular weights (MW) of all predicted NbFLAs were then determined by ExPASy[58] and their subcellular localization predicted by Plant-mPLoc[59].

Phylogenetic analysis and multiple sequence alignment

Sequences of AtFLA proteins were obtained from the NCBI protein database (http://www.ncbi.nlm.nih.gov/protein/). A neighbor-joining (NJ) phylogenetic tree of full-length sequences of AtFLAs and NbFLAs was constructed with 1000 bootstrap replicates using MEGA7.0. A multiple sequence alignment of all NbFLAs was also created by Clustal X 2.0[60].

Gene structure and conserved domain analysis

Gene structure and conserved domains were analyzed and visualized using NCBI Batch CD-Search[53, 54] and TBtools[52]. Conserved motifs of the genes were analyzed by the MEME program[61] with the following parameters: optimum motif width was set to 30-70, the number of repetitions was set to zero or one, the maximum number of motifs was set to identify 15 motifs.

Promoter cis-Acting elements and TFs prediction

The promoter cis-Acting elements were predicted by PlantCARE[62] and transcription factors were predicted by PlantRegMap[63], with *N. sylvestris* as the target species.

Plasmid construction and Agroinfection assays in *N. benthamiana*

Based on the sequences above, we cloned the CDS sequences of *NbFLA4/6/31/32* and constructed them into a transient expression vector with red fluorescent label. All primers used for plasmid construction are listed in Additional file 8: Table S5. Agroinfection assays were conducted as following: the constructs were transformed into *A. tumefaciens* (strain GV3101) by electroporation. The transformed bacterial cultures were grown and re-suspended using an inoculation buffer [10mM $MgCl_2$, 2mM acetosyringone, 100mM MES (pH 5.7)] for 3-5h at room temperature. The suspensions were adjusted to OD_{600} = 0.1 and were infiltrated into leaves of 4- to 6-week old *N. benthamiana* plants using needleless syringes.

Plant growth and pathogen inoculation

N. benthamiana seeds were donated by Dr. Yule Liu (Tsinghua University, China) and grown in mixed soil matrix (peat: vermiculite = 1:1) under a 16-h light (2,000 lux)/8-h dark photoperiod at 26±2°C with relative humidity 60 ± 5%. A TuMV infectious clone was kindly provided by Dr. Fernando Ponz (INIA, Laboratorio de Virologı´a Vegetal, Spain), a PVX infectious clone was kindly provided by Dr. Stuart MacFarlane (James Hutton Institute, UK) and a PMMoV infectious clone was created in our lab. The *Pst* DC3000 strain was kindly provided by Dr. Yule Liu (Tsinghua University, China). TuMV, PVX and PMMoV were inoculated onto the newly expanded leaves of *N. benthamiana*. Inoculum was obtained by homogenizing virus-infected leaves in phosphate buffer, and with phosphate buffer as mock control. The *Pst* DC3000 was cultured in King's B medium at 28°C. Leaves were syringe-infiltrated with a suspension of *Pst* DC3000 (OD_{600}=10^{-5}) in 10 mM of $MgCl_2$. Plants injected with 10 mM of $MgCl_2$ served as the corresponding control.

Expression analysis by RT-qPCR

RT-qPCR analysis was performed to confirm the expression of representative *NbFLA* genes. We used at least three independent biological replicates and three technical replicates. First-strand cDNA was synthesized from 0.5mg of RNA with PrimeScript RT reagent kit (TaKaRa). RT-qPCR was carried out by SYBR-green fluorescence using the Roche LightCycler®480 Real-Time PCR System. Relative gene expression levels were calculated according to the ΔΔCT method[64] and visualized in a heat map by Tbtools[52]. All primers used for RT-qPCR are listed in Additional file 8: Table S5.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its Additional files. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was financially supported by the National key research and development program(2018YFD0201200), China Agriculture Research System (CARS-24-C-04) and sponsored by K. C. Wong Magna Fund in Ningbo University. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' Contributions

WXY, ZHY initiated and designed the experiments. WXY, LYC, LLQ, JMF, HKL, YDK, LYW, PJJ and RSF performed the experiments and collected the data. WXY analyzed the data and wrote the manuscript. ZHY, YF and CJP revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Prof. M. J. Adams for correcting the English of the manuscript. We thank Dr. Fernando Ponz for providing the TuMV infectious clone, Dr. Stuart MacFarlane for providing the PVX infectious clone and Dr. Yule Liu for providing *Nicotiana benthamiana* seeds and the *Pst* DC3000 strain.

Authors' Information

College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310058, China

Xinyang Wu, Mengfei Ji, Jianping Chen

State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Key Laboratory of Biotechnology in Plant Protection of Ministry of Agriculture and Zhejiang Province, Institute of Plant Virology, Ningbo University, Ningbo, 315211, China

Xinyang Wu, Yuchao Lai, Lanqing Lv, Mengfei Ji, Kelei Han, Dankan Yan, Yuwen Lu, Jiejun Peng, Shaofei Rao, Fei Yan, Hongying Zheng, Jianping Chen

Correspondence to Hongying Zheng and Jianping Chen.

Abbreviations

FLAs: Fasciclin-like arabinogalactan proteins; AGPs: arabinogalactan proteins; GPI: glycosylphosphatidylinositol; TuMV: Turnip mosaic virus; *Pst* DC3000: *Pseudomonas syringae pv tomato* (*Pst*) strain DC3000; HRGPs: hydroxyproline-rich glycoproteins; TFs: Transcription factors; ABA: abscisic acid; MeJA: methyl jasmonate; TW: tension wood; PVX: Potato virus X; PMMoV: Pepper mottle mosaic virus; FAS1: Fasciclin 1;

References

1. De Lorenzo, G., et al., Cell wall traits that influence plant development, immunity, and bioconversion. Plant J. 2019;97(1):134-147.
2. Bacete, L., et al., Plant cell wall-mediated immunity: cell wall changes trigger disease resistance responses. Plant J. 2018;93(4):614-636.
3. Rui, Y. and J.R. Dinneny, A wall with integrity: surveillance and maintenance of the plant cell wall under stress. New Phytol. 2019;225(4):1428-1439.
4. Showalter, A.M., et al., A Bioinformatics Approach to the Identification, Classification, and Analysis of Hydroxyproline-Rich Glycoproteins. Plant Physiol. 2010;153(2):485-513.
5. Hijazi, M., et al., An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. Front Plant Sci. 2014;5:395.
6. Showalter, A.M., Arabinogalactan-proteins: structure, expression and function. Cell Mol Life Sci. 2001;58(10):1399-417.
7. Clout, N.J., D. Tisi and E. Hohenester, Novel fold revealed by the structure of a FAS1 domain pair from the insect cell adhesion molecule fasciclin I. Structure. 2003;11(2):197-203.

