Identification, evolution and alternative splicing profile analysis of Serine/Arginine-Rich Protein Splicing Factors (SR Proteins) in poplar, Arabidopsis, grape, and papaya

Hao Wang
Anhui Agricultural University
Hanwei Yan
Anhui Agricultural University
Ting He
Anhui Agricultural University
Yamei Wang
Anhui Agricultural University
Ruijia Wang
Anhui Agricultural University
Yan Xiang (xiangyanahau@sina.com)
Anhui Agricultural University

Research

Keywords: Intron retaining, Serine/Arginine-Rich, abiotic stresses, alternative splicing, gene duplication

Posted Date: December 30th, 2019

DOI: https://doi.org/10.21203/rs.2.19716/v1

Abstract

Alternative splicing (AS) regulates gene expression and produces proteome diversity. Serine/Arginine-Rich Protein Splicing Factors (SR Proteins) are important splicing factors that play significant roles in spliceosome assembly and splicing regulation, and play roles in regulating plant stress. In this report, we analyzed 30 SR genes in Populus trichocarpa, 18 genes in Arabidopsis thaliana, 14 genes in Vitis vinifera and 9 genes in Carica papaya. The SR proteins contained RRM and RS conserved domains, and based on different structural domain organization were divided into six subfamilies (SR, SC, SCL, RS, RSZ and RS2Z). Gene duplication analysis revealed 94 paralogs and 408 orthologs in the four species, and the SR genes had undergone strong purifying selection. A number of stress-related cis-elements (ABRE, LTR, MBS, TC-rich repeats cis-acting element) were identified in the promoters of the SR genes. Microarray and RNA-seq data showed that SR genes expression in different tissues of the four species responded differently to abiotic stress. Poplar, Arabidopsis and grape SR genes had many splice isoforms. Moreover, 26 of 30 poplar SR genes had intron retention (IR) events, and the relative IR rates of 27 intron sites in the poplar SR genes changed significantly under cold, heat, drought and salt stress conditions. This study provides valuable resources for the gene structure, function, and evolution of poplar SR proteins.

Background

Alternative splicing (AS) is the process of splicing pre-mRNA into different mature mRNAs by using different splicing types and splice sites [1]. The resulting multiple mature mRNAs are translated into different proteins, which increases the diversity of the proteome and the flexibility of gene expression regulation [2]. The pre-mRNAs in eukaryotes are spliced by the spliceosome. And the spliceosome is a large RNA-protein complexes, consisting of five small nuclear ribonucleoprotein particles (snRNPs), U1, U2, U4/U6 and U5, and a large number of non-snRNP proteins [3, 4, 5]. SR proteins are part of the splice complexes (non-snRNP proteins) and play a key role in constitutive splicing and pre-mRNA AS. All SR proteins contain one or two RNA recognition motifs (RRM) in the N-terminal region of the protein that interact with pre-mRNA. And the SR proteins have an RS domain rich in serine/arginine repeats, which plays an important role in the regulation of RNA splice assembly and AS [6, 7, 8]. In addition, the RS domain is related to nuclear and subnuclear localization of the SR genes [9]. Notably, translation of AT1G16610 (AtSR45) reveal that the protein contained two RS domains at the N- and C-termini, and an RRM in the middle, and thus are not classical SR proteins [10].

The plant SR proteins are first identified in Arabidopsis [11, 12]. There are 18, 20, 21, and 18 SR proteins in Arabidopsis, rice (Oryza sativa), maize (Zay mays), and sorghum (Sorghum bicolor) respectively. Based on different structural architectures [10], the SR proteins are divided into six subfamilies: SR, RSZ, SC, SCL, RS and RS2Z. Among these six subfamilies, the latter three are specific to plants [13]. The SC subfamily proteins have a single RRM. The SCL subfamily (SC35-like) proteins are similar to the SC subfamily but have an N-terminal extension that is rich in Arg, Pro, Ser, Gly and Tyr residues. The SR subfamily proteins have two RRMs with an evolutionarily conserved SWQDLKD motif in their second RRM followed. The RS subfamily proteins contain two RRMs (without the SWQDLKD motif). The RSZ

subfamily consist of SR proteins with one Zn knuckle. The RS2Z subfamily proteins have two Zn knuckles and an additional SP-rich region after the RS domain [10].

As components of the spliceosome, SR proteins play an important role in spliceosome assembly and splicing regulation. In animals, three SR proteins (SC35, SRp38 and ASF/SF2) interact with U1-70K (the subunits of U1 snRNP), resulting in 5' and 3' play a role in the selection of splice sites [14]. In Arabidopsis, a large number of SR proteins (AtRSZ21, AtRSZ22, AtSR34, AtSR33, AtSC35-like and AtSC35) interact with U1-70K and U2AF65a (the subunits of U1 and U2 snRNPs, respectively) [9, 15]. Overexpression of AtSRp30 results in changes in the AS of multiple genes including AtSRp30 itself [16]. Similarly, OsRSp29 and OsRSZp23 play a role in pre-mRNA splicing [17]. AtSC35 plays a key role in regulating the flowering locus C, which regulates flowering in Arabidopsis [9].

The SR genes are extensively alternatively spliced genes. In Arabidopsis, Palusa et al. find that 15 Arabidopsis SR genes produce 95 transcripts under hormone or abiotic stresses. The cold, heat and hormone stresses significantly alter the AS of several Arabidopsis SR genes, indicating that they are involved in response to environmental stresses [18]. At the seedling stage, 13 Arabidopsis SR genes produce 75 transcripts, 53 of which contain premature termination codons (PTCs) [19]. In maize and sorghum, 92 and 62 SR transcripts are detected respectively [20]. Arabidopsis SR34b is involved in cadmium resistance [21]. However, the detailed function of most plant SR genes and the importance of a large number of AS events remain unclear.

