Genome-wide identification and expression analysis of CYP450 gene family in *Cucumis sativus* L.

hongyu wang
Life Science and Technology College, Harbin Normal University
Pengfei Li
Life Science and Technology College, Harbin Normal University
Yu Wang
Life Science and Technology College, Harbin Normal University
Chunyu Chi
Life Science and Technology College, Harbin Normal University
Guohua Ding (hsddgh@hrbnu.edu.cn)
Life Science and Technology College, Harbin Normal University

Research Article

Keywords: Cytochrome P450, Genome-wide identification, Cucumber, Expression analysis, Salicylic acid

Posted Date: December 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1196334/v1

Abstract

The cytochrome P450 (CYP450) gene family plays a vital role in basic metabolism and enhances plant resistance to stress and pests. However, little information is available on the genome-wide characterization and evolutionary relationship of the CYP450 gene family in *Cucumis sativus* L. In the present study, a genome-wide bioinformatics analysis was performed, including gene structure, conserved motif, cis-acting promoter element, evolutionary analysis, collinearity, subcellular localization, and expression profile. The gene expression profile of CYP450 was verified using transcriptome sequencing and quantitative reverse transcription polymerase chain reaction. A total of 165 P450 genes were identified in the cucumber genome. These genes were classified into eight subfamilies and unevenly distributed on seven chromosomes. Subcellular localization predicted that most of P450 genes were located in chloroplasts and a few were located on the plasma membrane. CYP450 genes were differentially expressed in different tissues and in response to salicylic acid (SA) treatment. The sizes of all cucumber P450 proteins ranged from 317 to 1,056 aa, the theoretical isoelectric points ranged from 5.05 to 10.31, and the molecular weights ranged from 36,095 to 121,403 KD. This study provides a theoretical basis for further research on the biological functions of the P450 gene in cucumber plants.

Background

Cytochrome P450 (CYP450) is a monooxygenase encoded by the B group of cytochrome superfamily proteins with heme as an auxiliary group. It is the largest and oldest family in nature and is distributed throughout all branches of the evolutionary process of life [1]. CYP450s, named for its characteristic spectral property of an absorption peak at 450 nm when binding with CO, are characterized by a preserved heme-binding region with the sequence FxxGxRxCxG [2, 3]. P450s are involved in both basic metabolism and secondary metabolism of plants and play an essential role in enhancing plant resistance to stress and pests [4].

CYP450 plays an important role in plant development and growth. It has been confirmed that the CYP703, CYP704, and CYP86 subfamilies are essential for pollen wall formation and flower cuticle synthesis. In Arabidopsis, CYP703A2, CYP704B1, and CYP86C3 are responsible for the hydroxylation of fatty acids and development regulation of the pollen wall [5]. The CYP73A subfamily is involved in the hydroxylation of aromatic cinnamate rings, which is the core reaction of the phenylpropane pathway to synthesize heaps of phenolic compounds [6]. The CYP701 family has been shown to be necessary for gibberellic acid (GA) biosynthesis in rice [7]. In addition, the CYP450 gene family regulates the response to adverse stress through the synthesis and degradation of endogenous hormones in plants. For example, CYP94B3 and CYP94C1 can oxidize JA-Ile into 12-OH JA-Ile and 12-COOHC JA-Ile and results in severe depletion of JA-Ile, which can affect jasmonic acid metabolism pathway and regulate the expression of stress resistance genes in Arabidopsis [8, 9]. AtCYP82C2 can improve the expression of jasmine-induced defense-related gene expression and jasmonic acid-induced immunoglobulin content, and enhances resistance to *Botrytis cinerea* [10]. The fact that gene SSN (GhCYP82D) in cotton regulates systemic cell death by regulating octadecane pathway, while controlling the biosynthesis of jasmine acid to negatively

regulate resistance to Verticillium wilt, suggest a novel metabolic branch that might regulate the JA signaling pathway [11]. OsCYP71D81L in rice can participate in responses to various types of abiotic stress and the formation of some agronomic characteristics by affecting the homeostasis of cytokinin and gibberellic acid. Compared to wild types, the chlorophyll content of cyp71d8l and CYP71D8L-OE seedlings increased, but the level of reactive oxygen species (ROS) decreased [12].

Development and maturity of massively parallel sequencing (MPS). Members of the P450 gene family have been identified in many plants. Arabidopsis (*Arabidopsis thaliana*) has 245 CYP genes [13], soybean (*Glycine max*) has 332 CYP genes [14];[15], wheat (*Triticum aestivum*) has 1285 CYP genes [16], tomato (*Lycopersicon esculentum*) has 233 CYP genes [17], and rice has 326 CYP genes (*Oryza sativa* L.) [18]. According to the criteria of homology and phylogeny, plant CYP genes are grouped into six single families (CYP51, CYP74, CYP97, CYP710, CYP711, and CYP727) and four multi-families (CYP71, CYP72, CYP85, and CYP86) [19]. However, only a few CYP450 genes have been identified [20, 21]. To date, research on plant P450s has been focused on the genome-wide identification and expression profile analysis of the CYP450 gene family [22].

Cucumber (*Cucumis sativus* L.) is an annual herb belonging to the Cucurbitaceae family. Cucumbers are one of the main vegetable crops in China and have medical, therapeutic, and nutritional values. The analysis of gene families is beneficial for understanding the function and evolution of their members [23, 24]. The completion of cucumber genome sequencing has made gene family information and functions public [25–27]. Wang et al. studied the role of CsCYP85A1 encoding BR-C6-oxidase in the BR biosynthesis pathway to control internode length in cucumber [21], and Xu et al. revealed that Csa6G504590, encoding a cytochrome P450 monooxygenase, might be involved in cell growth by participating in ethylene-activated signaling pathways [28]. Pan et al. pointed out that CYP78A plays a significant role in the fluctuation of shape and cucumber fruit size [29]. However, there are few studies on CYP450 genes in cucumbers.

Here, we used bioinformatics analysis to identify the cytochrome P450 gene family in the cucumber genome and analyzed gene structure, conserved motif, cis-acting promoter element, evolutionary profile, collinearity, subcellular localization, and expression profile under salicylic acid (SA) treatment. The findings are intended to provide a theoretical basis for further exploration of the biological function of the cucumber P450 protein. In our previous study, transcriptome analysis was performed on cucumber leaves treated with salicylic acid, and differentially expressed genes were screened, including P450 genes, to provide reference information for this study.

Results And Analysis

Bioinformatics analysis of CsCYP450

Genome-wide identification of CYP450 gene family and physicochemical properties analysis of P450 protein in

cucumber

After removing the redundant genes, 165 P450 family members were identified in the cucumber genome (Table S1). The protein sequences encoded by these genes were further confirmed by NCBI and SMART screening to contain four conserved motifs, namely the FxxGxRxCxG heme-binding protein domain, I-helix, K-helix, and PERF/W motif. In accordance with the gene naming method used for Arabidopsis, the P450 genes in cucumber were named CsCYP450. The size of cucumber P450 protein was between 317 and 1,056 aa, the theoretical isoelectric point (pI) of the encoding protein was 5.05–10.31, and the molecular weight was between 36,095 and 121,403 kDa. The subcellular localization predicted that 70.3% of the CsCYP450 genes were located in chloroplasts, 12.7% of the CsCYP450 genes were located in plasma membrane, 7.8% of the CsCYP450 genes were located in the cytoplasm, and 4.8% of the CsCYP450 genes were located in the vacuole. The remaining CsCYP450 genes were located in the nucleus, endoplasmic reticulum, and mitochondria. These results suggest that CsCYP450 plays a significant role in cucumber production.

Analysis of gene structure, conserved domain, and motif of CYP450 family in cucumber

Based on the alignment results of the coding sequence to the genomic sequence of the cucumber P450 gene family, the exon numbers of these CsCYP450 genes varied from 1 to 15, revealing the structural diversity of the P450 gene family. Of the cucumber P450 gene family, the Csa3G115620 gene contained the most exons, whereas 17% harbored one exon without an intron. Further analysis showed that most of the same-family members had similar distribution characteristics in terms of exon length and number. For example, Csa3G115620, Csa6G448700, and Csa4G056760 belong to Clan97 and have more exons and introns. However, most CYP89, CYP94, and CYP96 family members have no introns. A few CsCYP450 members such as Csa5G158600 and Csa5G173490 have no typical genetic structure. Compared with other members of the same family, they have longer introns or shorter exons (Figure. 1B). The structure of CsCYP450s was greatly changed, suggesting that the cucumber genome underwent significant changes during its evolutionary history.

Conserved motif analysis of P450 in cucumbers is shown in Figure 2C. A total of 15 conserved motifs were identified and named motif1 to motif15, respectively (Table S2). Six motifs were found in more than 90% of the genes. All P450 gene members contained motif1: FGAGRRICPG (FxxGxRxCxG), which was annotated as the core functional domain (the heme-binding domain). The heme-binding domain contains critical cysteine residues, which are axial ligands of heme. This result shows that motif1 is highly conserved during cucumber development. The motif14 was annotated as a helix K domain (EXXR) that plays an important role in P450 protein folding. The motif3 was annotated as an I-helix conserved domain, which is involved in oxygen binding and oxygenation. The motif6 was annotated as a PERF domain and motif7 was annotated as a PKG motif. As shown in Figure 2, members of the same gene cluster shared similar conserved domains, whereas different gene clusters had specific conserved

domains. In conclusion, the highly conserved motif and similar gene structure of CsCYP450 in the same family further support the close evolutionary relationship and reliability of the phylogenetic analysis.

Chromosome distribution of CYP450 family members in cucumber

Figure 3 shows the chromosomal distribution of the P450 gene family in cucumber. P450 genes were unevenly distributed on seven chromosomes. Chromosome 3 had 48 P450 genes, which is the largest number. Chromosome 7 contained only seven P450 genes. There was no positive correlation between chromosome length and the number of P450 genes, and most were close to each other. In addition to chromosomes 5 and 7, the density of CsCYP450 on the short arm was higher. The gene distribution of each subfamily was not uniform and the similarity was high. Additionally, some genes were closely clustered on chromosomes 2, 3, 5, and 6. Whether there is a functional connection between these genes and the possible reasons for their formation still needs to be further analyzed. It is speculated that these may be related to the evolution of P450 genes and may affect their regulatory function.

Analysis of replication events of cucumber CYP450 genes family

Gene replication is one of the most essential evolutionary processes to produce genetic diversity and new functions and plays a vital role in adaptation and speciation. Results were divided into segmental, tandem, proximal, singleton, and dispersed genes. Among them, the frequency of tandem replication was 38.6% (76), fragment replication was 19.3% (38), proximal replication was 9.1% (18), decentralized replication was 33% (65), and no single-copy replication occurred (Figure 4). The results demonstrated that tandem reproduction was the main driving force for CsCYP450 gene family amplification. Tandem replication is required to maintain a large gene family, which can dilate and shrink rapidly in response to environmental alterations, resulting in increased genetic complexity under good conditions [30]. Additionally, the chromosomal distribution pattern of CsCYP450 strongly suggested that tandem replication promoted the amplification of CsCYP450 in cucumber. To further explore the possible evolutionary mechanism of the CsCYP450 gene family, Circos software was used to map collinear gene pairs in the cucumber genome (Figure 6).