8. Johnson, K.L., et al., The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. Plant Physiol. 2003;133(4):1911-25.
9. Ma, H. and J. Zhao, Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (Oryza sativa L.). J Exp Bot. 2010;61(10):2647-2668.
10. Faik, A., J. Abouzouhair and F. Sarhan, Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (Triticum aestivum) and rice (Oryza sativa): identification and bioinformatic analyses. Mol Genet Genomics. 2006;276(5):478-94.
11. Zang, L., et al., Genome-Wide Analysis of the Fasciclin-Like Arabinogalactan Protein Gene Family Reveals Differential Expression Patterns, Localization, and Salt Stress Response in Populus. Front Plant Sci. 2015;6:1140
12. Huang, G.Q., et al., Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. Physiologia Plantarum. 2008;134(2):348-359.
13. Jun, L. and W. Xiaoming, Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (Brassica rapa L.). Mol Bio Rep. 2012;39(12):10541-10555.
14. MacMillan, C.P., et al., The fasciclin-like arabinogalactan protein family of Eucalyptus grandis contains members that impact wood biology and biomechanics. New Phytol. 2015;206(4):1314-1327.
15. Guerriero, G., et al., Identification of fasciclin-like arabinogalactan proteins in textile hemp (Cannabis sativa L.): in silico analyses and gene expression patterns in different tissues. BMC Genomics. 2017;18(1):741.
16. Wang, H., et al., Fasciclin-like arabinogalactan proteins, PtFLAs, play important roles in GA-mediated tension wood formation in Populus. Sci Rep. 2017;7(1):6182-13.
17. Zhang, Z., et al., Xylem sap in cotton contains proteins that contribute to environmental stress response and cell wall development. Funct Integr Genomics. 2015;15(1):17-26.
18. Wang, H., et al., Antisense expression of the fasciclin-like arabinogalactan protein FLA6 gene in Populus inhibits expression of its homologous genes and alters stem biomechanics and cell wall composition in transgenic trees. J Exp Bot. 2015;66(5):1291-1302.
19. Qin, L.X., et al., The cotton β-galactosyltransferase 1 (GalT1) that galactosylates arabinogalactan proteins participates in controlling fiber development. Plant J. 2017;89(5):957-971.
20. Li, Y., et al., Suppression of GhAGP4 gene expression repressed the initiation and elongation of cotton fiber. Plant Cell Rep. 2010;29(2):193-202.
21. Yang, X., et al., Expression profile analysis of genes involved in cell wall regeneration during protoplast culture in cotton by suppression subtractive hybridization and macroarray. J Exp Bot. 2008;59(13):3661-3674.
22. MacMillan, C.P., et al., Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. Plant J. 2010;62(4):689-703.

23. Cagnola, J.I., et al., Reduced expression of selected FASCICLIN-LIKE ARABINOGALACTAN PROTEIN genes associates with the abortion of kernels in field crops of Zea mays (maize) and of Arabidopsis seeds. Plant Cell Environ. 2018;41(3):661-674.
24. Li, J., et al., The fasciclin-like arabinogalactan protein gene, FLA3, is involved in microspore development of Arabidopsis. Plant J. 2010;64(3):482-497.
25. Seifert, G.J., H. Xue and T. Acet, The Arabidopsis thaliana FASCICLIN LIKE ARABINOGALACTAN PROTEIN 4 gene acts synergistically with abscisic acid signalling to control root growth. Ann Bot. 2014;114(6):1125-33.
26. Johnson, K.L., et al., A fasciclin-like arabinogalactan-protein (FLA) mutant of Arabidopsis thaliana, fla1, shows defects in shoot regeneration. PloS One. 2011;6(9):e25154.
27. Basu, D., et al., Glycosylation of a Fasciclin-Like Arabinogalactan-Protein (SOS5) Mediates Root Growth and Seed Mucilage Adherence via a Cell Wall Receptor-Like Kinase (FEI1/FEI2) Pathway in Arabidopsis. PLoS One. 2016;11(1): e0145092.
28. Xue, H., et al., Arabidopsis thaliana FLA4 functions as a glycan-stabilized soluble factor via its carboxy-proximal Fasciclin 1 domain. Plant J. 2017;91(4):613-630.
29. Li, W., et al., Identification of early salt stress responsive proteins in seedling roots of upland cotton (Gossypium hirsutum L.) employing iTRAQ-based proteomic technique. Front Plant Sci. 2015;6:732.
30. Guerriero, G., et al., Textile Hemp vs. Salinity: Insights from a Targeted Gene Expression Analysis. Genes. 2017;8(10):242.
31. Shi, H., et al., The Arabidopsis SOS5 Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. Plant Cell. 2003;15(1):19-32.
32. Takahashi, D., Y. Kawamura and M. Uemura, Cold acclimation is accompanied by complex responses of glycosylphosphatidylinositol (GPI)-anchored proteins in Arabidopsis. J Exp Bot. 2016;67(17):5203-15.
33. Ge, P., et al., iTRAQ-based quantitative proteomic analysis reveals new metabolic pathways of wheat seedling growth under hydrogen peroxide stress. Proteomics. 2013;13(20):3046-58.
34. Bombarely, A., et al., A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. Molecular plant-microbe interactions : MPMI. 2012;25(12):1523.
35. Lafarguette, F., et al., Poplar Genes Encoding Fasciclin-Like Arabinogalactan Proteins Are Highly Expressed in Tension Wood. New Phytol. 2004;164(1):107-121.
36. Yu, Y., et al., The grapevine R2R3-type MYB transcription factor VdMYB1 positively regulates defense responses by activating the stilbene synthase gene 2 (VdSTS2). BMC Plant Biol. 2019; 19(1): 478.
37. Kurilla, A., et al., Nectar- and stigma exudate-specific expression of an acidic chitinase could partially protect certain apple cultivars against fire blight disease. Planta. 2019;251(1): 20.
38. Shan, T., et al., The wheat R2R3-MYB transcription factor TaRIM1 participates in resistance response against the pathogen Rhizoctonia cerealis infection through regulating defense genes. Sci Rep. 2016; 6: 28777

