Comparative anatomical and transcriptomic analyses of the color variation of leaves in *Aquilaria sinensis*

Jiaqi Gao [1, 2], **Tong Chen** [Corresp., 1], **Chao Jiang** [1], **Tielin Wang** [1], **Ou Huang** [3], **Xiang Zhang** [1], **Juan Liu** [Corresp. 1]

[1] National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

[2] School of Pharmacy, Jiangsu University, Zhenjiang, China

[3] Guangdong Shangzhengtang Group Co., Ltd, Dongguan, China

Corresponding Authors: Tong Chen, Juan Liu
Email address: chent@nrc.ac.cn, juanliu126@126.com

Color variation in plant tissues is a common phenomenon accompanied with a series of biological changes. In this study, a special-phenotype *Aquilaria sinensis* (GS) with color variation of leaf was firstly reported, and DNA barcode sequences showed GS samples could not be discriminated clearly with the normal *A. sinensis* sample (NS), which suggested that the variety was not the cause of the GS formation. To reveal the characteristics of GS compared to NS, the anatomical and transcriptome sequencing studies were carried out. In microscopic observation, the leaves of golden-vein-leaf sample (LGS) and normal-vein-leaf sample (LNS) showed significant differences including the area of the included phloem in midrib and the thickness parameters of palisade and spongy tissues; the stems of golden-vein-leaf sample (SGS) and normal-vein-leaf sample (SNS) were also different in many aspects such as the area of vessels and included phloem. In addition, the structure of chloroplast was more complete in the midrib of LNS than that of LGS, and some particles suspected as virus were found through transmission electron microscope as well. Genes upregulated in LGS in contrast with LNS were mainly enriched in photosynthesis. As for stems, most of the genes upregulated in SGS compared to SNS were involved in translation and metabolism processes. The pathways about photosynthesis and chlorophyll metabolism as well as some important transcription factors may explain the molecular mechanism of the unique phenotypes of leaves and the genes related to suberin biosynthesis may result in the difference of stems. In addition, the genes about defense response especially biotic stress associated with numerous pathogenesis-related (PR) genes upregulated in LGS compared to LNS indicated that the pathogen may be the internal factor. Taken together, our results reveal the macro- and micro-phenotype variations as well as gene expression profiles between GS and NS, which could provide valuable clues for elucidating the mechanism of the color variation of *Aquilaria*.

Comparative anatomical and transcriptomic analyses of the color variation of leaves in *Aquilaria sinensis*

Jiaqi Gao[1,2], Tong Chen[1*], Chao Jiang[1], Tielin Wang[1], Ou Huang[3], Xiang Zhang[1], Juan Liu[1*]

[1]National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
[2]School of Pharmacy, Jiangsu University, Zhenjiang 212013, China
[3]Guangdong Shangzhengtang Group Co., Ltd, Dongguan 523120, China

Corresponding Author:
Tong Chen[1]
Juan Liu[1]
No. 16, Dongzhimen S Alley, Beijing, 100700, China
Email address: Tong Chen, chent@nrc.ac.cn;
Juan Liu, juanliu126@126.com

Abstract

Color variation in plant tissues is a common phenomenon accompanied with a series of biological changes. In this study, a special-phenotype *Aquilaria sinensis* (GS) with color variation of leaf was firstly reported, and DNA barcode sequences showed GS samples could not be discriminated clearly with the normal *A. sinensis* sample (NS), which suggested that the variety was not the cause of the GS formation. To reveal the characteristics of GS compared to NS, the anatomical and transcriptome sequencing studies were carried out. In microscopic observation, the leaves of golden-vein-leaf sample (LGS) and normal-vein-leaf sample (LNS) showed significant differences including the area of the included phloem in midrib and the thickness parameters of palisade and spongy tissues; the stems of golden-vein-leaf sample (SGS) and normal-vein-leaf sample (SNS) were also different in many aspects such as the area of vessels and included phloem. In addition, the structure of chloroplast was more complete in the midrib of LNS than that of LGS, and some particles suspected as virus were found through transmission electron microscope as well. Genes upregulated in LGS in contrast with LNS were mainly enriched in photosynthesis. As for stems, most of the genes upregulated in SGS compared to SNS were involved in translation and metabolism processes. The pathways about photosynthesis and chlorophyll metabolism as well as some important transcription factors may explain the molecular mechanism of the unique phenotypes of leaves and the genes related to suberin biosynthesis may result in the difference of stems. In addition, the genes about defense response especially biotic stress associated with numerous pathogenesis-related (PR) genes upregulated in LGS compared to LNS indicated that the pathogen may be the internal factor. Taken together, our results reveal the macro- and micro-phenotype variations as well as gene

expression profiles between GS and NS, which could provide valuable clues for elucidating the mechanism of the color variation of *Aquilaria*.

Introduction

Agarwood is a resinous, fragrant wood, which is highly valued for its use in medicine, perfumes, and incense across Asia, Middle East and Europe (Takemoto et al. 2008). It is produced by species of tropical trees of the genus *Aquilaria*, whose population is dramatically declining due to overexploitation. Furthermore, all of the *Aquilaria* species, including *A. sinensis* (Lour.) Gilg which is the only certified source for agarwood listed in China Pharmacopoeia (Committee 2020), are conserved under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, http://www.cites.org). However, the source of wild agarwood is facing serious depletion drain owing to its slow and infrequent formation and uncontrolled collection in forests. Fortunately, *A. sinensis* has been widely cultivated in the Dongguan area of China, as early as the Tang and Song Dynasties. During long-term cultivation, some phenotypic differences of *A. sinensis* have been showed up, some of which could be seemed as potential varieties worth to be explored and bred. Fortunately, we firstly investigated the color variation of leaves existed in *A. sinensis*, which is also the first report about color variation in genus *Aquilaria*. However, there is no further information about this phenotype and potential value in agarwood industry.

In the plant kingdom, color variation has been widely reported in leaf, culm, flower, fruit, seed coat and other tissues derived from different plants (Ahmed et al. 2004; Barazani et al. 2019; Brand et al. 2014; Xia et al. 2015; Yang et al. 2010). Some phenotypes about color variation could be found in the whole tissue while others could show diversified phenotypes such as color stripes or banding mosaic (Wang et al. 2019; Xia et al. 2015). These phenotypes emerge with a series of changes in colored substances such as chlorophyll, carotenoid, anthocyanins, flavonoids or other pigments (Cheng et al. 2018; He et al. 2011; Khandagale & Gawande 2019; Xia et al. 2015). Accompanied with the changes in appearance, a battery of characteristics in botany or ecology could be different. For example, the flowers of *Eruca sativa* with different color display diverse attraction for their main pollinator because of the discrimination of the scents in flowers of different colors (Barazani et al. 2019). In *Capsicum annuum* L., the plant samples differing in leaf color are accompanied with different combinations of pigments which could influence the resistance of these plants against whitefly (Cheng et al. 2018) which is a kind of pest damaging plants seriously (Ullah & Lim 2016). In the studies about the color variation in fruit like tomato, the skin pigmentation, carotenoid biosynthesis and ripening-associated chlorophyll degradation were claimed as main factors in the formation mechanism of fruit color in tomato (Chattopadhyay et al. 2021). *SlMYB12*, *IDI1*, *PSY1*, *CrtISO*, *CYC-B*, *LCY-E*, and *SGR* were found to be important in the three pathways mentioned in color variation in tomato (Chattopadhyay et al. 2021). Regarding leaves, it was common that the color mutant would be used in studies about color variation (Zhu et al. 2015). Apart from the metabolism pathways like

chlorophyll metabolism (Zhu et al. 2015), other factors including the recessive nuclear gene like *SiYGL1* (Li et al. 2016), shock protein genes containing *HSP70*, *HSP90* (Wu et al. 2018) were also being studied.

Additionally, the plant with natural color variation is a kind of infrequent material which is worthwhile to be applied in ornamentation or scientific studies bringing considerable economic and practical value besides the influence and botanical functions to the plants themselves. For instance, the tea plants with green, yellowish, purplish leaves showed different active levels to the metabolism of catechins and the contents of theanine, caffeine and other chemicals with respect to the tea quality and characteristic suggesting the selection of breeding materials as new tea cultivars (Li et al. 2016). Besides, some researchers revealed that a purple-colored rice landrace was a good resource for rice breeding because of its disease resistance, stress tolerance, enhanced nutritional values, etc. after the whole genome sequencing and comparative analysis of the unique colored rice (Lachagari et al. 2019). In addition, the plants with color variation of tissues are also important materials used to study the biosynthesis of some colored substances (Gang et al. 2019b). Thus, the *A. sinensis* plant with color variation should be a worthy and potential material to be investigated.

Herein, a unique-phenotype *A. sinensis* whose leaves showing a deeper green compared to the normal *A. sinensis* accompanied with golden (yellow) veins was firstly found, and DNA barcode sequences showed that the variety was not the cause of the GS formation. Additionally, the differences between the unique-phenotype *A. sinensis* and normal *A. sinensis* were analyzed using anatomical observation and transcriptome sequencing. We performed functional analysis of differentially expressed genes and assessed Gene Ontology (GO) annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Furthermore, the candidate genes related to photosynthesis and defense responses enriched from functional analysis were verified using quantitative real-time PCR. Overall, our integrated analysis allows for a better understanding of the molecular changes occurring in color variation in *A. sinensis*.

Materials & Methods

Plant materials

The golden-vein-leaf and normal-vein-leaf sample (GS and NS) *A. sinensis* plants were obtained from Dalingshan town, Guangdong province, China (latitude 22º45′43″ N, longitude 113º48′45″ E). Those plants grow in close proximity with similar environment and at a similar altitude of 70-80 m. The samples were all collected from six *A. sinensis* trees in which three trees were GS and the others were NS. Then, three leaves and three stems of GS as well as those of NS were collected from each three branches of trees and mixed respectively, with three biological replicates used for each group in DNA barcode sequencing and RNA sequencing. All of the samples were frozen in liquid nitrogen and stored at -80 °C for the next studies.

DNA extraction and determination of chlorophyll content

The leaves were powdered in a mortar with a pestle in liquid nitrogen. The stems were sliced first, and then powered in liquid nitrogen according to the published methods (Jiao et al. 2014). Additionally, those samples were inserted into a 2 mL tube containing 1000 μL buffer AP1, 8 μL RNase A, and 1% polyvinylpyrrolidone with incubation at 65 °C for 6 h respectively. Cooling to room temperature, the 280 μL buffer P3 was added into the tubes and those samples were incubated for 2 h at -20 ℃. Subsequent steps were conducted as the direction supplied by the manufacturer of EASYspin Plus Plant DNA Kit (Aidlab Biotechnologies Co., Ltd, Beijing, China). The determination of chlorophyll content was performed based on the method published before (Sartory & Grobbelaar 1984).