Evolutionary analysis of related species

A phylogenetic tree was constructed according to the multiple sequence alignment results to investigate the genetic relationships among the members of the CYP450 gene family in cucumber. Based on phylogenetic analysis, 165 P450 genes were classified into A-type and non-A-type P450 gene family groups (Figure 7). The A-type P450 gene family was clustered into the CYP71 clan. The non-A-type P450 gene family contained seven clans, designated as CYP85, CYP86, CYP72, CYP97, CYP711, CYP51, and CYP710.

To investigate the evolutionary relationship between cucumber CYP450 genes, we constructed a phylogenetic tree using CYP450 proteins from cucumber and Arabidopsis (Figure 7). CYP450 proteins were classified into A-type (52.7%, 87/165) and non-A type (47.3%, 78/165). According to the criteria of phylogeny and homology, the CYP450 gene family was further divided into 40 families and eight clusters. Among these, the most extensive A-type gene family of cucumbers is CYP71, with 87 members. Studies have shown that the most prominent CYP450 gene family in most plants is CYP71, which contains more than half of the CYP450 genes and has rich and diverse functions, which is closely related to the metabolism of aromatic and aliphatic amino acid derivatives, some triterpenoid derivatives, fatty acids, alkaloids, and hormone precursors; the largest non-A type family in cucumber is CYP85, which is composed of nine gene families: CYP86, CYP707, CYP722, CYP716, CYP718, CYP87, CYP724, CYP85, and CYP90. It is worth mentioning that the CYP93, CYP701, CYP703, CYP51, CYP724, CYP718, and CYP710 families are all composed of single cucumber genes, suggesting that each gene has a unique, highly conserved function. In particular, the CYP51 family is an ancient and conservative family, with only one member in the study of gene families of all species to date. CYP51G and CYP710A encode 14α-demethylase and sterol 22-desaturase, respectively, which participate in sterol biosynthesis. In a survey of CYP703 family members, the Arabidopsis mutant cyp703a2 showed the phenotype of pollen development retardation and partial male sterility [31]. In addition, compared to the A-type P450 family, the non-A-type P450 gene family has a wide range of species and complex changes. The non-A-type P450 gene family may be evolutionarily older than the A-type P450 family, and the time for gene replication and rearrangement may be longer. This leads to a more diverse composition than that of the A-type P450.

We compared the P450 genes of cucumber with P450 gene families in Arabidopsis, tomato, soybean, maize, rice, poplar, grape, and moss (Tables S3 and S4). CYP736 has evolved in the cucumber genome compared with Arabidopsis. Some studies have shown that CYP736, CYP83, and CYP81 are very similar, and they strongly induce *Phytophthora sojae* to infect cucumber hypocotyls [14]. However, CYP82 has more P450 members in the cucumber, soybean, and grape genomes, whereas CYP82 is absent in the rice and moss genomes. Studies have shown that the CYP82 family makes cotton strongly resistant to disease and stress, and it is speculated that it has a similar role in cucumbers [11]. The CYP93 family is involved in the biosynthesis of soybean flavonoids, with 13 members in soybean and seven members in maize. Similar to Arabidopsis and rice, cucumber has fewer members of the CYP93 family, further confirming the specificity of the P450 gene family.

In addition, non-A-type CYP709, CYP702, CYP708, and CYP705 families exist in the Arabidopsis genome but not in cucumber, soybean, tomato, and other species. P450 genes from these families are specific to Arabidopsis and its closest relatives, such as *Brassica napus*, making them the only known cruciferous heterosexual CYP protein. It has been suggested that CYP705 family and other P450 genes congregate in different chromosome regions. The genes of the CYP702A, CYP705A, and CYP708A families may be involved in the modification of triterpenes. CYP705 is derived from the CYP712 family. The number of CYP705 family members in Arabidopsis is as high as 26, while they do not occur in other species [31]. The absence of these P450 families in cucumber may be related to their evolution.

Analysis of Ka/Ks

In genetics, Ka/Ks or dN/dS represents the ratio between the non-synonymous replacement rate (Ka) and synonymous replacement rate (Ks). This ratio can be used to determine whether there is a selective burden on the protein-coding genes. It is broadly trusted that non-synonymous modifications are influenced through natural selection, whereas synonymous mutations are not. In evolutionary analysis, it is important to understand the rate at which synonymous and non-synonymous mutations occur. The values of Ka, Ks, and Ka/Ks are based on the coding sequence alignment and are calculated using the Nei and Gojobori model of the KaKs_calculator software [32]. The typically applied indicators are as follows: synonymous mutation frequency (Ks), non-synonymous mutation frequency (Ka), and the ratio of the non-synonymous mutation rate to the synonymous mutation rate (Ka/Ks). Ka/Ks > 1 indicates positive selection, suggesting that natural selection has a changing consequence on protein, resulting in rapid fixation of mutant sites in the population and accelerated gene evolution; Ka/Ks=1 indicates neutral evolution, showing that natural selection has no effect on mutation; Ka/Ks < 1 indicates purification selection, which means that natural selection can effectively eliminate harmful mutations and maintain protein characteristics. These results showed that some P450 gene members in cucumbers experienced significant positive selection pressure.

Collinearity analysis

Collinearity was originally used to describe the locations of genes on the same chromosome. It now refers to the conservation of gene types and relative order in different species derived from the same ancestral type. Collinearity analysis can identify linear homologous genes among species, annotate protein-coding genes, and discover evolutionary events. The collinearity of species was constructed using McScanX and plotted using the Circos software (Figure 9). The CYP450 genes of cucumber, melon, and Arabidopsis genomes were jointly analyzed to study the collinear genetic relationships among them.

Collinearity analysis showed that 71 pairs of collinear genes were identified between the genomes of Arabidopsis and cucumber, indicating that the gene family was significantly amplified before the differentiation of the two species. There were 138 collinear gene pairs between cucumber and melon. These results showed that the CYP450 genes of cucumber and melon were highly evolutionarily conserved.

According to the collinear analysis of pairs between cucumber and Arabidopsis, in addition to one-to-one matching (e.g., Csa3G119520/AT1G01280 and Csa1G004040/AT3G48290), there is also many-to-one matching (e.g., Csa5G153010, Csa6G004550, Csa6G514850/AT1G01600); these results showed that CsCYP450 is relatively conserved and may have originated from the same ancestor and have differentiated functions during evolution. There are also many-to-one matches between cucumbers and melons, such as Csa3G127130, Csa3G903510, and Csa6G079220/MELO3C006237. These results suggest that the functional differentiation of these genes may have occurred in cucumbers and melons during evolution.

Analysis of cis-acting elements of CsCYP450 promoter

To better understand the potential regulatory mechanism of CsCYP450 during cucumber growth and development, the cis-regulatory element in the promoter regions of CsCYP450 was identified in this study. The upstream sequence (2.0kb) of all CsCYP450 translation initiation sites was scanned and the potential role of CsCYP450 expression elements was predicted using the Plant Care tool (Figure S1).

In addition to the TATA-box, CAT-box, and other specific elements, the CsCYP450 promoter contains various cis-regulatory elements that are related to light signal response, tissue and organ development, hormone response, defense, and stress. Many regulatory elements related to the light response were detected in the CsCYP450 gene, such as ACE, ATCT-motif, and CAG-motif. Most genes contain at least two box4 elements, and there are 11 Box4 elements in the promoter region of the Csa2G108690 gene, suggesting that some DNA modules may be involved in the light response.

Among the elements related to plant growth and development, nine specific regulatory elements were identified, including those related to cell differentiation (CAT-box), endosperm expression (GCN4-motif), and zein metabolism regulation (O2-site). Csa5G157290, Csa5G157320, Csa3G852560, and other genes were identified as cis-acting regulatory elements involved in regulating circadian rhythm, which means that their function may be affected by day length.

Among the elements related to plant hormones, 11 hormone response regulatory factors were identified, which were related to abscisic acid [33, 34], GA (GARE-motif, P-box, TATC-box), IAA (AuxRR-core, TGA-box, TGA-element), SA (TCA-element, SARE), and MeJA (CGTCA-motif, TGACG-motif). Among these, the ABA-related response elements are widely distributed. The promoter regions represented by the Csa6G088160 and Csa2G432220 genes were predicted to contain nine ABRE elements, which might be involved in the ABA signaling pathway, and SA and GA elements were enriched in most CsCYP promoters, indicating that they may be induced by SA and GA.

Among the biotic and abiotic stress-related elements, ARE elements were found in most of the gene promoter regions. Seven ARE elements were found in Csa3G818260, and six ARE elements were found in Csa1G595860. ARE is necessary for cis-acting element-based anaerobic induction and may directly affect the antioxidant capacity of the gene. ERE elements have been found in the promoter regions of some genes, such as Csa5G615280 and Csa4G641760. This could be related to the ethylene response. MBS cis-elements were identified in Csa6G366560, Csa1G039830, and other genes, and it was predicted that these genes were involved in MYB binding sites involved in drought induction. In addition, a small number of genes were rich in WUN-motif, LTR, and DRE response elements, suggesting that they may be interested in wound response, low temperature, and osmotic stress response mechanisms.

These results show that CsCYP450 may play a significant role in the response to stress, hormones, and light exposure.

Expression analysis of CsCYP450 gene

Salicylic acid is a significant signaling molecule and plant hormone that plays a significant role in the coordination of development and growth under environmental conditions. It has been demonstrated that SA played a significant role in the activation of defense responses of plants, such as inducing the expression of related genes and producing specific secondary metabolites for defense in plants [35]. At present, it has been confirmed in Arabidopsis that biological stress first induces the accumulation of SA, and then SA combines with SA-binding protein (SABP) to form the SA-SABP complex and initiate intracellular conduction. After NPR1 is activated, the activation and interaction of transcription factors, such as TGA and WRKY, can induce the expression of pathogen-related (PR) protein genes and generate stress resistance [36].

The CYP450 family plays a significant role in defense reactions. Zhang et.al have shown that CYP450 genes, antioxidant enzyme genes and ATP-binding cassette transporters genes, and other genes related to defense signals in *Salvia miltiorrhiza* are significantly overexpressed under SA induction, which can be used as a genetic tool to study disease resistance [37]. In cucumber, the relationship between CYP450 and SA signals is unclear. Therefore, we analyzed the expression patterns of some CYP450 family genes induced by SA, based on transcriptome analysis data to explore the relationship between CYP450 and SA signaling pathways. Among the genes annotated as P450 in the RNA-SEQ data, 35 were significantly upregulated or downregulated by SA (Table S5). According to the functional annotation of GO, the DEGs were mainly concentrated, especially the integral components of the membrane (GO:0016021) and ribosome (GO:0005840) in cellular components. The main enrichment items in molecular function were oxidoreductase activity (GO:0016705), monooxygenase activity (GO:0004497), iron ion binding (GO:0005506), and heme-binding (GO:0020037). In biological processes, brassinosteroid biosynthetic process (GO:0016132), sterol metabolic process (GO:0016125), RNA polymerase II promoter (GO:0006367), defense response to another organism (GO:0098542), biosynthetic process (GO:0016114), lignin biosynthetic process (GO:0009809), and fatty acid oxidation (GO:0019395) were the most differentially expressed enrichment items. According to the KEGG database comparison, the pathways of significant enrichment of differential genes included the phenylpropane biosynthesis pathway, phenylalanine metabolic pathway, flavonoid biosynthesis pathway, biosynthesis pathway of secondary metabolites, O-antigen nucleoside biosynthesis, and cutin, cork, and waxy biosynthesis pathways. Several studies have demonstrated the involvement of these pathways in plant disease tolerance.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was used to verify the differentially expressed genes in different tissues. In addition, some genes such as Csa1G002090 and Csa1G004040 were upregulated after 12 h of SA treatment, and the expression trend of qRT-PCR results was consistent with that of transcriptome sequencing, which indicated the reliability of transcriptome sequencing results. The slight difference between the two experimental methods may be related to differences in the practical methods and principles.