39. Zhang, H., et al., NbCZF1, a Novel C2H2-Type Zinc Finger Protein, as a New Regulator of SsCut-Induced Plant Immunity in Nicotiana benthamiana. Plant Cell Physiol. 2016: 57(12).2472-2484.
40. Yu, Y.H., et al., Grape (Vitis vinifera) VvDOF3 functions as a transcription activator and enhances powdery mildew resistance. Plant Physiol Biochem. 2019;143: 183-189.
41. Chezem, W.R., et al., SG2-Type R2R3-MYB Transcription Factor MYB15 Controls Defense-Induced Lignification and Basal Immunity in Arabidopsis. Plant Cell. 2017;29(8): 1907-1926.
42. Zhang, Y., et al., The R2R3 MYB transcription factor MdMYB30 modulates plant resistance against pathogens by regulating cuticular wax biosynthesis. BMC Plant Biol. 2019;19(1):362.
43. Cutler, S.R., et al., Random GFP: : cDNA Fusions Enable Visualization of Subcellular Structures in Cells of Arabidopsis at a High Frequency. Proc Natl Acad Sci USA. 2000;97(7):3718-3723.
44. Dahiya, P., et al., A fasciclin-domain containing gene, ZeFLA11, is expressed exclusively in xylem elements that have reticulate wall thickenings in the stem vascular system of Zinnia elegans cv Envy. Planta. 2006;223(6):1281-1291.
45. Fang, Q., et al., A salt-stress-regulator from the Poplar R2R3 MYB family integrates the regulation of lateral root emergence and ABA signaling to mediate salt stress tolerance in Arabidopsis. Plant Physiol Biochem. 2017;114:100-110.
46. Fang, Q., et al., AtDIV2, an R-R-type MYB transcription factor of Arabidopsis, negatively regulates salt stress by modulating ABA signaling. Plant Cell Rep. 2018;37(11):1499-1511.
47. Sun, B., et al., TaZFP1, a C2H2 type-ZFP gene of T. aestivum, mediates salt stress tolerance of plants by modulating diverse stress-defensive physiological processes. Plant Physiol Biochem. 2019;136:127-142.
48. Lorrai, R., et al., Genome-wide RNA-seq analysis indicates that the DAG1 transcription factor promotes hypocotyl elongation acting on ABA, ethylene and auxin signaling. Sci Rep. 2018;8(1): 15895-13.
49. Sultana, N., et al., Ascorbate deficiency influences the leaf cell wall glycoproteome in Arabidopsis thaliana. Plant Cell Environ. 2015;38(2):375-384.
50. Perdiguero, P., et al., Gene expression trade-offs between defence and growth in English elm induced by Ophiostoma novo-ulmi. Plant Cell Environ. 2018;41(1):198-214.
51. Scott Gens, J., M. Fujiki and B.G. Pickard, Arabinogalactan protein and wall-associated kinase in a plasmalemmal reticulum with specialized vertices. Protoplasma. 2000;212(1):115-134.
52. Chen, C., et al., TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. BioRxiv. 2018:289660.
53. Marchler-Bauer, A., et al., CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43(D1):D222-D226.
54. Marchler-Bauer, A., et al., CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45(D1):D200-D203.

55. Almagro, A.J., et al., SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–423.
56. Eisenhaber, F., B. Eisenhaber and P. Bork, Prediction of Potential GPI-modification Sites in Proprotein Sequences. J Mol Biol. 1999;292(3):741-758.
57. Morten Bo Johansen, Lars Kiemer, Søren Brunak, Analysis and prediction of mammalian protein glycation. Glycobiology.2006;16(9):844-853.
58. Wilkins, M.R., et al., Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531-52.
59. Chou, K. and H. Shen, Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. PloS One. 2010;5(6):e11335.
60. Larkin, M.A., et al., Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8.
61. Bailey, T.L., et al., MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(suppl_2):W202-W208.
62. Lescot, M., et al., PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325-327.
63. Jin, J., et al., PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45(D1):D1040-D1045.
64. Livak, K.J. and T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.