DNA barcode sequence amplification and phylogenetic tree analysis

ITS2 and *trn*L-*trn*F were used as barcode sequences to identify the species and the primers were adopted from a publishing paper (Lee et al. 2016). The PCR reaction system was also conducted as the previous report (Lee et al. 2016).

The ITS2 and *trn*L-*trn*F barcode sequences of GS and NS were aligned with *A. crassna*, *A. hirta, A. malaccensis, A. microcarpa, A. sinensis, A. yunnanensis, Gyrinops versteegii* and *Gonystylus bancanus* (**Table S1**) using Clustal W algorithm (Larkin et al. 2007). Subsequently, the phylogenetic tree was constructed based on neighbor-joining (NJ) method with 1000 bootstrap replications. All the above analyses were operated with MEGA-X (10.1.8) (Kumar et al. 2018).

Microscopic observation and measurement

The paraffin microsections were sliced by Leica CM1860 freezing microtome (Leica Microsystems Inc., Wetzlar, Germany) with Safranin O/Fast green staining subsequently. The microscopic observations of those microsections were conducted by ZEISS AX10 microscope (ZEISS corporation, Jena, Germany). ZEISS ZEN 2 lite software was applied to measure the parameters of the photomicrographs. The statistical differences were calculated by T-test using IBM SPSS Statistics 22. The ultrastructure was observed through transmission electron microscope H-7500 (Hitachi, Ltd., Japan).

RNA extraction and transcriptome sequencing

The RNA was extracted from the *A. sinensis* samples using the EASYspin Plus Plant RNA Kit (Aidlab Biotechnologies Co., Ltd, Beijing, China) according to the manufacturer's instructions. The quality of extracted RNA was examined by NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). Then, 0.5 μg RNA per sample was reverse-transcribed into single-stranded cDNA using cDNA Synthesis Kit (TaKaRa, Japan) as what the manufacturer's instructions suggested. Subsequently, the RNA template was eliminated and the second-strand cDNA synthesis was conducted with DNA polymerase, dNTPs, and RNase. After end repair, A-tailing and indexing ligation, the products were purified with PCR extraction kit, and amplified

to be cDNA libraries. Finally, the sequencing was conducted on Illumina NovaSeq 6000. The data have been deposited in NGDC's Genome Sequence Archive (Wang et al. 2017), under accession number CRA002670 that is publicly accessible at https://bigd.big.ac.cn/gsa/browse/CRA002670. The *de novo* assembly was conducted by Trinity method (Grabherr et al. 2011).

Annotation and differentially expressed gene analysis

The annotations of transcriptome data were conducted by Diamond (Buchfink et al. 2015), KAAS (Moriya et al. 2007) and Blast2GO (Gotz et al. 2008). The gene expression level was estimated with FPKM (Trapnell et al. 2010). Using DESeq2 (Love et al. 2014), when the parameters of genes expressed in different groups at $\log_2$(Fold Change) larger than 1 or less than -1, and false discovery rate (FDR) less than 0.05, they could be accepted as differentially expressed genes (DEGs). The GO and KEGG enrichments were conducted via R package clusterProfiler (Yu et al. 2012) with the custom data sets from the annotation results. For GO enrichment, 15,811 annotated genes were used as the background gene set. For KEGG, 4,865 annotated genes were used as the background gene set. In GO enrichment, the FDR screening threshold was set to be less than 0.05, while the value was set to be less than 0.1 in KEGG enrichment. The cluster analysis was conducted via the STEM clustering algorithm (Ernst et al. 2005) and the cluster whose P value was less than 0.05 was considered as significant difference. All the plots were generated using one public on-line tool BIC (www.ehbio.com/Cloud_Platform/front/) or R scripts.

Quantitative real-time PCR

The qRT-PCR was carried out with the TB Green@ Fast qPCR Mix (TaKaRa, Japan) and LightCycler@ 480 Real-Time PCR System (Roche, Switzerland). The reaction mixture contained 3 μL of nuclease-free water, 5 μL of TB Green Premix Ex Taq II (TaKaRa, Japan), 1 μL of cDNA, 0.5 μL of forward primer (10 μM), and 0.5 μL of reverse primer (10 μM). The program was set to run for 1 min at 94 ℃, 40 cycles of 10 s at 94 ℃, and 34 s at 60 ℃. Melt curves were obtained by heating the samples from 60 ℃ to 95 ℃ at a rate of 1.0 ℃/s. (**Table S9**). The primers of genes were listed in **Table S9** with a reference gene GAPDH (Xu et al. 2016). Each RNA sample was conducted with three biological repeats and three technical replicates. The relative gene expression levels were calculated through $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen 2001). The correlation analysis was conducted by Pearson correlation coefficient.

Results

DNA barcode sequencing and phylogenetic tree analysis

The samples used in this study were all collected from Guanxiang Intangible Cultural Heritage Protection Park, while their phenotypes differed from each other. The golden-vein-leaf sample (GS) showed a color variation of leaf vein compared to the normal-vein-leaf sample (NS) and had deeper green and harder texture in leaves (**Figs. 1A** and **1B**), and the contents of chlorophyll

including chlorophyll a, chlorophyll b and carotenoid in the leaves of GS (LGS) were all higher than those in the leaves of NS (LNS) (**Fig. 1C**). Though the selected samples were cultured as the same species in the cultivation place, the phenotypes differences were so remarkable that the DNA molecular identification was employed to confirm the phylogeny relationships between GS and NS (**Fig. S1**). According to the previous report, the combination of ITS2+*trn*L-*trn*F was used as DNA barcode for *Aquilaria* species identification (Lee et al. 2016). The phylogenetic tree was constructed based on ITS2+*trn*L-*trn*F DNA barcode sequences in 8 closely related species, and one distantly related species *Gonystylus bancanus* as outgroup according to the previous study (Lee et al. 2016) (**Fig. 1D**). Most of the selected reference species had their separate clades respectively indicating success of the phylogenetic tree construction. The high-level homology of GS and NS sequences with *A. sinensis* verified that they were all *A. sinensis*, but GS samples could not be discriminated clearly with NS samples, which suggested that the variety was not the cause of the GS formation.

Anatomical comparison of leaves and stems

The photomicrographs of the LGS and LNS were acquired after Safranin O/Fast green staining, of which the anatomical parameters were measured (**Fig. 2**; **Table S2**). In the transverse section of the midrib, the bicollateral bundles were apparent (**Figs. 2A** and **2C**). It was visible that the total area of included phloem in LGS midrib surpassed which in LNS midrib for more than 5 times (**Fig. 2E**). In addition, the palisade tissue was thicker in LNS whereas the spongy tissue was thicker in LGS (**Figs. 2B**, **2D**, **2F** and **2G**). Furthermore, the upper epidermis cell thickness in LGS is significantly smaller than LNS, while the difference of lower epidermis cell between LGS and LNS was not significant (**Figs. 2H** and **2I**; **Table S2**).

Using transmission electron microscopy, the ultrastructure of chloroplast was observed (**Fig. 3**). It was obvious that the structure of chloroplast in the midrib of LNS was more complete than that in LGS in which the thylakoids were seriously broken (**Fig. 3B**). Apart from the thylakoid, the starch could be observed in the midrib of LNS (**Fig. 3A**), however, it was seldom in the midrib of LGS. In addition, we found some particles which were suspected as virus (**Fig. 3C** and **3D**), and some vesicles (**Fig. 3C**).

With respect to the stems of GS and NS (SGS and SNS), the differences were also significant and the most obvious characteristics observed in transverse surface had been measured and digitized (**Fig. 4**; **Table S3**). One of the most obvious features in SGS was that the included phloem tended to connect with each other and their shapes were more irregular than that of SNS (**Figs. 4A** and **4E**) which was typical island-shaped (Liu et al. 2018). In addition, the area, length and width of included phloem of SNS were all greater than that of SGS (**Fig. 4I**; **Table S3**). Because most of the included phloem distributed in xylem, we calculated the percentage of included phloem in xylem of SGS and SNS and found that the number of the former was a little bigger though it was less than 2% (**Fig. 4J**), indicating that the difference of distribution level of

included phloem in SGS and SNS was not as significant as the area. In addition, many crystals could be observed clearly in the pith of SNS (**Fig. 4G**) while much less in the pith of SGS in which there were more colored grease-like granular substances (**Fig. 4C**).

The parameters about vessel of SGS were all smaller than SNS (**Figs. 4K** and **4L**; **Table S3**). In consideration of the features that single vessel area of SNS was larger than that of SGS and the difference between the percentage of vessel to xylem of SGS and SNS was less than 2%, the density of SGS vessel could be higher (**Figs. 4B** and **4F**). Furthermore, the single fiber cell of SGS was significantly smaller than that of SNS **(Fig. 4M**) and it could be observed intuitively (**Figs. 4D** and **4H**).

Transcriptome sequencing and differential gene expression profile

Totally 437 million high quality reads (65 G bases) were generated for four samples, LGS, LNS, SGS, and SNS (each with three biological duplicates) (**Table S4**). Due to the lack of gene annotation information, *de novo* assemble were performed to get 26,381 unigenes. The contig N50 of all transcripts was 2,168 nt, the average transcript length was 1418 nt, and the complete BUSCOs (Benchmarking Universal Single-Copy Orthologs) was more than half of the total (**Table S5**; **Fig. S2**). Functional annotation showed that 18,786 genes could be annotated to at least one of Pfam, SwissProt, TrEMBL, KEGG and GO databases (**Fig. S3**).

Differential gene expression analyses were performed between LGS and LNS, SGS and SNS using negative binomial models. There genes with multiple test corrected P-value less than 0.05 and absolute fold change no less than 2 were defined as DEGs. In summary, 1525 genes were upregulated while 752 genes were downregulated in LGS compared to LNS (**Fig. 5A**), and 808 genes were upregulated while 245 genes were downregulated in SGS compared to SNS (**Fig. 5B**). In addition, the samples from same superset could be clustered together using principal component analysis (PCA) of DEGs which were more varianced compared to the no-differential expressed genes (**Fig. S4**).

Photosynthesis related genes were enriched in color variation of *A. sinensis* by GO term enrichment analysis of DEGs

Genes upregulated in LGS compared to LNS were significantly enriched in photosynthesis related biological processes including photosynthesis, light harvesting in photosystem I, response to light stimulus (**Fig. 5C**). Besides, defense response and carbohydrate metabolic process were two other enriched items occupying the highest proportion of DEGs. Those upregulated photosynthesis related terms may explain the deeper green in LGS and carbohydrate metabolic process was another important term participating in photosynthesis (Atkin et al. 2000). In addition, some terms related to cell wall were also enriched, such as pectin catabolic process and plant-type cell wall organization. It was worth considering that the phenotype of midrib may in connection with pectin catabolism. As for cellular component, genes upregulated in LGS

enriched in components referred to photosynthesis. Except for “extracellular region”, the GO term “photosystem I” and “photosystem II” showed the most significance among these items. Meanwhile, “chloroplast thylakoid membrane” and “chloroplast envelope” covered a large amount of gene. GO terms “cell wall” and “plant-type cell wall” were also enriched which coincides with the terms about cell wall in biological process. For molecular function, the most significantly enriched term about upregulated genes in LGS were related to chlorophyll such as “chlorophyll binding” and “pigment binding” as well as many genes were enriched into glycosyl hydrolysis relevant term like “hydrolase activity, hydrolyzing O-glycosyl compounds”.