The results of the expression analysis showed that most CsCYP450 showed different expression patterns, but a few showed similar expression patterns. CsCYP450 was highly expressed in roots. Several genes (e.g., Csa3G852560 and Csa3G852630) were highly expressed in the three tissues, and their

expression levels increased with SA treatment time. However, other genes such as Csa6G514850 and Csa6G088710 were not expressed in any tissue. Some CsCYP450 genes show extremely high expression in certain tissues, such as Csa6G366560 in leaves, Csa3G651820 in stems, and Csa7G198310 in roots. In this study, CsCYP450 genes belonging to the CYP82 family were found to be highly expressed in roots, stems, and leaf tissues. In addition, most genes have different time- and tissue-specific expression patterns, which are speculated to be related to the differences in the biological functions of CsCYP450 (Figure 10).

Combined with the analysis of the potential function prediction results of the promoter cis-elements, no significant correlation was found with expression. For example, Csa7G198310 contained two TCA-element elements that were significantly upregulated under SA treatment; however, some genes with TCA-element or SARE elements, such as Csa3G852610 and Csa2G435520, showed a decreasing trend or remained unchanged under SA treatment.

Discussion

Comprehensive identification of P450 genes in cucumber

P450 (CYPs) is the third largest gene family involved in plant metabolism and the largest enzyme family. P450 genes account for approximately 1% of plant protein-coding genes, which is inferior to genes coding for transcriptional regulatory proteins [38]. CYP450 monooxygenase superfamily plays an important role in plant metabolism. Several studies have revealed the functions of CYP450 family genes. Our study is the first to report genome-wide identification and analysis of CYP450 in cucumbers. In total, 165 P450 genes were identified in the cucumber genome. The number of P450 genes varies widely among plants, including 343 in rice, 249 in Arabidopsis, 318 in maize, 416 in grape, and 1,476 in wheat [1]. The genome size of cucumber is 367 Mb, which is 2.7 times that of Arabidopsis (135 Mb), but the number of P450 genes in cucumber is 68.8% of that in Arabidopsis, suggesting that some P450 genes were lost during the evolution of cucumber. Studies have shown that the loss of the CYP450 family in different plants seems to be limited to a single species or taxon, while the expansion of different subfamilies may be related to plant species (Jiu et al. 2020). Based on the phylogenetic tree, the 165 CsCYP450 genes were divided into eight clusters. The distribution of P450 in Arabidopsis and cucumber showed a similar trend. Generally, CsCYP450 is grouped into non-A-type P450 and A-type P450. A-type P450 is plant-specific and is characterized by a remarkably consonant intron and a simple organization. Relevant A-type P450 genes are closely clustered in the genome; there are multiple genes on a minor segment of the chromosome (for instance, CYP71 subfamily genes). A-type P450 may be complicated in plant-specific biochemical pathways; other P450 genes (non-A type) form several different branches, characterized by many introns. In contrast, non-A-type P450 do not constitute a single phylogenetic group, and there are significant differences between them [39].

We compared P450 genes of cucumber with P450 gene families in Arabidopsis, tomato, soybean, maize, rice, poplar, grape, and moss. CYP736 has evolved in the cucumber genome compared with Arabidopsis. CYP82 has more P450 members in the cucumber, soybean, and grape genomes, whereas CYP82 is

absent in the rice and moss genomes. Studies have shown that the CYP82 family makes cotton strongly resistant to disease and stress, and it is speculated that it has a similar role in cucumbers. In addition, non-A-type CYP709, CYP702, CYP708, and CYP705 families exist in the Arabidopsis genome but not in cucumber, soybean, tomato, and other species.

The CsCYP450 arose mainly through tandem replication, accompanied by decentralized replication

Gene replication provides primitive and abundant genetic material for plant evolution. Cucumber has experienced a large number of tandem replications and fragment replication, and a large number of replication genes will mutate randomly, resulting in expression and functional differentiation under different selection pressures. This also explains why, compared with their original function, replicative genes have non-functionalization, sub functionalization, and informational functionalization. Overall, gene replication is a vital driving force in plant evolution. Selection pressure can promote retention of replication genes with new functions. For example, MADS-box transcription factors are involved in the structural evolution of floral organs, dioxygenase genes are involved in pigment variation incarnation, and so on. The CYP71 family accounts for more than half of the gene family in higher plants, and this burst of gene replication is likely to contribute to the adaptation to specific ecological niches and speciation [38]. During the evolution of cucumber, a rapid increase in CsCYP450s was due to the main mechanism of tandem replication, followed by dispersed replication, which is similar to the mechanism of grape expansion [40]. Tandem replication is mainly caused by the unequal exchange of alleles and occurs in all chromosomes except in chromosome 7 in cucumbers. Functional differentiation occurred in cucumbers during evolution.

Intron gain and loss events to investigate evolution of CsCYP450

There are relatively conserved motifs in the cucumber P450 gene and there is a heme-binding domain in each CsCYP450. CsCYP450 proteins from the same subfamily have similar motifs, indicating high similarity in their functions, which was further confirmed through gene structure analysis. Among the 295 members of the P450 superfamily screened in peaches, the number of specifically expressed genes in the root was the highest. These CYP genes have a long span and possess significant differences in the number of exons, which is similar to that in cucumber. Each subfamily shows a very similar gene structure in the number or length of exons, which is speculated to be associated with the rich biological functions of the CYP450 gene family [40]. Studies have shown that only a few P450 genes in Arabidopsis are intron-free, while P450 members without introns in cucumber account for 17% and more than in Arabidopsis [41]. The increase and decrease of introns are common phenomena during plant evolution, which enrich the complexity of gene structure [42]. There is an excellent correlation between phylogeny and intron conservation among members of the P450 subfamily [40].

Cis-elements and expression analysis in CsCYP450

In the process of differentiation, development, and growth, plants must integrate environmental and developmental signals into different tissues to regulate gene expression. As an important component of transcriptional regulation, cis-elements of the promoter are involved in the regulation of networks involved in many biological processes. Our study found that there were many repetitive areas in the promoter of CsCYP450s, which conformed to the results of previous studies [43, 44]. The identified regulatory elements were divided into four groups, excluding specific and general elements and including elements responsive to growth and development, hormones, pressure, and light. In light response detection, responsive elements related to circadian rhythm were identified, and the results were similar to those of previous studies [45]. Owing to the synergistic effect of transcription factors, such as WRKY, MYB, and NAC, stress resistance is the most complex of all plant response mechanisms.

These results demonstrate that CYP450 is involved in many biochemical pathways and plays an important role in plant growth and development. CYP82 exists only in dicotyledons and has been confirmed to perform critical biological functions in soybean and cotton. For instance, CYP82 in soybean is expressed in the stems, leaves, and roots, with the highest expression level, and SA possesses a strong induction effect. The results of this study were consistent with this. Some genes such as Csa1G00400 are expressed at low levels in cucumber, which may be due to their transcriptional or post-transcriptional silencing effect or functional redundancy. Studies have shown that phenomena and products are more common in the biological functions of gene families. For example, the IFS1 gene in soybean is preferentially expressed in roots and seeds, but not in stems, whereas the expression level of genes involved in fat hydroxylase is minimal. This may be because the position and function of P450 in the biochemical pathway are different; therefore, it is not surprising that the level of expression is significantly different [45]. Salicylic acid (SA) has been identified as a complex regulator of plant disease resistance. In addition, the CsCYP450 promoter also contains cis-regulatory elements related to hormonal responses and biological stress. Whether these findings indicate that CsCYP450 plays an intermediary role in activating the intracellular SA signaling pathway to regulate cucumber resistance needs to be further verified.

Materials And Methods

Plant materials

Cucumber 9930 cultivated in North China was selected as plant material (donated by Prof. Huang Sanwen, Institute of Vegetable Research, Chinese Academy of Agricultural Sciences). The 9930 seeds were placed in a clean culture dish with wet filter paper. After the cotyledons are opened, they were moved to the soil (vermiculite : nutrient soil = 1 : 1), in the greenhouse at 22 ± 4°C with a photoperiod of 16 h light and 8 h dark. The seedlings were treated with salicylic acid (concentration of 10mmol/L and controlled by spraying distilled water) during the four-leaf one-heart period. Time gradients are 0h, 3h, 12h, and 24h. All materials including roots, stems, and leaves were immediately frozen in liquid nitrogen and stored at -80℃ until total RNA isolation.

Identification of CYP450 gene family members in Cucumis sativus L.

The whole genome and protein sequence data of cucumber were downloaded from a public database (http://cucurbitgenomics.org/; Cucumber (Chinese Long) genome v2. The Hidden Markov Model (HMM) analysis and a Simple Modular Architecture Research Tool (SMART) were used [46]. The HMM profile was downloaded from Pfam protein family database (http://pfam.xfam.org/) to obtain the Pfam number of protein sequence. The domain contained in the target protein sequence was detected using the hmmsearch command in the HMMER package, with an e-value≦1e-3. The results of the HMMER sequence alignment were screened to remove protein sequences that were 45% longer than the length of the HMM model domain, while retaining the longest protein sequence in the variable shear. All non-redundant protein sequences were retrieved and further analyzed with SMART (a Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de/) to examine the results. The same genes were confirmed as family members. The subcellular location prediction of candidate genes was carried out with WoLF PSORT (https://wolfpsort.hgc.jp/) [45]. To indentify all potential CYP450 genes in cucumber genome, a BLAST search against the cucumber genome was conducted using 245 AtCYP450s genome sequence downloaded from Cytochrome P450 database (http://drnelson.uthsc.edu/CytochromeP450). Based on the 230 cucumber P450 gene sequences obtained from P450 homepage and the downloaded Arabidopsis thaliana sequences to find a sequence with higher consistency, it is classified and named according to international standards. The consistency of amino acid sequences between 40% and 55% belong to the same family, the consistency of amino acid sequences between 55% and 97% belong to the same subgroup, and the consistency of amino acid sequences greater than 97% are alleles gene.

Conserved domain, motif identification, and gene structure analysis

The conserved motif of cucumber P450 was analyzed using the online tool MEME (http://meme-suite.org/tools/meme). The maximum number of motifs was 15, and the optimized motif width was 10-100 amino acid residues. The rest parameters are default [47]. The location information of the gene and the conservative area was confirmed and then plotted using the Perl SVG package.