Figures

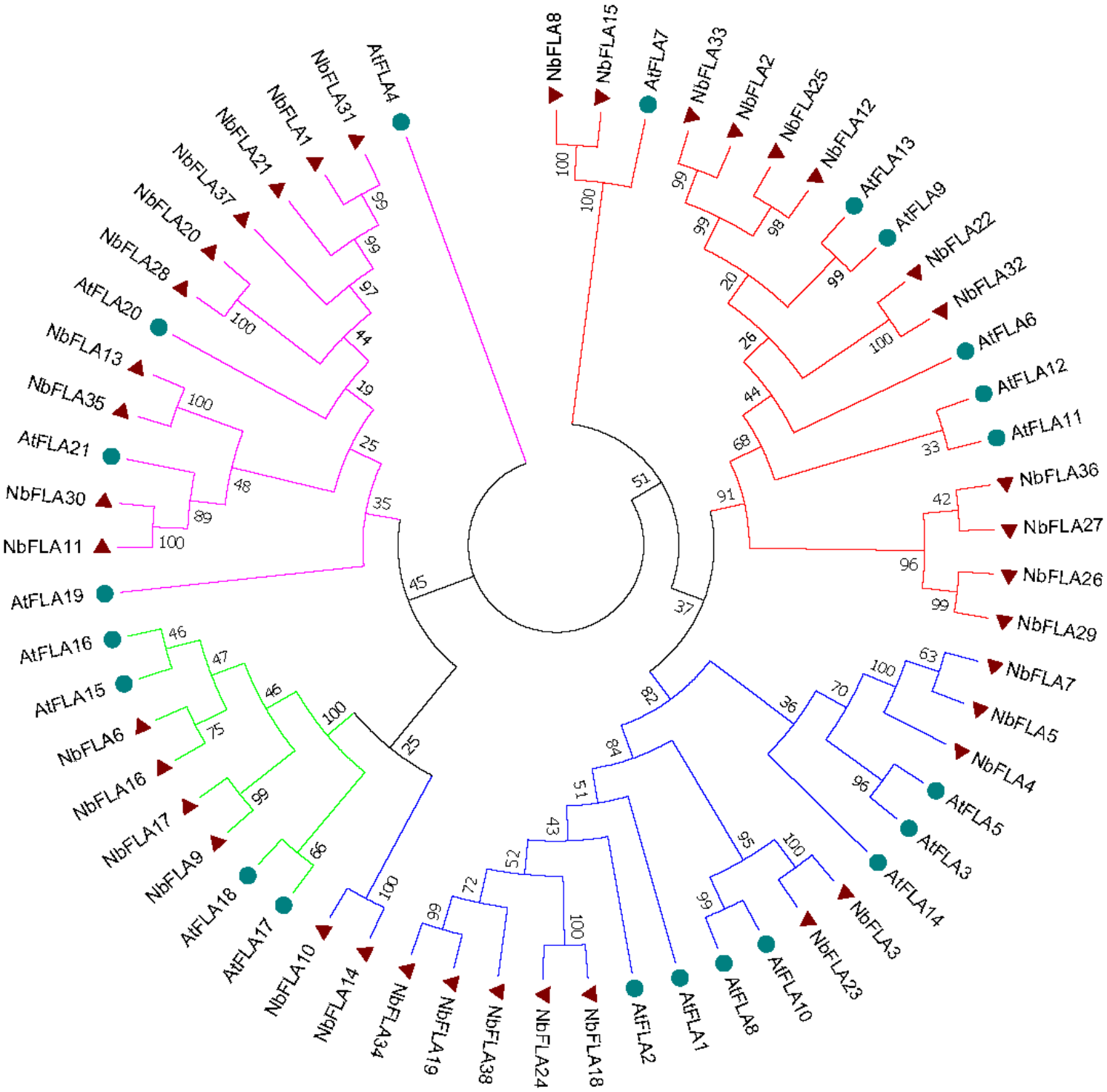

Figure 1

Unrooted phylogenetic tree representing relationships among FLA proteins of N. benthamiana and A. thaliana. All FLA proteins were divided into four subclasses represented by different colored clusters. Red, green, blue and pink clusters represent subclasses I, II, III and IV, respectively. The phylogenetic tree was constructed by the neighbor-joining method using MEGA7 software with 1000 bootstrap replicates.

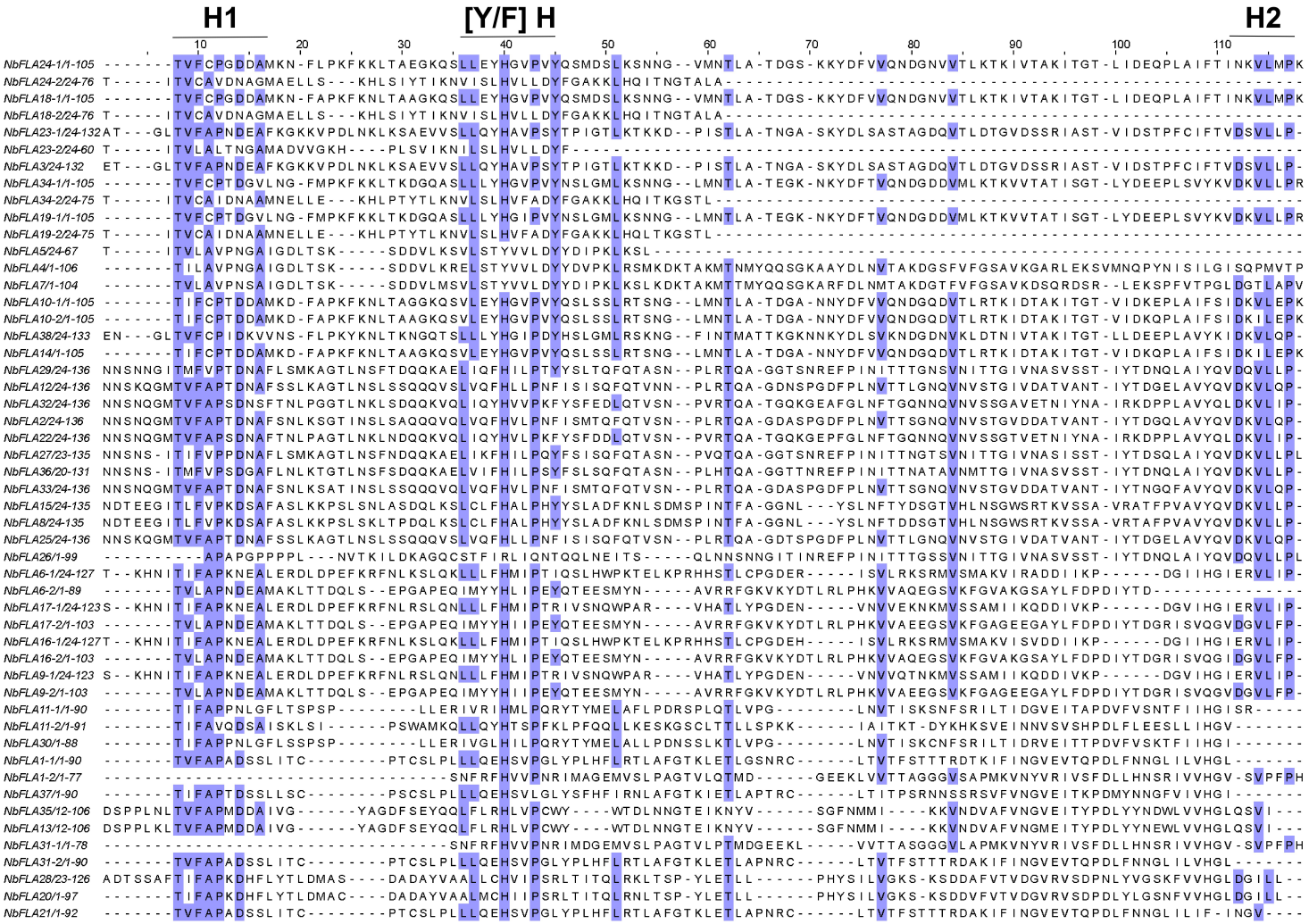

Figure 2

Phylogenetic relationship, gene structure and architecture of the conserved protein motifs in NbFLAs. a The phylogenetic tree was constructed based on the full-length sequences of NbFLA proteins. b Exon-intron structure of NbFLAs. Pink boxes indicate untranslated 5’- and 3’-regions; green boxes indicate exons; and black lines indicate introns. The fasciclin domains are shown by yellow boxes. c The motif composition. The motifs, numbered 1-20, are displayed in different colored boxes. The sequence information for each motif is provided in Additional file 1: Table S2.

