Master's Thesis of Science in Agriculture

Hairs absent is Involved in Type I Trichome Initiation in Tomato

토마토에서 type I 모상체 발달에 관여하는

*Hairs absent*에 대한 연구

August 2018

Heejin Kim

Department of International Agricultural Technology

Graduate School of International Agricultural Technology

Seoul National University

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

35S	Cauliflower mosaic virus 35S promoter
AC	Ailsa Craig
bp	base pair
cDNA	Complementary DNA
DMS	Dissecting microscope
ESEM	Environmental scanning electron microscopy
gDNA	Genomic deoxyribonucleic acid
H (h)	*Hairs absent (hairs absent)*
ITAG	International Tomato Annotation Group
MYB	myeloblastosis transcription factor
MYC	myelocytomatosis transcription factor
PCR	Polymerase Chain Reaction
qRT-PCR	Quantitative Real-Time PCR
RNA	Ribonucleic Acid
WT	wild type
ZFP	Zinc finger protein

ABSTRACT

Hairs absent is involved in type I trichome initiation in tomato

Heejin Kim

Department of International Agricultural Technology

Graduate School of International Agricultural Technology

Seoul National University

Trichomes are specialized epidermal structures and can be categorized into two types: glandular and non-glandular trichomes. Trichomes play an important role in plant defense against biotic and abiotic stresses by serving as physical barriers or chemical factories. Arabidopsis has one type of unicellular trichomes, while tomato has seven types of multicellular trichomes. In the Rosids including Arabidopsis, the initiation pathway of trichomes has been well studied. However, in the Asterids including tomato, only a few genes related to trichome development were identified so far. To identify genes involved in trichome initiation of tomato, we characterized the tomato trichome mutant, *hairs absent* (*h*), which does not have the longest glandular trichomes (type I) on stems and sepals. Using a map-based cloning, we identified that *H* encodes a zinc finger protein homologous to Arabidopsis *GLABROUS*

INFLORESCENCE STEMS, which is a key regulator in trichome development on inflorescence stems. The *h* mutation showed whole deletion of *Solyc10g078970* gene. In addition, we identified additional *h* mutation alleles from wild species of tomato, *S. pimpinellifolium* and *S. pennellii*, both of which also do not have type I trichomes. These results indicate that *Solyc10g078970* is *H* gene. By developing knock-out mutant of *Solyc10g078990* gene, which is the closest homolog of *Solyc10g078970*, using the CRISPR-Cas9 system, we revealed that *Solyc10g078990* was involved in trichome development of leaves in tomato. RNA-seq comparison between wild-type and *h* mutant showed that 1570 genes were differentially regulated by *H* gene. These genes will be further studied for the identification of new trichome developmental genes and secondary metabolite biosynthetic genes.

Keywords

trichome, *hairs absent*, zinc finger protein, tomato, map-based cloning

Student Number: 2016-25982

INTRODUCTION

As sessile organisms, plants cannot escape from unfavorable growth conditions or stimuli throughout their lifetime. Consequence of this nature, plants have developed abilities to detect and respond to dynamic environment. The very first barrier in plants that interacts with its changing environments is epidermal surface.

Trichomes are hair-like structures differentiated from epidermal cells, and are getting attention as an excellent model for cell differentiation study due to its accessibility and functional importance [1]. Trichomes can be categorized into either non-glandular or glandular type, and generally non-glandular trichomes play a role in physical defense and glandular trichomes are in charge of chemical defense [1, 2]. Yet, trichomes are not crucial for plant growth, they serve diverse functions including protecting the plant against insects and UV irradiation, reducing water loss from transpiration, and increasing tolerance to abiotic stress conditions such as extreme temperatures [3-6].

The model plant Arabidopsis has one type of unicellular and non-glandular trichome consisted of three branches. The Arabidopsis trichome initiation pathway has been well studied and over 40 different genes were known to control the developmental processes [7, 8]. Genetic analysis revealed that groups of transcription factors are involved in the pathway. Zinc finger proteins (ZFP), MYB, basic helix-loop-helix (bHLH), WD40 repeat (WDR), and homeodomain leucine zipper (HD-Zip) proteins are known so far, and MYB, bHLH, WDR proteins assembled to form MBW complex. R2R3 MYBs including GLABRA1 (GL1), and MYB23 are involved as positive regulators, while R3 MYBs, which lack the activation domain, are involved as negative regulators. GLABRA3 (GL3),

ENHANCER OF GLABRA3 (EGL3), TRANSPARENT TESTA 8 (TT8), and MYC1 are involved as bHLH transcription factors, and TRANSPARENT TESTA GLABRA 1 (TTG1) participate as WDR [7]. The MBW complex activates HD-Zip protein GLABRA2 (GL2), which determines epidermal cell fate to become a trichome. On the upstream of the initiation pathway, a number of ZFPs are revealed to integrate hormonal signals and regulate MBW complex. GLABROUS INFLORESCENCE STEMS (GIS), GIS2, GIS3, ZFP8, ZFP5, and ZFP6 are all C2H2-type ZFPs and reported as a positive regulators [9].

In contrast to Arabidopsis, cultivated tomato (*Solanum lycopersicum*) has seven types of multicellular trichomes. The seven types of trichomes can be divided into two groups, four glandular trichomes (type I, IV, VI, VII) and three non-glandular trichomes (type II, III, V) [2]. Among these seven trichomes, the most studied trichomes are type I and VI. Type I trichomes are characterized by the longest stalk (~2 mm) with a multicellular base and small glandular head on top. Type VI trichome has a short (~0.1 mm) stalk and a four-celled glandular head. Type IV trichomes have a unicellular base, a multicellular stalk shorter than type I (~0.3 mm), and a small glandular tip. The glandular head of type IV produce acyl sugar and this trichome type is highly abundant in wild species *S. pennellii* but are reported to be absent in cultivated tomato (*S. lycopersicum*) [10, 11]. The great diversity of trichome types and densities, as well as chemical compositions within tomato species has been revealed to date [12]. Numerous studies have shown that trichomes play an important role in resistance of tomato species to herbivores. Non-glandular trichomes may contribute to resistance by mechanically obstructing the movement of herbivores across the plant surface [13]. More importantly, resistance is also mediated by glandular trichome-borne metabolites that exert toxic effects on insect herbivores, or physically entrap the insect upon rupture of the trichome gland. Among the broad classes of tomato

leaf compounds acyl sugars, methyl ketones, terpenes, and alkaloids are known to deter insect herbivores.

Tomato belongs to the Asterids whereas Arabidopsis belongs to the Rosids, and whether these two groups share a similar trichome initiation pathway is not clearly verified. In the clade of Asterids, the R2R3 MYB-related transcription factor MIXTA is known to be involved in trichome development. However, expression of MIXTA in Arabidopsis did not affect trichome phenotype, and expression of GL1 in tobacco also had no effect on trichome development. The results imply that trichome initiation pathway in the Rosids and the Asterids might not be conserved and independently evolved [14]. Only a few genes controlling the initiation of multicellular trichomes have been discovered. In tomato, *Woolly* (*Wo*) and *CUTIN DEFICIENT2* (*CD2*), which encodes HD-Zip IV, are reported to regulate trichome development. Silencing of *Wo* showed trichome-less phenotype on the epidermis of all aerial parts, and the mutant with a premature stop codon on *CD2* showed over 3-fold decreased number of type VI glandular trichomes [15, 16].

hairs absent (*h*) is a spontaneous tomato mutant that lacks type I trichomes on stems and sepals [17]. In previous studies, F_1 plants obtained from crossing *h* mutants with wild-type plants showed wild-type phenotype, and F_2 progenies from selfing the F_1 plants showed the ratio of 3:1 for wild type to mutant phenotype, implying that the phenotype is caused by a single recessive mutation. Fine-mapping using the map-based cloning for *H* locus narrowed the region down to a 109 kb interval on chromosome 10. Here, we selected a C2H2 type zinc finger transcription factor, *Solyc10g078970* (hereafter referred as *ZFP970*), as a candidate gene for *H*, and conducted functional studies. Transcriptomes obtained from wild-type and *h* mutant stems were compared to examine which genes were regulated by *H*, and 1570 differentially expressed genes were identified. Go and MapMan

analyses showed groups belonging to transcription factors, secondary metabolite biosynthetic pathways, and immune response related genes were significantly different between wild-type and *h* stems.

MATERIALS AND METHODS

1. Plant materials and growth conditions

Tomato (*Solanum lycopersicum*) cv Alisa Craig (AC) (LA2838A) was used as the wild type for all experiments. Seeds for *hairs absent* (LA3172), its parental cultivar Alisa Craig, M-82 (LA3475), and introgression line (IL)10-2 (LA4089), and wild tomato species seeds for *S. pennellii* (LA0716) and *S. pimpinellifolium* (LA1589) were obtained from C.M. Rick Tomato Genetics Resource Center (University of California, Davis, CA, USA). Tomato seeds were sown on half strength Murashige and Skoog (MS) and Gamborg's B5 medium (#M0231, Duchefa, Haarlem, Netherlands) containing 0.8% agar and 3% sucrose and placed in a growth room at 25℃ under a 16h-light/8h-dark cycle with the light intensity of 50 μE m^{-2} s^{-1}. After 6 to 7 days, seedlings were transferred to Jiffy soil pellets (Hummert International, Earth City, MO, USA) and grown in a greenhouse under natural light. After 3 weeks, plants were transferred to 25-cm-diameter pots.

For the morphology analysis, one-week-old young seedlings were transferred into Jiffy soil pellets (Hummert International, Earth City, MO, USA) and grown in a greenhouse up to 4 weeks. *Nicotiana benthamiana* seeds were sown on a standard soil mixture (#16GA11014, SeoulBio, Korea) and grown in a growth chamber at 25℃ with 16h-light/8h-dark cycle. Four- to five-week-old plants were moved to dark condition and watered sufficiently a day before agroinfiltraton. *Arabidopsis thaliana gis* mutant seeds were purchased from the Arabidopsis Biological Resource Center (ABRC, Columbus, OH, USA). Arabidopsis seeds were germinated on half strength MS basal medium (#M0221, Duchefa, Haarlem, Netherlands) containing 0.6% agar, 0.05% MES and 1% sucrose, and

grown in a growth chamber at 22℃ with 16h-light/8h-dark cycle and light intensity of 150 μE m^{-2} s^{-1}.

2. Morphology analysis

A dissecting microscope (DMS) (Leica CH-M205A, Wetzlar, Germany) was used to image tomato trichomes on hypocotyls, main stems, leaves, and sepals. For Arabidopsis, internodes of main stems, the first branches, the first to third cauline leaves, and sepals were imaged. Environmental scanning electron microscopy (ESEM) (Hitachi TM3030plus, Tokyo, Japan) equipped with DEBEN coolstage (Deben, London, UK) was used to fix and examine epidermal tissues of tomato and Arabidopsis. Images were captured using 15kV to minimize surface charging of trichomes.

3. Map-based cloning of *H* and RT-PCR

Previous research has found that *h* locus is located within a 109 kb interval between SGN-U578758 and SGN-U573045 on chromosome 10. Within this interval, 14 candidate genes were predicted based on ITAG 3.2 gene models offered by the Sol Genomics Network (https://solgenomics.net/jbrowse_solgenomics/). The RNA expression patterns of the 14 candidates were examined between wild-type and *h* plants through Reverse-Transcription (RT)-PCR. Total RNA was extracted from the young stems in about 2 cm length of right below the apical bud of wild-type and *h* plants with Trizol reagent (#15596018, Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocol. cDNA was synthesized with 2ug of total RNA using oligo(dT) primed RTase (#EP0442, Thermo

Fischer Scientific, Waltham, MA, USA) according to manufacturer's protocol. The PCR reaction was performed with Solg™ 2X Taq PCR Smart mix 2 (#SEF02, Solgent, Daejeon, Korea) in a 20 uL using an GeneAmp PCR system 9700 (#805S1241207, Applied Biosystems, Foster City, CA, USA) with the following cycling program: 2 min at 95℃, 30 or 35 cycles of 20 s at 95℃ 40 s at 58℃, 30 s at 72℃, and one cycle of 5 min at 72℃. Primers used to amplify each gene are listed in Table 1.

4. Genetic analysis of wild tomato species

Genomic DNA and total RNA was extracted, and cDNA was synthesized from two cultivated wild-type tomatoes (M-82, Ailsa Craig), two wild species (LA0716, LA1589), introgression line (IL) 10-2, and *h* mutant. Genomic DNA PCR and RT-PCR analyses of *ZFP970* and *ZFP990* were performed with the six tomato samples using the primers listed in Table 1.

The *hairs absent* mutant was crossed to IL 10-2 and LA1589, respectively, by collecting and transferring pollens from the open flower of *h* mutants to the stigma of IL 10-2 and LA1589. The resulting F_1 plants were grown and imaged as described above.

5. Subcellular localization

pBCo-DC-YFP [18] destination vector and pBCo-*Myc2*:CFP control vector were received from Prof. Choonkyun Jung (Graduate school of international agricultural technology, Seoul National University, Seoul, Korea). pBCo-*ZFP970*:YFP and pBCo-

ZFP990:YFP vectors were constructed to tag yellow fluorescent protein at C-terminal in frame with the following method. *Each of ZFP970 and ZFP990* gene was amplified without stop codon and cloned into pGEM-T easy vector. Cloned gene was cut with *BamH*I and *Xho*I restriction enzymes and ligated into pENTR3cDual entry vector (#A10464, Invetrogen, Carlsbad, California, USA). The genes were transferred from the entry vector to pBCo-DC-YFP destination vector via Gateway LR reaction system (#11791-020, Invitrogen, Carlsbad, California, USA). Primer sequences used for vector construction are listed in Table 1. pBCo-*ZFP970*:YFP and pBCo-*Myc2*:CFP vectors were transformed into *Agrobacterium tumefaciens* GV3101 strain, respectively. The single colony of *A. tumefaciens* containing each construct was inoculated into 5mL LB medium supplemented with spectinomycin (100mg/L), and grown at 28℃, 220 rpm for 2 days. 1mL of the culture is transferred to 50mL LB liquid medium with 10mM MES (pH 5.6) and 40mM acetosyringone. The bacteria were cultured in a shaking incubator at 28℃ until reached the $O.D._{600}$ value of 0.8 and spun down at 4000 rpm for 10 min. Pellets were resuspended in 50mL of 10mM $MgCl_2$ with 100mM of acetosyringone and kept at room temperature for 3 hours. Infiltration was performed with a 1mL syringe without needle. Third or fourth flat leaves were chosen to be injected. Confocal microscopy images were obtained after 40 hours of injection. Transiently transformed tobacco leaves were cut into 5*5mm size and mounted on slide glasses. YFP and CFP fluorescent signals were visualized with Leica SP8x gSTED confocal laser scanning microscope using 60x water immersion objective. Excitation/emission conditions were 514nm/520-540nm for YFP, and 405nm/460-485nm for CFP. Localization experiment for ZFP990 was performed with the same protocol used for ZFP970.

6. Total RNA extraction and qRT-PCR for expression pattern analysis

Expanding terminal leaflet of second to third emerging compound leaf, individual peel and peeled-off core of stem right below the shoot apical meristem, premature floral buds of 1 cm in length, and 5 cm length of root right above the root apical meristem tissues of wild-type tomato samples were collected by snap-freezing in liquid nitrogen. Total RNA extraction and cDNA synthesis were performed as mentioned above. Real-time PCR analysis was performed using Solg^{TM} 2x real time PCR smart mix and EvaGreen (#SRH71, Solgent, Daejeon, Korea) with a Mx 3000P Real-time PCR system (Agilent technologies, Santa Clara, CA, USA). *SlACT7* (Solyc03g078400) was used as an internal standard. Three biological replicates were analyzed for every experiment.

7. Phylogenetic analysis and conserved sequence analysis

The homologs of ZFP970 protein in tomato and *Arabidopsis* were obtained by BLAST search using non-redundant protein sequences database on National Center for Biotechnology Information (https://blast.ncbi.nlm.nih.gov). Phylogenetic tree was constructed based on the amino acid sequences of homologs by using MEGA7 with Neighbor-joining method and bootstrap test of 1000 replicates.