With regard to the downregulated genes (**Table S6**), the biological process GO terms “response to abscisic acid” and “ethylene biosynthetic process” coincided with phytohormone playing important roles in plant development (Emenecker & Strader 2020; Iqbal et al. 2017). Some terms in molecular function involved transporting, such as “xenobiotic transmembrane transporting ATPase activity” and “ATPase activity, coupled to transmembrane movement of substances”. In summary, the enrichment terms of downregulated genes were much less than upregulated genes so that there were only three molecular function terms, and two cellular component terms enriched into “plasmodesma” and “plant-type vacuole”.

Comparing the GO enrichment analysis for DEGs between SGS and SNS, the significantly enriched GO terms of upregulated genes in SGS were shown in **Fig. 5D**. For biological process, a lot of upregulated genes in SGS were categorized into the terms which were bound up with translation, energy transporting and metabolism such as “translation”, “cytoplasmic translation”, “ATP synthesis coupled proton transport” as well as some terms related to ribosome. Focusing on cellular component, many upregulated genes were about ribosome among which the term “ribosome” was the most significant whose gene number was also the most. Moreover, some terms about mitochondrion were enriched too. For molecular function, the most significantly enriched term “structural constituent of ribosome” occupied a dominant quantitative advantage over others echoing with the situation in biological process and cellular component.

The number of downregulated genes which could be enriched was small (**Table S7**). The terms in cellular component category were not enriched because of the lack of significance and only one term “nitrate transport” was enriched in biological process. The three most significant terms in molecular function were “ADP binding”, “sucrose alpha-glucosidase activity” and “double-stranded RNA binding”.

Photosynthesis related pathways were characterized in color variation of *A. sinensis* by KEGG pathway enrichment analysis of DEGs

To analyze the gene functional pathways of DEGs in leaves and stems, the KEGG pathway enrichment analyses were conducted in leaves and stems samples respectively. It was found that only upregulated genes of GS compared to NS were enriched successfully together with the

small amount (**Fig. 5E**). Although the number of genes enriched was much smaller than that of GO enrichment, it also attached importance to analyze the meaningful pathways such as the "photosynthesis - antenna proteins", "photosynthesis", "porphyrin and chlorophyll metabolism" pathways which were related to photosystem obviously. Furthermore, the other pathways associated with saccharide and steroid were worth observing too.

Compared with SNS, 74 genes were enriched in "ribosome" pathway which occupied the largest number in upregulated genes of SGS. The remaining three pathways enriched were "oxidative phosphorylation", "cutin, suberin and wax biosynthesis" and "glyoxylate and dicarboxylate metabolism" respectively.

Analysis of transcription factors in DEGs

Transcription factors play important roles in plant development in response to biotic or abiotic stimulus through regulating gene expression of abundant genes (Singh et al. 2002). To explore the expression of transcription factors, the transcription factors expressed differentially in leaves or stems of samples were screened as shown in **Fig. 5F**. There were 46 transcription factors expressed differentially in leaf samples and the number in stem samples was 18. In total transcription factors, *bHLH* took up the largest number, and they are important in phytochrome signaling according to the previous studies (Duek & Fankhauser 2005). With respect to other transcription factor family genes, they also played roles in the formation of the phenotype of our samples. For example, the HB family gene *HAT5* (*ATHB1*) could emit its effect in the conversion of palisade to spongy tissue cells in the leaves of *Arabidopsis thaliana* (Aoyama et al. 1995). In addition, the G2-like gene *GLK1* was another important transcription factor which could regulate the chloroplast development (Gang et al. 2019a) involved in the leaf etiolation (Xie et al. 2018). Additionally, since suberin biosynthetic process in SGS could be in connection with the formation of the phenotypes in **Fig. 4** because wood fiber tissue accounted for the majority of stem and cutin and suberin play important roles in the formation of wood (Kolattukudy 2001), some *NAC* and *MYB* family genes may be active in the regulation of those process (Zhong & Ye 2007).

Pathogenesis-related (PR) genes significantly upregulated in golden-vein leaves and stems

The yellow vein symptom of leaves is widely distributed in the plants infected with pathogens. For instance, the genus *Begomovirus* is the largest geminivirus genus containing more than 200 species or members (Fauquet et al. 2003; Fauquet et al. 2008) of which some viruses have infected a lot of plants with yellow vein symptoms in south China (Jiao et al. 2013) which is the place we got our samples from. Dissimilar to the common phenotypes of the infected showing only yellow vein symptom or accompanied with fading, LGS showed a deeper green beside the yellow vein symptom compared to LNS. In consideration of the enrichment of "defense response" and "response to biotic stimulus" especially the latter in LGS, most of the genes in "response to biotic stimulus" belonged to pathogenesis-related (PR) gene family (**Table 1**) which

was associated to infection and defense (van Loon et al. 2006). Using the annotation information, 25 PR family genes expressed differentially in leaves or stems were screened. In those genes, 19 genes were upregulated while one gene was downregulated in LGS compared to LNS as well as five upregulated and two downregulated genes in SGS compared to SNS (**Fig. 6A**; **Table S8**). Meanwhile, as shown in **Fig. 6B**, most of the genes distributed in the third quadrant so that it was clear that most of the PR genes were upregulated in LGS and SGS. Using cluster analysis, 7 genes (*PRB1*, *PR-1, E137*, *PER3*, *PER47*, *STH-2*, *NLTP*) were grouped under one cluster in which the gene expressions were both higher in LGS and SGS than their corresponding groups (**Fig. 6C**). The upregulation of those genes in same expression pattern may be in connection with the infection and the formation of the phenotypes in our samples.

Quantitative real-time PCR verification of transcriptome data

The expression of those genes in different tissues were identified via qRT-PCR (**Figs. 6E** and **6F**). Based on the enrichment data, those 23 selected genes used in qRT-PCR referred to photosynthesis, cell wall constituent, and some about secondary metabolism (**Fig. 6D**). Our data revealed that most of the gene expression data in qRT-PCR matched well with transcriptome results (**Fig. S5**), which verified that photosynthesis related pathways were involved in the color variation phenotypes of *A. sinensis*.

Discussion

A lot of colored substances are natural colorants showing different colors such as chlorophyll in green, curcuminoid in yellow, carotenoid in yellow to orange to red (Sigurdson et al. 2017). In photosynthesis, the sun-light is captured by chromophores the light-absorbing chemical structure (Mirkovic et al. 2017). Chlorophyll, produced from chloroplast which is important in plant development and plant defense (Lu & Yao 2018), could determine the color of plant tissues. When the chlorophyll is broken down, the color of plant tissues would change accompanied with the metabolites showing green, red, or other colors even colorless (Hortensteiner & Krautler 2011). Now that golden-vein-leaf sample of *A. sinensis* in this study showing a deeper green is one of the strongest features, which suggested that the genes and pathways related to photosynthesis should be important. As expected, a lot of GO terms and pathways about photosystem were enriched successfully in LGS, among which the upregulated pathways, including “photosynthesis - antenna proteins”, “photosynthesis” and “porphyrin and chlorophyll metabolism”, should be important to the formation of the dark green leaves in GS (**Figs. S6A**, **S6B** and **S6C**). In the study about birch, the genes related to photosynthesis downregulated in yellow-green leaf whose chlorophyll was less than the normal wild birch, and the parameters including the thicknesses of adaxial epidermis, palisade parenchyma, spongy parenchyma of them were different (Gang et al. 2019b). Similar to the condition in the study of birch, our samples also exhibited the anatomical difference in leaves, and it might involve in photosynthesis. Now that transcription factor is a kind of crucial regulator participating in the regulation of gene expression, differentially expressed transcription factors in different leaf

sample groups were screened. Specially, a G2-like gene *GLK1* was upregulated significantly in LGS (**Fig. 6F**). *GLKs* including *GLK1* and *GLK2* are important transcription factors in leaf growth such as chloroplast development (Powell et al. 2012) and leaf senescence (Rauf et al. 2013). A research in *Arabidopsis thaliana* has revealed that the induction of *GLK1* could increase the expression of the photosystem and chlorophyll biosynthesis related genes, including *LHCB4*, *PSBQ*, *CHLH*, *CHLM* (Waters et al. 2009), all of which also upregulated in our LGS group (**Figs. S6A**, **S6B** and **S6C**). Besides the herbaceous plants, the studies on *GLK* were also conducted in woody plants showing similar results (Gang et al. 2019a). In addition, *GLK* genes show positive effects in pathogen resistance in plants (Han et al. 2016). A study about *Fusarium graminearum* suggests the *Arabidopsis* with *GLK1* overexpression could encourage the expression of some pathogenesis-related genes and positive in the resistance of fungal pathogen (Savitch et al. 2007). According to the properties of *GLK1*, it may be a crucial upstream regulator in the regulation of photosystem and chlorophyll and then leaf-colors.

Chlorophyll breakdown is as important as its synthesis, and the chlorophyll catabolites could be effective in some aspects such as antioxidants (Muller et al. 2007) or internal signals sources (Mur et al. 2010). Genes upregulated in LGS enriched in the chlorophyll metabolism related pathway and in which the gene like *CLH1* (chlorophyllase) is also active in other chlorophyll breakdown pathway (Hortensteiner & Krautler 2011). Another obvious features in LGS samples are yellow veins. It was a common feature in plants infected with pathogens especially virus like the yellow vein symptoms in *Oxalis debilis* infected with *begomovirus* (Herrera et al. 2015). Regarding the pictures from transmission electron microscope, we observed some particles (**Figs. 3C** and **3D**) suspected as virus because of their shape and the structure of vesicles which is common in the structures beside the virus (Grangeon et al. 2012; Grangeon et al. 2013; Lerich et al. 2011). In consideration of the GO enrichment results showing the hints of biotic stimulus and most of the PR genes which are closely related to pathogens defense (van Loon et al. 2006) upregulated in LGS, as well as the particle suspected as virus, it is assumed that external factors causing the series of changes could be some kind of pathogens which may be virus.