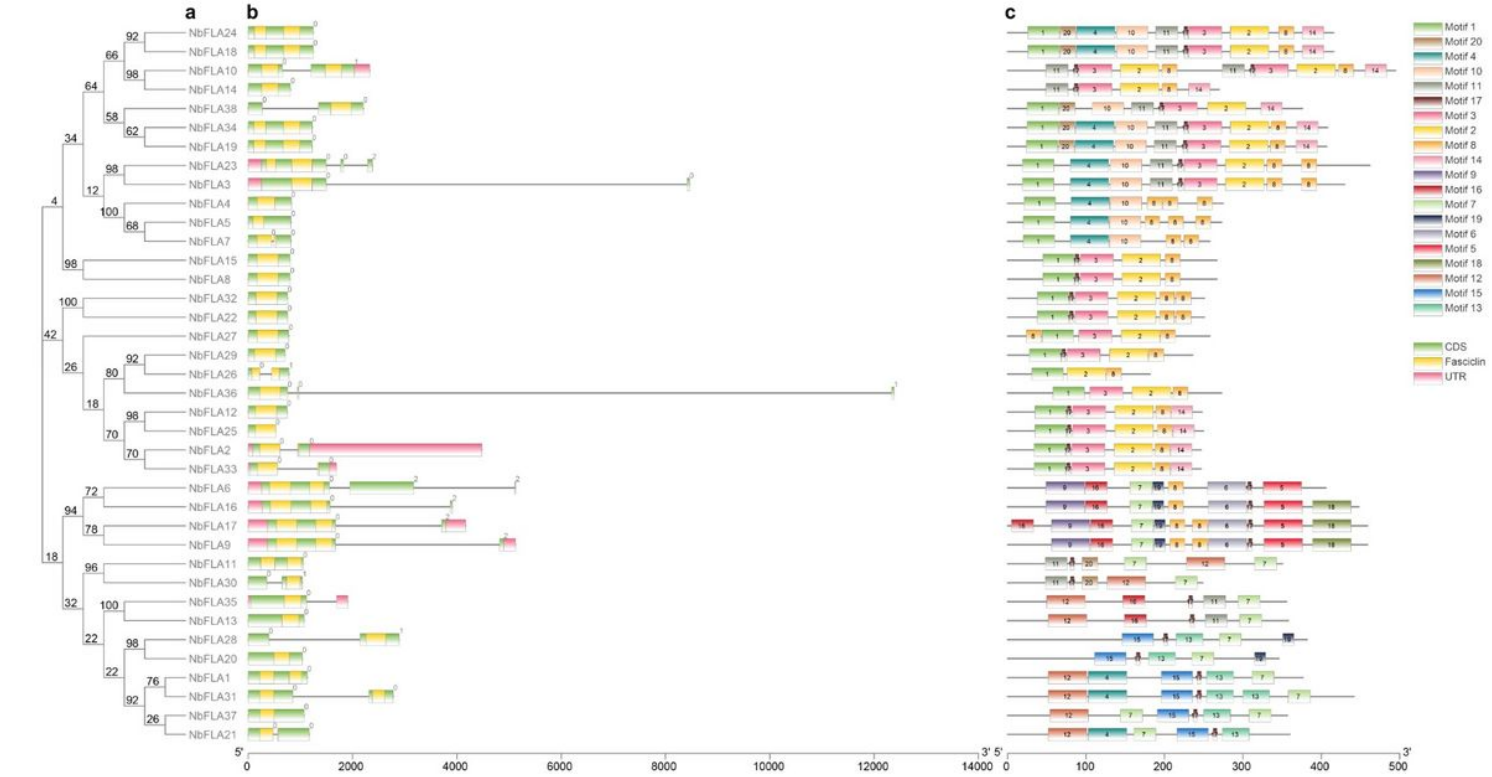

Figure 3

Phylogenetic relationship, gene structure and architecture of the conserved protein motifs in NbFLAs. a The phylogenetic tree was constructed based on the full-length sequences of NbFLA proteins. b Exon-intron structure of NbFLAs. Pink boxes indicate untranslated 5'- and 3'-regions; green boxes indicate exons; and black lines indicate introns. The fasciclin domains are shown by yellow boxes. c The motif composition. The motifs, numbered 1-20, are displayed in different colored boxes. The sequence information for each motif is provided in Additional file 1: Table S2.