Poplar, Arabidopsis, grape, and papaya belong to the dicot. Poplar, grape and papaya are important economic species. Arabidopsis is a classic model plant. In this report, we analyzed 30 SR genes in Populus trichocarpa (poplar), 18 genes in Arabidopsis thaliana (Arabidopsis), 14 genes in Vitis vinifera (grape) and 9 genes in Carica papaya (papaya). Then we conducted comprehensive analysis of the SR genes from poplar, Arabidopsis, grape, and papaya, including basic characterization, phylogenetic analysis, conserved motifs, gene duplication analysis, promoter analysis, expression patterns, and splice isoforms. To analyze the relationship between AS and abiotic stress, we analyzed IR rates and the relative IR rates in poplar SR genes under cold, heat, drought and salt stresses. This study provides valuable resources for the gene structure, evolution, and function of poplar SR proteins.

Methods

SR genes identification and data collection

According to the method described by Barta and Kalyna [10], 18 Arabidopsis and 22 rice SR protein sequences were used as the query in a BLASTp (e-value cutoff = 1 e − 10) search to identify poplar, grape, and papaya SR protein sequences from the phytozome database (http://www.phytozome.net) [22], and grape genome database (http://genomes.cribi.unipd.it/grape/) [23]. SMART (http://smart.embl-heidelberg.de) [24] and PFAM (http://pfam.janelia.org/) [25] were used to confirm whether the candidate sequence had one or two N-terminal RRMs (RBD; PF00076), and the sequences were manually confirmed to have at least 50 amino acids in the downstream sequence with 20% of the RS or SR dipeptide [10].

Fifty-three sequences were identified from poplar, grape and papaya. Based on different structural architectures [10], the SR proteins were subdivided into six subfamilies (Figure S1). For genes with different transcripts, we used the primary transcript specified by phytozome (http://www.phytozome.net) [22] and grape genome database (http://genomes.cribi.unipd.it/grape/) [23] for analysis. The physicochemical characteristics of each SR gene, including gene ID, NCBI accession number, gene length, intron number, transcript number, CDS length, coding protein length, protein number, molecular weight (MW) and isoelectric point (pI) of the coded protein were acquired from the phytozome database (http://www.phytozome.net) [22], grape genome database (http://genomes.cribi.unipd.it/grape/) [23], National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) and ExPASy (http://www.expasy.ch/tools/pi_tool.html) [26]. Prediction information for subcellular localization was obtained from WoLF PSORT (https://wolfpsort.hgc.jp) [27].

Phylogenetic tree and conserved motifs analysis in poplar, Arabidopsis, grape, and papaya

The amino acid sequences of the aforementioned 71 SR genes were used for phylogenetic analysis. For genes with different transcripts, the primary transcript specified by phytozome and grape genome database was selected. Multiple sequence alignments of all SR proteins were carried out in ClustalX 1.83 with default settings [28]. A phylogenetic tree for all of the complete SR protein sequences was built using MEGA 7.0 with the neighbor-joining (NJ) method and 1000 bootstrap replicates [29]. In addition, the full-length protein sequences were submitted to Multiple Expectation Maximization for Motif Elicitation (MEME) (http://meme.sdsc.edu/meme/itro.html) [30] with the following parameters: an optimum width of 6–200 residues and a maximum number of 10 motifs. The function of each motif was verified in PFAM (https://www.ebi.ac.uk/Tools/hmmer/results/) [25] and SMART (http://smart.embl-heidelberg.de) [24].

Gene structure analysis in poplar, Arabidopsis, grape, and papaya and AS profile analysis in poplar, Arabidopsis, and grape

Exon/intron position information for all SR gene available alternative transcripts and proteins were obtained from the phytozome database (http://www.phytozome.net) [22], and grape genome database (http://genomes.cribi.unipd.it/grape/) [23]. The exon/intron structures were determined with the Gene Structures Display Server (GSDS) (http://gsds.cbi.pku.edu.cn/) [31]. Import the protein sequences of all available alternative transcripts into the MEME website (http://meme-suite.org/tools/meme) [30] to analyze the conserved motifs. And the same protein isoform conserved motif analysis was only shown once.

Paralogs and orthologs of SR genes in poplar, Arabidopsis, grape, and papaya

To identify paralogs and orthologs of four dicots, the method was as follows: all local BLASTP was first searched for all potential duplicated gene pairs using the four dicots protein sequence (E < 1e-10, top 5 matches and m8 format output). As a result, orthologs and paralogs were obtained. The paralogs were then classified using MCScanX-transposed, including the WGD, TD, PD, TRD, and DSD. PAML 4.0 [32] was

used to calculate Ka and Ks for duplicated gene pairs. Then the duplicated genes were divided into three types by calculating the rates of nonsynonymous substitution (Ka) and the rates of Ks. Ka/Ks > 1 was positively selected genes (PSGs), Ka/Ks = 1 was neutral genes, and Ka/Ks < 1 was negatively (or purifying) selected genes (NSGs).

Cis-acting element analysis in poplar, Arabidopsis, grape, and papaya

Information on the 2.0-kbp sequences upstream of the transcriptional initiation sites of the SR genes of the four dicots was obtained from the phytozome database (http://www.phytozome.net/populus.php) [22], and grape genome database (http://genomes.cribi.unipd.it/grape/) [23]. The cis-acting regulatory elements were then predicted among the sequences of putative promoter regions using Plant CARE (http://bioinformatics. psb.ugent.be/webtools/plantcare/html/) [33]. Motifs with specific functions associated with abiotic stress were selected in this study.