Stem is the fundamental source part of *A. sinensis* to acquire agarwood resin the precious product (Liu et al. 2018). As shown in **Fig. 4**, the most obvious features in the transverse surface of SGS compared to SNS were the denser wood fiber cell tissue and the thicker cell wall. According to previous studies, cell wall plays an important role in disease resistance (Underwood 2012; Molina et al. 2021), it may explain the phenotype of SGS. In the outer layer, cell wall is covered with pectin and cuticle, and there are suberized wall, plasma membrane, cytoplasm, vacuole from outside of cell wall to inside respectively (Kolattukudy 2001). According to previous researches, some genes from *NAC* or *MYB* families, which were upregulated in SGS (**Fig. 5F**), have shown their abilities in the regulation of plant cell wall development (Zhong & Ye 2007). Genes upregulated in SGS were mainly enriched in suberin biosynthetic process, and cutin and wax biosynthesis pathways, which suggested that the formation of suberin might change in SGS

(**Fig. S6D**). Taking up half of the above upregulated genes, *CYP86* family genes exhibit important functions in cutin and suberin and biosynthesis (Pinot & Beisson 2011) in which suberin plays a role in respond to physical, chemical or biological stresses as barriers (Kolattukudy 2001). In consideration of the transcription factors screened, *NAC073* (*SND2*) and *NAC075* are related to the regulation of cell wall according to the previous studies (Sakamoto & Mitsuda 2015; Zhong et al. 2008). The phenotype of the *SND2* overexpression in *A. thaliana* showed that *SND2* could increase the secondary wall thickening in fiber cells (Zhong et al. 2008) which was similar to the photomicrograph of SGS (**Figs. 4B** and **4F**). In addition, the components of the cell walls would even change under the regulation of these *NAC* genes including *SND2* and *NAC075* (Sakamoto & Mitsuda 2015). Therefore, the irregular shapes of include phloem may also be created by the change of components of cell wall leading to the variation of its supporting ability. However, another research claimed that the secondary cell wall thickness of *Arabidopsis* would reduce with the overexpression of *SND2* (Hussey et al. 2011). Thus, the roles of transcription factors are still complex to research.

The development of most plants infected with pathogens showing yellow vein symptoms would be influenced and sometimes the plants would die finally. There have been many studies reported that the plants infected with yellow vein related pathogens would grow accompanied with a series of changes like the damage to photosystem in leaves (Palanisamy et al. 2009), necrotic lesions and veinal necrosis (Ravelo et al. 2007). However, if the infected plants can still grow, they could become new sources like ornamental plants (Valverde et al. 2012). According to our observation, the trees would not die but grow slower than the normals, so that they could be a new resource applied in breeding. For the moment, the grafting of the trees with unique phenotype is being carried out.

In consideration of the defense response related genes and the pathways about photosynthesis, the pathogen which is responsible for these symptoms may enhance photosystem and promote the metabolism of chlorophyll in tissue specificity. Furthermore, the gene expression variations of stems especially a large number of genes about translation and ribosome suggest that changes in stems are significant and complex. In addition, the grease-like granular substances in the pith in SGS may also hint that the factor causing the symptoms could be potential as a tool to explore the method promoting the formation of agarwood. In the next step, the pathogen, no matter virus, bacterium, fungus or others, should be worthy of isolation and identification, which could be a useful tool to investigate the phenytype formation of *A. sinensis* and expand the potential economic values in the future.

Conclusions

Anatomical comparison between golden-vein-leaf sample and normal-vein-leaf sample was conducted in this study, revealing the differences at the microscopic level. In the observation of ultrastructure, the chloroplast was more complete in the midrib of LNS than that of LGS, and

some particles suspected as virus were observed. Meanwhile, transcriptome sequencing was performed to analyze the differences of those samples at the molecular level. The genes about photosynthesis and its regulation such as *LHCB4*, *PSBQ*, *CHLH*, *CHLM* and *GLK1* should be important to the color variation of leaves in *A. sinensis* samples. Analyzing the pathogenesis-related genes in DEGs of all samples, the internal factor leading to the changes might be pathogen infection and 7 genes (*PRB1*, *PR-1, E137*, *PER3*, *PER47*, *STH-2*, *NLTP*) which were clustered together could be important to it. In summary, our research discussed the internal variation of *A. sinensis* with different phenotypes, and laid the foundation for the studies of the unique phenotype.

Conflict of Interest

Ou Huang is employed by Guangdong Shangzhengtang Group Co., Ltd. The remaining authors declare that they have no competing interests.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Public Welfare Research Institutes (ZZ13-YQ-093-C1, ZZ13-YQ-095, ZZXT202112), and Funds for Basic Resources Investigation Research of the Ministry of Science and Technology (2018FY100800).

References

Ahmed EU, Hayashi T, and Yazawa S. 2004. Leaf color stability during plant development as an index of leaf color variation among micropropagated *Caladium*. *HortScience* 39(2):328-332. DOI 10.21273/hortsci.39.2.328

Aoyama T, Dong CH, Wu Y, Carabelli M, Sessa G, Ruberti I, Morelli G, and Chua NH. 1995. Ectopic expression of the *Arabidopsis* transcriptional activator *Athb-1* alters leaf cell fate in tobacco. *Plant Cell* 7(11):1773-1785. DOI 10.1105/tpc.7.11.1773

Atkin OK, Millar AH, Gardeström P, and Day DA. 2000. Photosynthesis, carbohydrate metabolism and respiration in leaves of higher plants. In: Leegood RC, Sharkey TD, and von Caemmerer S, eds. *Photosynthesis: Physiology and Metabolism*. Dordrecht: Springer Netherlands, 9:153-175. DOI 10.1007/0-306-48137-5_7

Barazani O, Erez T, Ogran A, Hanin N, Barzilai M, Dag A, and Shafir S. 2019. Natural variation in flower color and scent in populations of *Eruca sativa* (Brassicaceae) affects pollination behavior of honey bees. *Journal of Insect Science* 19(3):6. DOI 10.1093/jisesa/iez038

Brand A, Borovsky Y, Hill T, Rahman KAA, Bellalou A, Van Deynze A, and Paran I. 2014. *CaGLK2* regulates natural variation of chlorophyll content and fruit color in pepper fruit. *Theoretical and Applied Genetics* 127(10):2139-2148. DOI 10.1007/s00122-014-2367-y

Buchfink B, Xie C, and Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12(1):59-60. DOI 10.1038/nmeth.3176

518 Chattopadhyay T, Hazra P, Akhtar S, Maurya D, Mukherjee A, and Roy S. 2021. Skin colour,
519 carotenogenesis and chlorophyll degradation mutant alleles: genetic orchestration behind
520 the fruit colour variation in tomato. *Plant Cell Reports* 40:767-782. DOI 10.1007/s00299-
521 020-02650-9
522 Cheng GX, Li RJ, Wang M, Huang LJ, Khan A, Ali M, and Gong ZH. 2018. Variation in leaf
523 color and combine effect of pigments on physiology and resistance to whitefly of pepper
524 (*Capsicum annuum* L.). *Scientia Horticulturae* 229:215-225. DOI
525 10.1016/j.scienta.2017.11.014
526 Committee CP. 2020. *Pharmacopoeia of People's Republic of China*. Beijing: China Medical
527 Science Press.
528 Duek PD, and Fankhauser C. 2005. *bHLH* class transcription factors take centre stage in
529 phytochrome signalling. *Trends in Plant Science* 10(2):51-54. 10.1016/j.tplants.2004.12.005
530 Emenecker RJ, and Strader LC. 2020. Auxin-abscisic acid interactions in plant growth and
531 development. *Biomolecules* 10(2):281. DOI 10.3390/biom10020281
532 Ernst J, Nau GJ, and Bar-Joseph Z. 2005. Clustering short time series gene expression data.
533 *Bioinformatics* 21(Suppl 1):i159-168. DOI 10.1093/bioinformatics/bti1022
534 Fauquet CM, Bisaro DM, Briddon RW, Brown JK, Harrison BD, Rybicki EP, Stenger DC, and
535 Stanley J. 2003. Revision of taxonomic criteria for species demarcation in the family
536 *Geminiviridae*, and an updated list of begomovirus species. *Archives of Virology*
537 148(2):405-421. DOI 10.1007/s00705-002-0957-5
538 Fauquet CM, Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini M, and Zhou X. 2008.
539 Geminivirus strain demarcation and nomenclature. *Archives of Virology* 153(4):783-821.
540 DOI 10.1007/s00705-008-0037-6
541 Gang H, Li R, Zhao Y, Liu G, Chen S, and Jiang J. 2019a. Loss of *GLK1* transcription factor
542 function reveals new insights in chlorophyll biosynthesis and chloroplast development.
543 *Journal of Experimental Botany* 70(12):3125-3138. DOI 10.1093/jxb/erz128
544 Gang H, Liu G, Chen S, and Jiang J. 2019b. Physiological and transcriptome analysis of a
545 yellow-green leaf mutant in Birch (*Betula platyphylla* × *B. Pendula*). *Forests* 10(2):120.
546 DOI 10.3390/f10020120
547 Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M,
548 Dopazo J, and Conesa A. 2008. High-throughput functional annotation and data mining
549 with the Blast2GO suite. *Nucleic Acids Research* 36(10):3420-3435. DOI
550 10.1093/nar/gkn176
551 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
552 Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F,
553 Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, and Regev A. 2011. Full-length
554 transcriptome assembly from RNA-Seq data without a reference genome. *Nature*
555 *Biotechnology* 29(7):644-652. DOI 10.1038/nbt.1883

Grangeon R, Agbeci M, Chen J, Grondin G, Zheng H, and Laliberte JF. 2012. Impact on the endoplasmic reticulum and Golgi apparatus of turnip mosaic virus infection. *Journal of Virology* 86(17):9255-9265. DOI 10.1128/JVI.01146-12

Grangeon R, Jiang J, Wan J, Agbeci M, Zheng H, and Laliberte JF. 2013. 6K2-induced vesicles can move cell to cell during turnip mosaic virus infection. *Front Microbiol* 4:351. DOI 10.3389/fmicb.2013.00351

Han XY, Li PX, Zou LJ, Tan WR, Zheng T, Zhang DW, and Lin HH. 2016. *GOLDEN2-LIKE* transcription factors coordinate the tolerance to *Cucumber mosaic virus* in *Arabidopsis*. *Biochemical and Biophysical Research Communications* 477(4):626-632. DOI 10.1016/j.bbrc.2016.06.110

He QL, Shen Y, Wang MX, Huang MR, Yang RZ, Zhu SJ, Wang LS, Xu YJ, and Wu RL. 2011. Natural variation in petal color in *Lycoris longituba* revealed by anthocyanin components. *PloS One* 6(8):e22098 DOI 10.1371/journal.pone.0022098

Herrera F, Aboughanem-Sabanadzovic N, and Valverde RA. 2015. A begomovirus associated with yellow vein symptoms of *Oxalis debilis*. *European Journal of Plant Pathology* 142(1):203-208. DOI 10.1007/s10658-015-0594-y

Hortensteiner S, and Krautler B. 2011. Chlorophyll breakdown in higher plants. *Biochimica Et Biophysica Acta-Bioenergetics* 1807(8):977-988. DOI 10.1016/j.bbabio.2010.12.007

Hussey SG, Mizrachi E, Spokevicius AV, Bossinger G, Berger DK, and Myburg AA. 2011. *SND2*, a *NAC* transcription factor gene, regulates genes involved in secondary cell wall development in *Arabidopsis* fibres and increases fibre cell area in *Eucalyptus*. *BMC Plant Biology* 11(1):1-17. DOI 10.1186/1471-2229-11-173

Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, and Khan MIR. 2017. Ethylene role in plant growth, development and senescence: interaction with other phytohormones. *Frontiers in Plant Science* 8:475. DOI 10.3389/fpls.2017.00475

Jiao L, Yin Y, Cheng Y, and Jiang X. 2014. DNA barcoding for identification of the endangered species *Aquilaria sinensis*: comparison of data from heated or aged wood samples. *Holzforschung* 68(3):487-494. DOI 10.1515/hf-2013-0129

Jiao X, Gong H, Liu X, Xie Y, and Zhou X. 2013. Etiology of ageratum yellow vein diseases in South China. *Plant Disease* 97(11):1497-1503. DOI 10.1094/pdis-01-13-0074-re

Khandagale K, and Gawande S. 2019. Genetics of bulb colour variation and flavonoids in onion. *Journal of Horticultural Science and Biotechnology* 94(4):522-532. DOI 10.1080/14620316.2018.1543558

Kolattukudy PE. 2001. Polyesters in higher plants. In: Babel W, and Steinbüchel A, eds. *Biopolyesters.* Berlin, Heidelberg: Springer Berlin Heidelberg, 7:1-49. DOI 10.1007/3-540-40021-4_1

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. 35(6):1547-1549. DOI 10.1093/molbev/msy096

Lachagari VBR, Gupta R, Lekkala SP, Mahadevan L, Kuriakose B, Chakravartty N, Katta A, Santhosh S, Reddy AR, and Thomas G. 2019. Whole genome sequencing and comparative genomic analysis reveal allelic variations unique to a purple colored rice landrace (*Oryza sativa ssp. indica* cv. *Purpleputtu*). *Frontiers in Plant Science* 10:513. DOI 10.3389/fpls.2019.00513

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, and Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948. DOI 10.1093/bioinformatics/btm404

Lee SY, Ng WL, Mahat MN, Nazre M, and Mohamed R. 2016. DNA barcoding of the endangered *Aquilaria* (Thymelaeaceae) and its application in species authentication of agarwood products traded in the market. *PloS One* 11(4):e0154631. DOI 10.1371/journal.pone.0154631

Lerich A, Langhans M, Sturm S, and Robinson DG. 2011. Is the 6 kDa tobacco etch viral protein a bona fide ERES marker? *Journal of Experimental Botany* 62(14):5013-5023. DOI 10.1093/jxb/err200

Li W, Tang S, Zhang S, Shan JG, Tang CJ, Chen QN, Jia GQ, Han YH, Zhi H, and Diao XM. 2016. Gene mapping and functional analysis of the novel leaf color gene *SiYGL1* in foxtail millet *Setaria italica* (L.) P. Beauv. *Physiologia Plantarum* 157(1):24-37. DOI 10.1111/ppl.12405

Li YC, Chen CS, Li YS, Ding ZT, Shen JZ, Wang Y, Zhao L, and Xu M. 2016. The identification and evaluation of two different color variations of tea. *Journal of the Science of Food and Agriculture* 96(15):4951-4961. DOI 10.1002/jsfa.7897

Liu P, Zhang X, Yang Y, Sui C, Xu Y, and Wei J. 2018. Interxylary phloem and xylem rays are the structural foundation of agarwood resin formation in the stems of *Aquilaria sinensis*. *Trees* 33(2):533-542. DOI 10.1007/s00468-018-1799-4

Livak KJ, and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402-408. DOI 10.1006/meth.2001.1262

Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12):1-21. 10.1186/s13059-014-0550-8

Lu Y, and Yao J. 2018. Chloroplasts at the crossroad of photosynthesis, pathogen infection and plant defense. *International Journal of Molecular Sciences* 19(12):37. DOI 10.3390/ijms19123900

Mirkovic T, Ostroumov EE, Anna JM, van Grondelle R, Govindjee, and Scholes GD. 2017. Light absorption and energy transfer in the antenna complexes of photosynthetic organisms. *Chemical Reviews* 117(2):249-293. DOI 10.1021/acs.chemrev.6b00002

Molina A, Miedes E, Bacete L, Rodriguez T, Melida H, Denance N, Sanchez-Vallet A, Riviere MP, Lopez G, Freydier A, Barlet X, Pattathil S, Hahn M, and Goffner D. 2021. *Arabidopsis*

cell wall composition determines disease resistance specificity and fitness. *Proceedings of the National Academy of Sciences* 118(5). DOI 10.1073/pnas.2010243118

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, and Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35 (suppl_2):W182-W185. DOI 10.1093/nar/gkm321

Muller T, Ulrich M, Ongania KH, and Krautler B. 2007. Colorless tetrapyrrolic chlorophyll catabolites found in ripening fruit are effective antioxidants. *Angewandte Chemie International Edition* 46(45):8699-8702. DOI 10.1002/anie.200703587

Mur LAJ, Aubry S, Mondhe M, Kingston-Smith A, Gallagher J, Timms-Taravella E, James C, Papp I, Hortensteiner S, Thomas H, and Ougham H. 2010. Accumulation of chlorophyll catabolites photosensitizes the hypersensitive response elicited by *Pseudomonas syringae* in Arabidopsis. *New Phytologist* 188(1):161-174. DOI 10.1111/j.1469-8137.2010.03377.x

Palanisamy P, Michael PI, and Krishnaswamy M. 2009. Physiological response of yellow vein mosaic virus-infected bhendi [*Abelmoschus esculentus*] leaves. *Physiological and Molecular Plant Pathology* 74(2):129-133. DOI 10.1016/j.pmpp.2009.10.003

Pinot F, and Beisson F. 2011. Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. *FEBS Journal* 278(2):195-205. DOI 10.1111/j.1742-4658.2010.07948.x

Powell ALT, Nguyen CV, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, Ashrafi H, Pons C, Fernandez-Munoz R, Vicente A, Lopez-Baltazar J, Barry CS, Liu YS, Chetelat R, Granell A, Van Deynze A, Giovannoni JJ, and Bennett AB. 2012. Uniform ripening Encodes a *Golden 2-like* Transcription Factor Regulating Tomato Fruit Chloroplast Development. *Science* 336(6089):1711-1715. DOI 10.1126/science.1222218

Rauf M, Arif M, Dortay H, Matallana-Ramirez LP, Waters MT, Gil Nam H, Lim PO, Mueller-Roeber B, and Balazadeh S. 2013. *ORE1* balances leaf senescence against maintenance by antagonizing G2-like-mediated transcription. *EMBO Reports* 14(4):382-388. DOI 10.1038/embor.2013.24

Ravelo G, Kagaya U, Inukai T, Sato M, and Uyeda I. 2007. Genetic analysis of lethal tip necrosis induced by clover yellow vein virus infection in pea. *Journal of General Plant Pathology* 73(1):59-65. DOI 10.1007/s10327-006-0324-6

Sakamoto S, and Mitsuda N. 2015. Reconstitution of a secondary cell wall in a secondary cell wall-deficient Arabidopsis mutant. *Plant and Cell Physiology* 56(2):299-310. DOI 10.1093/pcp/pcu208

Sartory DP, and Grobbelaar JU. 1984. Extraction of chlorophyll a from freshwater phytoplankton for spectrophotometric analysis. *Hydrobiologia* 114(3):177-187. DOI 10.1007/BF00031869

Savitch LV, Subramaniam R, Allard GC, and Singh J. 2007. The *GLK1* 'regulon' encodes disease defense related proteins and confers resistance to *Fusarium graminearum* in Arabidopsis. *Biochem Biophys Res Commun* 359(2):234-238. DOI 10.1016/j.bbrc.2007.05.084

673 Sigurdson GT, Tang PP, and Giusti MM. 2017. Natural colorants: food colorants from natural
674 sources. In: Doyle MP, and Klaenhammer TR, eds. *Annual Review of Food Science and*
675 *Technology,* Palo Alto: Annual Reviews, 8:261-280. DOI 10.1146/annurev-food-030216-
676 025923
677 Singh KB, Foley RC, and Onate-Sanchez L. 2002. Transcription factors in plant defense and
678 stress responses. *Current Opinion in Plant Biology* 5(5):430-436. DOI 10.1016/s1369-
679 5266(02)00289-3
680 Takemoto H, Ito M, Shiraki T, Yagura T, and Honda G. 2008. Sedative effects of vapor
681 inhalation of agarwood oil and spikenard extract and identification of their active
682 components. *Journal of Natural Medicines* 62(1):41-46. DOI 10.1007/s11418-007-0177-0
683 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold
684 BJ, and Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals
685 unannotated transcripts and isoform switching during cell differentiation. *Nature*
686 *Biotechnology* 28(5):511-U174. DOI 10.1038/nbt.1621
687 Ullah MS, and Lim UT. 2016. Within-greenhouse and within-plant distribution of greenhouse
688 whitefly, *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae), in strawberry greenhouses.
689 *Applied Entomology and Zoology* 51(2):333-339. DOI 10.1007/s13355-016-0394-7
690 Underwood W. 2012. The plant cell wall: a dynamic barrier against pathogen invasion. *Frontiers*
691 *in Plant Science* 3:85. DOI 10.3389/fpls.2012.00085
692 Valverde RA, Sabanadzovic S, and Hammond J. 2012. Viruses that enhance the aesthetics of
693 some ornamental plants: beauty or beast? *Plant Disease* 96(5):600-611. DOI 10.1094/pdis-
694 11-11-0928-fe
695 van Loon LC, Rep M, and Pieterse CM. 2006. Significance of inducible defense-related proteins
696 in infected plants. *Annual Review of Phytopathology* 44:135-162. DOI
697 10.1146/annurev.phyto.44.070505.143425
698 Wang Y, Yang Z-Y, Tian Y-P, Geng C, Yuan X-F, and Li X-D. 2019. Role of *Tobacco vein*
699 *banding mosaic virus* 3′-UTR on virus systemic infection in tobacco. *Virology* 527:38-46.
700 DOI 10.1016/j.virol.2018.11.001
701 Wang YQ, Song FH, Zhu JW, Zhang SS, Yang YD, Chen TT, Tang BX, Dong LL, Ding N,
702 Zhang Q, Bai ZX, Dong XN, Chen HX, Sun MY, Zhai S, Sun YB, Yu L, Lan L, Xiao JF,
703 Fang XD, Lei HX, Zhang Z, and Zhao WM. 2017. GSA: genome sequence archive.
704 *Genomics Proteomics & Bioinformatics* 15(1):14-18. DOI 10.1016/j.gpb.2017.01.001
705 Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, and Langdale JA. 2009. *GLK*
706 transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis.
707 *The Plant Cell* 21(4):1109-1128. DOI 10.1105/tpc.108.065250
708 Wu H, Shi N, An X, Liu C, Fu H, Cao L, Feng Y, Sun D, and Zhang L. 2018. Candidate genes
709 for yellow leaf color in common wheat (*Triticum aestivum* L.) and major related metabolic
710 pathways according to rranscriptome profiling. *International Journal of Molecular Sciences*
711 19(6):1594. DOI 10.3390/ijms19061594