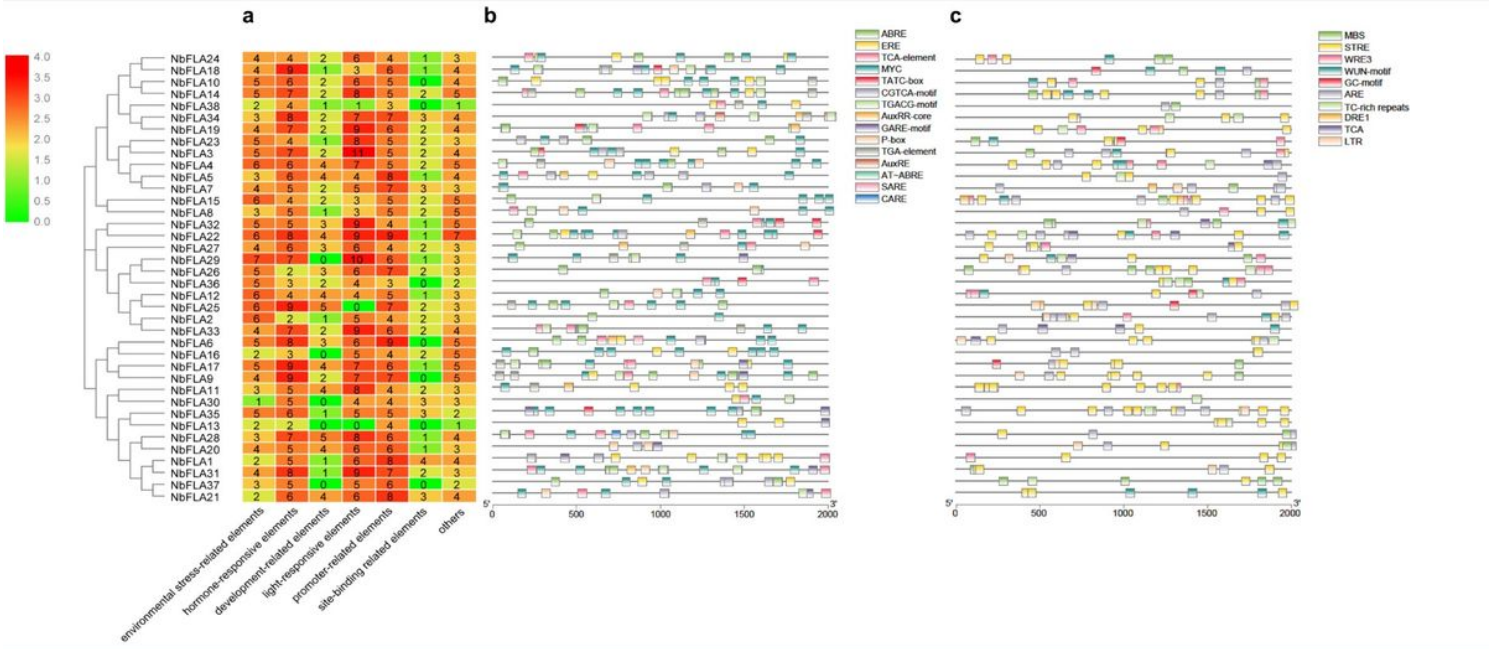

Figure 4

Prediction of cis-acting elements in NbFLAs. a numbers of cis-acting elements detected in the promoter region of each NbFLA gene. All cis-acting elements were divided into seven types. b Kind, quantity and position of environmental stress-related elements in NbFLAs. c Kind, quantity and position of hormone responsive elements in NbFLAs.

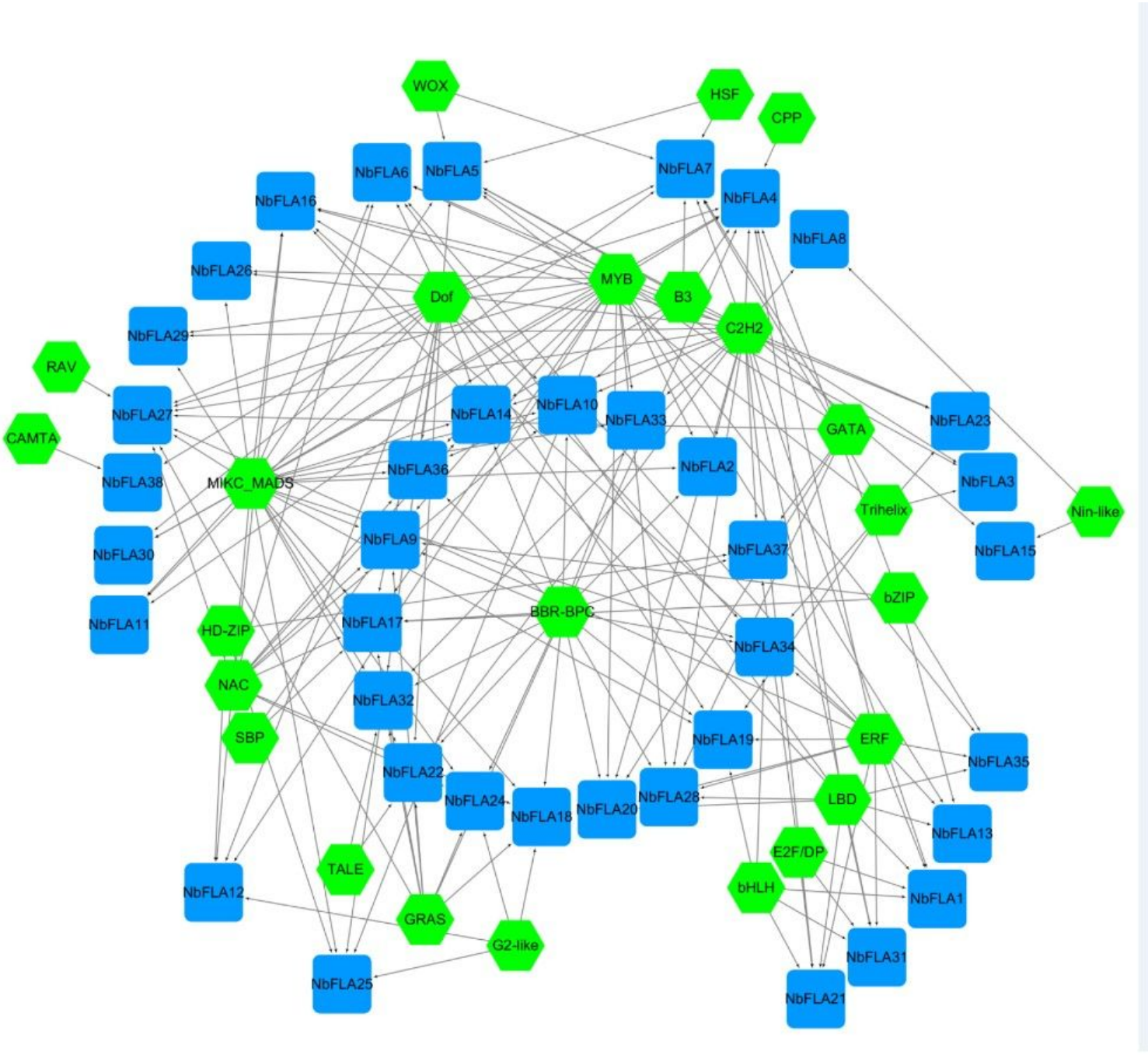

Figure 5

Regulation network between NbFLAs and potential TFs. Green hexagons represent transcription factors, blue rectangles represent NbFLAs, and black lines represent potential regulatory relationships.