The amino acid sequences of ZFP970 homologs are aligned by using ClustalW and displayed with Color Align Conservation tool (http://www.bioinformatics.org/sms2/color_align_cons.html)

8. Generation of *ZFP990* knock-out plants

To make *ZFP990* knockout vector, single-guide RNA (sgRNA) candidates for *ZFP990* were designed by using CRISPR RGEN Tools (http://rgenome.ibs.re.kr), and pHAtC::*ZFP*990-sgRNA vectors were constructed as described previously [23]. The sequence information of sgRNAs are listed in Table 1. pHAtC binary vector was received from Dr. Sang-Gyu Kim (Institute for Basic Science, Daejeon, Korea). Constructed vector was introduced into *Agrobacterium tumefaciens* strain LBA4404 and used to transform wild-type cotyledon explants as described previously [22]. 432 bp region encompassing gRNA target sequence was amplified using the primer set listed in Table 1.

9. Arabidopsis transgenic plants

The pGEM-T easy-*ZFP970* vector was digested with *BamH*I and *Xho*I restriction enzyme and cloned into pENTR3cdual entry vector. pBCo-DC destination vector [18], which has duplicated *35S* promoter, attL1, and attL2 recombination sites, was used to construct *ZFP970* overexpression vector. *ZFP970* gene was transferred from the entry vector to the destination vector by using Gateway LR Clonase II enzyme mix (#11791020, Invitrogen, Carlsbad, California, USA) according to the manufacturer's protocol. The constructed pBCo-*ZFP970* vector was transformed into *Agrobacterium tumefaciens* strain GV3101 to overexpress *ZFP970* in Arabidopsis. Arabidopsis thaliana *gis* mutant was transformed by using floral dip method and transgenic plants were screened as described previously [24].

10. cDNA library construction and massively parallel sequencing

RNA-Seq paired end libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit v2 (#RS-122-2001, Illumina, San Diego, CA). Starting with total RNA, mRNA purified using poly (A) selection or rRNA depleted, then RNA was chemically fragmented and converted into single-stranded cDNA using random hexamer priming. Next, the second strand was generated to create double-stranded cDNA. Library construction began with generation of blunt-end cDNA fragments from ds-cDNA. Then A-base was added to the blunt-end in order to make them ready for ligation of sequencing adapters. After the size selection of ligates, the ligated cDNA fragments which contain adapter sequences were enhanced via PCR using adapter specific primers. The library was quantified with KAPA library quantification kit (#KK4854, Kapa biosystems) following the manufacturer's instructions. Each library was loaded on Illumina Hiseq2000 platform for high-throughput sequencing to ensure that each sample meets the desired average sequencing depth.

11. Preprocessing and short read mapping,

Sequence data in which quality of base pair is upper than $Q \geq 20$ were extracted by SolexaQA [25]. Trimming resulted in reads with a mean length of 98.05 bp across all samples and a minimum length of 25 bp. Trimmed reads were mapped using the RNA-seq mapping algorithm implemented in bowtie2 (v2.1.0) software [26, 27] to the reference transcripts of *Solanum lycopersicum* (iTAGv2.3) downloaded from the Phytozome database (http://phytozome.jgi.doe.gov/) allowing all aligning with a maximum of two mismatches. The number of mapped clean reads for each gene was counted and then normalized with DESeq package in R [28] to avoid bias due to difference of sequencing

amount.

12. Identification of differentially expressed genes (DEG)

Differential expression analysis was performed using the DESeq R package, which provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. After statistical analysis, the DEGs were identified using significance analysis by DESeq R library and at least two-fold changes (either up- or down-regulation) being considered significant. All correlation analysis, hierarchical clustering was performed using AMAP library in R [29].

13. Functional enrichment analysis

For functional annotation of each DEG gene list, GO (Gene Ontology) analysis was carried out based on the sequence similarity (e-value cut off≤ 1e-10) of protein in Gene Ontology database [30]. The number of genes assigned in each GO term was counted using in-house scripts of SEEDERS Co. In addition, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway was studied using the sequence similarity of tomato protein in KEGG database [31]. A BLASTx search (cut-off e-value 1e-50, identity > 80%) was performed against the Transcription Factor (TF) databases of tomato downloaded from PlantTFDB v3.0 [32]. MapMan (v3.6.0RC1) is used for data visualization and functional annotation of transcriptome [33] (http://mapman.gabipd.org/). Reference transcripts with log2-fold-change values extracted from the RNA-seq data were cross referenced with *Solanum lycopersicum* (ITAG v2.3) provided by MapMan and mapped into the pathways offered in

the software.

RESULTS

1. The *h* mutant is impaired in the development of type I trichomes on stems and sepals

Light microscopy was used to compare the morphology of trichomes on leaves, stems, hypocotyls, and sepals of the h mutant with its wild-type parent (cv Alisa Craig) (Fig. 1). *h* mutant had all types of trichomes on leaves and hypocotyls but did not have type I trichomes on stems and sepals. To examine the structure of all types of trichomes in detail, we used SEM (Fig. 2). In agreement with the light microscopic observations, the most obvious *h* phenotype was the absence of type I trichomes on stems and sepals. However, other phenotypes related to plant growth or fertility were not observed.

2. Map-based cloning of *H*

The *h* locus was previously mapped between TG406 and PGAL markers on chromosome 10 [34]. Fine mapping of the locus allowed us to position h to a 109 kb region flanked by markers SGN-U578758 and SGN-U573045 on chromosome 10 (Fig. 3). Among the 14 hypothetical genes in this region, the tomato genome locus *Solyc10g078970* and *Solyc10g078990* were predicted to encode transcription factors that are highly homologous to Arabidopsis GLABROUS INFLORESCENCE STEMS (GIS) family proteins, which encode transcription factors and regulate trichome initiation on the inflorescence organs in Arabidopsis thaliana [35]. As *gis* mutant in Arabidopsis shows no trichomes on stems like *h* mutant, we considered *Solyc10g078970* or *Solyc10g078990* as a good candidate gene for

H. Reverse transcription (RT)-PCR analysis showed that *Solyc10g078990* transcripts accumulated in both wild-type and *h* mutant, whereas *Solyc10g078970* transcripts accumulated only in the wild-type but not in the *h* mutant (Fig. 4), suggesting that *Solyc10g078970* is *H*. PCR reaction of full-length genomic DNA of *Solyc10g078970* yielded the DNA fragment of 1 kb from wild-type (AC) but failed to amplify any fragment from *h* mutant (Fig. 5A). Further genomic DNA PCR with a primer set, which are from 5' and 3' non-coding regions of *Solyc10g078970*, showed a ~1.1 kb polymorphism between wild-type and *h* plants (Fig. 5B). Sequencing of *Solyc10g078970* genomic DNA clone from *h* showed a 1.4 kb deletion including whole coding region and 0.3 kb addition in the 3' non-coding region (Fig. 6). These findings further supported the inference that *Solyc10g078970* is the gene responsible for the *h* phenotype.

3. Identification of new alleles from tomato wild species

h phenotype is also found in wild tomato species *S. pennellii* (LA0716), *S. pimpinellifolium* (LA1589) and *S. pennellii* introgression line (IL) 10-2 (LA4089) (Fig. 7). IL 10-2 has tomato genome background except some part of chromosome 10 that is substituted to *S. pennellii* genome [36]. To verify whether the *h* phenotype observed in the wild species is caused by the same gene that is mutated in *h*, complementation tests were conducted, F_1 progenies were obtained by crossing the *h* mutant with the wild species, respectively. All F_1 plants showed *h*-like phenotype (Fig. 8), indicating that the *h* phenotype of the wild species is caused by the same gene of *h* mutant.

RT-PCR analysis showed that *ZFP970* transcripts were amplified only in the wild-type but not in the wild species (Fig. 9). Comparison of deduced amino acid sequence of

ZFP970 from M-82 (wild-type of IL10-2) and IL10-2 showed 96% homology (Fig. 10A). However, there was no mutation causing premature stop codon (nonsense mutation) in IL10-2. Further characterization of promoter region of *Solyc10g078970* showed several polymorphisms between M-82 and IL10-2 (Fig. 10B). Thus, no expression of *ZFP970* in IL10-2 might be due to structural modification of the promoter region which is important for the regulation of transcription. Sequencing of *ZFP970* from LA1589 revealed a single nucleotide deletion at base 286, generating a frame shift and a premature stop codon at base 349 in exon 1 (Fig. 11). All together, these results demonstrated that *ZFP970* is *H*.

4. ZFP970 is a nuclear localization transcription factor

Previously, one of the GIS family protein GIS3 was known to localize into nucleus [37]. In addition to the reference, as a transcription factor and inferred from the predicted nuclear localization signal on its sequence, ZFP970 is assumed to be a nuclear localized protein. To verify where the ZFP970 proteins are expressed in a cellular level, yellow fluorescent protein tagged ZFP970 protein was transiently expressed in tobacco and visualized by confocal microscopy. To confirm the location of the signal, MYC2, which is a nuclear localized protein, fused with cyan fluorescent protein were co-transformed [38]. Yellow fluorescent signals and cyan fluorescent signals were co-localized on the nucleus of tobacco cells. This result indicated that ZFP970 protein is localized to nucleus in cells. (Fig. 12)

5. Expression pattern of *ZFP970*

To examine the expression pattern of *ZFP970* gene from wild-type plants, leaves, epidermal stems, peeled-off stems, floral buds, and root tissue were used. The expression levels of RNA were analyzed by quantitative RT-PCR. The qRT-PCR result showed the highest expression levels in floral buds followed by leaves, epidermal stems, and peeled-off stems and the lowest expression in root (Fig. 13).

6. Phylogenetic analysis and conserved sequence analysis

ZFP970 homologs in Arabidopsis and tomato are used to construct phylogenetic tree (Fig. 14). ZFP970 and eight tomato zinc finger proteins were clustered into group I with six Arabidopsis zinc finger proteins. Arabidopsis genes in this clade are known to be involved in trichome initiation pathway [35, 37-39]. Nine tomato genes and eight Arabidopsis genes were clustered in the group II. In this group, Arabidopsis genes are known to participate in seedling or flower development, while the function of tomato genes are not studied yet [40-42]. Arabidopsis and tomato genes in the group I were used for conserved sequence analysis to find common motifs they share. The genes in the group I have three highly conserved motifs, which are RLFGV-like motif, C2H2-type zinc finger motif followed by nuclear localization signal, and EAR motif (Fig. 15).

7. Overexpression of *ZFP970* affect trichome phenotype in Arabidopsis

Several reports showed that MBW complex genes involved in the trichome initiation of Arabidopsis did not affect trichome development in the plant species of the Asterids clade. However, phylogenetic analysis showed that Arabidopsis GIS family and

ZFP970 homologs are highly conserved. This result arouse an inquiry about genetic interchangeability between Arabidopsis and tomato on the zinc finger proteins that are positioned on the upstream of MBW complex in the trichome regulation network. The phenotype of Arabidopsis *gis* (N342188) mutant is described as decrease of trichome densities on inflorescence stem, branches, and sepal compare to wild-type. To verify whether *ZFP970* from tomato could replace the function of *AtGIS*, Arabidopsis transgenic plant expressing *ZFP970* under the constitutive promoter had been generated. Overexpression of *ZFP970* in *gis* caused increased trichome density in sepal and developed ectopic trichomes in silique, but did not show recovery of wild-type phenotype on stem and paraclades (Fig. 16). In addition, most of siliques were short and sterile, suggesting *ZFP970* might have other functions in Arabidopsis.

8. De novo transcriptome assembly

RNA sequencing with the Illumina Hiseq 2000 produced 32,134,595 and 33,968,189 paired-end 101 bp reads on average for epidermal tissue of AC and *h* respectively, corresponding to 3 billion base pairs of sequence. For whole stem of AC and *h*, 17,432,487 and 20,508,317 paired-end 126 bp reads were produced on average. The total number of high-quality reads were 29,827,645 for the epidermal tissue of AC, 31,535,500 for the epidermal tissue of *h*, 19,080,073 for the whole stem of AC, and 16,194,990 for the whole stem of *h*. In total, 82.71% reached a strict quality score threshold of Q≥20 bases and read length ≥25 bp, and these were used for de novo assembly (Table 2). Trimmed reads were mapped to the reference transcripts of Solanum lycopersicum (iTAGv2.3) provided by Phytozome v10.1. Among 34,727 reference genes, transcripts matches were

found for 30,657 genes with 85.27% of mapping rate, of which 26,649 genes have functional description (Table 3).

9. Identification of differentially expressed genes and functional annotation

RNA-Seq data were used for the identification of differentially expressed genes (DEGs) in stem epidermal tissue of AC and *h*. DEGs were identified using the following filters: adjusted p-value<0.01, and log2 (fold change) ≤-1 and ≥1. From the 30,657 genes, which mapped by the RNA-seq data, 1570 DEGs were identified between the stem epidermal tissue of AC and *h* (Table 4, Fig. 17). GO enrichment analysis was performed to investigate the distribution of DEGs in in biological process (BP), cellular component (CC) and molecular function (MF). 64 GO terms (BP: 47 terms, CC: 4 terms, MF: 13 terms) were enriched. Transport, multicellular organismal development, and nucleic acid binding were the most significantly enriched terms in BP and MF (Table 5). KEGG pathway enrichment analysis showed that DEGs were enriched in 17 pathways, and 'biosynthesis of other secondary metabolites' was the most significantly enriched pathway (Table 6). DEGs were also mapped and visualized with MapMan, and cellular response and receptor-like kinase pathway marked that 67 biotic stress related genes and immune response related genes were noticeably down-regulated in *hairs absent* mutant (Fig. 18, 19). From the DEGs, genes in the groups of transcription factors that function in the trichome initiation network of Arabidopsis were categorized (Table 8, Fig. 20). Among the identified 40 transcription factors, 3 genes were up-regulated and the rest of 37 genes were down-regulated in *h*.

10. *ZFP990*, the closest homolog of *H*, regulates trichome development in leaves.

Phylogenetic analysis on H homologs revealed that there is a great homology between ZFP970 and ZFP990. Since *ZFP970* has a function in trichome development, it was rational to assume that *ZFP990* might also participate in trichome development. To validate this assumption, functional studies on *ZFP990* were performed.

To verify the *ZFP990* loss-of-function phenotype, multiple transgenic lines in which *ZFP990* gene was knocked out by CRISPR/Cas9 system were generated. Transgenic lines were genotyped and homozygous or heterozygous biallelic mutants were identified. Most of these lines showed an apparent phenotype on leaves, on which long trichomes were absent (Fig. 21). However, knockout of *ZFP990* did not affect trichomes on stem.

To examine the subcellular localization of ZFP990, the proteins fused with YFP were transiently expressed in tobacco. pBCo-MYC2:CFP vector was co-transformed as a nuclear marker. ZFP990:YFP signal was colocalized with MYC2:CFP signal, indicating ZFP990 is a nuclear localized protein (Fig. 22). The expression pattern of *ZFP990* was analyzed from wild-type plants, leaves, epidermal stems, peeled-off stems, floral buds, and root tissues by using quantitative RT-PCR. The qRT-PCR result showed *ZFP990* is most highly expressed in leaf tissue (Fig. 23).

Table 1. List of primers.