712 Xia XW, Gui RY, Yang HY, Fu Y, Wei F, and Zhou MB. 2015. Identification of genes involved
713 in color variation of bamboo culms by suppression subtractive hybridization. *Plant*
714 *Physiology and Biochemistry* 97:156-164. DOI 10.1016/j.plaphy.2015.10.004
715 Xie F, Yuan JL, Li YX, Wang CJ, Tang HY, Xia JH, Yang QY, and Wan ZJ. 2018.
716 Transcriptome analysis reveals candidate genes associated with leaf etiolation of a
717 cytoplasmic male sterility line in Chinese cabbage (*Brassica Rapa* L. ssp. *Pekinensis*).
718 *International Journal of Molecular Sciences* 19(4):922. DOI 10.3390/ijms19040922
719 Xu Y-H, Liao Y-C, Zhang Z, Liu J, Sun P-W, Gao Z-H, Sui C, and Wei J-H. 2016. Jasmonic
720 acid is a crucial signal transducer in heat shock induced sesquiterpene formation in
721 *Aquilaria sinensis*. *Scientific Reports* 6(1):1-9. DOI 10.1038/srep21843
722 Yang K, Jeong N, Moon JK, Lee YH, Lee SH, Kim HM, Hwang CH, Back K, Palmer RG, and
723 Jeong SC. 2010. Genetic Analysis of genes controlling natural variation of seed coat and
724 flower colors in soybean. *Journal of Heredity* 101(6):757-768. DOI 10.1093/jhered/esq078
725 Yu GC, Wang LG, Han YY, and He QY. 2012. clusterProfiler: an R package for comparing
726 biological themes among gene clusters. *Omics: a Journal of Integrative Biology* 16(5):284-
727 287. DOI 10.1089/omi.2011.0118
728 Zhong R, Lee C, Zhou J, McCarthy RL, and Ye ZH. 2008. A battery of transcription factors
729 involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell*
730 20(10):2763-2782. DOI 10.1105/tpc.108.061325
731 Zhong R, and Ye ZH. 2007. Regulation of cell wall biosynthesis. *Curr Opin Plant Biol*
732 10(6):564-572. DOI 10.1016/j.pbi.2007.09.001
733 Zhu GF, Yang FX, Shi SS, Li DM, Wang Z, Liu HL, Huang D, and Wang CY. 2015.
734 Transcriptome characterization of *Cymbidium sinense* 'Dharma' using 454 pyrosequencing
735 and its application in the identification of genes associated with leaf color variation. *PloS*
736 *One* 10(6):e0128592. DOI 10.1371/journal.pone.0128592

Table 1(on next page)

Genes enriched in GO term “response to biotic stimulus”

Negative Log_2FC represents the upregulated genes in LGS while the positive value represents the downregulated genes. High absolute value of Log_2FC represents a high fold change value. FDR < 0.05 suggests the significant statistical difference in DEGs estimation.

Table 1. Genes enriched in GO term "response to biotic stimulus"

Gene symbol	Description	Log_2FC	FDR
GNS1	Glucan endo-1,3-beta-glucosidase, basic isoform	-9.8857	<0.0001
E13B	Glucan endo-1,3-beta-glucosidase	-6.9149	<0.0001
PRB1	Basic form of pathogenesis-related protein 1	-5.2072	0.0030
MLP328	MLP-like protein 328	-4.9729	0.0008
STH-2	Pathogenesis-related protein STH-2	-4.4692	0.0027
PRR1	Pathogenesis-related protein R major form	-4.4122	0.0286

Note: Negative Log_2FC represents the upregulated genes in LGS while the positive value represents the downregulated genes. High absolute value of Log_2FC represents a high fold change value. FDR < 0.05 suggests the significant statistical difference in DEGs estimation.

Figure 1

Plant materials, content of chlorophyll, and the phylogenetic tree for origin identification.

(A) Golden-vein-leaf sample (GS); (B) Normal-vein-leaf sample (NS); (C) The content of chlorophyll of the leaves of GS and NS (LGS and LNS); (D) Phylogenetic tree of GS, NS, 8 closely related species, and one distantly related species *Gonystylus bancanus* based on DNA barcode ITS2+*trn*L-*trn*F. The gene accessions of sequences for each species used to construct the phylogenetic tree were listed in Table S1. The scale bar represents 0.01 nucleotide substitutions per site and the numbers next to the nodes are percentages of confidence from 1000 replicates. The numbers above the different nucleotides indicate aligned positions. Nucleotides specifically exist in outgroup species are not listed.

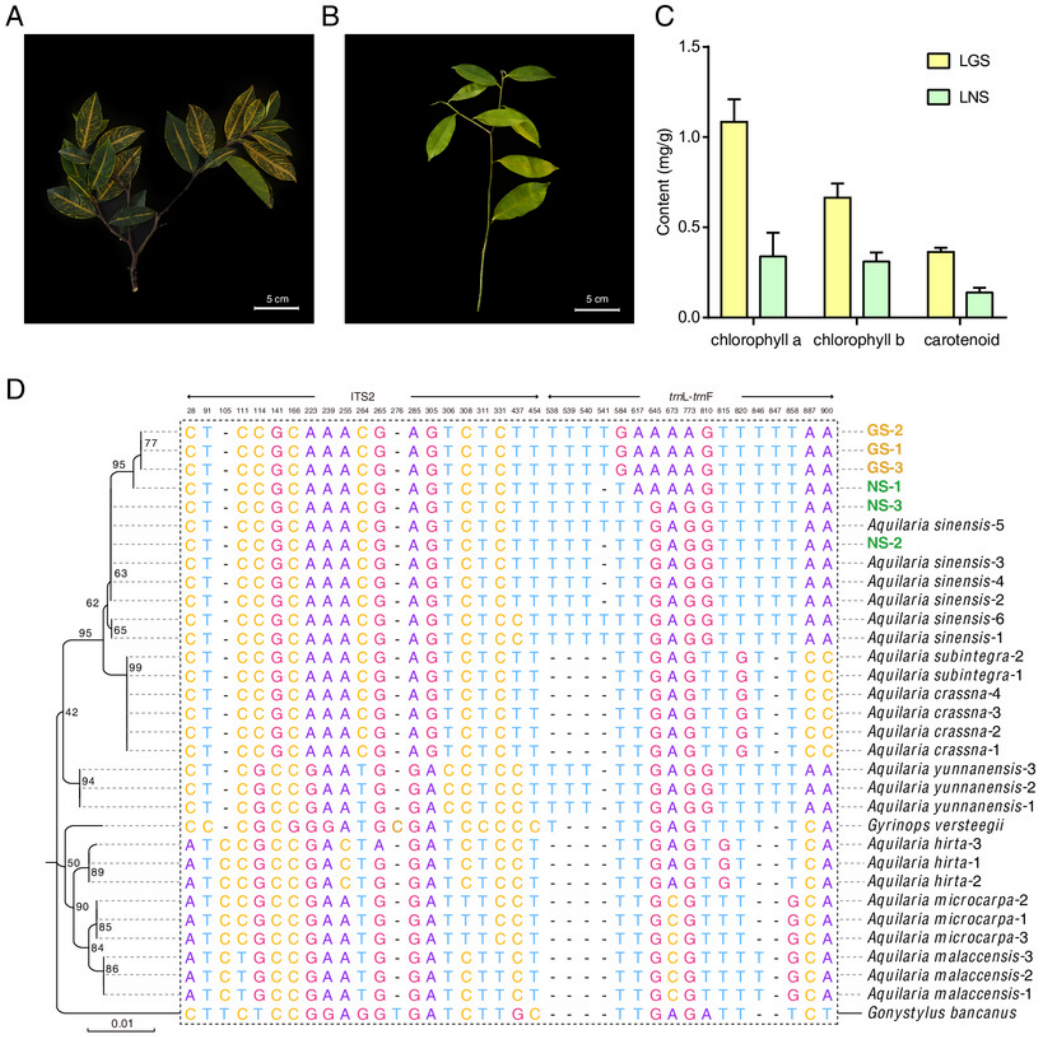

A
B
C
5 cm
5 cm
1.5
1.0
0.5
0.0
Content (mg/g)
LGS
LNS
chlorophyll a
chlorophyll b
carotenoid
D
ITS2
trnL-trnF
GS-2
GS-1
GS-3
NS-1
NS-3
Aquilaria sinensis-5
NS-2
Aquilaria sinensis-3
Aquilaria sinensis-4
Aquilaria sinensis-2
Aquilaria sinensis-6
Aquilaria sinensis-1
Aquilaria subintegra-2
Aquilaria subintegra-1
Aquilaria crassna-4
Aquilaria crassna-3
Aquilaria crassna-2
Aquilaria crassna-1
Aquilaria yunnanensis-3
Aquilaria yunnanensis-2
Aquilaria yunnanensis-1
Gyrinops versteegii
Aquilaria hirta-3
Aquilaria hirta-1
Aquilaria hirta-2
Aquilaria microcarpa-2
Aquilaria microcarpa-1
Aquilaria microcarpa-3
Aquilaria malaccensis-3
Aquilaria malaccensis-2
Aquilaria malaccensis-1
Gonystylus bancanus
0.01

Figure 2

Anatomical structures and statistics of main measured parameters of LGS and LNS.