IR in poplar SR genes

We obtained IR information about poplar SR genes under hot, cold, drought, salt in leaf, root and xylem from RNA-seq data [34]. Under certain stress, if the relative IR rate (stress treated samples vs. untreated control samples) (IRR_ratio_diff) of an intron site of some SR genes changed by more than 30%, we mapped the relative IR rate of this intron site into a heat map using R (R is a free software environment for statistical computing and graphics) (Fig. 8).

Expression pattern analysis of SR genes in poplar, Arabidopsis, grape, and papaya

To explore the expression level characteristics of SR genes under different abiotic stresses in different tissues, we obtained the expression level data of the poplar and papaya SR genes from RNA-seq data [34, 35], and Arabidopsis and grape chip data GSE5620, GSE5621, GSE5623, GSE5624, GSE5628 and GSE31594 were obtained from the NCBI GEO database. The log2-transformed fold-change values were used for creating the heatmap.

Statistical analyses

The Pearson correlation tested determined the correlation between two independent variables. A one-way ANOVA test was performed to analyze the significance in the correlation analysis. Among all statistical tests, a P-value of < 0.05 was considered to be significant, and a P-value of < 0.01 was considered to be extremely significant. And if the correlation coefficient was < 0.1, there was no correlation, even if the correlation was significant. Note: * indicated significance at $P < 0.05$; ** indicated significance at $P < 0.01$.

Results

Identification and characterization of SR genes in poplar, Arabidopsis, grape, and papaya

According to the method described by Barta and Kalyna [10], 18 Arabidopsis and 22 rice SR protein sequences were used as the query in a BLASTp (e-value cutoff = 1 e − 10) search to identify poplar, grape, and papaya SR protein sequences from the phytozome database (http://www.phytozome.net) [22], and grape genome database (http://genomes.cribi.unipd.it/grape/) [23]. SMART (http://smart.embl-heidelberg.de) [24] and PFAM (http://pfam.janelia.org/) [25] were used to confirm whether the candidate sequence had one or two N-terminal RRMs (RBD; PF00076), and the sequences were manually confirmed to have at least 50 amino acids in the downstream sequence with 20% of the RS or SR dipeptide [10]. Finally, 30 genes in Populus trichocarpa, 14 genes in Vitis vinifera and 9 genes in Carica papaya were identified (Table S1). Basic information describing the primary transcript of each genes from poplar, Arabidopsis, grape, and papaya were listed in Table S1, and include gene ID, NCBI accession number, gene length, CDS length, coding protein length, pI, MW and the prediction of their subcellular location (Table S1). In poplar, the SR genes were located on 14 of 19 chromosomes with no distribution on chromosomes 4, 7, 9, 11 and 17. Among them, the most poplar SR genes distributed on chromosome 2. Arabidopsis SR genes were distributed on all chromosomes. Grape SR genes were distributed on 11 of 19 chromosomes (i.e., 1, 4, 6, 7, 8, 12, 13, 14, 15, 16 and 18) (Table S1). The statistical data showed that VvSCL34 had the longest gene length (18119 bp). CpRS50 coded for the longest protein (447 aa), whereas PtRSZ20 (180 aa) coded for the shortest protein among the 71 SR proteins identified. The pI values of six proteins in the 71 SR protiens were less than 10, whereas that of the remaining proteins were more than 10. Furthermore, the MW of these proteins ranged from 20.46 to 50.42 kDa with an average of 30.46 kDa. Predicted subcellular localization of the SR proteins indicated that most of these proteins were located in the nuclear, which was consistent with their putative roles as splicing factors.

Phylogenetic tree and conserved motifs analysis in poplar, Arabidopsis, grape, and papaya

Based on different structural architectures (Figure S1), the SC subfamily contained proteins with a single RRM followed by an RS domain. The SR subfamily proteins had two RRMs with an evolutionarily conserved SWQDLKD motif in their second RRM followed by an RS domain with a characteristic SR dipeptide. The RSZ subfamily consisted of SR proteins with one Zn knuckle. The plant-specific SCL subfamily (SC35-like) was similar to the SC subfamily but had an N-terminal extension that is rich in Arg, Pro, Ser, Gly and Tyr residues. Proteins of the plant-specific RS2Z subfamily had two Zn knuckles and an additional SP-rich region after the RS domain. The plant-specific RS subfamily contained two RRMs (without the SWQDLKD motif) followed by an RS-rich with many RS dipeptides [10].

And the protein sequences of the 71 putative SR genes were used to construct a phylogenetic tree to study the evolutionary relationship between these proteins (Fig. 1a). For genes with different transcripts, the primary transcript specified by phytozome and grape genome database was selected. The phylogenetic tree showed a similar pattern of evolutionary divergence among poplar, Arabidopsis, papaya and grape, which reflected conserved evolution and function in different SR genes. The SCL subfamily constituted the largest clade, each containing 18 members and accounting for 25.4% of the total SR genes, whereas RS2Z had the lowest number of members (seven members of each group) and accounted for 9.8% of the total SR genes. SR genes in poplar, Arabidopsis, grape and papaya were distributed in all

six groups. In poplar, seven genes belonged to the SR subfamily, eight to the SCL subfamily, six to the RS subfamily, three to the RSZ subfamily, two to the RS2Z subfamily and four to the SC subfamily.

We analyzed the conserved motifs of SR proteins using the online MEME software, as shown in Fig. 1b. Ten specific motifs were detected and detail information was provided in Table S2. Each assumed pattern was annotated by searching PFAM and SMART. Motifs 1, 2, and 5 were found to encode the N-terminal RRMs (RBD; PF00076), whereas motif 3 encoded proteins of RRM with an evolutionarily conserved SWQDLKD motif. Since the RRM containing an evolutionarily conserved SWQDLKD motif was only in the SR subfamily, we found that only the SR subfamily had motif 3. Motif 7 and 9 were the RS domain of each SR gene. Other motifs identified had unknown functions. As expected, members of the same subfamily had highly similar motifs, for example, the 3, 6, and 10 Motifs existed only in the SR subfamily, and the motif 5 existed only in the RS subfamily. It was suggested that SR proteins of the same subfamily had functional similarities.