Phylogenetic analysis

The full-length amino acid sequences of gene family members derived from closely related species were used for phylogenetic analysis. The phylogenetic tree was constructed using the Neighbor-Joining (NJ) method of MEGA (version 7.0) with the bootstrap test of 1000 times [48].

Gene replication

The potential replication genes were identified using the duplicate MCScanX software's gene classifier program [49]. The genome's all coding gene protein sequences were aligned consuming blastP, and the alignment consequences were used as input files for MCScanX Software for reproduction gene projection. The gene was identified as a reproduction gene pursuant to e-value<1e-5 or e-value<1e10.

Closely related species analysis

The gene family members of associated species were respectively identified by Pfam and SMART. The full-length amino acid sequences of gene family members derived from closely related species were used for phylogenetic analysis. The phylogenetic tree was constructed with the neighbor-joining algorithm in MEGA with the bootstrap test of 1000 times. In genetics, the Ka/Ks ratio indicates of selective pressure (the force applied by natural selection) acting on a protein-coding gene. It is calculated as the ratio of the number of non-synonymous substitutions per non-synonymous site (Ka), to the number of synonymous substitutions per synonymous site (Ks), in each period. The values of Ka, Ks, and Ka/ Ks are based on coding sequence alignment and calculated by Nei and Gojobori model of Ka Ks calculator software [50]. The collinearity of multiple species was constructed by McScanX and drawn with Circos Software [49].

Promoter cis element analysis

We intercepted the upstream sequence (2.0kb) of each CsCYP450 gene annotation file in Cucurbitaceae database (http://cucurbitgenomics.org/) and scanned the cis-acting elements from those sequences. Cis-acting elements prevalent in genes were identified including CAAT-box, TATC-box, TATA-box, and cis-acting elements with specific functions were only shown [27, 51, 52].

Transcriptome analysis

Total RNA was extracted from each samples using RNApure Plant Kit (TaKaRa, Dalian, China) according to the instructions of the kit. Magnetic beads with Oligo (DT) were used to enrich mRNA. Then fragmentation buffer was added to break mRNA into short fragments. Using mRNA as template, the first strand cDNA was synthesized with six base random hexamers, then the second strand cDNA was synthesized by adding buffer, dNTPs and DNA polymerase I, and then double strand cDNA was purified by AMPure XP beads. The purified double-stranded cDNA was repaired by terminal repair, A tail was added and sequenced, then the fragment size was selected by AMPure XP beads, and finally the final c-DNA library was enriched by PCR. After the construction of the library, the library was initially quantified with Qubit 2.0 and diluted to 1ng μL-1. Then the insert size of the library was detected by Agilent 2100. After insert size met the expectations, the effective concentration of the library was accurately quantified by QPCR method (the effective concentration of the library > 2nM) to ensure the quality of the library. After passing the library inspection, different libraries were sequenced by HiSeq 2000 after pooling according to the effective concentration and the target amount of data off the machine.

Gene Ontology (GO) functional significance enrichment analysis was carried out by using Gene Ontology data base (http://www. Geneontology. Org /) to map differentially expressed genes to each term in geneontology database. Calculate the number of genes per term. The corrected P value is less than 0.05 is significantly enriched GO term. Pathway significant enrichment analysis (pathway enrichment analysis) uses KEGG database to map differentially expressed genes to KEGG database, and statistics on the enrichment degree of genes in each pathway (also need to meet the correction $P \leq 0.05$).

Quantitative reverse transcription PCR (qRT-PCR) expression analysis

Primers for quantitative reverse transcription PCR (qRT-PCR) were designed by using Primer Premier 5.0 software (Table S6). The 18s rRNA gene of cucumber was selected as a reference gene. Total RNA was extracted from each samples using RNApure Plant Kit (Cwbio, Beijing, China) according to the manufacturer's instructions. Total RNA was extracted from each samples using RNApure Plant Kit (Cwbio, Beijing, China) according to the manufacturer's instructions.The RNA quality was determined by Nanodrop2000 (Thermo Science) at A260 and A280. First-strand cDNA synthesis was performed using an oligo (dT) primer and 2 μg of total RNA in a 20-μl reaction volume,according to the manufacture's instruction for the SuperRT cDNA Synthesis Kit (Cwbio, Beijing, China). The integrity of RNA was assessed by 1% agarose gel electrophoresis, qRT-PCR was carried out using a Ul'traSYBRMixture (WithRox) kit (Cwbio, Beijing, China) with a Fast Real-Time PCR System (Applied Biosystems 7500, USA). The recommended conditions for PCR was used as follows: 95 ℃ for 10 m, n,followed by 40 cycles of 95℃ for 30s, Tm for 30s and 72℃ for 30s, All reactions were performed in triplicate for each sample. The relative expression analysis of each gene were calculated using the $2^{-\triangle\triangle CT}$ method [16].

Abbreviations

P450, CYP450
cytochrome P450
SA
salicylic acid
GA
gibberellic acid
ROS
reactive oxygen species
MPS
massively parallel sequencing
pI
isoelectric point
PR
pathogen-related
SABP
SA-binding protein
GO
Gene ontology
CsCYP450
cucumber CYP450
KEGG
Kyoto Encyclopedia of Genes and Genomes

Declarations

Ethics approval and consent to participate

Cucumber 9930 seeds used in our study were donated by Prof. Huang Sanwen, Institute of Vegetable Research, Chinese Academy of Agricultural Sciences.The seedings grew in plant growth chamber in Harbin Normal University.The authors declared the compliance with institutional, national and international guidelines.

Consent for publication

Not applicable.

Availability of data and material

All the datasets supporting the results of this article are included within the article and its Additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Harbin Normal University Postgraduate Innovation Project (HSDBSCX2021-02)

Authors` contributions

GHD and HYW conceptualized, designed the project. HYW, PFL, and YW carried out computational analysis, and expression profiling. HYW performed the experiments. HYW, CYC, and GHD wrote the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The following format is required when data are included as additional files:

All data generated or analysed during this study are included in this published article [and its supplementary information files].

The whole genome and protein sequence data of cucumber were downloaded from a public database (http://cucurbitgenomics.org/; Cucumber (Chinese Long);

The HMM profile was downloaded from Pfam protein family database (http://pfam.xfam.org/);

AtCYP450s genome sequence downloaded from Cytochrome P450 database (http://drnelson.uthsc.edu/CytochromeP450);

Gene Ontology (GO) functional significance enrichment analysis was carried out by using Gene Ontology data base (http://www. Geneontology. Org /);

Uses KEGG database to map differentially expressed genes to KEGG database,https://www.kegg.jp/;

References

1. Guengerich FP: **Cytochrome P450 research and The Journal of Biological Chemistry**. *J Biol Chem* 2019, **294**(5):1671–1680.
2. Nelson DR: **A world of cytochrome P450s**. *Philos Trans R Soc Lond B Biol Sci* 2013, **368**(1612):20120430.
3. Tu L, Su P, Zhang Z, Gao L, Wang J, Hu T, Zhou J, Zhang Y, Zhao Y, Liu Y *et al*: **Genome of Tripterygium wilfordii and identification of cytochrome P450 involved in triptolide biosynthesis**. *Nat Commun* 2020, **11**(1):971.
4. Aubert Y, Widemann E, Miesch L, Pinot F, Heitz T: **CYP94-mediated jasmonoyl-isoleucine hormone oxidation shapes jasmonate profiles and attenuates defence responses to Botrytis cinerea infection**. *J Exp Bot* 2015, **66**(13):3879–3892.
5. Shi J, Cui M, Yang L, Kim YJ, Zhang D: **Genetic and Biochemical Mechanisms of Pollen Wall Development**. *Trends Plant Sci* 2015, **20**(11):741–753.
6. Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W: **A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis**. *Plant Cell* 2012, **24**(9):3506–3529.
7. Wang Q, Hillwig ML, Wu Y, Peters RJ: **CYP701A8: a rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism**. *Plant Physiol* 2012, **158**(3):1418–1425.
8. Koo AJ, Cooke TF, Howe GA: **Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine**. *Proc Natl Acad Sci U S A* 2011, **108**(22):9298–9303.
9. Heitz T, Widemann E, Lugan R, Miesch L, Ullmann P, Desaubry L, Holder E, Grausem B, Kandel S, Miesch M *et al*: **Cytochromes P450 CYP94C1 and CYP94B3 catalyze two successive oxidation steps of plant hormone Jasmonoyl-isoleucine for catabolic turnover**. *J Biol Chem* 2012, **287**(9):6296–6306.
10. Liu F, Jiang H, Ye S, Chen WP, Liang W, Xu Y, Sun B, Sun J, Wang Q, Cohen JD *et al*: **The Arabidopsis P450 protein CYP82C2 modulates jasmonate-induced root growth inhibition, defense gene expression and indole glucosinolate biosynthesis**. *Cell Res* 2010, **20**(5):539–552.
11. Sun L, Zhu L, Xu L, Yuan D, Min L, Zhang X: **Cotton cytochrome P450 CYP82D regulates systemic cell death by modulating the octadecanoid pathway**. *Nat Commun* 2014, **5**:5372.
12. Ma Y, Cui G, Chen T, Ma X, Wang R, Jin B, Yang J, Kang L, Tang J, Lai C *et al*: **Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in Salvia miltiorrhiza**. *Nat Commun* 2021, **12**(1):685.