Primer name	Primer sequence (5'→3')
Primers for *ZFP970* gene amplification	
BamHI-*ZFP970*_F	ACTGGATCC**ATG**GAGAAGATTGGAAGAGAAGC
XhoI-*ZFP970*_R	AGTCTCGAGCTATGATCC**TTA**TAAGTGCAAATCTAAACTC
XhoI-*ZFP970*-gw-R	ACGCTCGAGCGTAAGTGCAAATCTAAACTCACATGATT
Primers used to amplify full length of *H* gDNA from wild-type, *h*, LA1589, and LA0716	
*Sl-H*_F	TCACACACACACAAAATCAAAATTAAATCC
*Sl-H*_R	TACATTCAATCATTCTTCAAAAGCAAATC
*Sl-h*_F	ATGGCAATACAAATATTGTTTCTTATTACATGTC
*Sl-h*_R	TAGGATACCATTGACAGCCTTAAGG
LA1589-*h*_F	TCACACACACACAAAATCAAAATTAAATCC
LA1589-*h*_R	TGTATAACTACATTCAATCATTCTTCAAAAGC
LA0716-*h*_F	ACACACACACACAAAATCAAAATTAAATCC
LA0716-*h*_R	TACATTCAATCATTCTTCAAAAGCAAATC
Primers used to amplify *ZFP970*, *ZFP990* and *GAPDH* for RT-PCR	
ZFP970-RT_F	TGGTGATAGTACTAACATGTC
ZFP970-RT_R	TCCCTCTTCTCTTCATCATTAG
ZFP990-RT_F	AAGCAACTGTATCTACAAGAAC
ZFP990-RT_R	TGCTAATATTAACCATTGGAGG
GAPDH-RT_F	GACAAGGCTGCTGCTCACTT
GAPDH-RT_R	GCTTGACCTGCTGTCACCAAC
Gene specific primers	
ACT-F	GGGATGGAGAAGTTTGGTGGTGG
ACT-R	CTTCGACCAAGGGATGGTGTAGC
*Solyc10g078890*_F	ATTTGGACGTGCCCAATGTC
*Solyc10g078890*_R	TCTAATCTAATTTCTCCATTAGTCCCTG
*Solyc10g078900*_F	GGACGCAGAGAACAGAATCCT
*Solyc10g078900*_R	GGTGCACGTGAATCTCCACT
*Solyc10g078910*_F	GGAATTCAGGTGGAGCTTGTG
*Solyc10g078910*_R	ATGGACCCGACCCTCTACG
*Solyc10g078920*_F	ATTGGATGGCAGCTTGGTG
*Solyc10g078920*_R	TGTGCCCTCCAATCACCTC
*Solyc10g078930*_F	TGGATGAGAATAGGTGGAAGGG
*Solyc10g078930*_R	CAGCCCGTACCCTGTTGAAG
*Solyc10g078940*_F	ATGAGTTCGCGACAATGGGA
*Solyc10g078940*_R	TCGCGATGAACTACAGGTGG
*Solyc10g078950*_F	AACACACCAGCTCCCATTGG

*Solyc10g078950*_R	TCATCCTTTGCAGTGAGACCC
*Solyc10g078960*_F	GGGTTCATGGTTGAAGGTGC
*Solyc10g078960*_R	TTGCACAAGAAGACAATCGTATCA
*Solyc10g078970*_F	TCAGTCTTCAATAGTGCATGAGAC
*Solyc10g078970*_R	CAAAATGCTTCATTGTAATTAGGACTATAAAC
*Solyc10g078980*_F	CCCAGTATTCACTTCAGGCCC
*Solyc10g078980*_R	TACACAGATCATGAGTAATCGCCC
*Solyc10g078990*_F	TATTAGTCCTATAAATGGAAATCCATTGA
*Solyc10g078990*_R	TGTGAGTATAAACCAAAGGGAGAAGT
*Solyc10g079000*_F	AAGGAGTCAACCCGGGATCT
*Solyc10g079000*_R	GGAATAACCCAGCTCCAGCA
*Solyc10g079010*_F	CCAATGTGTTGGTTTCTCCC
*Solyc10g079010*_R	CCGTCCGATTCGATTCAGA
gRNA oligos for knock-out experiment	
*ZFP990*_gRNA_F	GATTGAGGGTAATAGGAAGTTGAG
*ZFP990*_gRNA_R	AAACCTCAACTTCCTATTACCCTC
Primers used for genotyping of *ZFP990* knock-out plants	
*ZFP990*_5’_F	TTTAGGCTCCGGACTCGTG
*ZFP990*_q3_R	TTCTTTAACAATATTTGGTTTGTCATAA

Start and stop codons are in bold. Restriction enzyme sites are underlined. F, forward primer; R, reverse primer.

Figure 1. DMS images of trichomes on the leaves, stems, hypocotyls, and sepals of wild-type and *h* mutant. All photos were taken from four-week-old plants. The arrows indicate type I trichomes. Scale bars = 2mm.

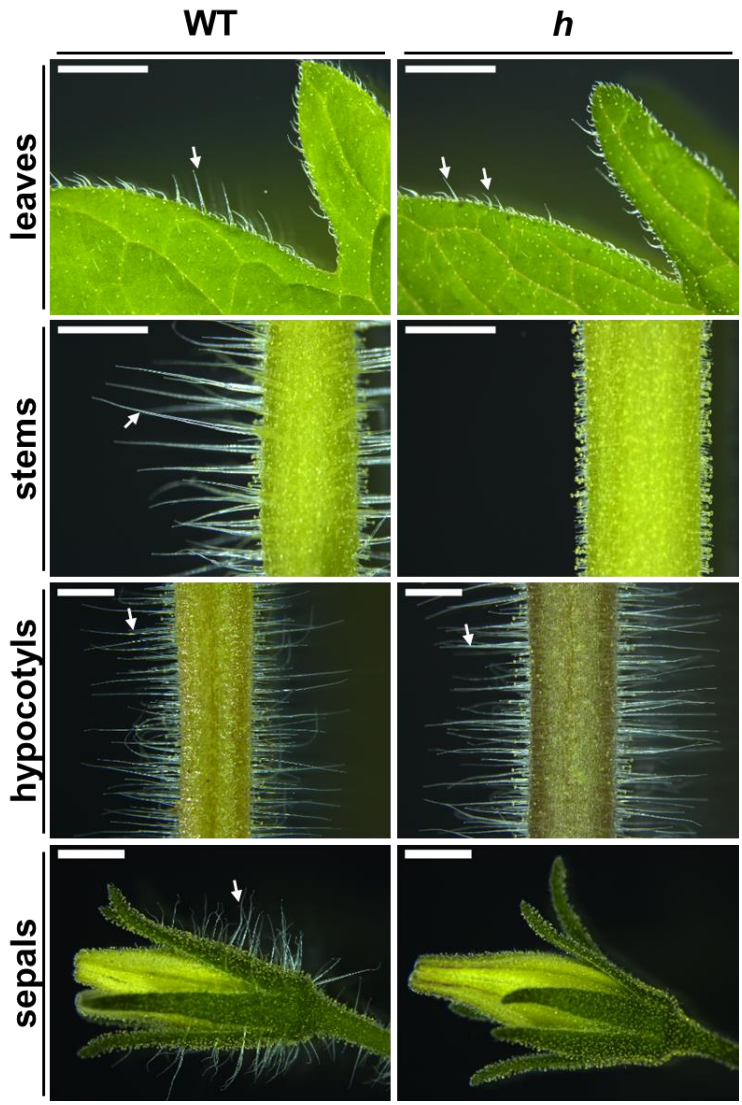

WT
h
leaves
stems
hypocotyls
sepals

Figure 2. SEM images of trichomes on leaves, stems, hypocotyls, and sepals of wild-type and *h* mutant. Four-week-old plants were used for all images. The arrows indicate type I trichomes. Scale bars = 200μm.

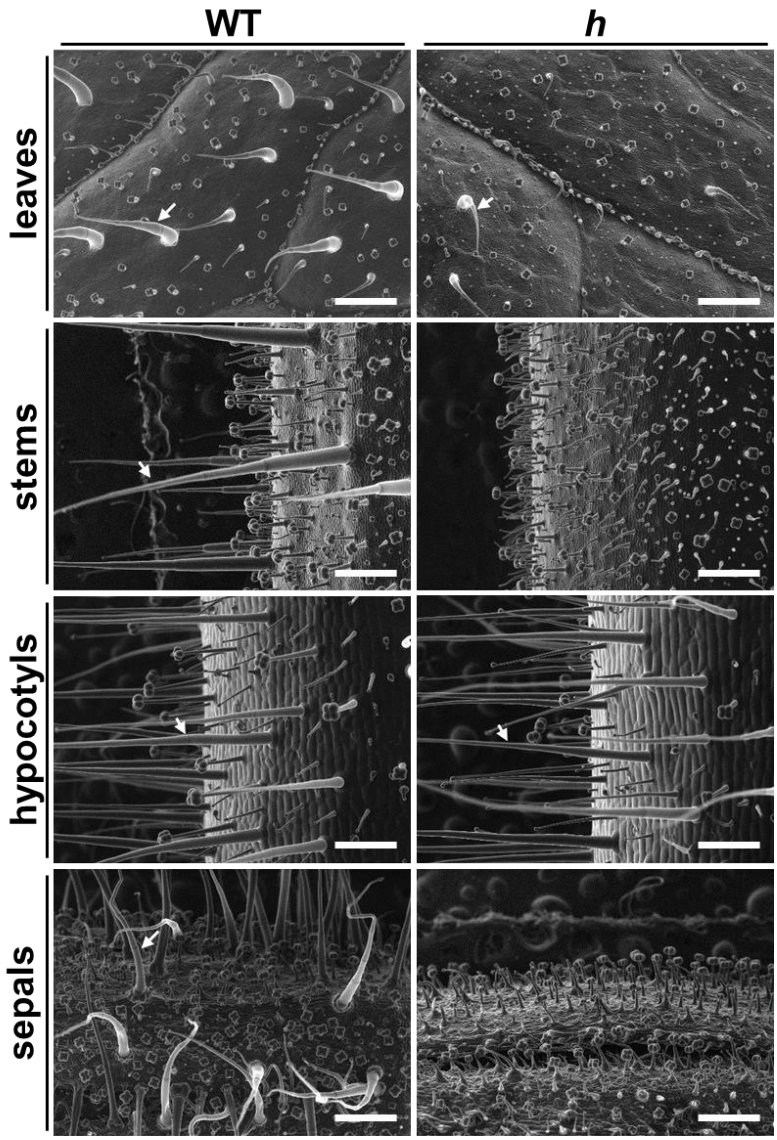

WT
h
leaves
stems
hypocotyls
sepals

Figure 3. **Map-based cloning of *H* gene.** Numbers enclosed by parentheses are the number of recombinants between the markers. *H* was located on chromosome 10 between SGN-U578758 and SGN-U573045 markers spanning 109 kb region. 14 predicted genes were annotated in the region according to the *Solanum lycopersicum* ITAGv3.1

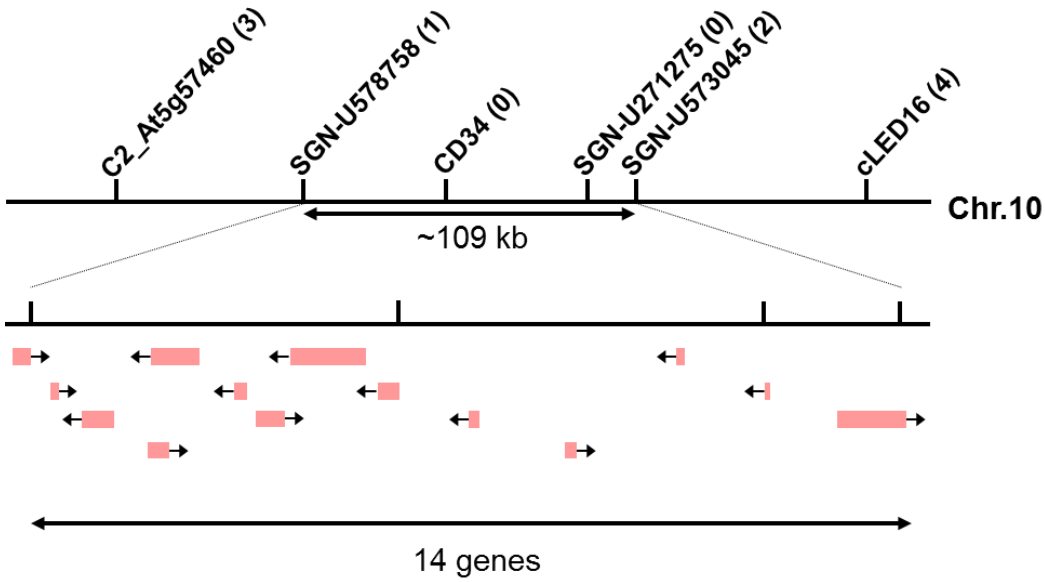

C2_At5g57460 (3)
SGN-U578758 (1)
CD34 (0)
SGN-U271275 (0)
SGN-U573045 (2)
cLED16 (4)
Chr.10
~109 kb
14 genes

Figure 4. RT-PCR analysis of candidate genes. The expression of each gene in wild-type and *h* plants was confirmed by RT-PCR with gene specific primers. All the genes except *Solyc10g078970* were expressed in both wild-type and *h* plants. *Solyc10g078970* was not amplified from *h* mutant at all. *890*: *Solyc10g078890*, *900*: *Solyc10g078900*, *910*: *Solyc10g078910*, *920*: *Solyc10g078920*, *930*: *Solyc10g078930*, *935*: *Solyc10g078935*, *940*: *Solyc10g078940*, *950*: *Solyc10g078950*, *960*: *Solyc10g078960*, *970*: *Solyc10g078970*, *980*: *Solyc10g078980*, *990*: *Solyc10g078990*, *000*: *Solyc10g079000*, *010*: *Solyc10g079010*, *ACT*: *Actin7 (Solyc03g078400)*

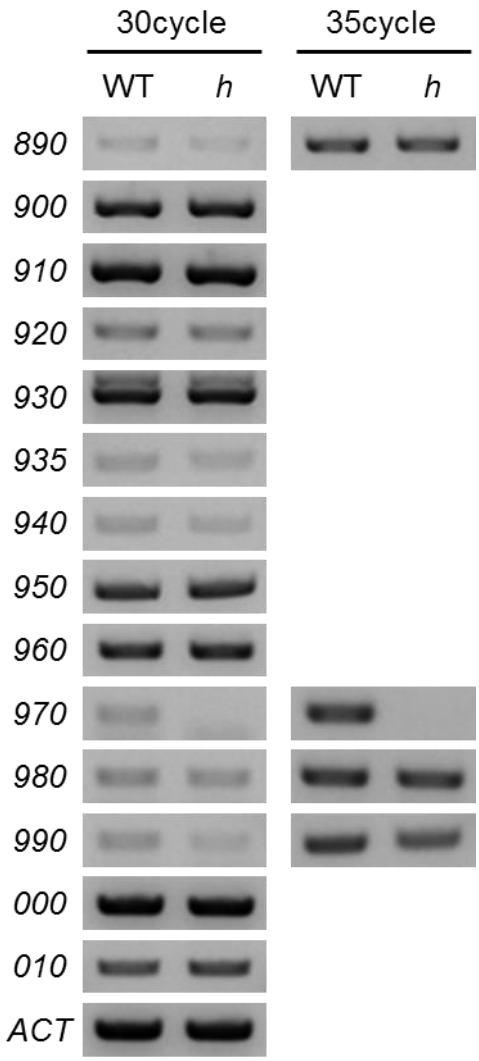
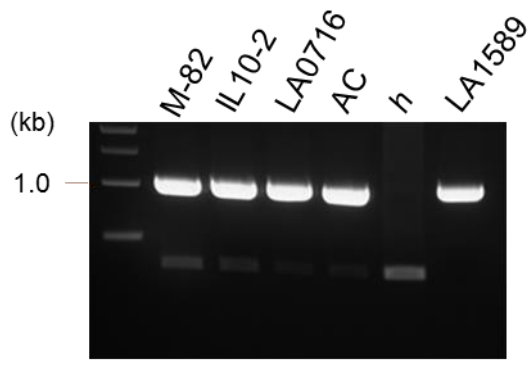

30cycle
35cycle
WT
h
WT
h
890
900
910
920
930
935
940
950
960
970
980
990
000
010
ACT

Figure 5. gDNA PCR for full length of *ZFP970*. (A) PCR amplification of full length of *ZFP970* gDNA in wild-type (M-82 and AC) and loss-of-function plants (IL10-2, LA716, *h*, and LA1589). (B) PCR amplification in AC and *h* with primers from 5' and 3' non-coding region of *ZFP970*.