Gene structure in poplar, Arabidopsis, grape, and papaya and AS profile analysis in poplar, Arabidopsis, and grape

For intron/exon structure analysis (Fig. 2b, Fig. 2c), we found that the intron number of SR genes was 4 ~ 13 in general, and the SR subfamily had more introns, the distribution of intron number was 10 ~ 13. Second, the number of introns in the SC subfamily was between 6 and 9. We found that 10, 7 and 5 SR genes had 5 introns in the RS, SCL and RSZ subfamilies, respectively, accounting for 76.9%, 38.8%, and 55.5% of all genes in the subfamily. There were 5 genes in the RS2Z subfamily with 6 introns, accounting for 71.4% of all RS2Z subfamily genes. It was worth noting that some genes contain introns in their 5' or 3' untranslated (UTR) regions. An extreme example was that AtSR34b contained 4 introns in the 3' untranslated region.

Some researches have reported that AS profile is common in SR genes [18]. For each gene, the primary transcript was placed at the top and the other alternative transcripts were listed below using phytozome database and grape genome database annotation files. And the same protein isoform conserved motif analysis was only shown once (Fig. 3, Figure S2, Figure S3). Since papaya had no annotated multi-transcript information, it was not analyzed here. The results showed that a total of 79 transcripts were detected in 30 genes from poplar (Fig. 2, Fig. 3). PtSR34b and PtSCL29a had the most transcripts, i.e., 5. A total of 68 transcripts were detected in 18 genes from Arabidopsis thaliana with AtRS40 having the highest number of transcripts, i.e., 8 (Figure S2). A total of 65 transcripts were detected in 14 genes in grape (Figure S2). VvSR28 was found to have the highest number (16) of transcripts (Figure S3). Multiple transcripts of genes could produce multiple different protein isoforms. For example, AtRS41 had seven transcripts that could produce four protein isoforms (Figure S2), and VvRS26 had ten transcripts that produce five protein isoforms (Figure S3). PtRS29b had three transcripts that produce two protein isoforms (Fig. 3). Two transcripts of PtRS29b had undergone AS in the 5' untranslated (UTR) region, indicating potential mutations that controlled their transcriptional or translational efficiency (Fig. 2, Fig. 3).

Paralogs and orthologs of SR genes in poplar, Arabidopsis, grape, and papaya

To further investigate the evolution of the SR family, duplicated gene pairs (paralogs and orthologs) analysis was used to investigate gene duplication events within the poplar, Arabidopsis, grape, and papaya. We identified 94 paralogs in four dicots (Fig. 4, Table S3). Specifically, there were 51 paralogs within poplar, 22 within Arabidopsis, 15 within grape, and 6 within papaya. In addition, 68 SR genes including 30 poplar SR genes, 14 grape SR genes, 18 Arabidopsis SR genes and six papaya SR genes were shown to have a duplicated relationship, accounting for 95.8% of all SR gene family members. Transposed duplication, and dispersed duplication were found in the paralogs of four dicots, and whole-genome duplication (WGD) was found in the paralogs of poplar, grape, and Arabidopsis. Tandem duplication and proximal duplication were not found in the four dicots (Fig. 4, Table S3). We next identified 408 orthologs among four dicots (Table S4). Specifically, there were 147 orthologs between Arabidopsis and poplar, 66 between Arabidopsis and grape, 44 between Arabidopsis and papaya, 67 between poplar and grape, 45 between poplar and papaya, and 39 between grape and papaya.

To study evolutionary selection process, K_a value, K_s value, and K_a/K_s ratios of all duplicated gene pairs were listed in Table S3 and Table S4. K_s value of poplar paralogs varied from 0.0155 to 4.9695 (Table S3). The K_s value of Arabidopsis, grape and papaya paralogs varied from 0.6421 to 2.9420, 0.9462 to 4.9368 and 1.3729 to 2.4849, respectively. The K_a/K_s ratios of poplar, Arabidopsis, grape and papaya duplicated gene pairs varied from 0.0555 to 1.4837, 0.0863 to 0.3738, 0.0644 to 0.4607 and 0.1128 to 0.4715, respectively. The results showed that all K_a/K_s ratios were smaller than 0.5 except that the ratio of PtSR34/PtSR33 pair was greater than 1, indicating that the SR genes had undergone strong purifying selection. In addition, for orthologs, we found that all K_a/K_s ratios were less than 1, orthologs had undergone positive selection (Table S4).

To further analyze the evolutionary relationship of the SR genes, we analyzed it by the K_s value of the duplicated gene pairs (K_s value could be used to determine the separation time of the duplicated genes). In poplar, according to Tang et al., the median K_s of the duplicated genes associated with the γ triplication event was 1.54, and the total K_s associated with P-WGD was 0.27 [36]. We detected 22 WGDs associated with amplification of the SR gene in the poplar genome. The values of K_s in different WGDs showed two different ranges (Table S3): 0.2249–0.4493, and 0.9350–1.7387. The former K_s range of duplicated genes might come from P-WGD event, the latter K_s range of duplicated genes might come from γ triplication event. In Arabidopsis, the median K_s value associated with β-WGDs and γ triplication event was close to the saturation median K_s value of 2.00 [36]. Therefore, based on the K_s value, β-WGDs and γ triplication event were indistinguishable. The median K_s associated with α-WGD was reported to be 0.86 [36]. The values of K_s in WGD showed one ranges (Table S3): 0.6421-1.040. This might indicate that these Arabidopsis WGD only experience α-WGD. In grape and papaya, the overall median K_s values of γ triplication event were 1.76 and 1.22, respectively [36]. Then, according to the K_s value of the orthologous gene, it could be divided into a paralogous gene formed before polyploidization and a paralogous gene