13. Morikawa T, Mizutani M, Aoki N, Watanabe B, Saga H, Saito S, Oikawa A, Suzuki H, Sakurai N, Shibata D *et al*: **Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato**. *Plant Cell* 2006, **18**(4):1008–1022.
14. Guttikonda SK, Trupti J, Bisht NC, Chen H, An YQ, Pandey S, Xu D, Yu O: **Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases**. *BMC Plant Biol* 2010, **10**:243.
15. Yan Q, Cui X, Lin S, Gan S, Xing H, Dou D: **GmCYP82A3, a Soybean Cytochrome P450 Family Gene Involved in the Jasmonic Acid and Ethylene Signaling Pathway, Enhances Plant Resistance to Biotic and Abiotic Stresses**. *PLoS One* 2016, **11**(9):e0162253.
16. Duan X, Song X, Wang J, Zhou M: **Genome-Wide Identification and Characterization of Fusarium graminearum-Responsive lncRNAs in Triticum aestivum**. *Genes* 2020, **11**(10).
17. Vasav AP, Barvkar VT: **Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in Solanum lycopersicum L. suggested tissue specific promoters**. *BMC Genomics* 2019, **20**(1):116.
18. Waseem M, Huang F, Wang Q, Aslam MM, Abbas F, Ahmad F, Ashraf U, Hassan W, Fiaz S, Ye X *et al*: **Identification, methylation profiling, and expression analysis of stress-responsive cytochrome P450 genes in rice under abiotic and phytohormones stresses**. *GM Crops Food* 2021:1–13.
19. Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S: **Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot**. *Plant Physiol* 2004, **135**(2):756–772.
20. Hazman M, Suhnel M, Schafer S, Zumsteg J, Lesot A, Beltran F, Marquis V, Herrgott L, Miesch L, Riemann M *et al*: **Characterization of Jasmonoyl-Isoleucine (JA-Ile) Hormonal Catabolic Pathways in Rice upon Wounding and Salt Stress**. *Rice (N Y)* 2019, **12**(1):45.
21. Wang H, Li W, Qin Y, Pan Y, Wang X, Weng Y, Chen P, Li Y: **The Cytochrome P450 Gene CsCYP85A1 Is a Putative Candidate for Super Compact-1 (Scp-1) Plant Architecture Mutation in Cucumber (Cucumis sativus L.)**. *Frontiers in Plant Science* 2017, **8**.
22. Koo AJ, Thireault C, Zemelis S, Poudel AN, Zhang T, Kitaoka N, Brandizzi F, Matsuura H, Howe GA: **Endoplasmic reticulum-associated inactivation of the hormone jasmonoyl-L-isoleucine by multiple members of the cytochrome P450 94 family in Arabidopsis**. *J Biol Chem* 2014, **289**(43):29728–29738.
23. Bartholomew ES, Black K, Feng Z, Liu W, Shan N, Zhang X, Wu L, Bailey L, Zhu N, Qi C *et al*: **Comprehensive Analysis of the Chitinase Gene Family in Cucumber (Cucumis sativus L.): From Gene Identification and Evolution to Expression in Response to Fusarium oxysporum**. *Int J Mol Sci* 2019, **20**(21).
24. Chen C, Chen X, Han J, Lu W, Ren Z: **Genome-wide analysis of the WRKY gene family in the cucumber genome and transcriptome-wide identification of WRKY transcription factors that respond to biotic and abiotic stresses**. *BMC Plant Biol* 2020, **20**(1):443.

25. Chen Y, Wang G, Pan J, Wen H, Du H, Sun J, Zhang K, Lv D, He H, Cai R *et al*: **Comprehensive Genomic Analysis and Expression Profiling of the C2H2 Zinc Finger Protein Family Under Abiotic Stresses in Cucumber (Cucumis sativus L.)**. *Genes (Basel)* 2020, **11**(2).
26. Gu R, Song X, Liu X, Yan L, Zhou Z, Zhang X: **Genome-wide analysis of CsWOX transcription factor gene family in cucumber (Cucumis sativus L.)**. *Sci Rep* 2020, **10**(1):6216.
27. Wang X, Zhang H, Nyamesorto B, Luo Y, Mu X, Wang F, Kang Z, Lagudah E, Huang L: **A new mode of NPR1 action via an NB-ARC-NPR1 fusion protein negatively regulates the defence response in wheat to stem rust pathogen**. *New Phytol* 2020, **228**(3):959–972.
28. Xu X, Ji J, Xu Q, Qi X, Chen X: **Inheritance and quantitative trail loci mapping of adventitious root numbers in cucumber seedlings under waterlogging conditions**. *Mol Genet Genomics* 2017, **292**(2):353–364.
29. Pan Y, Wang Y, McGregor C, Liu S, Luan F, Gao M, Weng Y: **Genetic architecture of fruit size and shape variation in cucurbits: a comparative perspective**. *Theor Appl Genet* 2020, **133**(1):1–21.
30. Yu J, Tehrim S, Wang L, Dossa K, Zhang X, Ke T, Liao B: **Evolutionary history and functional divergence of the cytochrome P450 gene superfamily between Arabidopsis thaliana and Brassica species uncover effects of whole genome and tandem duplications**. *BMC Genomics* 2017, **18**(1):733.
31. Mizutani M, Ohta D: **Diversification of P450 genes during land plant evolution**. *Annu Rev Plant Biol* 2010, **61**:291–315.
32. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J: **KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging**. *Genomics, Proteomics & Bioinformatics* 2006, **4**(4):259–263.
33. Jayakannan M, Bose J, Babourina O, Shabala S, Massart A, Poschenrieder C, Rengel Z: **The NPR1-dependent salicylic acid signalling pathway is pivotal for enhanced salt and oxidative stress tolerance in Arabidopsis**. *J Exp Bot* 2015, **66**(7):1865–1875.
34. Martin Y, Gonzalez YV, Cabrera E, Rodriguez C, Siverio JM: **Npr1 Ser/Thr protein kinase links nitrogen source quality and carbon availability with the yeast nitrate transporter (Ynt1) levels**. *J Biol Chem* 2011, **286**(31):27225–27235.
35. Pu GB, Ma DM, Chen JL, Ma LQ, Wang H, Li GF, Ye HC, Liu BY: **Salicylic acid activates artemisinin biosynthesis in Artemisia annua L.**. *Plant Cell Rep* 2009, **28**(7):1127–1135.
36. Chen Q, Zhang R, Li D, Wang F: **Integrating Transcriptome and Coexpression Network Analyses to Characterize Salicylic Acid- and Jasmonic Acid-Related Genes in Tolerant Poplars Infected with Rust**. *Int J Mol Sci* 2021, **22**(9).
37. Zhang X, Dong J, Liu H, Wang J, Qi Y, Liang Z: **Transcriptome Sequencing in Response to Salicylic Acid in Salvia miltiorrhiza**. *PLoS One* 2016, **11**(1):e0147849.
38. Nelson D, Werck-Reichhart D: **A P450-centric view of plant evolution**. *Plant J* 2011, **66**(1):194–211.
39. Ma B, Luo Y, Jia L, Qi X, Zeng Q, Xiang Z, He N: **Genome-wide identification and expression analyses of cytochrome P450 genes in mulberry (Morus notabilis)**. *J Integr Plant Biol* 2014, **56**(9):887–901.

40. Jiu S, Xu Y, Wang J, Wang L, Liu X, Sun W, Sabir IA, Ma C, Xu W, Wang S *et al*: **The Cytochrome P450 Monooxygenase Inventory of Grapevine (Vitis vinifera L.): Genome-Wide Identification, Evolutionary Characterization and Expression Analysis**. *Front Genet* 2020, **11**:44.
41. Paquette SM: **Intro-exon organization and phylogeny in a large superfamily,the paralogous cytochrome P450 genes of Arabidopsis thaliana**. 2000
42. Wei K, Chen H: **Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice**. *BMC Genomics* 2018, **19**(1):35.
43. Wen H, Chen Y, Du H, Zhang L, Zhang K, He H, Pan J, Cai R, Wang G: **Genome-Wide Identification and Characterization of the TCP Gene Family in Cucumber (Cucumis sativus L.) and Their Transcriptional Responses to Different Treatments**. *Genes (Basel)* 2020, **11**(11).
44. Zhang S, Zhou Q, Chen F, Wu L, Liu B, Li F, Zhang J, Bao M, Liu G: **Genome-Wide Identification, Characterization and Expression Analysis of TCP Transcription Factors in Petunia**. *Int J Mol Sci* 2020, **21**(18).
45. Ai G, Zhang D, Huang R, Zhang S, Li W, Ahiakpa JK, Zhang J: **Genome-Wide Identification and Molecular Characterization of the Growth-Regulating Factors-Interacting Factor Gene Family in Tomato**. *Genes (Basel)* 2020, **11**(12).
46. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K *et al*: **The Pfam protein families database**. *Nucleic Acids Res* 2010, **38**(Database issue):D211-222.
47. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server**. *Bioinformatics* 2015, **31**(8):1296–1297.
48. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. *Mol Biol Evol* 2016, **33**(7):1870–1874.
49. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity**. *Nucleic Acids Res* 2012, **40**(7):e49.
50. Hao L, Qiao X: **Genome-wide identification and analysis of the CNGC gene family in maize**. *PeerJ* 2018, **6**:e5816.
51. Lv D, Wang G, Xiong LR, Sun JX, Chen Y, Guo CL, Yu Y, He HL, Cai R, Pan JS: **Genome-Wide Identification and Characterization of Lectin Receptor-Like Kinase Gene Family in Cucumber and Expression Profiling Analysis under Different Treatments**. *Genes (Basel)* 2020, **11**(9).
52. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation**. *Nat Biotechnol* 2010, **28**(5):511–515.

Figures

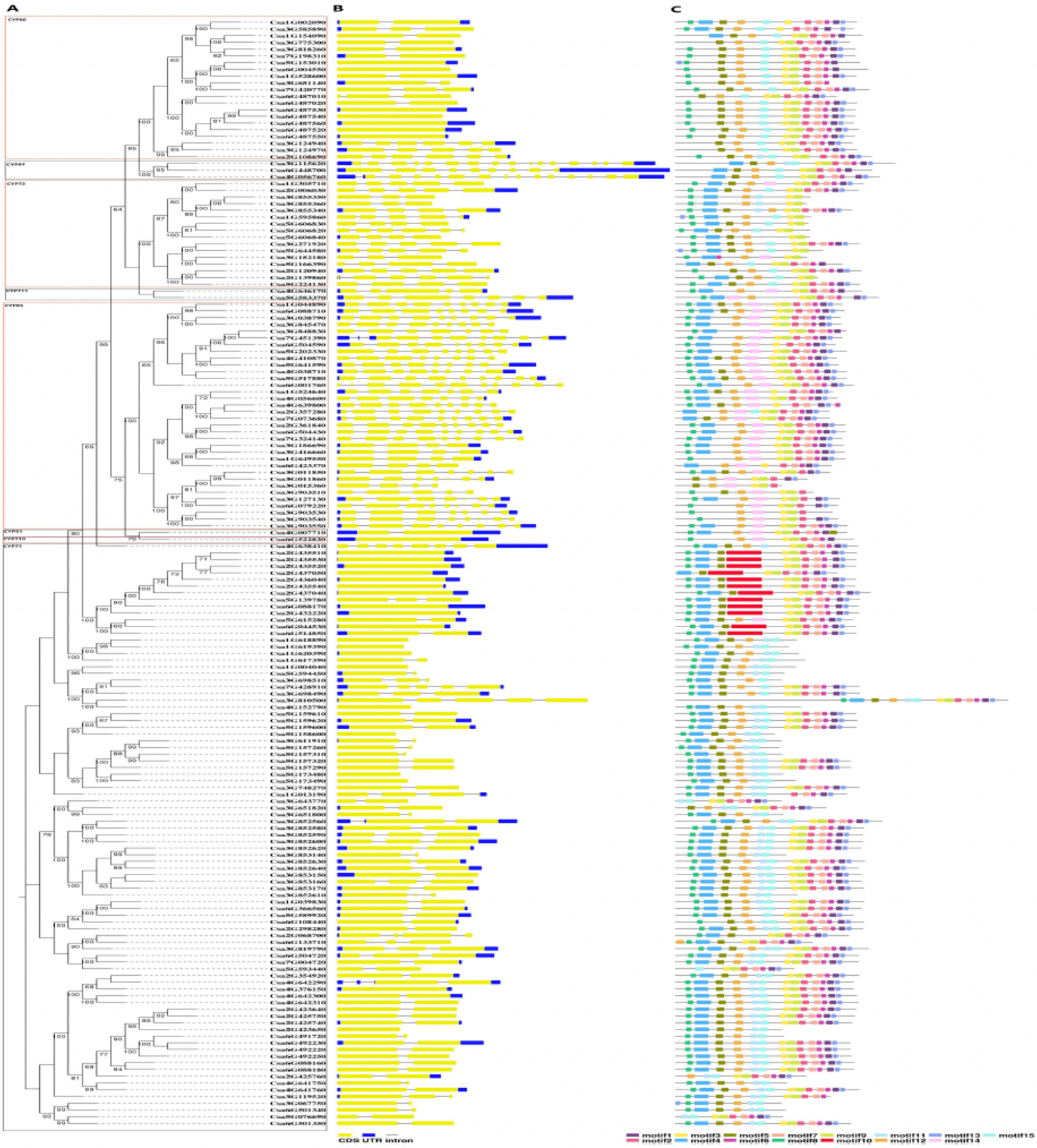

Figure 1

Phylogenetic analysis (A), motif analysis (B) and conservative motif (C) analysis of cucumber P450 members.