A



B

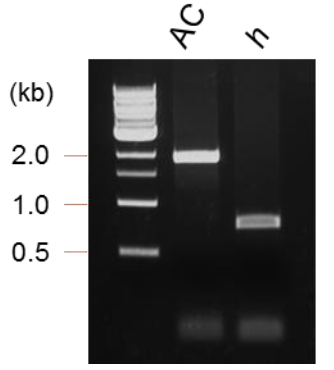


Figure 6. gDNA structure of *ZFP970* in wild-type and *h* mutant. 1.4 kb region including the whole gene was deleted from *h*, and 309 bp fragment was inserted at 3' non-coding region.

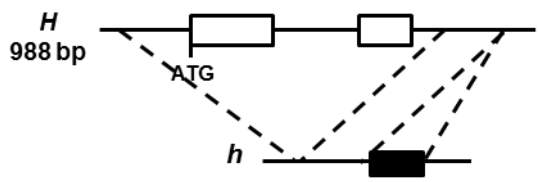

H
988 bp
ATG
h

Figure 7. DMS images of cultivated tomato M-82 (wild-type) and other *h*-like wild species. Introgression line (IL) 10-2 and wild tomato species LA0716 (*S.pennellii*) and LA1589 (S. *pimpinellifolium*) do not have type I trichomes on stem and sepal. Scale bars = 1mm (stem), 2mm (sepal)

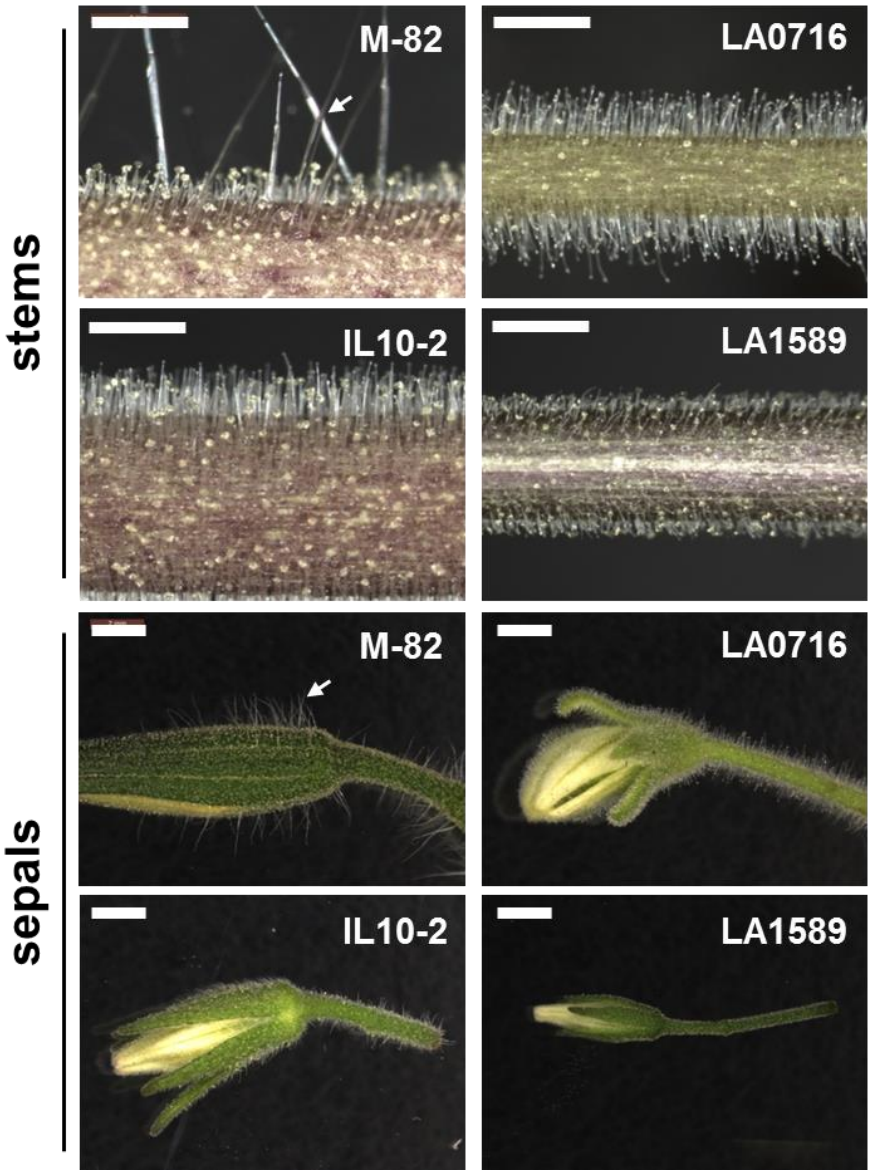
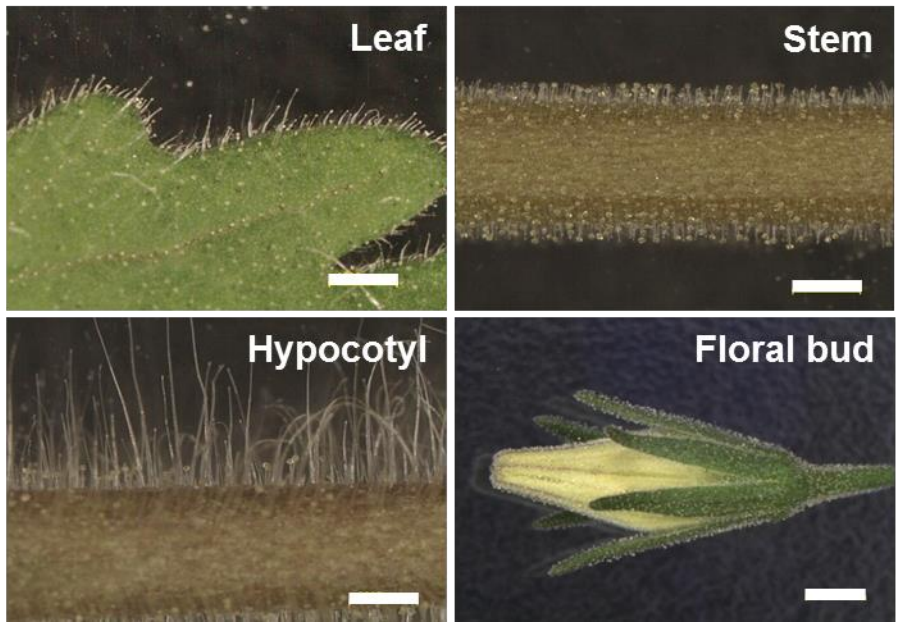

stems
sepals
M-82
LA0716
IL10-2
LA1589
M-82
LA0716
IL10-2
LA1589

Figure 8. Allelism test of two *h*-like tomato, IL 10-2 and LA1589, with *h*. F_1 plants obtained by crossing IL10-2 with *h* (A), and LA1589 with *h* (B) did not complement the type I trichome-less phenotype. Scale bars on leaf, stem, hypocotyl = 1mm, floral bud = 2mm.

A


Leaf
Stem
Hypocotyl
Floral bud

B

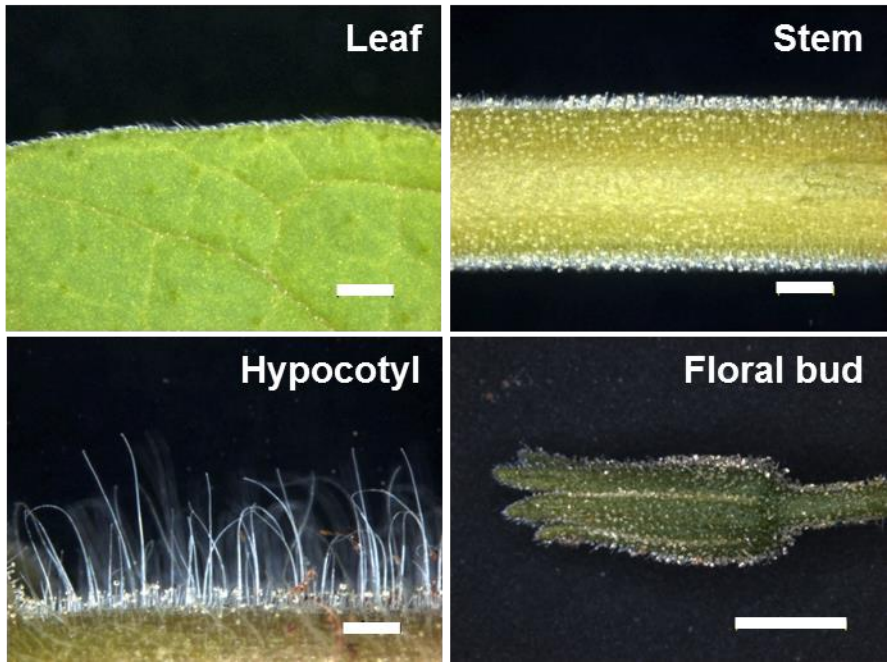

Leaf
Stem
Hypocotyl
Floral bud

Figure 9. RT-PCR analysis of *ZFP970* and *ZFP990* expression in M-82 (wild-type) and other *h*-like wild species. *ZFP970* was expressed only from the wild-type tomatoes, M-82 and Ailsa Craig (AC). IL 10-2, LA0716, *h*, and LA1589 tomatoes, which do not have type I trichomes, did not expressed *ZFP970*, whereas *ZFP990* was expressed in all tomatoes.

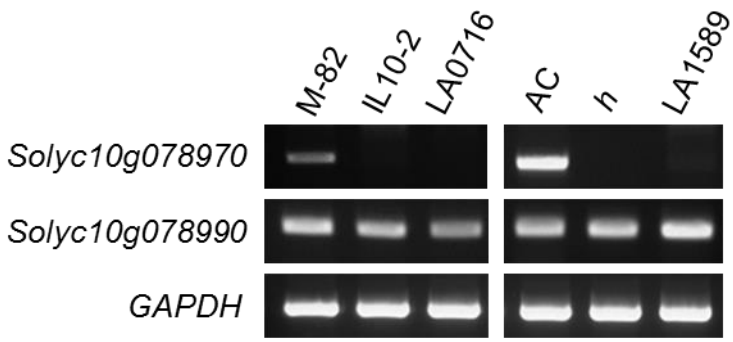

M-82
IL10-2
LA0716
AC
h
LA1589
Solyc10g078970
Solyc10g078990
GAPDH

Figure 10. *ZFP970* gene comparison between wild-type (AC) and *S. pennellii.* (A) Deduced amino acid sequences from wild-type and *S.pennellii* were aligned. (B) Promoter sequences (~500 bp length) from wild-type and *S.pennellii* were compared. Predicted cis-elements are marked in yellow. Start codons (ATG) are colored in red.

A

```
AC          MEKIGREAVDYMNMKSFSQP-LRKKSIRLFGKEFSVGDSTNMSESTDKNPLHHEPKPNTM 59
IL10-2      MEKIGREAVDYMNVKSFSQPPLRKKSIRLFGKEFSVGDSTNMSESTDKNPLHHEPKPNTM 60
            *************:****** ****************************************

AC          SISANRIDKTGHVDEISRKYECYYCFRSFPTSQALGGHQNAHKKERQNAKLSHLQSSIVH 119
IL10-2      SISANRIDKRGHVDEISRKYECYYCFRSFPTSQALGGHQNAHKKERQNAKLSHLQSSIVH 120
            ********* **************************************************

AC          ETNRNRFGEPSTAATRLTHYHSTWSNINNNNVYSPNYNEAFWQIPPTIHHYQNNINPPSS 179
IL10-2      ETNRHRFGEPSAAATRLTHHHSTWSNIKNNNVYSPNYNEAFWQIPPTIHHYQNNINPSSS 180
            ****:******:*******:*******:*****************************.**

AC          FSHDSFFPNDEEKREVQNHVSLDLHL 205
IL10-2      FSHDSFFPNDEEKREFQNHVSLDLHL 206
            ***************.**********
```

B

```
AC          TACCCCACAGTCTGGCACCGAGATCGATTGTCAAGTTTTATATCATATAAATTTTAAAAG 60
LA0716      TACCCCACAGTCTGGCACCAAGATCGATTGTCATGTTTTATATCATATAAATTTTAAAAG 60
            ****************** ************* **************************

                                                          CACTFTPPCA1
AC          TTTTATGGCAATACAAATATTGTTTCTTATTACATGTCACAAATTTTATCGTTACTTATA 120
LA0716      TTTTATAGCAATACAAATATTGTTACTTATTACATG----------------------- 96
            ****** ***************** ***********

                                  CPBCSPOR
AC          TTACATGTCACAAATTTCAAAAGTTTTATATTAGTACAAATATTACTATAACATGTTATA 180
LA0716      -------TCACAAATTTCAAAAGTTCTATATCAGTATAAATATTACTATAACATGTTATA 149
                   ****************** ***** **** ************************

AC          TATCTAAATTT-CAAAAAAAAAATTCTTTTTTTGTTGAAATTTTATATTGAGTTAAATTA 239
LA0716      TATCTAAATTTTTTTAAAAAAAATTCTTTTTCGACTTAAATATTATACTGAGTCAAATTA 209
            ***********    ****************    * **** ***** ***** ******

                           MYBGAHV
AC          TGTCGGATAATTAATTAATCTTAACAAAAGTAGTAAGTTACTAGCTAATTCGATATTGTA 299
LA0716      TGTTAGATAATTAATTAATCGCAACAAAAGTAGTAAGTTACTAGCTAATTCGATATTGTA 269
            ***  ***************  ************************************

                                  TATABOX4
AC          TTTTAAGTGGTTCACAAATAGTTGTTTTATATAAGTAATCTGATGTTTTAAATTGTTAAC 359
LA0716      TTTTAAGTGGTTCACAAATAGTTGTTTTATATATGTGTTCTGATGTTTTAAATTGTTAAC 329
            ********************************** **  **********************

AC          TTATACTAAAGATCATTAAATAGACCACTTCCTATATTGGT-CCATTAGACCTTCCCAAA 418
LA0716      TTATACTAAAGATCATTAAATAGACCACTTCCTATATTGGTCCCATTAGACCTTCCCAAA 389
            ****************************************** ******************

AC          ACCCTAATATAAACAAAAATATTCTTCAATACCTCCAAATACCACTTATAACTTTAGAAA 478
LA0716      ACCCTAATATAAACAAAAATATTCTTCAATTCCTCCAAATACCACTTATAACTTTAGAAA 449
            ****************************** ******************************

               CACTFTPPCA1
AC          ACACAACACACTCACACACACACAAAATCAAAATTAAATCCATG 522
LA0716      ACACAACACAC--ACACACACACAAAATCAAAATTAAATCCATG 491
            ***********  ********************************
```

Figure 11. Alignment of ZFP970 amino acids from wild-type (AC) and *S. pimpinellifolium.* Wild-type amino acid sequence of ZFP970 was compared with deduced amino acid sequence of ZFP970 from *S. pimpinellifolium*, and premature stop codon was found at position 117 in *S. pimpinellifolium.*

```
AC          MEKIGREAVDYMNMKSFSQPLRKKSIRLFGKEFSVGDSTNMSESTDKNPLHHEPKPNTMS 60
LA1589      MEKIGREAVDYMNMKSFSQPLRKKSIRLFGKEFSVGDSTNMSESTDKNPLHHEPKPNTMS 60
            ************************************************************

AC          ISANRIDKTGHVDEISRKYECYYCFRSFPTSQALGGHQNAHKKERQNAKLSHLQSSIVHE 120
LA1589      ISANRIDKKGHVDEISRKYECYYCFRSFPTSQALGAIKMHTRKKDKMPNYLIFSLQ-----116
            ********.**************************. :   :*: :

AC          TNRNRFGEPSTAATRLTHYHSTWSNINNNNVYSPNYNEAFWQIPPTIHHYQNNINPPSSF 180
LA1589      ------------------------------------------------------------


AC          SHDSFFPNDEEKREVQNHVSLDLHL 205
LA1589      -------------------------
```

Figure 12. Subcellular localization of ZFP970. *A. tumefaciens* strain LBA4404 were transformed with each of pBCo-*MYC2*-CFP and pBCo-*ZFP970*-YFP vectors, and infiltrated into tobacco leaves. 2 days after infiltration, CFP fused MYC2 (a), and YFP fused ZFP970 (b) were visualized under the confocal microscope. (c) Overlay, and (d) transmitted light image. Scale bars, 20μm.