(A) Transverse section of the midrib of LGS (left) and (B) zoomed parts of transverse section of LGS (right); (C) Transverse section of the midrib of LNS and (D) zoomed part of transverse section of LNS; (E) Boxplot of total area of IP in midrib; (F) Boxplot of PT thickness; (G) Boxplot of ST thickness; (H) Boxplot of LE thickness; (I) Boxplot of UE thickness. The mark "**" above two boxes indicates that the *P-value* is less than 0.0001 and the single mark "*" suggests that the *P-value* is less than 0.05. Those without asterisks mean no statistical significance. (IP: included phloem, UE: upper epidermis cell, LE: lower epidermis cell, Cr: crystal, PT: palisade tissue, ST: spongy tissue)

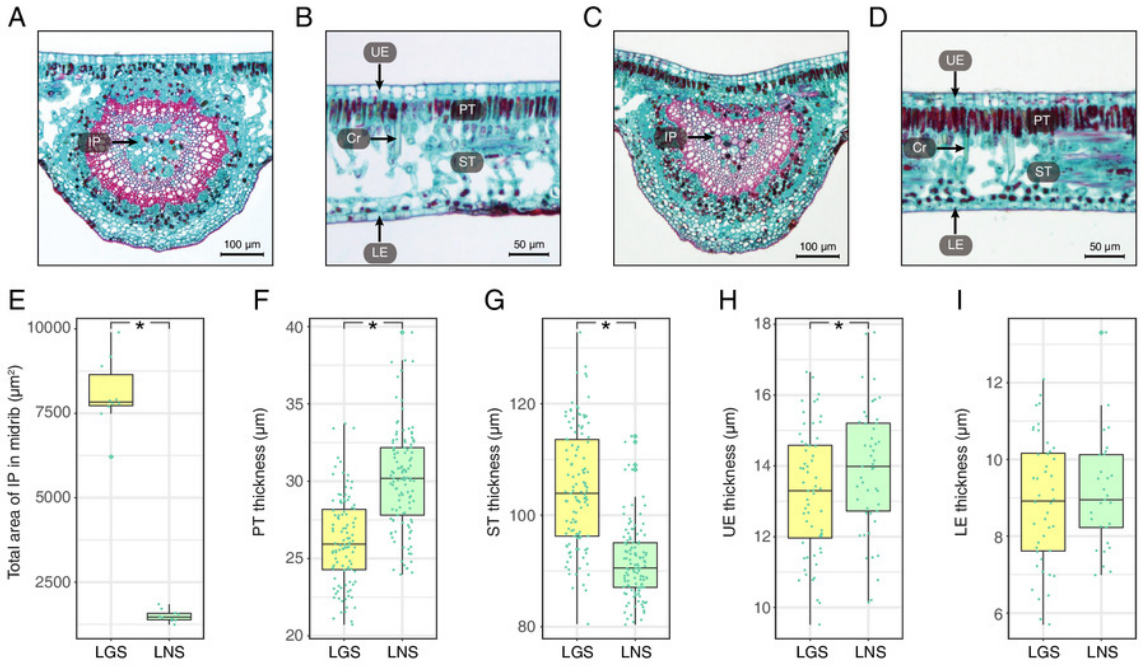


Figure 3

The ultrastructures of the midribs of LGS and LNS.

(A) The ultrastructure of the midrib of LNS; (B-D) The ultrastructure of the midrib of LGS. (Chl: chloroplast; M: mitochondria; Thy: thylakoid; VLP: virus-like particle)

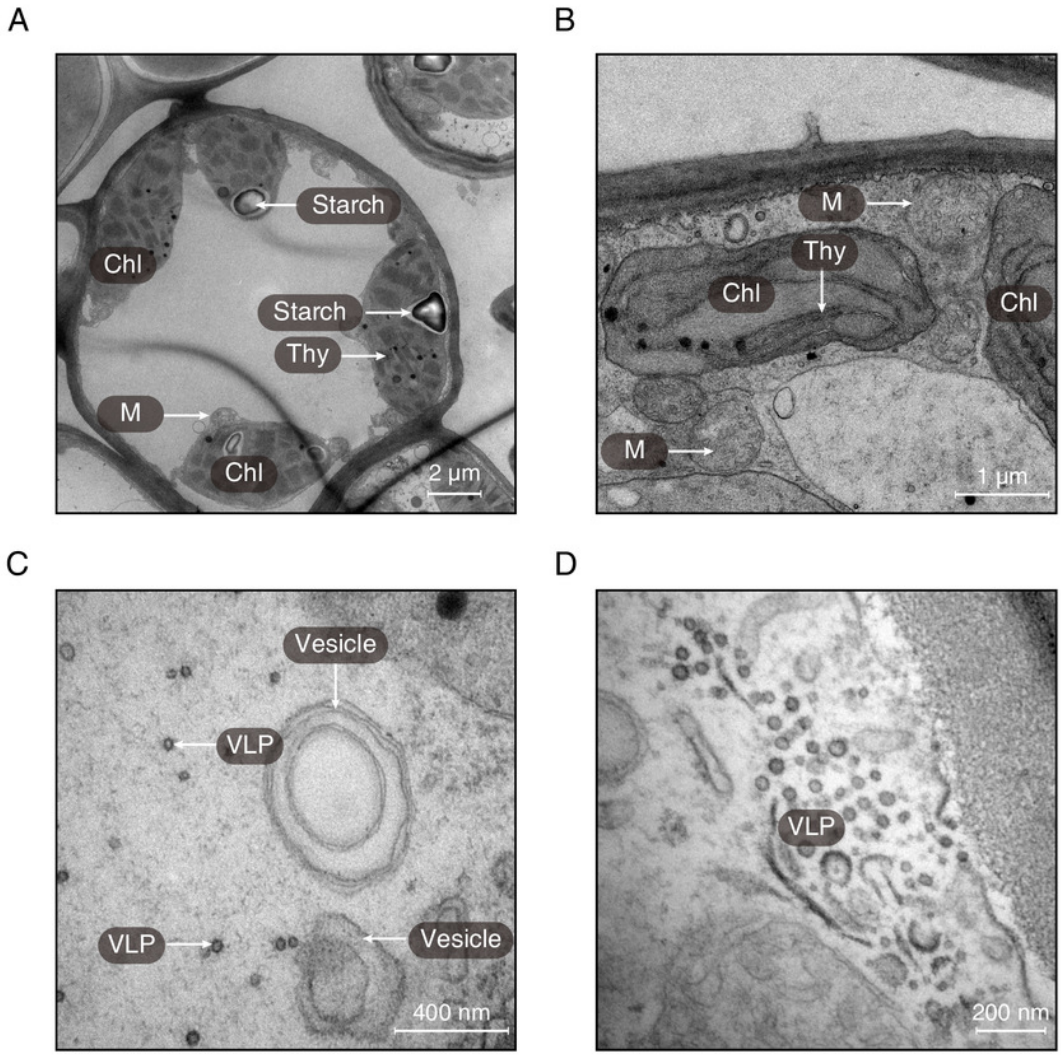

A
Starch
Chl
Starch
Thy
M
Chl
2 µm
B
M
Thy
Chl
Chl
M
1 µm
C
Vesicle
VLP
VLP
Vesicle
400 nm
D
VLP
200 nm

Figure 4

Anatomical structure and statistics of main measured parameters of SGS and SNS.

(A) Overview of SGS transverse section; (B) A part of transverse section of SGS; (C) Transverse section of pith of SGS; (D) Radial section of SGS; (E) Overview of SNS transverse section; (F) A part of transverse section of SNS; (G) Transverse section of pith of SNS; (H) Radial section of SNS; (I) Boxplot of single IP area; (J) Boxplot of percentage of IP in xylem; (K) Boxplot of single Ve area; (L) Boxplot of percentage of Ve to xylem; (M) Boxplot of single WF area. The mark "**" above two boxes indicates that the *P-value* is less than 0.0001 and the single mark "*" suggests that the *P-value* is less than 0.01 and larger than 0.001. (Pi: pith, IP: included phloem, WF: wood fiber cell, Ve: vessel, XR: xylem ray, Cr: crystal)

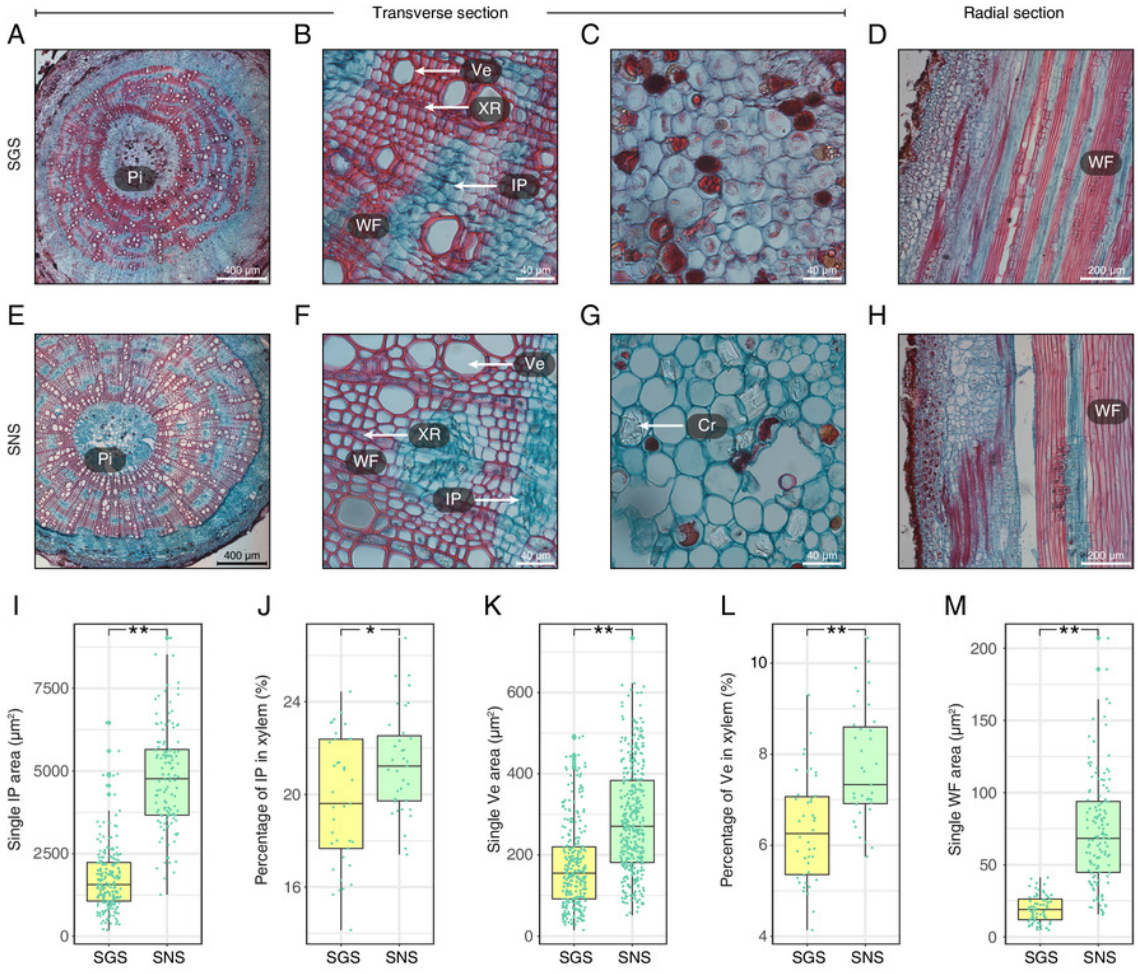

Transverse section
Radial section
A
B
C
D
SGS
Pi
Ve
XR
IP
WF
Cr
400 μm
40 μm
200 μm
E
F
G
H
SNS
I
J
K
L
M
Single IP area (μm²)
Percentage of IP in xylem (%)
Single Ve area (μm²)
Percentage of Ve in xylem (%)
Single WF area (μm²)
SGS
SNS

Figure 5

Distribution and analyses of DEGs.