A phylogenetic tree was constructed using MAGA7 for the full-length amino acid sequences of 197 CsCYP450 proteins, and Bootstrap was set to 1000 times; b. Exon/intron analysis. Yellow squares

represent exons, black lines represent introns, and green squares represent upstream/downstream regions of the gene; c. CsCYP450 protein motif, numbered 1-15, are shown in squares of different colors.The sequence information of each motif is shown in Table S2.

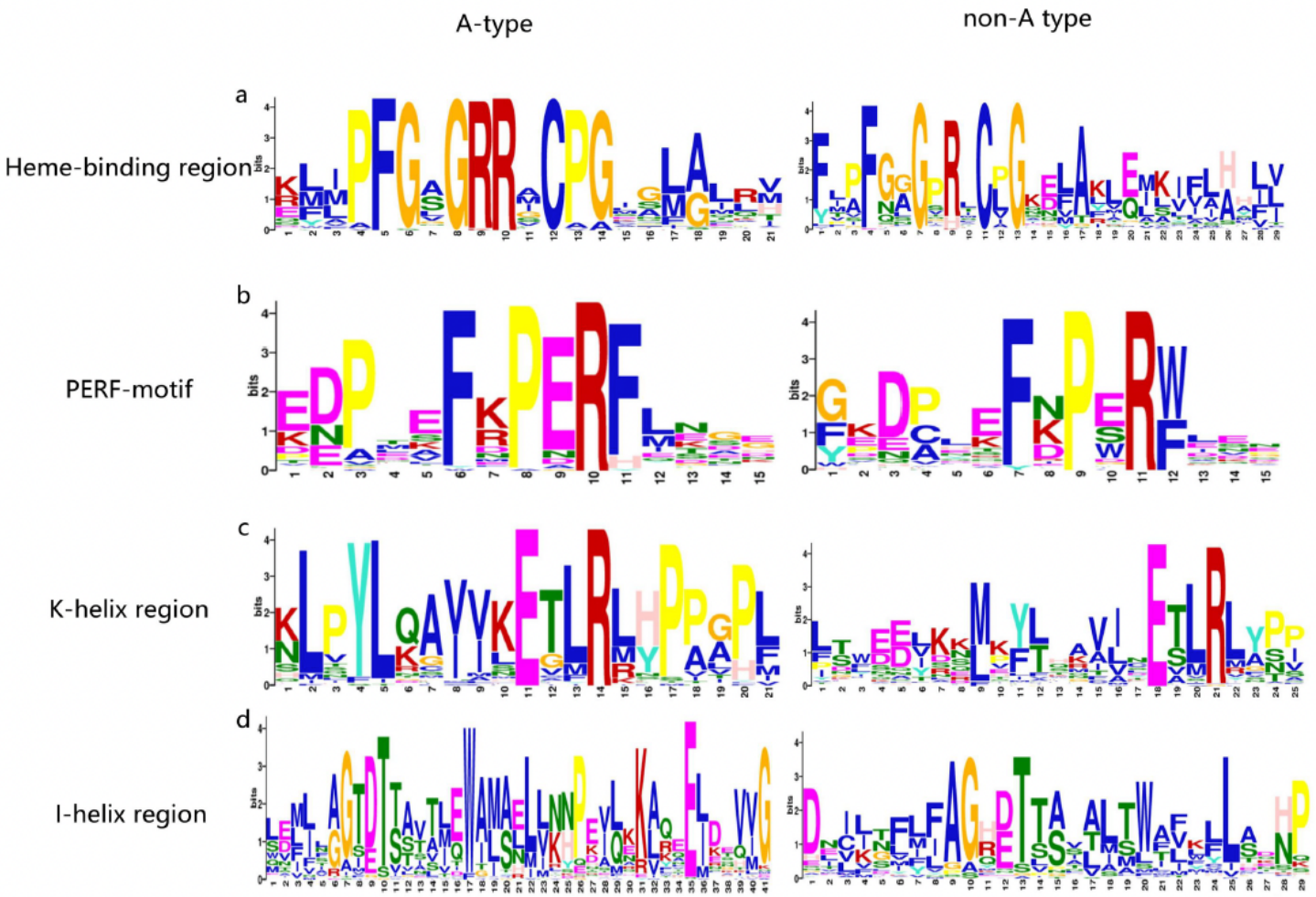

Figure 2

Comparative Analysis of A-type P450 Gene and non-A-type P450 motif patterns in cucumber

a. Heme-binding region, b. PERF/W motif, c. K-helix region(KETLR), d. I-helix region(KETLR)