formed after polyploidization. So we mapped the evolutionary relationship of the SR genes of poplar, Arabidopsis, grape, and papaya (Fig. 5). The SR gene, which was not in the same line, indicated that the isolation of these genes was before the γ triplication event. For each row of SR genes of the corresponding species, they were produced by different WGDs. The results showed that most of the SR genes were lost after WGD. For example, in the γ triplication event, CpSC30 did not produce two other corresponding SR genes. Then the presence of two SR genes in the box indicated that the two genes were genes produced by other duplicated types after the WGD. PtSR29 and PtSR34b were SR genes produced by other duplicated types after WGD. In addition, we found that the duplicated genes generated by WGD and the duplicated genes generated by other recent duplicated types had highly similar gene structures, for example PtSR33, PtSR34, PtSR34c, which had the same number of transcripts, protein numbers, and similar intron numbers.

Cis-acting elements analysis in poplar, Arabidopsis, grape, and papaya

The key roles of cis-acting elements in the promoter region have affected the tissue-specific or stress-responsive expression patterns. In this study, we identified four cis-acting elements of environmental stress type, including those directly related to the ABA response element (ABRE), cis-acting element involved in defense and stress responsiveness (TC-rich repeats), low temperature responsive element (LTR) and MYB binding site involved in drought-inducibility (MBS). In poplar, six genes contained ABRE cis-acting elements, 17 contained LTR cis-acting elements, 15 contained MBS cis-acting elements and nine contained TC-rich repeats cis-acting elements. In Arabidopsis, papaya and grape, 32 contained ABRE cis-acting elements, 20 contained LTR cis-acting elements, 16 contained MBS cis-acting elements and 14 contained TC-rich repeats cis-acting elements (Fig. 6). These findings could aid further investigations into the stress-regulatory mechanisms of SR genes in plant.

Expression pattern analysis of SR genes in poplar, Arabidopsis, grape, and papaya

Since abiotic stress may adversely affect plant growth and development, stress-tolerant gene studies of plants are important. We obtained microarray (Arabidopsis and grape) and RNA-seq (poplar and papaya) data under various stress conditions from different tissues of the four dicots [34, 35]. The results showed that only the CpRSZ20 gene was not expressed, and the rest were expressed (Fig. 7). In poplar, we found that three poplar SR genes (PtSCL25, PtSR34a, PtSCL23) were significantly down-regulated (the value of log2-foldchange was larger than 2) in cold stress, whereas under heat stress three poplar SR genes (PtSR35, PtSCL25, PtSCL23) were up-regulated (the value of log2-foldchange was larger than 2) and the expression level of the two poplar SR genes (PtSR33, PtSR34) were declined. No significant changes were observed under drought and salt stress conditions (Fig. 7). Two papaya genes (CpSCL25, CpSCL35a) were significantly upregulated under drought stress. In Arabidopsis, the expression value of AtSCL28 was much lower than that of other Arabidopsis SR genes, but its expression value was significantly up-regulated under heat stress, and other Arabidopsis SR genes were not significantly changed (the value of log2-foldchange was larger than 2). In grape, no significant changes were observed in the expression levels of SR genes (Fig. 7).

In addition, we studied the correlation between transcript number, protein number, intron number, four cis-acting elements, and the expression values of SR genes (Figure S4). The results showed that no correlation was found in Arabidopsis, grape, and papaya. In poplar, transcript number, protein number, intron number showed a negative correlation with partial heat stress. The four cis-acting elements showed no correlation with expression.

Intron retaining in poplar SR genes

In humans, ES accounts for 35.2% of AS, whereas IR represents only 0.01% of AS [37]. In contrast, IR is the most common type in plants [38]. And we obtained information on the IR events of the poplar SR genes from the RNA-seq data (Table S5). The results showed that 26 poplar SR genes underwent IR events; seven IR sites in PtSR34b, six IR sites in PtSR34c and 5 IR sites in PtRS29, PtRS29a, PtRS2Z33 and PtRS2Z34 (Fig. 3, Table S5). These IR events greatly increased the complexity of the SR genes. In addition, PtSCL29b-2, PtSR34c-4, PtSCL29c-4 and PtRS29a-5 were without IR rates under standard conditions, whereas they had IR events under hot and cold stress conditions (Table S5). Under cold stress, the relative IR rates (treated tissues vs. untreated control samples) of PtSR29-1, PtSR29-1, PtSR29-3, PtSCL23-1, PtSCL23-2, PtSCL25-1 and PtSCL25-2 increased significantly (significant change indicated that relative IR rates changed more than 30% under a certain stress), whereas the relative IR rates of PtSR34b-3, PtRS29a-1 and PtRS29a-2 decreased significantly (Fig. 8, Table S5). Under heat stress, the relative IR rates of PtSC27-1 and PtSCL23-1 decreased significantly, whereas the relative IR rates of PtSR34b-3, PtSR34-2 and PtSR34c-2 increased significantly. Under drought stress, the relative IR rates of PtSCL31-1 in the leaf increased, whereas the ratio in xylem and root decreased. Under salt stress, the relative IR rates of PtRS29-2 in the leaf decreased whereas the ratio in the roots increased (Fig. 8, Table S5).

Discussion

In the previous study by Richardson DN et al, 20 poplar SR genes, 18 Arabidopsis SR genes, 9 grape SR genes were identified [39]. Here, we analyzed 71 SR genes from poplar, Arabidopsis, grape, and papaya (30 poplar SR genes, 18 Arabidopsis SR genes, 14 grape SR genes and 9 papaya SR genes) by combining bioinformatics. The results showed that the number of poplar SR genes in our study was 10 more than the previous study, and the grape SR gene was 5 more than the previous study. Previous studies had shown that there are no SC subfamily genes in poplar, and we found that the SC subfamily had four genes (Table S1). Previous research used the old version (v2) of P. trichocarpa and V. vinifera genome information [39], while our research used the version (v3.1) of P. trichocarpa genome information [22] and version (v2) of V. vinifera genome information [23]. The update of genome information might be the cause of the change in the number of SR genes.