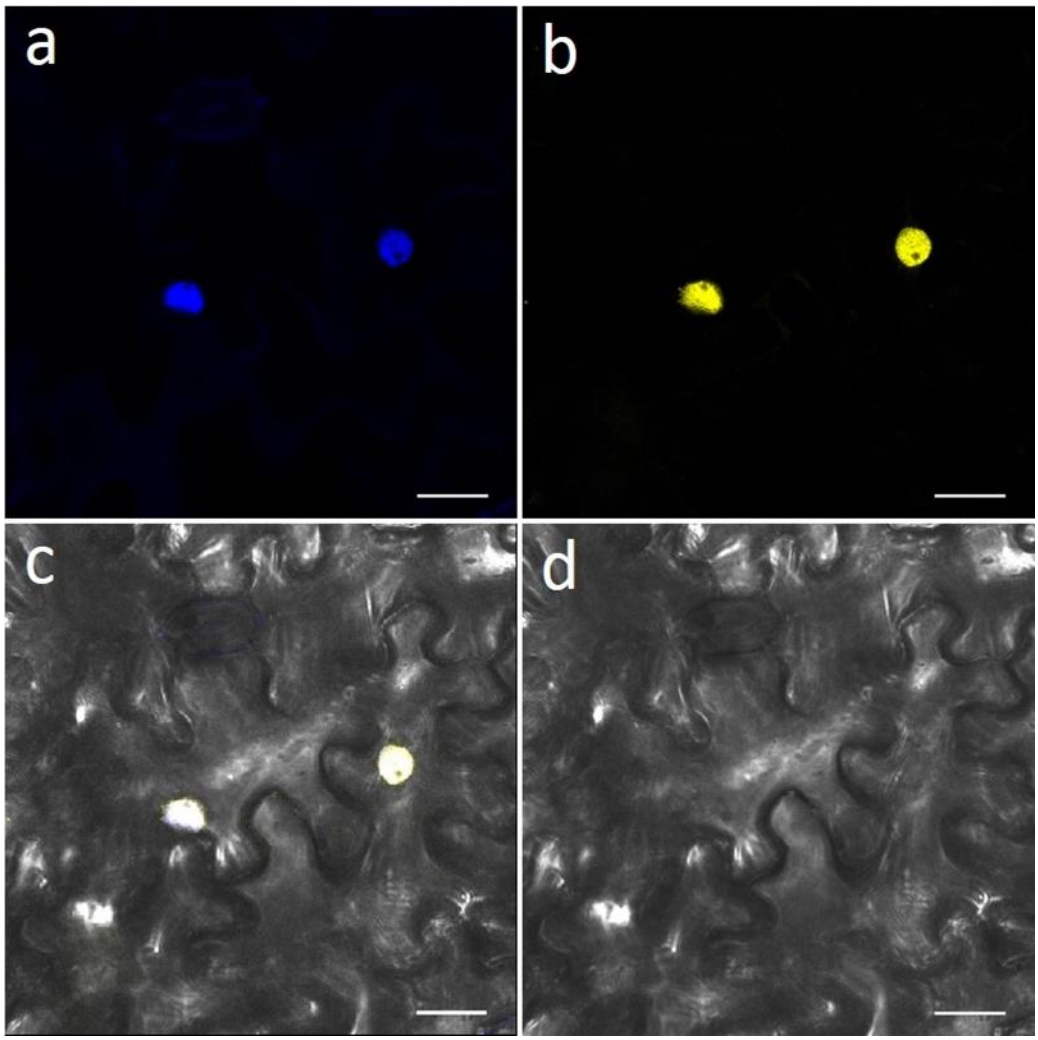

a
b
c
d

Figure 13. Quantitative RT-PCR analysis of *ZFP970* expression levels in different tissues of wild-type plants. Data is means of three independent biological replicates and *SlACT7 (Solyc03g078400)* was used to normalize the expression levels. Error bars indicate standard error of three biological replicates. The expression level of *ZFP970* in leaf tissue has been calibrated to 1.

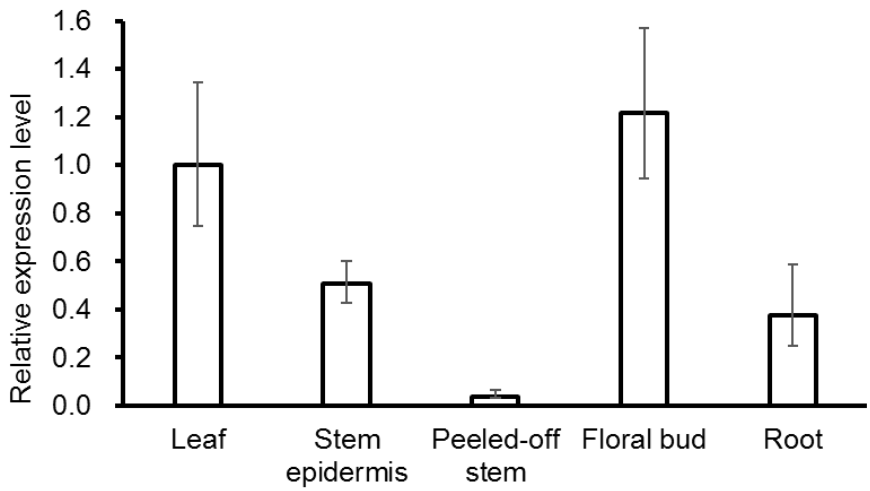

Relative expression level
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Leaf
Stem epidermis
Peeled-off stem
Floral bud
Root

Figure 14. Phylogenetic tree of H protein homologs in tomato and Arabidopsis. The unrooted phylogenetic tree was constructed with MEGA7 using the Neighbor-joining method based on the amino acid sequences of homologs aligned by ClustalW. Bootstrap values in 1000 replicates are shown next to the branches. H is highlighted. Bar = 0.1 amino acid substitution per site.

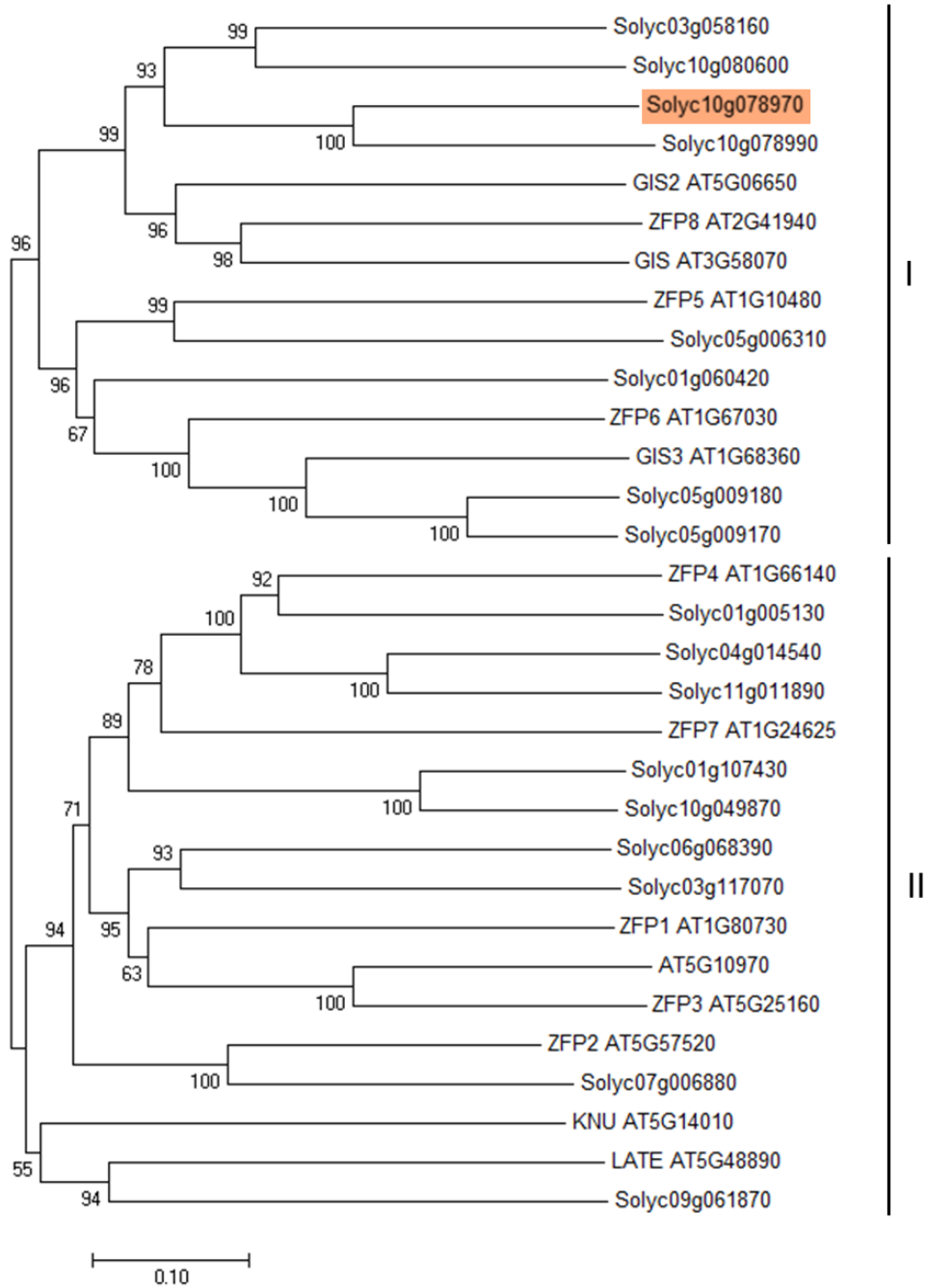

Solyc03g058160
Solyc10g080600
Solyc10g078970
Solyc10g078990
GIS2 AT5G06650
ZFP8 AT2G41940
GIS AT3G58070
ZFP5 AT1G10480
Solyc05g006310
Solyc01g060420
ZFP6 AT1G67030
GIS3 AT1G68360
Solyc05g009180
Solyc05g009170
ZFP4 AT1G66140
Solyc01g005130
Solyc04g014540
Solyc11g011890
ZFP7 AT1G24625
Solyc01g107430
Solyc10g049870
Solyc06g068390
Solyc03g117070
ZFP1 AT1G80730
AT5G10970
ZFP3 AT5G25160
ZFP2 AT5G57520
Solyc07g006880
KNU AT5G14010
LATE AT5G48890
Solyc09g061870
I
II
0.10

Figure 15. Amino acid comparison of ZFP970 homologs in tomato and Arabidopsis. Identical amino acids are shaded in black, and conserved (90%) amino acids are shaded in grey. Conserved motifs are underlined with colors: RLFGV-like motif (blue), C2H2 zinc finger motif (red), NLS (nuclear localization signal) (grey), EAR motif (green).

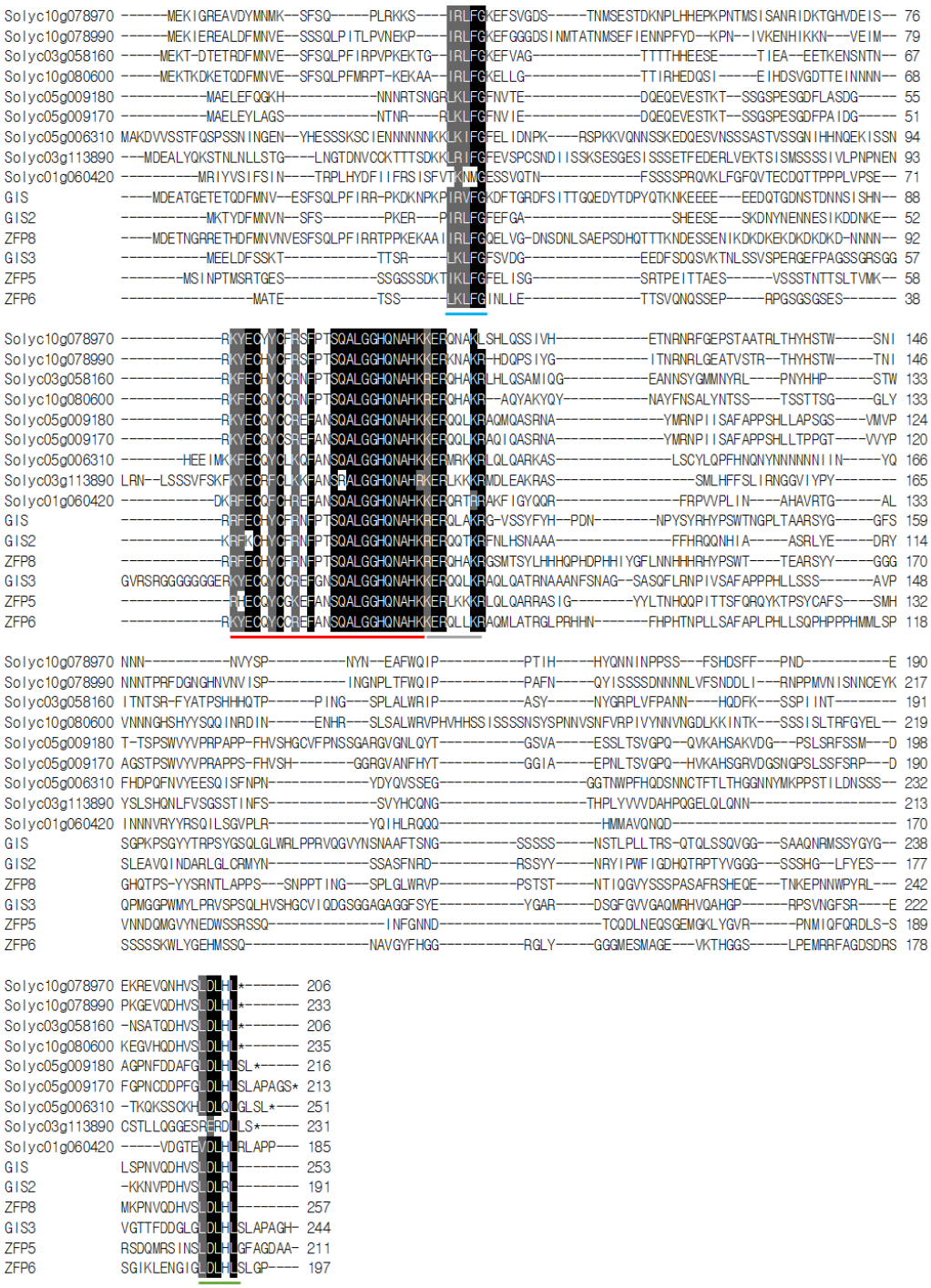
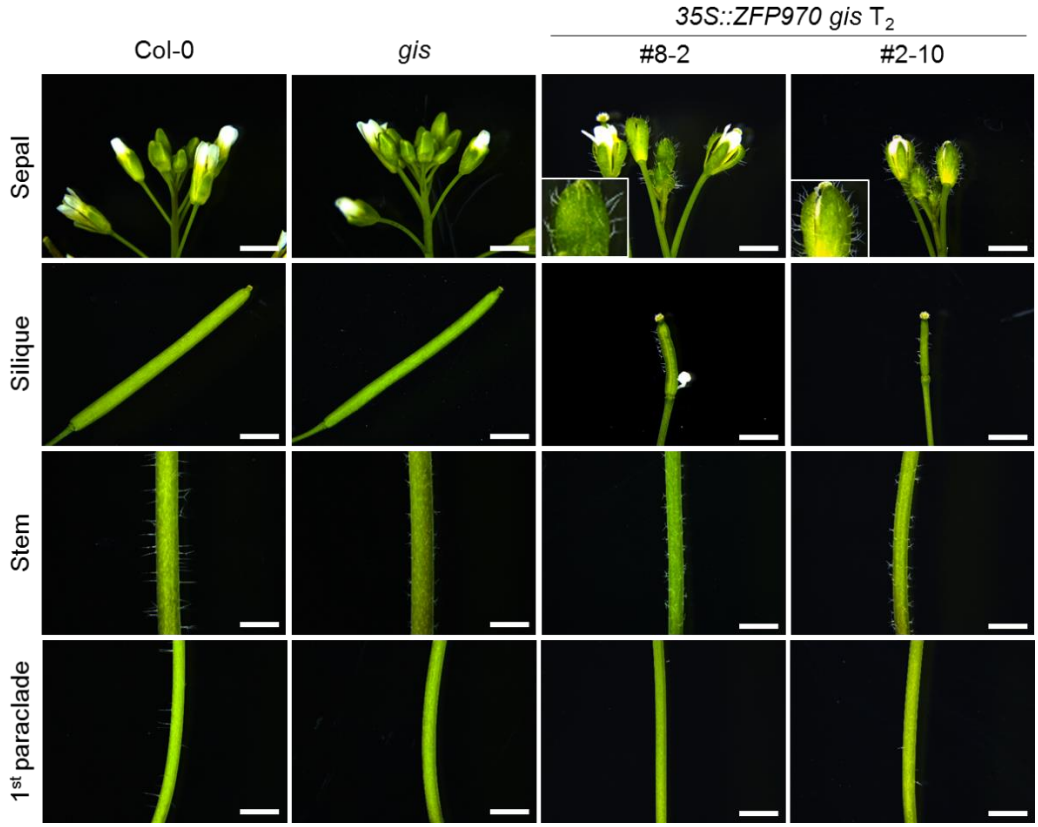


Figure 16. Phenotype of wild-type, *gis* mutant, and *35S::ZFP970* transgenic lines. (A) Transgenic lines showed high densities of trichomes on sepal (first row). Ectopic development of trichomes on siliques were found, and most of siliques were sterile (second row). *ZFP970* overexpression in *gis* background did not recovered the wild-type phenotype on stem (third row), and no significant differences were noticed on the first paraclade (fourth row). Scale bars = 2 mm. (B) Relative expression of wild-type, *gis* mutant, *35S*::*ZFP970* overexpressing lines were analyzed by using qRT-PCR. All values are normalized using a common internal standard (*AtACT2*). #8-2, #2-10 represent *35S::ZFP970* transgenic lines.