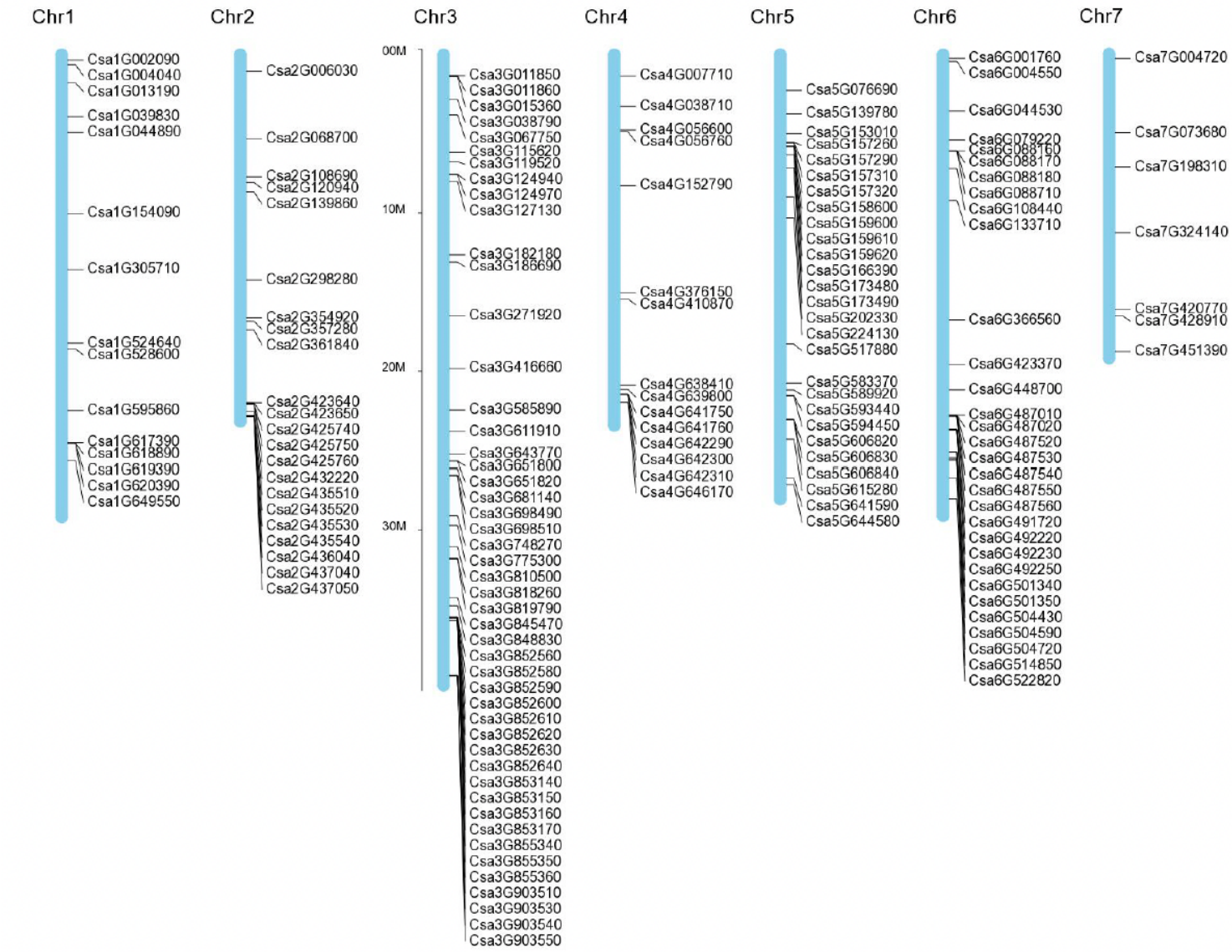

Figure 3

Chromosome mapping of P450 genes in cucumber

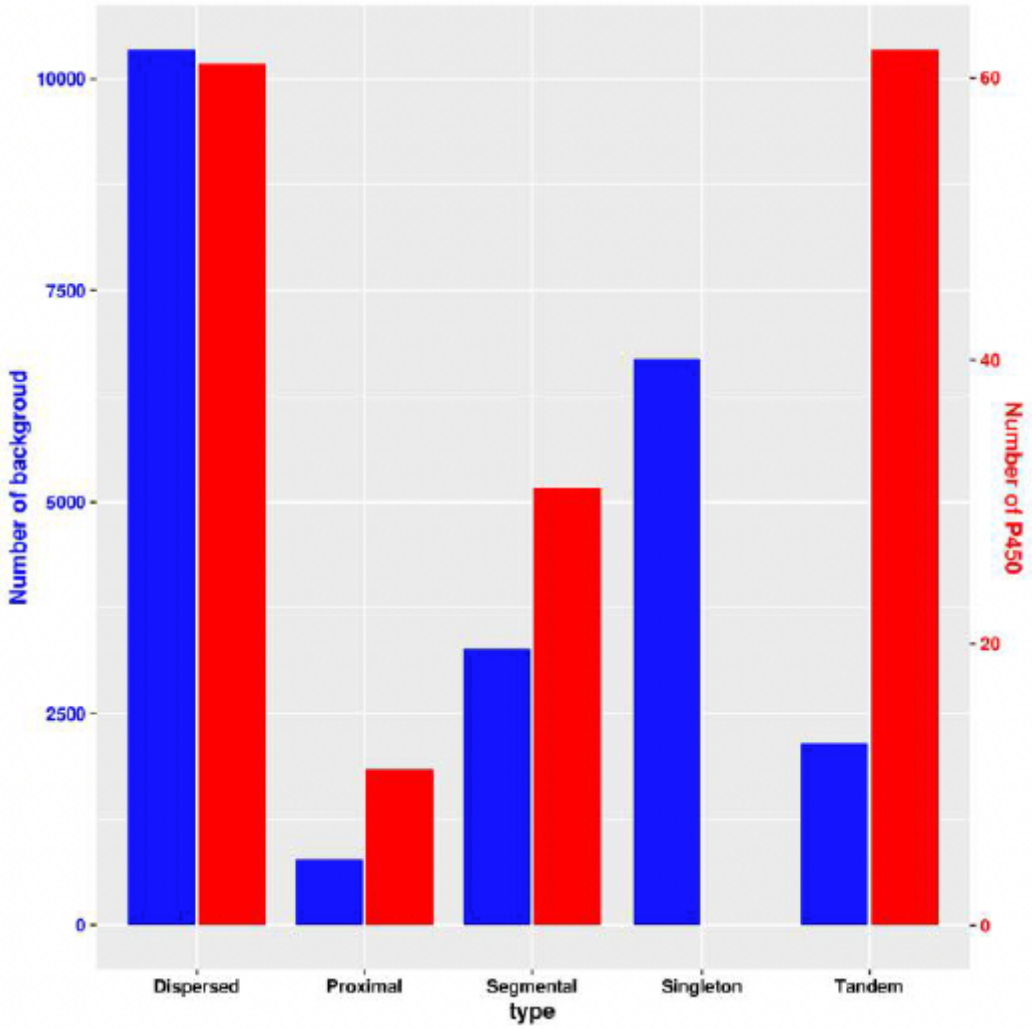

Figure 4

Statistics on the number of each type of replication genes

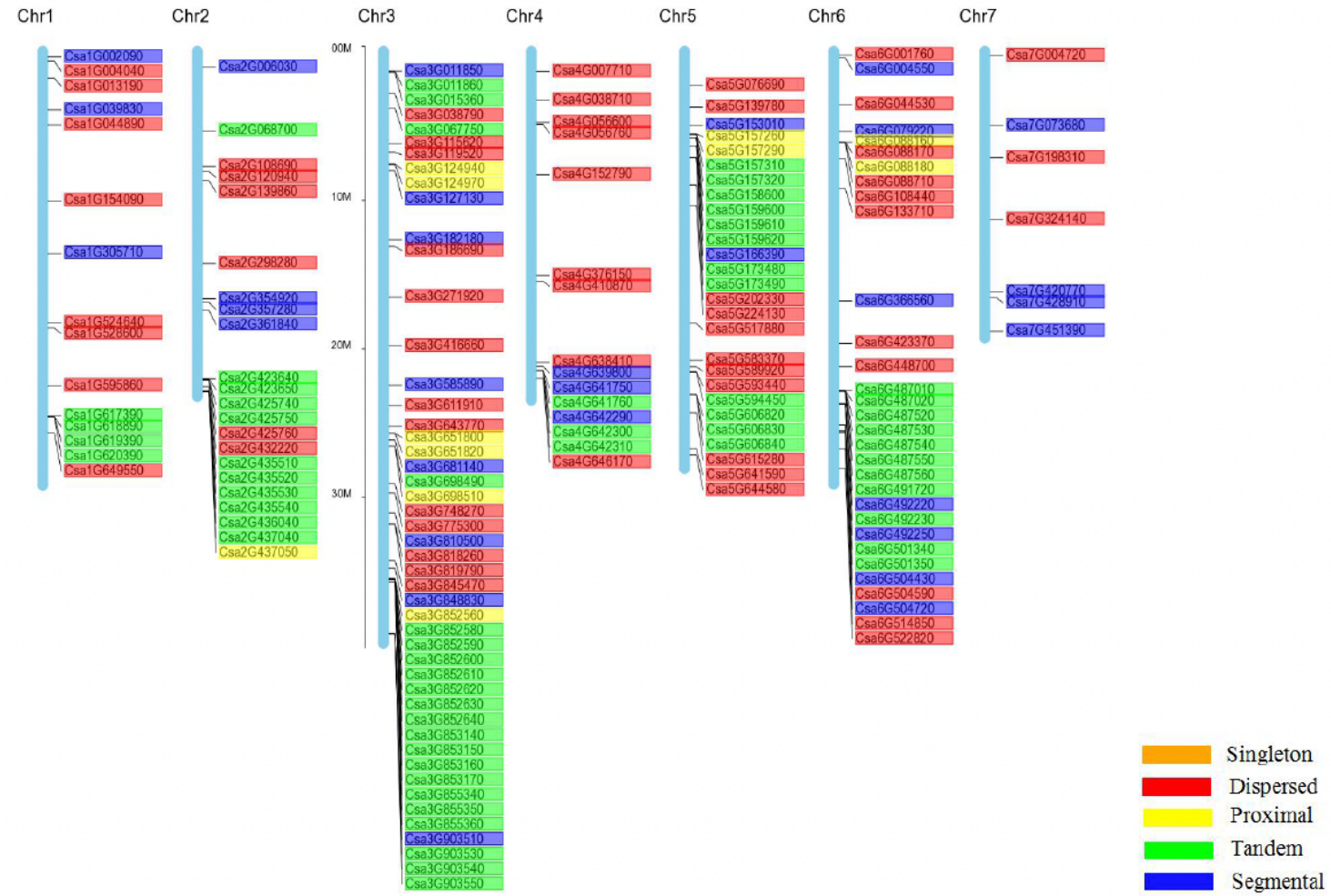

Figure 5

Distribution map of replication gene types of P450protein family on chromosomes in cucumber