Evolutionary assessment of SR genes in poplar, Arabidopsis, grape, and papaya

At present, the evolutionary veins of these four dicots have been studied very clearly, and they originate from common ancestors. Previous studies have shown that Arabidopsis has experienced three rounds of

polyploidization events (γ triplication event, α- and β-WGDs). Poplar experienced two rounds of polyploidy duplication events (γ triplication event and P-WGD) [40]. Papaya and grapes only experienced a γ triplication event. Therefore, in theory, after a number of evolutionary events, the ratio of the four dicots should be 4 (Arabidopsis): 2 (poplar): 1 (papaya): l (grape) [41]. In this study, the ratio of the number of SR genes in the four dicots obtained was 18: 30: 14: 9. This suggested that the ratio of the SR genes was not the same as the theoretical ratio. In our study, we found that all of the four dicots except VvSR31 did not produce a new SR gene after the γ triplication event, which might mean that after the γ triplication event, the plant experienced a large the gene of the fragment was lost (Fig. 5). For the recent the recent P-WGD, we found that all poplar SR genes were genome-wide duplicated occurred in the P-WGD and was retained except PtSR29, PtSR34b, PtSC30, PtSC31, and PtRSZ20. And after the P-WGD, PtSR33, PtSR29, and PtSC30 had other types of duplicated events. In Arabidopsis, five SR genes did not produce a new SR gene in the recent α-WGDs, and only AtSCL30a and AtSCL30 were produced by other duplicated events. So the final poplar SR gene was more than Arabidopsis. The grape SR gene had a small number of gene duplications after the γ triplication event. Papaya had many SR genes lost before the γ triplication event, so in the end we only found 9 SR genes in papaya. It could be seen that the SR genes of the four dicots might have different degrees of gene expansion or deletion in the long-term evolution. The might be an important reason why their actual SR genes number ratio was different from the theoretical SR genes ratio.

AS events in the SR genes

Whether transcripts produced by AS are functional has long been a subject of debate. Here, we constructed AS profile of the poplar, Arabidopsis, grape SR genes from the phytozome database and grape genome database (Fig. 3, Figure S2, Figure S3). The result showed that 77.4% (48 of the 62 genes) had multiple splicing isoforms and 48 genes produced 198 transcripts. These transcripts tripled the transcriptome complexity of the poplar, Arabidopsis, and grape SR genes. In addition, AS of many transcripts occured in the untranslated region, and their appearance did not result in a change in the putative protein sequence. The splicing isoforms might have RNA level functions. These 48 SR genes with multiple splicing isoforms produced 114 putative protein sequences. For conserved motif analysis, some transcripts produced truncated proteins, such as the fourth putative protein isoform of PtSCL29b lacking RS domain, which might cause functional changes in the protein.

Then, for the overall expression level of SR genes (Fig. 7), the results showed that only a small number of SR gene expression changed significantly under heat, cold, drought and salt stresses, and most of the SR genes expression were relatively stable. In previous studies [18], it was found that in the various hormonal and abiotic stresses tested in the Arabidopsis SR genes, temperature stresses (cold and heat) significantly changed the AS of the pre-mRNA of several SR genes, while the hormone changed only splicing of three SR genes. In our study, since the poplar SR gene contained a large number of introns, and IR events were the most common type of AS in plants [38]. We obtained information on the IR events of the poplar SR genes from the RNA-seq data. We found 26 of 30 poplar SR genes had intron retention (IR) events. It was found that different intron sites had changes in the IR ratio under heat, cold, drought

and salt stresses (Table S5). Especially under heat and cold stresses, and the relative IR ratio changed more significantly (Fig. 8). This reflected that SR genes responsed stresses might be mainly from changing the AS of itself, rather than changing the overall expression value of the genes. In a sense, these intron-rich SR genes could provide raw materials for IR events and help plants to face heat and cold stresses. However, the transcripts produced by these IR events were mostly degraded by the nonsense-mediated mRNA decay (NMD) pathway due to the incomplete translation of the premature stop codon. The functional significance of these splicing isoforms was currently unclear.

In summary, we compared AS profile of the poplar, Arabidopsis, grape SR genes from the phytozome and grape genome database. And the IR events of the poplar SR genes under heat, cold, drought and salt stresses were also analysis and compared from RNA-seq data. However, our research wass limited to currently available transcriptome data. In the future, the loss of functional mutants of SR gene is constructed, for example in the model plants Arabidopsis and poplar, to examine their biological importance to plants.

Conclusions

We had identified 71 SR genes from poplar, Arabidopsis, grape, and papaya, and comprehensively analyzed the conserved motifs, evolutionary, promoter analysis, splice isoforms, and expression patterns among the SR protein family members. In view of the key role of SR proteins as essential splicing factors in plants, functional studies of SR genes from plants are important. This study provides valuable resources for further studies of the stress effect of SR proteins in plants.

Abbreviations

AS
Alternative splicing
K_a
The nonsynonymous substitution rate
K_s
The synonymous substitution rate
IR
Intron Retaining
RRM
RNA recognition motifs
SR
Serine/arginine-rich
MW
molecular weight
pI
isoelectric point

WGD
whole-genome duplication

Declarations

Ethical Approval and Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of supporting data

The datasets are obtained from XX.

Competing interests

The authors have no conflicts of interest.