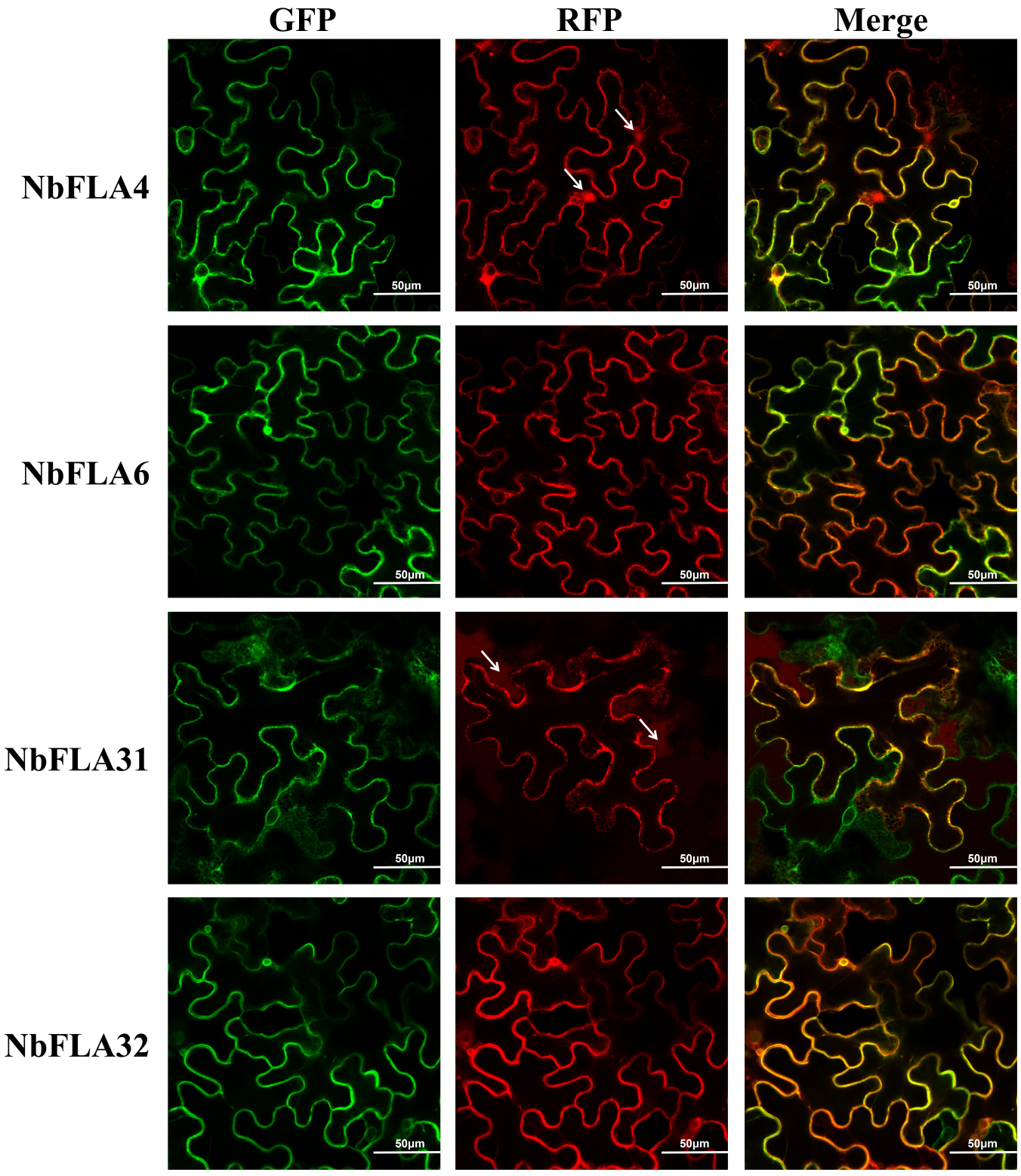

Figure 6

Subcellular localization of NbFLA4/6/31/32. Confocal microscopy images of N. benthamiana epidermal leaf cells co-expressing the membrane marker AtP1P2A-GFP (left panels) with NbFLA4-mCherry, NbFLA6-mCherry, NbFLA31-mCherry and NbFLA32-mCherry (middle panels), respectively. Merged images are shown in the right panels. Scale bars=50 μm. Arrows in the panel of NbFLA4-mCherry indicate red

fluorescence in the nucleus. Arrows in the panel of NbFLA31-mCherry indicate red fluorescence in the cytoplasm.