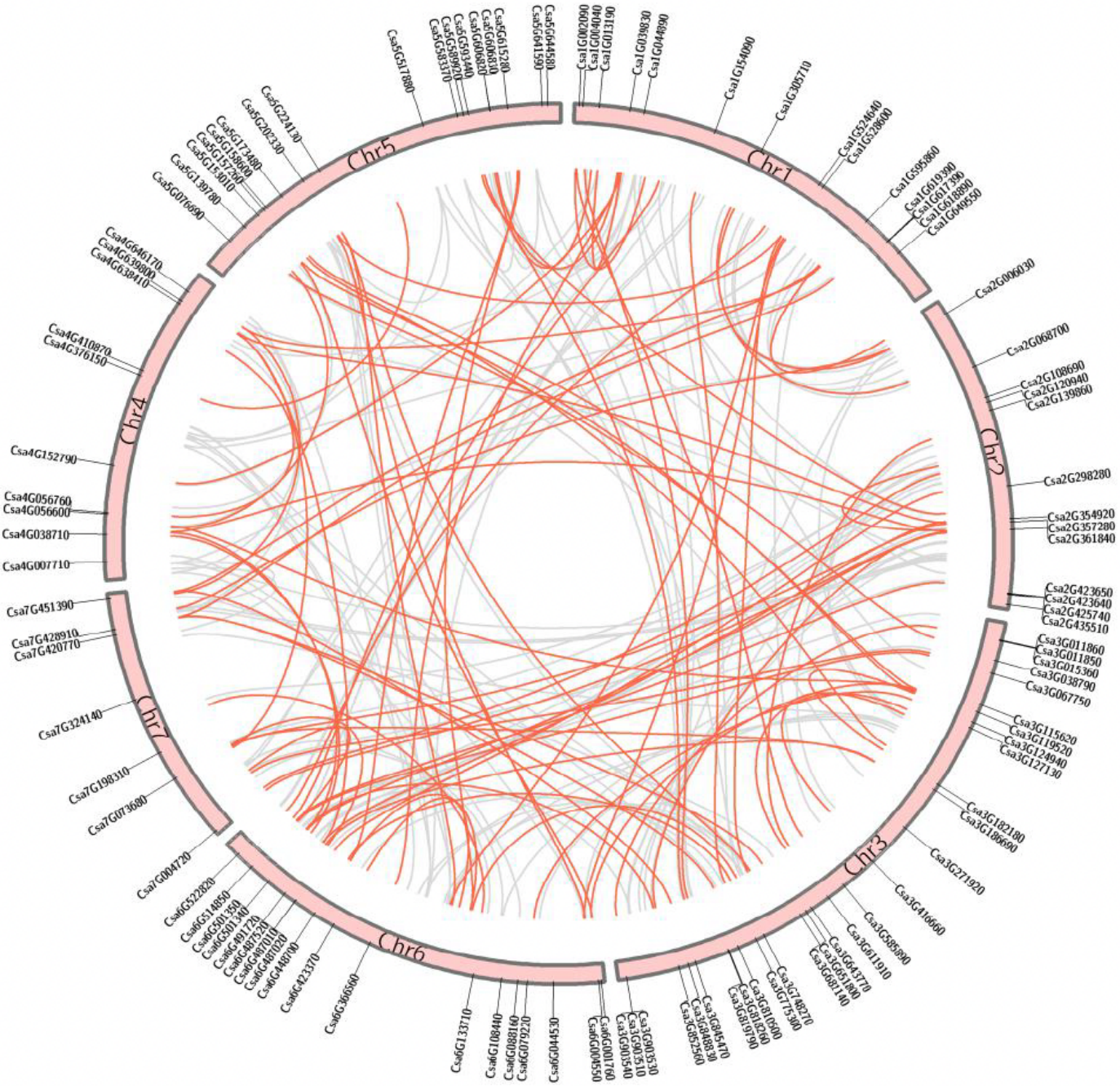

Figure 6

The collineatity of replication gene of p450 protein family in cucumber

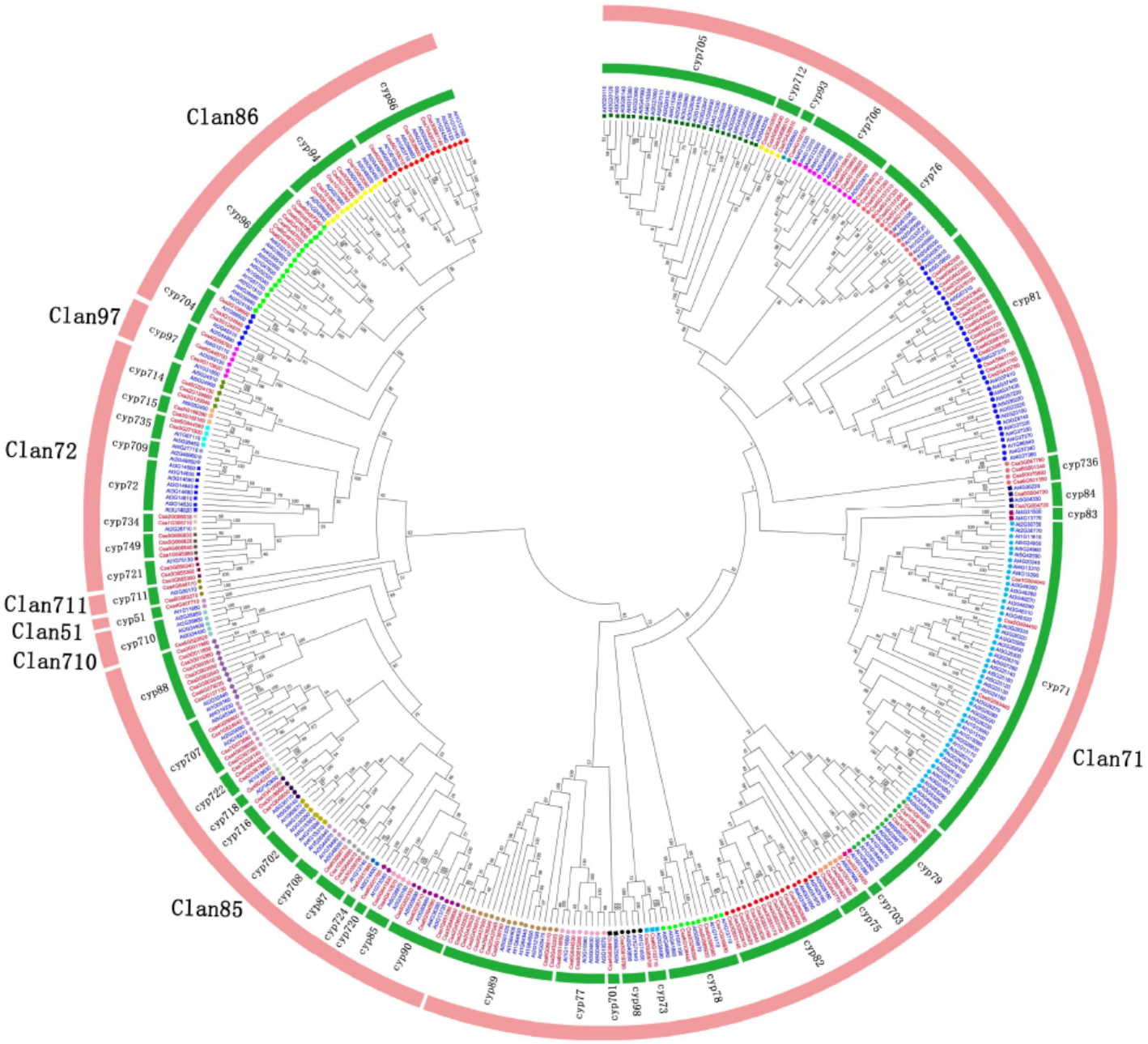

Figure 7

Phylogenetic tree of P450 gene family in cucumber and Arabidopsis

Note: Cucumber genes ID are shown in red, Arabidopsis genes ID are shown in blue.

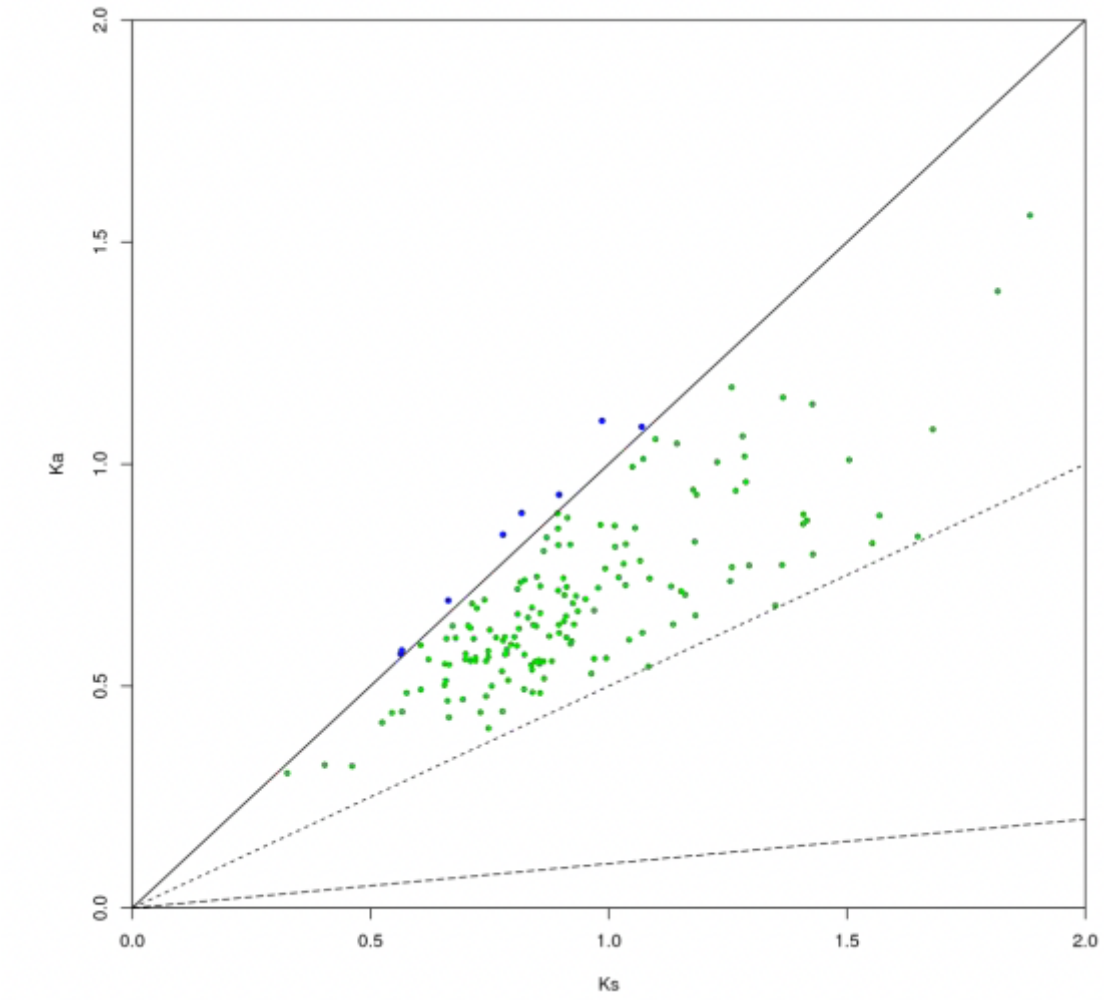

Figure 8

Ka/Ks scatter plot

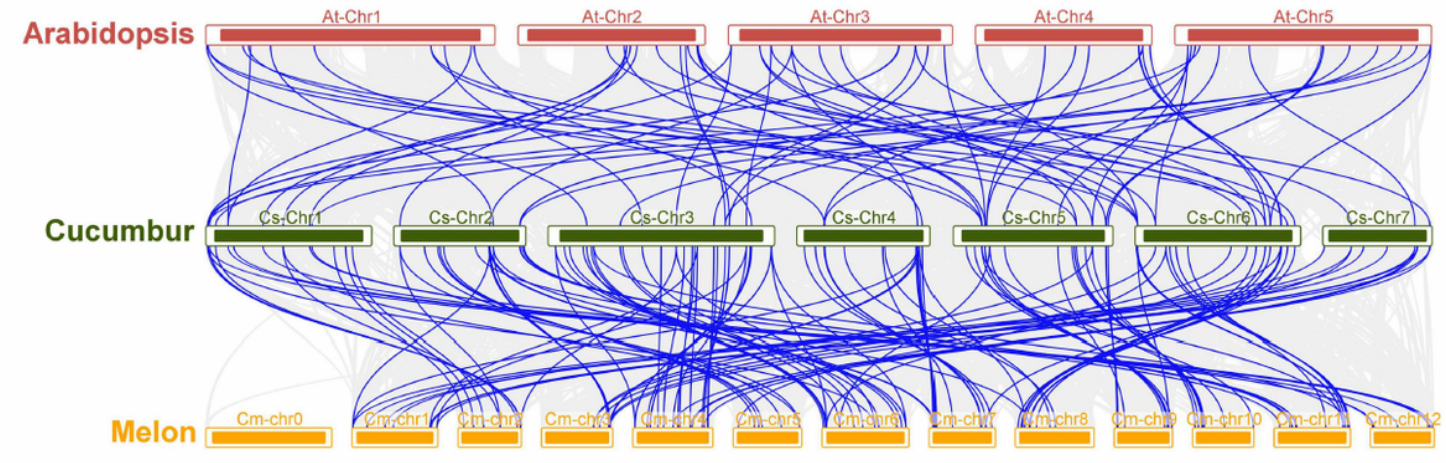

Figure 9

Collinear analysis diagram

Note: The gray line in the background represents the collinear block between cucumber and Arabidopsis, and the blue line represents the collinear CsCYP gene pair.

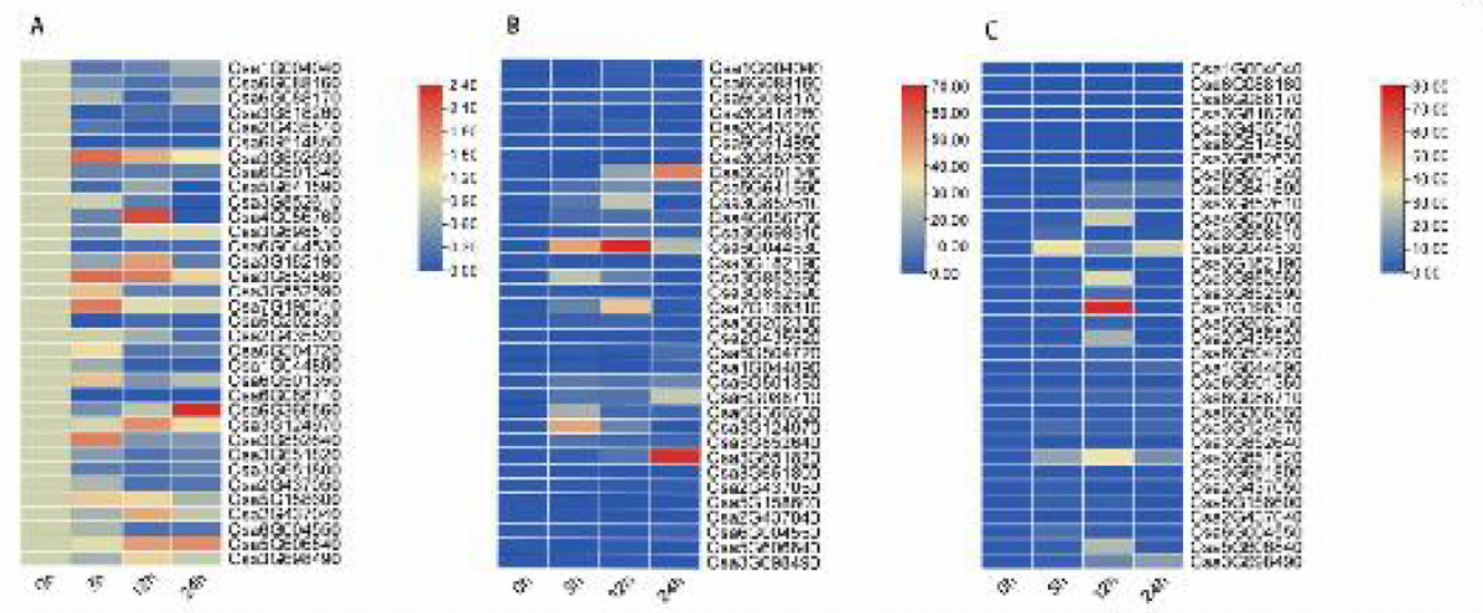

Figure 10

The expression pattern of 35 CsCYP450 genes in leaf, stem and root in cucumber by qRT-PCR

Supplementary Files

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

- Appendix.docx
- Fig.S1.png