(A) Volcano plot showing the distribution of DEGs between LGS and LNS; (B) Volcano plot showing the distribution of DEGs between SGS and SNS. Both in (A) and (B), the yellow and green dots in volcano plots indicate the DEGs with log2FC less than -1 or larger than 1 and FDR (false discovery rate) less than 0.05. The numbers above the volcano plots represent the number of upregulated genes for each group. (C) Gene Ontology enrichment analysis for DEGs upregulated in LGS compared to LNS; (D) Gene Ontology enrichment analysis for DEGs upregulated in SGS compared to SNS. The green, blue, yellow gradient color bars indicate the enrichment significance for biological process, cellular component, molecular function respectively in (C) and (D). The length of each bar represents the number of DE genes in each GO term. The color saturation of the bars represents the enrichment significance for each GO term. (E) KEGG pathway enrichment analysis for DEGs upregulated in LGS compared to LNS and the genes upregulated in SGS compared to SNS. (F) Distribution of differentially expressed transcription factors. The color of each cell is on behalf of the Z-score indicating the relative gene expression.

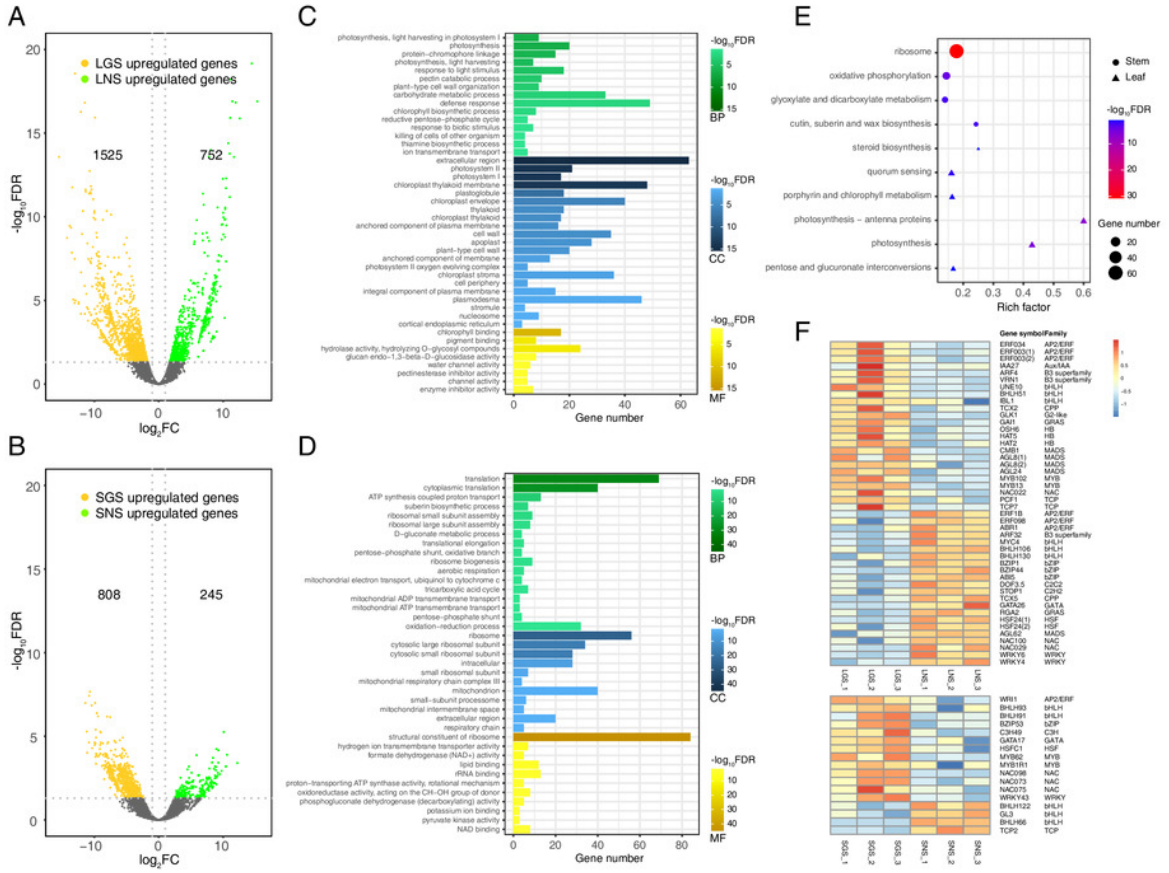
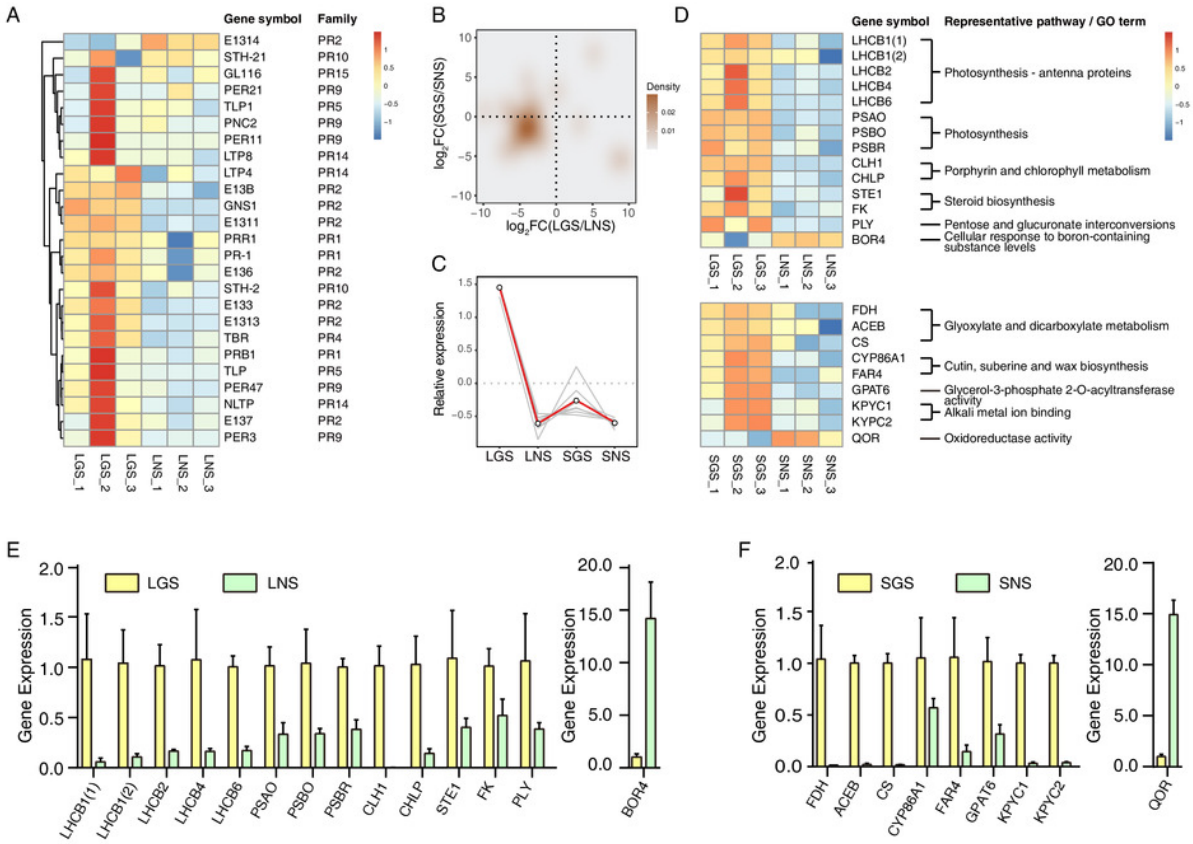

A
LGS upregulated genes
LNS upregulated genes
1525
752
-log10FDR
log2FC
B
SGS upregulated genes
SNS upregulated genes
808
245
-log10FDR
log2FC
C
-log10FDR
BP
CC
MF
Gene number
D
-log10FDR
BP
CC
MF
Gene number
E
Stem
Leaf
-log10FDR
Gene number
ribosome
oxidative phosphorylation
glyoxylate and dicarboxylate metabolism
cutin, suberin and wax biosynthesis
steroid biosynthesis
quorum sensing
porphyrin and chlorophyll metabolism
photosynthesis – antenna proteins
photosynthesis
pentose and glucuronate interconversions
Rich factor
F
Gene symbol
Family

Figure 6

qRT-PCR verification of selected genes and the heatmaps of selected genes and pathogenesis-related (PR) genes.

(A) Heatmap of pathogenesis-related (PR) genes in DEGs. Color saturation represents Z-score scaled gene expression FPKM value. (B) Density map of PR genes. The shade in picture represents the density of gene distribution. (C) Cluster of PR genes with same expression pattern. The gray lines represent the relative expression of each gene and the red line represents the average value for all genes. (D) Transcriptome expression of genes used to be verified with qRT-PCR as well as their annotation and enrichment details. Color saturation represents Z-score scaled gene expression FPKM value. (E-F) Relative expression changes of genes detected using qRT-PCR in leaves (E) and stems (F). In (E) and (F), the genes upregulated and downregulated in the golden samples were divided into two parts. The error bars represent standard deviations (SDs).


A
Gene symbol
Family
E1314 PR2
STH-21 PR10
GL116 PR15
PER21 PR9
TLP1 PR5
PNC2 PR9
PER11 PR9
LTP8 PR14
LTP4 PR14
E13B PR2
GNS1 PR2
E1311 PR2
PRR1 PR1
PR-1 PR1
E136 PR2
STH-2 PR10
E133 PR2
E1313 PR2
TBR PR4
PRB1 PR1
TLP PR5
PER47 PR9
NLTP PR14
E137 PR2
PER3 PR9
LGS_1 LGS_2 LGS_3 LNS_1 LNS_2 LNS_3
B
log2FC(SGS/SNS)
log2FC(LGS/LNS)
Density
C
Relative expression
LGS LNS SGS SNS
D
Gene symbol
Representative pathway / GO term
LHCB1(1)
LHCB1(2)
LHCB2
LHCB4
LHCB6
Photosynthesis - antenna proteins
PSAO
PSBO
PSBR
Photosynthesis
CLH1
CHLP
Porphyrin and chlorophyll metabolism
STE1
FK
Steroid biosynthesis
PLY
Pentose and glucuronate interconversions
BOR4
Cellular response to boron-containing substance levels
LGS_1 LGS_2 LGS_3 LNS_1 LNS_2 LNS_3
FDH
ACEB
CS
Glyoxylate and dicarboxylate metabolism
CYP86A1
FAR4
Cutin, suberine and wax biosynthesis
GPAT6
Glycerol-3-phosphate 2-O-acyltransferase activity
KPYC1
KYPC2
Alkali metal ion binding
QOR
Oxidoreductase activity
SGS_1 SGS_2 SGS_3 SNS_1 SNS_2 SNS_3
E
LGS
LNS
Gene Expression
LHCB1(1) LHCB1(2) LHCB2 LHCB4 LHCB6 PSAO PSBO PSBR CLH1 CHLP STE1 FK PLY BOR4
F
SGS
SNS
Gene Expression
FDH ACEB CS CYP86A1 FAR4 GPAT6 KPYC1 KPYC2 QOR