A


35S::ZFP970 gis T2
Col-0
gis
#8-2
#2-10
Sepal
Silique
Stem
1st paraclade

B

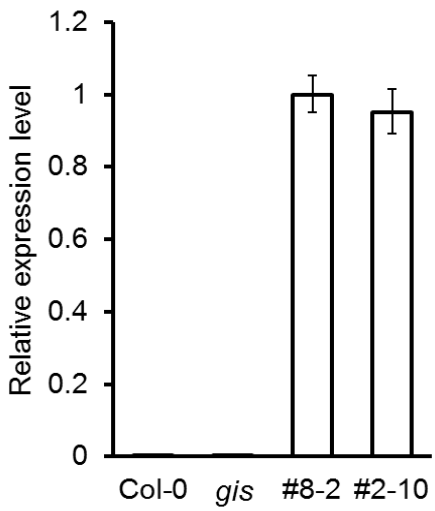

Relative expression level
1.2
1
0.8
0.6
0.4
0.2
0
Col-0
gis
#8-2
#2-10

Table 2. DynamicTrim and LengthSort data. Sequence data in which quality of base pair is upper than 20 were extracted by SolexaQA, and read length lower than 25 bp after trimming were excluded.

Index	AC-ET			*h*-ET		
	sample 1	sample 2	sample 3	sample 1	sample 2	sample 3
No. of trimmed reads	32,087,262	31,860,045	25,535,629	33,907,413	29,099,837	31,599,252
No. of mapped reads	26,728,314	26,521,773	21,508,760	28,688,370	24,616,016	26,380,831

Table 3. Statistics of annotation. Among the 34,727 reference genes of tomato, the expression of 30,657 genes were detected from the RNA-seq analysis, and 26,649 genes had functional description.

	Number of genes used in analysis	Number of genes expressed in a transcriptome	Number of genes with functional description
Tomato reference gene	34,727	30,657	26,649

Table 4. **Number of identified DEGs.**

	Regulation pattern	Num. of DEGs	Annotated DEGs
AC_ET vs *h*_ET	Up regulation	253	204
	Down regulation	1317	1117
	Total	1570	1321

Figure 17. Number of identified DEGs.

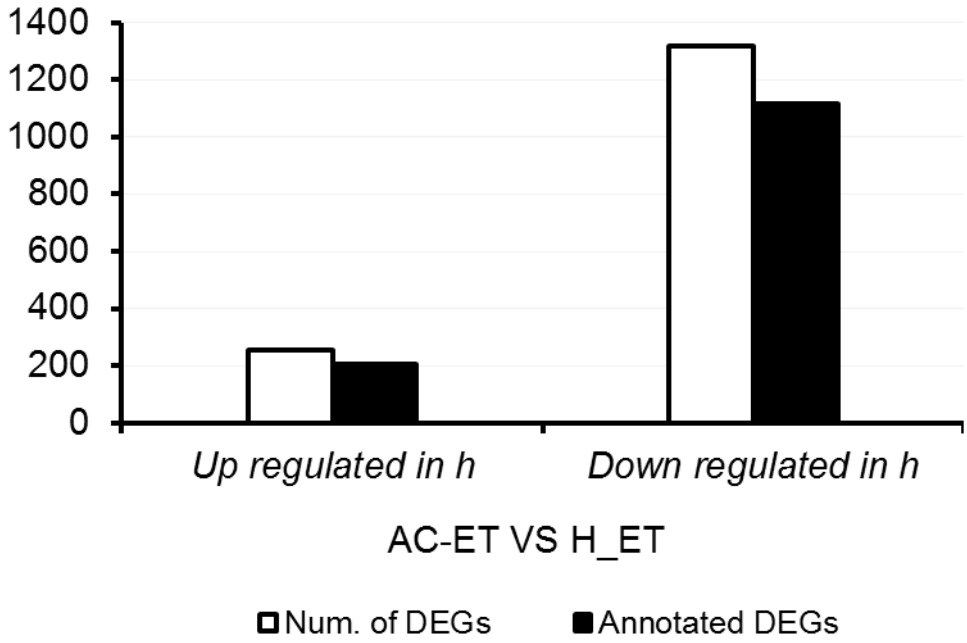

1400
1200
1000
800
600
400
200
0
Up regulated in h
Down regulated in h
AC-ET VS H_ET
Num. of DEGs
Annotated DEGs

Table 5. GO enrichment analysis.

Category	GO ID	GO Term	Gene count
BP	GO:0001666	response to hypoxia	1
BP	GO:0005975	carbohydrate metabolic process	26
BP	GO:0006281	DNA repair	2
BP	GO:0006486	protein glycosylation	1
BP	GO:0006520	cellular amino acid metabolic process	8
BP	GO:0006629	lipid metabolic process	10
BP	GO:0006810	transport	75
BP	GO:0006952	defense response	28
BP	GO:0006970	response to osmotic stress	21
BP	GO:0006979	response to oxidative stress	8
BP	GO:0007017	microtubule-based process	1
BP	GO:0007275	multicellular organismal development	50
BP	GO:0008219	cell death	9
BP	GO:0009408	response to heat	1
BP	GO:0009409	response to cold	1
BP	GO:0009414	response to water deprivation	1
BP	GO:0009595	detection of biotic stimulus	1
BP	GO:0009617	response to bacterium	24
BP	GO:0009620	response to fungus	25
BP	GO:0009624	response to nematode	2
BP	GO:0009698	phenylpropanoid metabolic process	13
BP	GO:0009790	embryo development	2
BP	GO:0009793	embryo development ending in seed dormancy	3
BP	GO:0009826	unidimensional cell growth	2
BP	GO:0009845	seed germination	1
BP	GO:0010073	meristem maintenance	4
BP	GO:0010162	seed dormancy process	1
BP	GO:0010228	vegetative to reproductive phase transition of meristem	1
BP	GO:0010431	seed maturation	1
BP	GO:0015979	photosynthesis	1
BP	GO:0016131	brassinosteroid metabolic process	1
BP	GO:0016458	gene silencing	2
BP	GO:0016570	histone modification	1
BP	GO:0023014	signal transduction by protein phosphorylation	7
BP	GO:0040008	regulation of growth	4
BP	GO:0042546	cell wall biogenesis	7
BP	GO:0045087	innate immune response	29
BP	GO:0048438	floral whorl development	2
BP	GO:0048444	floral organ morphogenesis	2
BP	GO:0048513	organ development	20
BP	GO:0048608	reproductive structure development	12
BP	GO:0048868	pollen tube development	2
BP	GO:0051301	cell division	2
BP	GO:0051607	defense response to virus	2
BP	GO:0055085	transmembrane transport	16
BP	GO:0055114	oxidation-reduction process	6
BP	GO:0071555	cell wall organization	5
CC	GO:0000151	ubiquitin ligase complex	3
CC	GO:0005618	cell wall	2
CC	GO:0005737	cytoplasm	68
CC	GO:0031225	anchored component of membrane	3
MF	GO:0000166	nucleotide binding	24
MF	GO:0003676	nucleic acid binding	44
MF	GO:0004497	monooxygenase activity	5
MF	GO:0008233	peptidase activity	2
MF	GO:0015144	carbohydrate transmembrane transporter activity	2
MF	GO:0015238	drug transmembrane transporter activity	1

MF	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	8
MF	GO:0016746	transferase activity, transferring acyl groups	4
MF	GO:0016757	transferase activity, transferring glycosyl groups	10
MF	GO:0016788	hydrolase activity, acting on ester bonds	8
MF	GO:0016798	hydrolase activity, acting on glycosyl bonds	11
MF	GO:0042802	identical protein binding	1
MF	GO:0046983	protein dimerization activity	3

Table 6. KEGG pathways found in DEGs.

Major Classification	Sub Classification	Gene count
Cellular Processes	Transport and catabolism	7
Environmental Information	Membrane transport	5
Environmental Information	Signal transduction	18
Genetic Information Processing	Folding, sorting and degradation	16
Genetic Information Processing	Replication and repair	2
Genetic Information Processing	Translation	18
Metabolism	Amino acid metabolism	20
Metabolism	Biosynthesis of other secondary metabolites	31
Metabolism	Carbohydrate metabolism	21
Metabolism	Energy metabolism	16
Metabolism	Glycan biosynthesis and metabolism	1
Metabolism	Lipid metabolism	22
Metabolism	Metabolism of cofactors and vitamins	6
Metabolism	Metabolism of other amino acids	22
Metabolism	Metabolism of terpenoids and polyketides	12
Metabolism	Nucleotide metabolism	4
Organismal systems	Environmental adaptation	19

Table 7. List of annotated Arabidopsis homologs in RNA-seq data. Genes annotated as homologs of Arabidopsis genes that are reported to be involved in trichome development are listed in the table. Genes highlighted in yellow represent zinc finger proteins homologous to ZFP970. White-to-blue and white-to-red colors indicate down-regulation and up-regulation in *h*, respectively.

Reference transcripts	Log2FC	Annotation	
		TAIR gene ID	TAIRsymbol
Solyc10g078990	-1.67	AT2G41940	ZFP8
Solyc10g078970	-1.98	AT2G41940	ZFP8
Solyc01g060420	-1.06	AT1G67030	ZFP6
Solyc03g113890	0.20	AT1G67030	ZFP6
Solyc10g080600	0.36	AT3G58070	GIS
Solyc03g058160	0.19	AT2G41940	ZFP8
Solyc05g009170	-0.29	AT1G68360	GIS3
Solyc05g006310	0	AT1G68360	GIS3
Solyc05g009180	0.34	AT1G68360	GIS3
Solyc07g052490	0.49	AT2G46410	CPC
Solyc12g005800	-0.04	AT1G01380	ETC1
Solyc01g095640	-0.90	AT5G53200	TRY
Solyc08g081140	0.30	AT5G41315	GL3,MYC6.2
Solyc03g081210	-0.03	AT5G24520	ATTTG1,TTG,TTG1,URM23
Solyc03g081200	0.05	AT5G24520	ATTTG1,TTG,TTG1,URM23
Solyc03g097340	-0.09	AT5G24520	ATTTG1,TTG,TTG1,URM23
Solyc08g062530	-0.02	AT1G79840	GL2
Solyc03g120620	0.17	AT1G79840	GL2
Solyc05g015020	0.22	AT3G61150	HD-GL2-1,HDG1
Solyc09g010180	0.02	AT3G11540	SPY

Table 8. List of four groups of transcription factors in DEGs. *ZFP970* and *ZPF990* are highlighted in yellow. The red-to-blue label in Log2FC (Fold Change) column represents up-regulation and down-regulation in *h*, respectively.

Symbol	Reference transcript	Log2FC	Annotation TAIR gene ID
bHLH	Solyc07g039570	-3.29	AT5G56960
	Solyc07g005400	-1.95	AT2G24260
	Solyc04g007300	-1.40	AT1G25330
	Solyc05g014590	-2.24	AT5G57150
	Solyc03g006910	1.08	AT4G25410
	Solyc10g049720	1.02	AT2G14760
	Solyc06g069600	-1.49	AT4G00050
C2H2	Solyc11g073060	-3.20	AT2G37430
	Solyc05g054650	-2.01	AT2G37430
	Solyc11g066420	-1.27	AT5G22890
	Solyc10g078990	-1.67	AT2G41940
	Solyc01g060420	-1.06	AT1G67030
	Solyc06g075780	-1.52	AT5G59820
	Solyc04g077980	-1.11	AT1G27730
	Solyc05g009770	-1.04	AT1G13290
	Solyc10g078970	-1.98	AT2G41940
	Solyc12g088390	-2.89	AT1G27730
	Solyc01g081630	-1.17	AT1G05577
	Solyc08g078590	-1.39	AT1G02030
HD-ZIP	Solyc07g062790	-1.55	AT5G53980
	Solyc02g085630	-1.25	AT4G36740
	Solyc02g062960	-1.08	AT2G18550
	Solyc06g050160	-1.03	AT4G00730
MYB	Solyc04g064540	-1.54	AT1G17950
	Solyc03g093890	-4.90	AT3G23250
	Solyc04g077260	-1.25	AT5G65790
	Solyc05g055030	-1.17	AT3G23250
	Solyc10g081320	-1.58	AT1G69560
	Solyc07g054980	-1.73	AT3G23250
	Solyc04g079360	-1.11	AT3G50060
	Solyc09g090130	-1.35	AT3G23250
	Solyc10g086250	1.02	AT1G66370
	Solyc07g053240	-2.38	AT3G23250
	Solyc07g055000	-2.40	AT5G16770
	Solyc03g093930	-3.68	AT3G23250
	Solyc03g093940	-1.47	AT4G21440
	Solyc07g053230	-2.51	AT3G23250
	Solyc03g005570	-1.41	AT3G23250
	Solyc10g044680	-1.01	AT1G09540
	Solyc03g119370	-1.72	AT1G68320

Figure 18. Alignment of DEGs on cellular response pathway by using MapMan. DEGs (log2 fold change of normalized expression values) involved in cellular response pathway is visualized with up-regulated genes in *h* as red color and down-regulated in *h* labeled in blue.