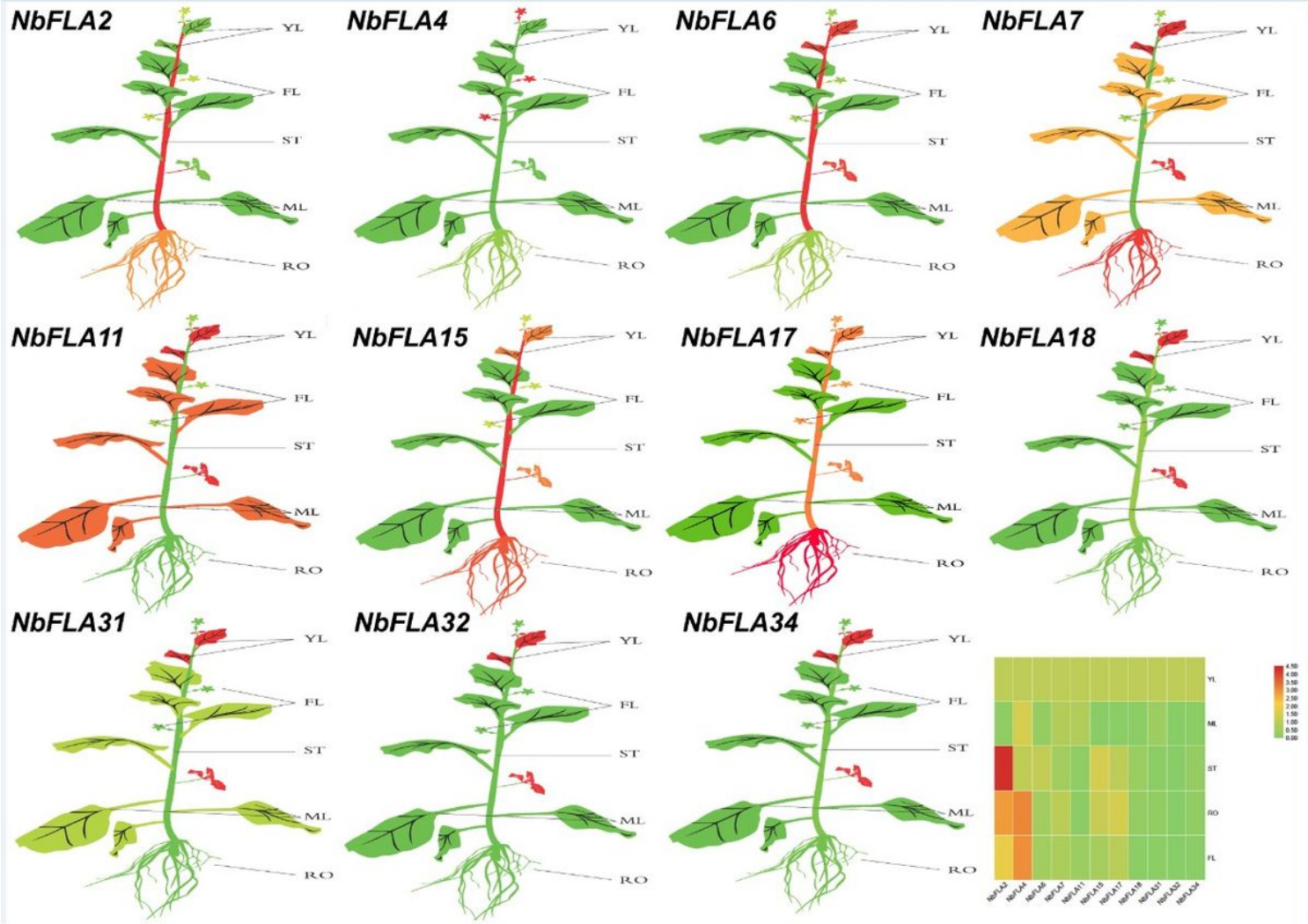

Figure 7

The differential expression of representative NbFLA genes in different tissues by RT-qPCR. YL: young leaf; MF: mature leaf; ST stem: RO root; FL: flower. The mean expression value was calculated from three independent biological replicates relative to that in young leaves. The mean expression values were visualized by Tbtools; red represents high expression level and green represents low expression level. The raw data of relative expression values and standard errors is provided in Additional file 6: Fig. S2.

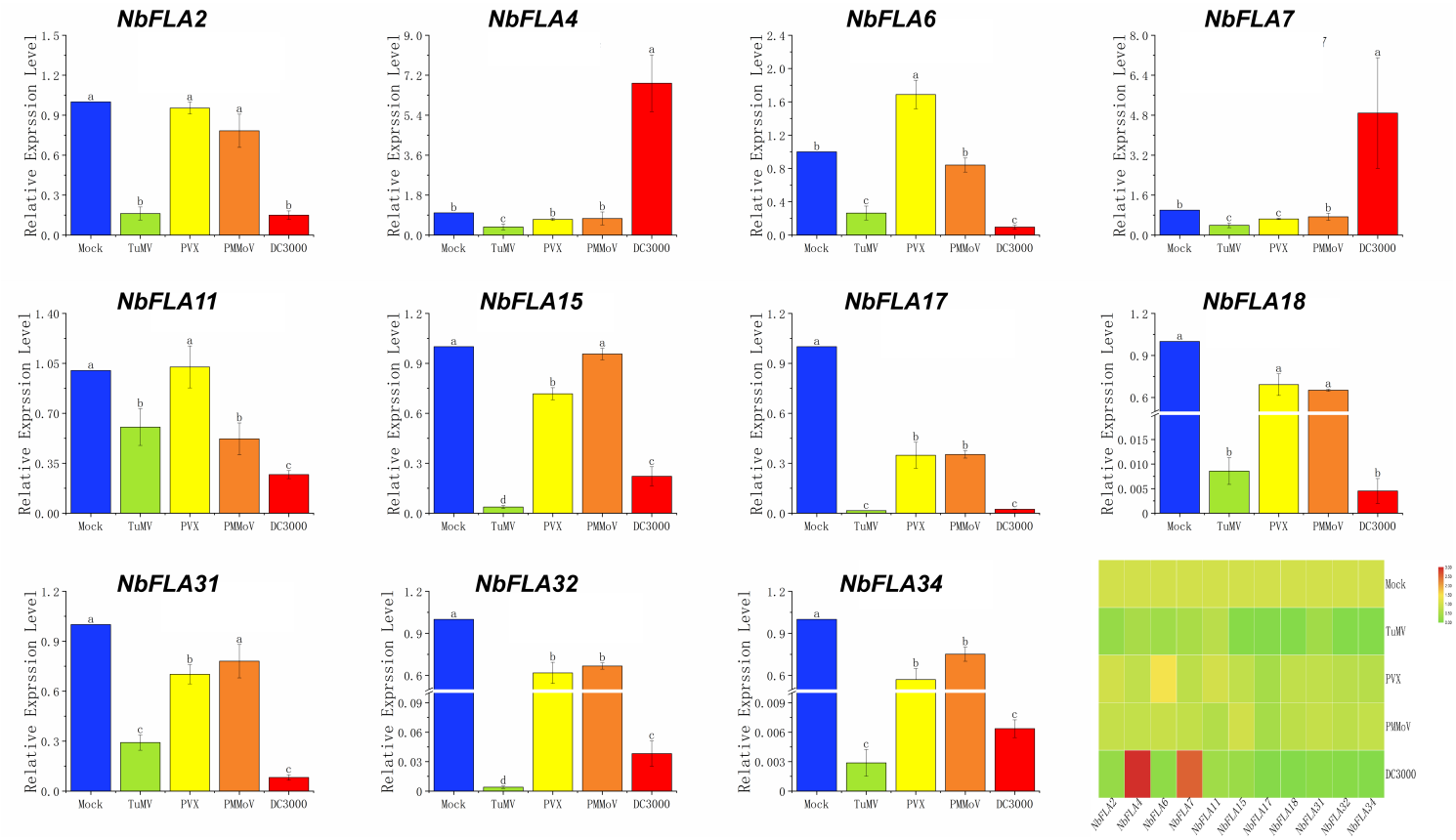

Figure 8

Expression analysis of representative NbFLA genes infected with different pathogens by RT-qPCR. The mean expression values were calculated from three independent biological replicates and are relative to mock-inoculated controls.

Supplementary Files

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

- supplement1.tif
- TableS1ListofNbFLACDSandproteinsequence.xls
- TableS2TheMEMEMotifSequenceandLengthofNbFLAs.xls
- TableS4potentialtranscriptionfactorsofNbFLAs.xls
- TableS5Primersusedinthisstudy.xls
- TalbeS3cisactingelementsinNbFLAs.xls
- Additionalfile2.jpg
- Additionalfile6.tif
- Additionalfile7.tif