Funding

This work was supported by the Sub-project I under National Science and Technology Support Program [2015BAD07B070104]; the National Natural Science Foundation of China [31800557]. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

HW collected all the data, and wrote the manuscript. HY provided scientific suggestions and criticisms for improving the manuscript. YX supervised the whole project. YX, as the correspondence author, provided financial support for the article. All authors read and approved the final manuscript.

Acknowledgements

Special thanks go for the anonymous reviewers for their enlightening comments in reformulating this paper.

Authors' information

Laboratory of Modern Biotechnology, School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei 230036, China.

References

1. Auinash, K.; Cooper, T. A., Functional consequences of developmentally regulated alternative splicing. Nature Reviews Genetics 2011, 12, (10), 715-29.
2. Reddy, A. S. N.; Yamile, M.; Maria, K.; Andrea, B., Complexity of the alternative splicing landscape in plants. Plant Cell 2013, 25, (10), 3657-3683.
3. Wang, B. B.; Brendel, V., The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. Genome Biology 2004, 5, (12), R102-R102.
4. Zhaolan, Z.; Licklider, L. J.; Gygi, S. P.; Robin, R., Comprehensive proteomic analysis of the human spliceosome. Nature 2002, 419, (6903), 182-185.
5. Koncz, C.; Dejong, F.; Villacorta, N.; Szakonyi, D.; Koncz, Z., The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. Frontiers in plant science 2012, 3, 9-9.
6. Wu, J. Y.; Maniatis, T., Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell 1993, 75, (6), 1061-70.
7. Valcárcel, J.; Green, M. R., The SR protein family: pleiotropic functions in pre-mRNA splicing. Trends in Biochemical Sciences 1996, 21, (8), 296-301.
8. Cáceres, J. F.; Misteli, T., .; Screaton, G. R.; Spector, D. L.; Krainer, A. R., Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. Journal of Cell Biology 1997, 138, (2), 225-238.
9. Yan, Q.; Xia, X.; Sun, Z.; Fang, Y., Depletion of Arabidopsis SC35 and SC35-like serine/arginine-rich proteins affects the transcription and splicing of a subset of genes. Plos Genetics 2017, 13, (3), e1006663.
10. Barta, A.; Kalyna, M.; Reddy, A. S. N., Implementing a Rational and Consistent Nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR Proteins) in Plants. Plant Cell 2010, 22, (9), 2926-2929.
11. Lazar, G.; Schaal, T.; Maniatis, T.; Goodman, H. M., Identification of a Plant Serine-Arginine-Rich Protein Similar to the Mammalian Splicing Factor SF2/ASF. Proceedings of the National Academy of Sciences of the United States of America 92, (17), 7672-7676.
12. Lopato, S.; Waigmann, E.; Barta, A., Characterization of a Novel Arginine/Serine-Rich Splicing Factor in Arabidopsis. The Plant Cell 8, (12), 2255.
13. Reddy, A. S.; Shad, A. G., Plant serine/arginine-rich proteins: roles in precursor messenger RNA splicing, plant development, and stress responses. Wiley Interdisciplinary Reviews Rna 2011, 2, (6), 875-889.
14. Roth, M. B.; Murphy, C., .; Gall, J. G., A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. Journal of Cell Biology 1990, 111, (1), 2217-23.
15. Golovkin, M., .; Reddy, A. S., The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. Plant Cell 1998, 10, (10), 1637-1647.

16. Lopato, S.; Kalyna, M.; Dorner, S.; Kobayashi, R.; Krainer, A. R.; Barta, A., atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. Genes Dev 13, (8), 987-1001.
17. Isshiki; M., The Serine/Arginine-Rich Protein Family in Rice Plays Important Roles in Constitutive and Alternative Splicing of Pre-mRNA. Plant Cell 18, (1), 146-158.
18. Saiprasad Goud, P.; Gul Shad, A.; Reddy, A. S. N., Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant Journal 2010, 49, (6), 1091-1107.
19. Palusa, S. G.; Reddy, A. S. N., Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. 185, (1), 83-89.
20. Rauch, H. B.; Patrick, T. L.; Klusman, K. M.; Battistuzzi, F. U.; Mei, W.; Brendel, V. P.; Lal, S. K., Discovery and Expression Analysis of Alternative Splicing Events Conserved among Plant SR Proteins. Molecular Biology & Evolution 2013, (3), 3.
21. Zhang, W.; Du, B.; Liu, D.; Qi, X., Splicing factor SR34b mutation reduces cadmium tolerance in Arabidopsis by regulating iron-regulated transporter 1 gene. Biochemical & Biophysical Research Communications 2014, 455, (3-4), 312-317.
22. Goodstein, D. M.; Shengqiang, S.; Russell, H.; Rochak, N.; Hayes, R. D.; Joni, F.; Therese, M.; William, D.; Uffe, H.; Nicholas, P., Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 2012, 40, (Database issue), D1178-D1186.
23. Vitulo, N.; Forcato, C.; Carpinelli, E. C.; Telatin, A.; Campagna, D.; D'Angelo, M.; Zimbello, R.; Corso, M.; Vannozzi, A.; Bonghi, C., A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biology,14,1(2014-04-17) 2014, 14, (1), 99.
24. Letunic, I.; Bork, P., 20 years of the SMART protein domain annotation resource. Nucleic Acids Research 2018, 46, (Database issue), D493-D496.
25. Finn, R. D.; Alex, B.; Jody, C.; Penelope, C.; Eberhardt, R. Y.; Eddy, S. R.; Andreas, H.; Kirstie, ; Liisa, H.; Jaina, M., Pfam: the protein families database. Nucleic Acids Research 2014, 42, (Database issue), 222-30.
26. Panu, A.; Manohar, J.; Konstantin, A.; Delphine, B.; Gabor, C.; Edouard, D. C.; Séverine, D.; Volker, F.; Arnaud, F.; Elisabeth, G., ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research 2012, 40, (Web Server issue), W597.
27. Paul, H.; Keun-Joon, P.; Takeshi, O.; Naoya, F.; Hajime, H.; Adams-Collier, C. J.; Kenta, N., WoLF PSORT: protein localization predictor. Nucleic Acids Research 2007.
28. Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Multiple Sequence Alignment Using ClustalW and ClustalX. Curr Protoc Bioinformatics 2002, Chapter 2, (Unit 2), Unit 2.3.
29. Kumar, S.; Stecher, G.; Tamura, K., MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology & Evolution 2016, 33, (7), 1870.