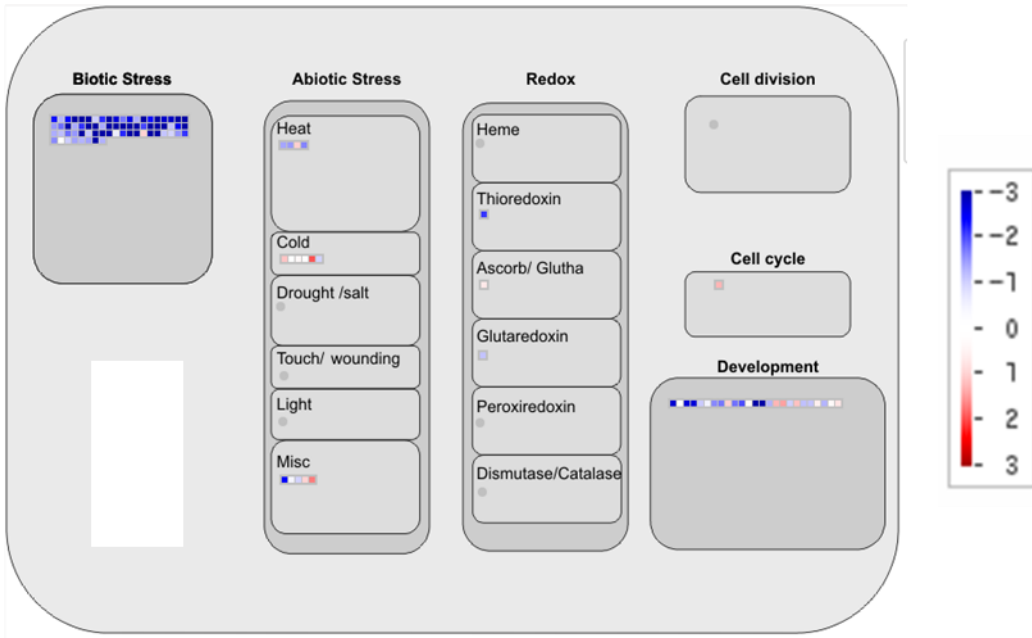

Biotic Stress
Abiotic Stress
Heat
Cold
Drought /salt
Touch/ wounding
Light
Misc
Redox
Heme
Thioredoxin
Ascorb/ Glutha
Glutaredoxin
Peroxiredoxin
Dismutase/Catalase
Cell division
Cell cycle
Development
-3
-2
-1
0
1
2
3

Figure 19. Alignment of DEGs in receptor-like kinase pathway with MapMan.

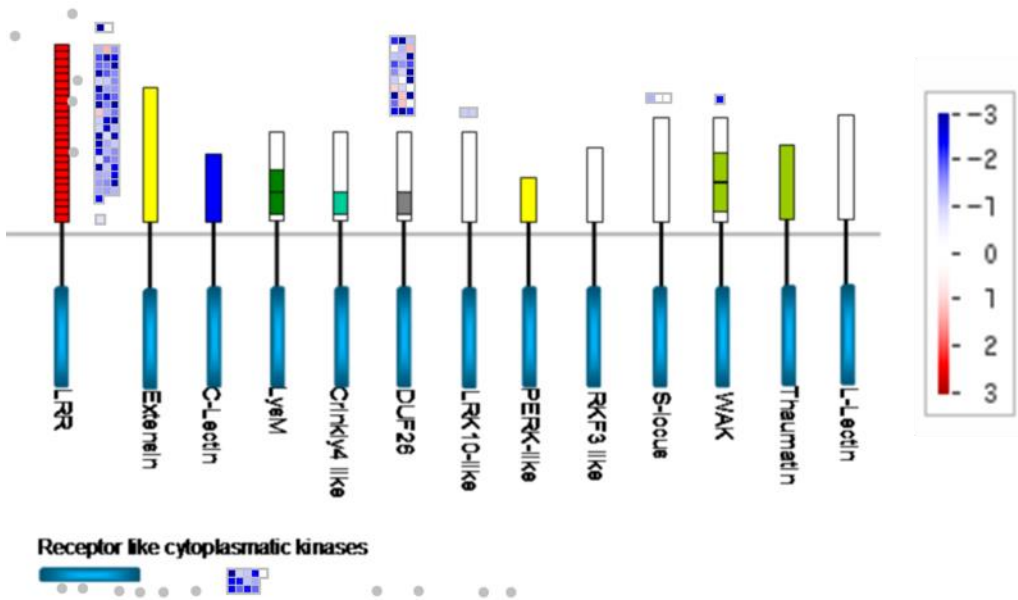

LRR
Extensin
C-Lectin
LysM
Crinkly4 like
DUF26
LRK10-like
PERK-like
RKF3 like
S-locus
WAK
Thaumatin
L-Lectin
3
2
1
0
1
2
3
Receptor like cytoplasmatic kinases

Figure 20. Number of four groups of transcription factors identified in DEGs. Number of DEGs in each group is labeled on the bars.

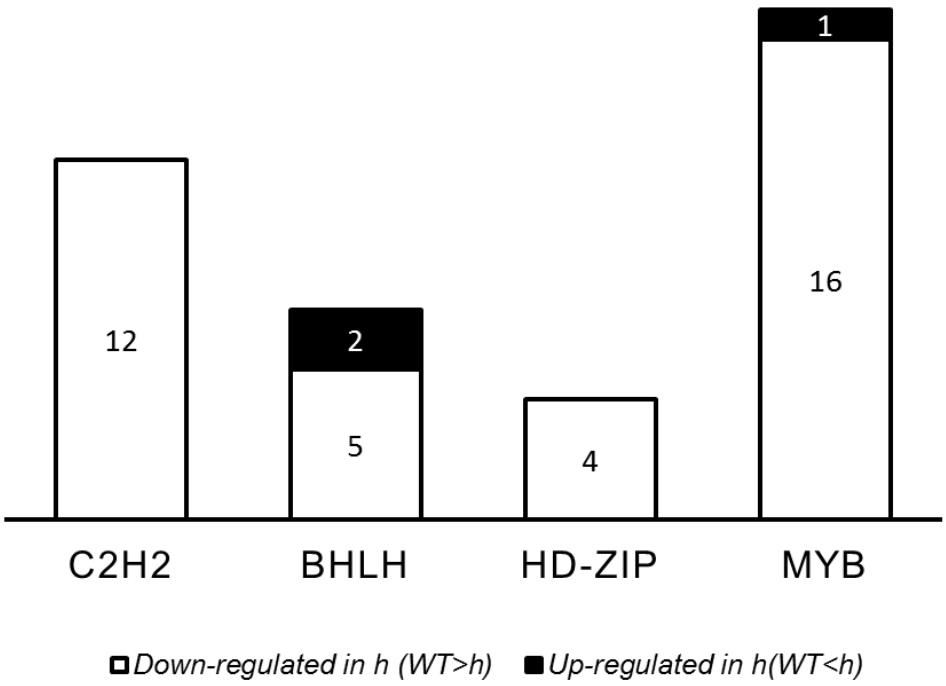
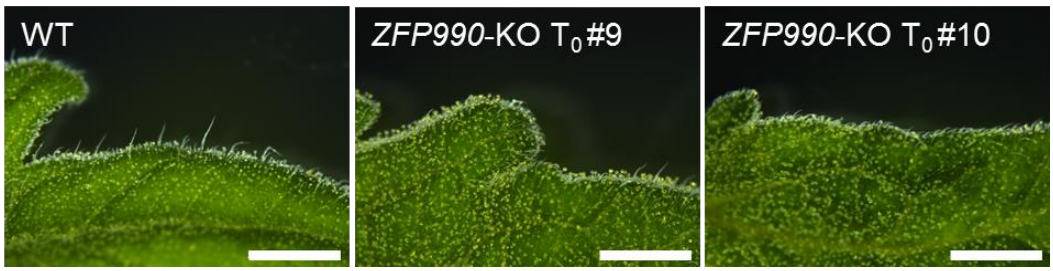

1
16
12
2
5
4
C2H2
BHLH
HD-ZIP
MYB
Down-regulated in h (WT>h)
Up-regulated in h(WT<h)

Figure 21. Phenotype of *ZFP990* overexpression and knockout lines. (A) Leaves of wild-type, and *ZFP990* knockout transgenic lines were imaged with DMS. Long trichomes were not observed on the leaves of knock-out plants (right). Scale bars = 2mm. (B) Wild-type ZFP990 amino acid sequence compared with deduced amino acid sequences of *ZFP990* knockout lines. Premature stop codons were identified from the amino acid position 43 and 45 of *ZFP990* knock-out line 9 and 10, respectively.

A



B

```
Solyc10g078990      MEKIEREALDFMNVESSSQLPITLPVNEKPIRLFGKEFGGGDSINMTATNMSEFIENNPF
990KO-#9            MEKIEREALDFMNVES--SSQLPYPLTKSLLGFLVKNLVVVIVLI--------------
990KO-#10           MEKIEREALDFMNVESSFHFLLPYPLTKSLLGFLVKNLVVVIVLI--------------
                    ***************     :. *:.:. : :: *::     :
Solyc10g078990      YDKPNIVKENHIKKNVEIMRKYECHYCFRSFPTSQALGGHQNAHKKERKNAKRHDQPSIY
990KO-#9            ------------------------------------------------------------
990KO-#10           ------------------------------------------------------------

Solyc10g078990      GITNRNRLGEATVSTRTHYHSTWTNINNNTPRFDGNGHNVNVISPINGNPLTFWQIPPAF
990KO-#9            ------------------------------------------------------------
990KO-#10           ------------------------------------------------------------

Solyc10g078990      NQYISSSSDNNNNLVFSNDDLIRNPPMVNISNNCEYKPKGEVQDHVSLDLHL
990KO-#9            ---------------------------------------------------
990KO-#10           ---------------------------------------------------
```

Figure 22. Subcellular localization of ZFP990. *A. tumefaciens* strain LBA4404 were transformed with each of pBCo-*MYC2*-YFP and pBCo-*ZFP990*-YFP vectors, and infiltrated into tobacco leaves. 2 days after infiltration, (a) CFP fused MYC2, and (b) YFP fused ZFP990 were visualized under confocal microscope. (c) Overlay, and (d) transmitted light image. Scale bars, 20μm.

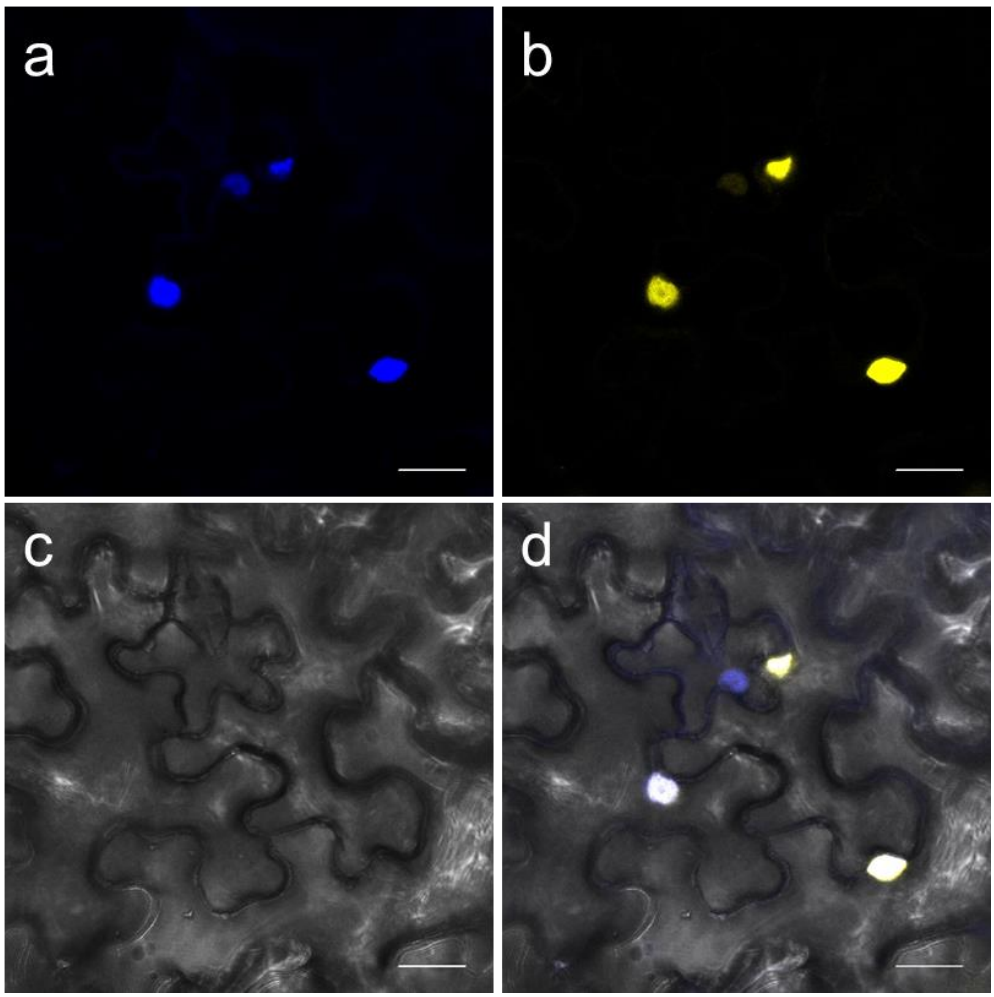
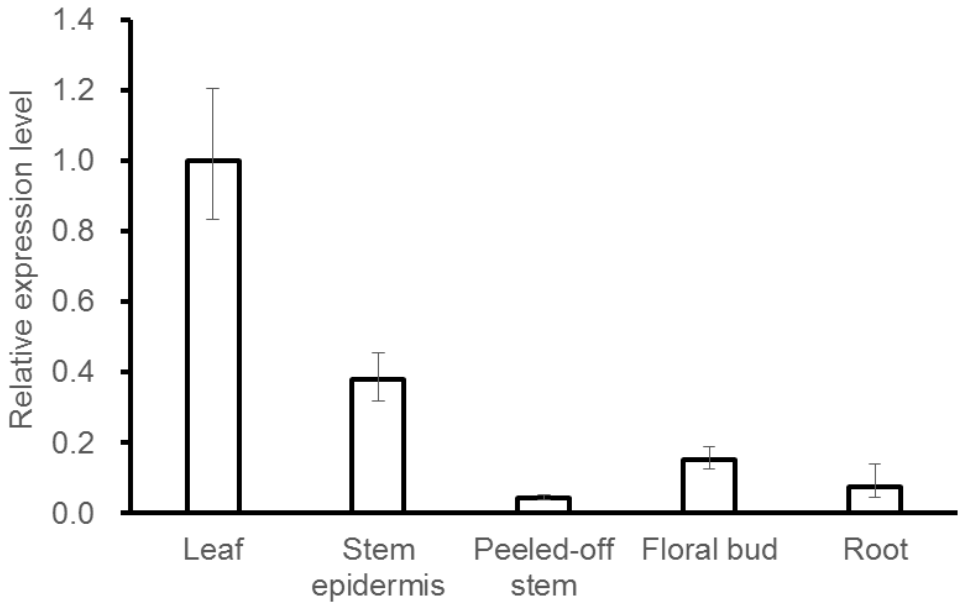

a
b
c
d

Figure 23. Quantitative RT-PCR analysis of *ZFP990* expression levels in different tissues of wild-type plants. Data is means of three independent biological replicates and *SlACT7* was used to normalize the expression levels. Error bars indicate standard error of three biological replicates. The expression level of *ZFP990* in leaf tissue has been calibrated to 1.


Relative expression level
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Leaf
Stem epidermis
Peeled-off stem
Floral bud
Root

DISCUSSION

Wide range of plant species have trichomes but not all of them shares conserved molecular and genetic pathways to generate trichomes. Trichomes have number of known functions mostly conferring resistance on plants against biotic and abiotic stresses. Thus, understanding how trichomes are differentiated from epidermal cells will provide insights into diverse and comparative pathways that plants have evolved between species. In addition, it will provide deeper knowledge on how plants develop and activate its defense system in response to changing environment condition. Genetic researches on trichome differentiation have been done extensively with species in the Rosids including Arabidopsis and cotton, which has unicellular trichomes. Researches to fine trichome differentiation controlling genes in the Asterids had been performed by using the homologs of the Rosids. However, the results showed that the clades of the Rosids and the Asterids do not share the trichome regulatory network, thus, we used forward genetics to find genes involved in the pathway. In this study, *hairs absent* tomato mutant, which long trichomes are absent on stem and sepal, was used to identify genes, with two wild tomato species that have *h*-like phenotype. Complementation test showed that *h* and the two wild species had a mutation on the same gene, which was *ZFP970*. Phylogenetic analysis revealed that ZFP970 and ZFP990 are the closest homologs in tomato, and are grouped into the same clade with AtGIS and other Arabidopsis zinc finger transcription factors reported to be participate in trichome initiation.

Overexpression of *ZFP970* in Arabidopsis affected trichome formation on sepal and caused ectopic trichome formation on siliques. This result, together with the evidence from the recent study that *AtGIS* could regulate glandular trichome development in tobacco, contradicts with the previous reports that trichome initiation pathway in the two clades, the Rosids and the Asterids, are not homologous and functionally independent [48]. Since the previous reports were mostly studied with the downstream part of the trichome initiation pathway, where MBW complex and HD-Zip proteins are involved in, it is assumed that upstream of the trichome formation network might be conserved between the Rosids and the Asterids. It can be inferred that the Rosids and the Asterids are divided after zinc finger proteins are evolved. The acquisition of downstream genes involved in trichome initiation occurred after the division of the two clades, and have evolved independent pathways. To further support this assumption, whether the expression of Arabidopsis genes in tomato could affect the trichome formation is remained to be verified. Additionally, *AtGIS* mainly function in developing stem, however, *ZFP970* did not affect trichomes on stem. It is supposed that sequence variation between AtGIS and ZFP970 led to interact with different proteins in a tissue specific manner. Whether this different tissue specificity comes from the diversified regulatory network is needed to be elucidated.

There are six of Arabidopsis and nine of tomato zinc finger transcription factors in the clade I. Among the six Arabidopsis genes, *ZFP8, GIS2*, and *GIS* are functionally equivalent, but partly they function in a tissue specific way [9]. Alike to the Arabidopsis zinc fingers, tomato zinc fingers are also assumed to function redundantly but partly in a tissue specific way. In this study, we revealed that *ZPF970*

mainly function in stem and sepal, while *ZPF990* function in leaf tissue. Both *ZFP970* and *ZFP990* are involved in type I trichome initiation, but appeared to be function interdependently in a separate tissues. It would be interesting to investigate if the expression of *ZFP990* under the control of *ZFP970* promoter can restore the wild-type phenotype in *h*. Furthermore, how the type I trichome-less phenotype of *hairs absent* and *ZFP990* knockout plants changed the other types of trichomes would be clarified by conducting a trichome density analysis.

Protein prediction and motif analysis showed that all the zinc fingers in the tree have a single C2H2-type zinc finger motif with a conserved QALGGH sequence and an EAR motif on the C-terminal end. The EAR motif is a the consensus sequence, representatively described with a patterns of either LxLxL or DLNxxP [49]. EAR motif-containing proteins, especially transcription factors, are identified as dominant repressors in plant, which lead to function as negative regulators in developmental stages. For instance, R2R3-MYB transcription factors are positive regulators of trichome initiation. Overexpression of chimeric MYB protein fused with EAR motif in Arabidopsis suppressed trichome development and produced glabrous leaves [50, 51]. In this research, however, *ZFP970* and *ZFP990* were found to be positive regulators in trichome differentiation. It is unclear whether the motif actually functions or not, thus additional experiment such as pull-down assay or yeast two-hybrid should be performed.

According to the RNA-seq data, genes related to biotic stress and immune response have been predominantly down-regulated in *hairs absent* mutant. It is assumed that the function of type I trichome is important in conferring resistance

against pathogen and herbivore attacks. Existing studies have shown that both non-glandular and glandular trichomes function in resistance to herbivores either in physical or chemical ways [44-47]. Reported studies on the role of trichomes in the resistance of tomato to herbivores are mostly based on experiments performed with wild tomato species. Moreover, our knowledge on the importance of trichomes on tomato resistance against pathogen is narrow and limited. This functional annotation result showing DEG enrichment in secondary metabolite biosynthesis, biotic stress, and immune response groups will suggest a connection between defense mechanism against insect and pathogen with the function of type I trichomes.

In conclusion, the identification of type I trichome differentiation regulatory gene *H* will provide the base of researches on multicellular trichome differentiation network. Moreover, revealing trichome development pathway will offer a cornerstone for plant breeding in aspects of metabolic engineering and strengthening plant stress tolerance.

REFERENCES

1. Kang, J.H., et al., *The Tomato odorless-2 Mutant Is Defective in Trichome-Based Production of Diverse Specialized Metabolites and Broad-Spectrum Resistance to Insect Herbivores.* Plant Physiology, 2010. **154**(1): p. 262-272.
2. Kang, J.H., et al., *Distortion of trichome morphology by the hairless mutation of tomato affects leaf surface chemistry.* J Exp Bot, 2010. **61**(4): p. 1053-1064.
3. Wagner, G.J., *Secreting Glandular Trichomes: More than Just Hairs.* Plant Physiology, 1991. **96**(3): p. 675-679.
4. Wagner, G.J., E. Wang, and R.W. Shepherd, *New Approaches for Studying and Exploiting an Old Protuberance, the plant trichome.* Ann Bot, 2004. **93**(1): p. 3-11.
5. Wester, K., et al., *Trichome diversity and development.* Development, 2009. **136**(9): p. 1487-96.
6. LEVIN, D.A., *The role of trichomes in plant defense.* The Quarterly Review of Biology, 1973. **48**(1): p. 3-15.
7. Pattanaik, S., et al., *An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis.* Front Plant Sci, 2014. **5**: p. 259.
8. Hulskamp, M., A. Schnittger, and U. Folkers, *Pattern formation and cell differentiation: trichomes in Arabidopsis as a genetic model system.* International Review of Cytology - a Survey of Cell Biology, Vol 186, 1999. **186**: p. 147-178.
9. Yan, A., et al., *Involvement of C2H2 zinc finger proteins in the regulation of epidermal cell fate determination in Arabidopsis.* J Integr Plant Biol, 2014. **56**(12): p. 1112-7.
10. Luckwill, L.C., *The Genus Lycopersicon : an Historical, Biological, and Taxonomic Survey of the Wild and Cultivated Tomatoes.* University of Aberdeen: The Genus Lycopersicon, 1943(120): p. 44.
11. Antonious, G.F., *PRODUCTION AND QUANTIFICATION OF METHYL*

KETONES IN WILD TOMATO ACCESSIONS. J Environ Sci Health B, 2001. **36**(6): p. 835-48.

12. Schilmiller, A.L., R.L. Last, and E. Pichersky, *Harnessing plant trichome biochemistry for the production of useful compounds.* Plant J, 2008. **54**(4): p. 702-11.
13. Kennedy, G.G., *Tomato, pests, parasitoids, and predators: tritrophic interactions involving the genus Lycopersicon.* Annu Rev Entomol, 2003. **48**: p. 51-72.
14. Serna, L. and C. Martin, *Trichomes: different regulatory networks lead to convergent structures.* Trends Plant Sci, 2006. **11**(6): p. 274-80.
15. Yang, C., et al., *A regulatory gene induces trichome formation and embryo lethality in tomato.* Proc Natl Acad Sci U S A, 2011. **108**(29): p. 11836-41.
16. Nadakuduti, S.S., et al., *Pleiotropic Phenotypes of the sticky peel Mutant Provide New Insight into the Role of CUTIN DEFICIENT2 in Epidermal Cell Function in Tomato.* Plant Physiology, 2012. **159**(3): p. 945-960.
17. Rick, C.M. and L. Butler, *Cytogenetics of the Tomato.* Advances in Genetics Incorporating Molecular Genetic Medicine, 1956. **8**: p. 267-382.
18. Wu, H.W., et al., *Discriminating mutations of HC-Pro of zucchini yellow mosaic virus with differential effects on small RNA pathways involved in viral pathogenicity and symptom development.* Mol Plant Microbe Interact, 2010. **23**(1): p. 17-28.
19. Eric Richards, M.R., and Sharon Rogers, *Preparation of Genomic DNA from Plant Tissue.* Current Protocols in Molecular Biology, 1994.
20. Koo, A.J., et al., *Identification of a Peroxisomal Acyl-activating Enzyme Involved in the Biosynthesis of Jasmonic Acid in Arabidopsis.* J Biol Chem, 2006. **281**(44): p. 33511-20.
21. Schilmiller, A.L., A.J. Koo, and G.A. Howe, *Functional Diversification of Acyl-Coenzyme A Oxidases in Jasmonic Acid Biosynthesis and Action.* Plant Physiol, 2007. **143**(2): p. 812-24.
22. McCORMICK, S., *Transformation of tomato with Agrobacterium tumefaciens* Plant tissue culture manual 1991. **B6**: p. 1-9.
23. Kim, H., et al., *A simple, flexible and high-throughput cloning system for plant genome editing via CRISPR-Cas system.* J Integr Plant Biol, 2016. **58**(8): p. 705-12.

24. Zhang, X.R., et al., *Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method.* Nature Protocols, 2006. **1**(2): p. 641-646.
25. Murray P Cox, D.A.P.a.P.J.B., *AtSolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. .* BMC Bioinformatics, 2010. **11**: p. 485.
26. Ben Langmead, C.T., Mihai Pop, Steven L Salzberg, *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biology, 2009. **10**(3): p. R25.
27. Ben Langmead, S.L.S., *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
28. Simon Anders, W.H., *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
29. Lucas, A., *amap: Another Multidimensional Analysis Package.* 2014.
30. Michael Ashburner, C.A.B., Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin & Gavin Sherlock, *Gene Ontology: tool for the unification of biology.* Nature Genetics, 2000. **25**(1): p. 25-9.
31. Minoru Kanehisa, S.G., *KEGG: Kyoto encyclopedia of genes and genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
32. Jinpu Jin, H.Z., Lei Kong, Ge Gao, Jingchu Luo, *PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors.* Nucleic Acids Research, 2014. **42**(Database issue): p. D1182-7.
33. Thimm, O., et al., *MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.* Plant J, 2004. **37**(6): p. 914-39.
34. Tanksley, S.D., et al., *High density molecular linkage maps of the tomato and potato genomes.* Genetics, 1992. **132**(4): p. 1141-60.
35. Gan, Y., et al., *GLABROUS INFLORESCENCE STEMS modulates the regulation by gibberellins of epidermal differentiation and shoot maturation in Arabidopsis.* Plant Cell, 2006. **18**(6): p. 1383-1395.
36. Yuval Eshed, D.Z., *A genomic library of Lycopersicon pennellii in L.*

esculentum: A tool for fine mapping of genes. Euphytica, 1994. **79** p. 175-179.

37. Sun, L., et al., *GLABROUS INFLORESCENCE STEMS3 (GIS3) regulates trichome initiation and development in Arabidopsis.* New Phytol, 2015. **206**(1): p. 220-30.
38. Zhou, Z., et al., *Zinc Finger Protein 6 (ZFP6) regulates trichome initiation by integrating gibberellin and cytokinin signaling in Arabidopsis thaliana.* New Phytol, 2013. **198**(3): p. 699-708.
39. Zhou, Z.J., et al., *Zinc Finger Protein5 Is Required for the Control of Trichome Initiation by Acting Upstream of Zinc Finger Protein8 in Arabidopsis.* Plant Physiology, 2011. **157**(2): p. 673-682.
40. Joseph, M.P., et al., *The Arabidopsis ZINC FINGER PROTEIN3 Interferes with Abscisic Acid and Light Signaling in Seed Germination and Plant Development.* Plant Physiol, 2014. **165**(3): p. 1203-1220.
41. Payne, T., S.D. Johnson, and A.M. Koltunow, *KNUCKLES (KNU) encodes a C2H2 zinc-finger protein that regulates development of basal pattern elements of the Arabidopsis gynoecium.* Development, 2004. **131**(15): p. 3737-3749.
42. Weingartner, M., C. Subert, and N. Sauer, *LATE, a C2H2 zinc-finger protein that acts as floral repressor.* Plant Journal, 2011. **68**(4): p. 681-692.
43. Choi, H.S., M. Seo, and H.T. Cho, *Two TPL-Binding Motifs of ARF2 Are Involved in Repression of Auxin Responses.* Front Plant Sci, 2018. **9**: p. 372.
44. Simmons, A.T. and G.M. Gurr, *Trichomes of Lycopersicon species and their hybrids: effects on pests and natural enemies.* Agricultural and Forest Entomology, 2005. **7**(4): p. 265-276.
45. Hartman, J.B. and D.A. St Clair, *Combining ability for beet armyworm, Spodoptera exigua, resistance and horticultural traits of selected Lycopersicon pennellii-derived inbred backcross lines of tomato.* Plant Breeding, 1999. **118**(6): p. 523-530.
46. Mutschler, S.L.B.G.A.C.M.A., *Identification of quantitative trait loci associated with acylsugar accumulation using intraspecific populations of the wild tomato, Lycopersiconpennellii.* Theor Appl Genet, 1998. **96**: p. 458-467.
47. Juvik, J.A., et al., *Acylglucoses from Wild Tomatoes Alter Behavior and*

Reduce Growth and Survival of Helicoverpa zea and Spodoptera exigua (Lepidoptera: Noctuidae). Journal of Economic Entomology, 1994. **87**(2): p. 482-492.

48. Yihua Liu, D.L., Rui Hu, Changmei Hua, Imran Ali, Aidong Zhang, and M.W. Bohan Liu, Linli Huang, Yinbo Gan, *AtGIS, a C2H2 zinc-finger transcription factor from Arabidopsis regulates glandular trichome development through GA signaling in tobacco.* Biochemical and Biophysical Research Communications, 2017.
49. Kagale, S. and K. Rozwadowski, *EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression.* Epigenetics, 2011. **6**(2): p. 141-6.
50. Liang, G., et al., *MYB82 functions in regulation of trichome development in Arabidopsis.* J Exp Bot, 2014. **65**(12): p. 3215-23.
51. Matsui, K., et al., *A Chimeric AtMYB23 Repressor Induces Hairy Roots, Elongation of Leaves and Stems, and Inhibition of the Deposition of Mucilage on Seed Coats in Arabidopsis.* Plant Cell Physiol, 2005. **46**(1): p. 147-55.

ABSTRACT IN KOREAN

토마토에서 type Ⅰ 모상체 발달에 관여하는 *Hairs absent*에 대한 연구

김 희 진

서울대학교 국제농업기술대학원 국제농업기술학과

지도교수 강 진 호

모상체는 식물의 표피세포에서 분화되는 구조로 크게 glandular과 non-glandular 두 종류로 나눌 수 있다. 모상체는 화학물질을 만들거나 물리적 방어를 통해 생물학적 또는 비생물학적 스트레스로부터 식물을 보호하는 역할을 한다. 모델식물인 애기장대에는 단일세포로 이루어진 한 종류의 모상체가 있고 토마토에는 다세포로 이루어진 7종류의 모상체가 있다. 애기장대가 속한 장미군에서는 모상체의 발달기작이 잘 밝혀져 있지만 토마토가 속한 국화군에서는 거의 밝혀진 유전자가 없다. 본 연구에서는 모상체 발달에 관여하는 유전자를 동정하기 위해 줄기와 꽃받침에

type Ⅰ 모상체가 없는 *hairs absent* (*h*)라는 돌연변이 토마토를 이용하였다. *H* 유전자는 애기장대의 화서 줄기에서 모상체 발달을 조절하는 *GLABROUS INFLORESCENCE STEMS*과 상동유전자로, C2H2 type의 아연 손가락 단백질을 암호화하고 있었으며 돌연변이에서는 이 유전자의 소실이 일어난 것으로 나타났다. Type Ⅰ 모상체가 없는 토마토 야생종인 *S. pimpinellifolium*과 *S. pennellii*에서 *H*의 다른 돌연변이 대립유전자를 확인하였고, 이를 통해 *ZFP970* 유전자가 *H*임을 밝혔다. H와 가장 가까운 상동단백질인 Solyc10g078990 또한 모상체 발달에 관여하는 것을 확인하였다. Wild-type과 h 돌연변이의 RNA-seq 비교 분석을 통해 1570개의 차별발현유전자를 선발하였다. 이 유전자들은 이후 모상체 발달을 조절하는 유전자와 이차대사물질 생합성에 관여하는 유전자 동정을 위해 연구될 것이다.

주요어: 모상체, *hairs absent*, 아연 손가락 단백질, 토마토, 순행유전학

학번: 2016-25982