30. Bailey, T. L.; Nadya, W.; Chris, M.; Li, W. W., MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research 2006, 34, (Web Server issue), 369-73.
31. Guo, A. Y.; Zhu, Q. H.; Xin, C., GSDS:a gene structure display server. Hereditas 2007, 29, (8), 1023-1026.
32. Yang, Z., PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology & Evolution 2007, 24, (8), 1586-1591.
33. Magali, L.; Patrice, D.; Gert, T.; Kathleen, M.; Yves, M.; Yves, V. D. P.; Pierre, R.; Stephane, R., PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 2002, 30, (1), 325-327.
34. Filichkin, S. A.; Hamilton, M.; Dharmawardhana, P. D.; Singh, S. K.; Jaiswal, P., Abiotic Stresses Modulate Landscape of Poplar Transcriptome via Alternative Splicing, Differential Intron Retention, and Isoform Ratio Switching. Frontiers in Plant Science 2018, 9, 5-.
35. Gamboa-Tuz, S. D.; Pereira-Santana, A.; Zamora-Briseño, J. A.; Castano, E.; Rodríguez-Zapata, L. C., Transcriptomics and co-expression networks reveal tissue-specific responses and regulatory hubs under mild and severe drought in papaya (Carica papaya L.). Scientific Reports 2018, 8, (1).
36. Tang, H. B.; Wang, X. Y.; Bowers, J. E.; Ming, R.; Alam, M.; Paterson, A. H., Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Research 2008, 18, (12), 1944-1954.
37. Markus, M. A.; Yang, Y. H. J.; Morris, B. J., Transcriptome-wide targets of alternative splicing by RBM4 and possible role in cancer. Genomics 2016, 107, (4), 138-144.
38. Yamile, M.; Brown, J. W. S.; Craig, S.; Andrea, B.; Maria, K., Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Research 2012, 22, (6), 1184-1195.
39. Richardson, D. N.; Rogers, M. F.; Labadorf, A.; Ben-Hur, A.; Guo, H.; Paterson, A. H.; Reddy, A. S. N., Comparative Analysis of Serine/Arginine-Rich Proteins across 27 Eukaryotes: Insights into Sub-Family Classification and Extent of Alternative Splicing. Plos One 2011, 6.
40. Tuskan, G. A.; Difazio, S., .; Jansson, S., .; Bohlmann, J., .; Grigoriev, I., .; Hellsten, U., .; Putnam, N., .; Ralph, S., .; Rombauts, S., .; Salamov, A., . The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 2006, 313, (5793), 1596-604.
41. Olivier, J.; Jean-Marc, A.; Benjamin, N.; Alberto, P.; Christian, C.; Alberto, C.; Nathalie, C.; Sébastien, A.; Nicola, V.; Claire, J., The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 2007, 449, (7161), 463.

Supplementary Information

Figure S1. Domain Architecture of the Arabidopsis, poplar, grape and papaya SR protein subfamilies. The SC subfamily contained proteins with a single RRM followed by an RS domain. The SR subfamily proteins had two RRMs with an evolutionarily conserved SWQDLKD motif in their second RRM followed

by an RS domain. The RSZ subfamily consisted of SR proteins with one Zn knuckle. The plant-specific SCL subfamily (SC35-like) was similar to the SC subfamily but had an N-terminal extension that is rich in Arg, Pro, Ser, Gly and Tyr residues. Proteins of the plant-specific RS2Z subfamily had two Zn knuckles and an additional SP-rich region after the RS domain. The plant-specific RS subfamily contained two RRMs (without the SWQDLKD motif) followed by an RS-rich with many RS dipeptides [10]. According to the method described by Barta and Kalyna [10], we used the following a standardized nomenclature for plant SR proteins: (1) a species identifier based on the Latin binomial (e.g., At for Arabidopsis thaliana; Pt for Populus trichocarpa); three-letter prefixes can be used in ambiguous cases; (2) an abbreviation of the subfamily; (3) a calculated molecular weight of the longest protein isoform; and (4) a suffix (a, b, c, etc.) where required to distinguish paralogous proteins with the same calculated molecular weight belonging to the same subfamily.

Figure S2. Alternative splicing profile of Arabidopsis SR genes. a Alternative transcripts detected from 18 Arabidopsis SR genes were listed in the left panel. For each gene, the isoform designated as the primary transcript by the phytozome database and was placed at the top and the other alternative transcripts were listed below. Exons and introns were represented by yellow rectangles and thin lines, respectively. The untranslated area (UTR) was represented by a thick blue line. b The conserved motifs of the protein encoded by each primary transcript were listed to the right of the corresponding transcript. The conserved motifs of the protein encoded by other alternative transcripts were listed below, and identical protein isoforms were only shown once.

Figure S3. Alternative splicing profile of grape SR genes. a Alternative transcripts detected from 14 grape SR genes were listed in the left panel. For each gene, the isoform designated as the primary transcript by the grape genome database and was placed at the top and the other alternative transcripts were listed below. Exons and introns were represented by yellow rectangles and thin lines, respectively. The untranslated area (UTR) was represented by a thick blue line. b The conserved motifs of the protein encoded by each primary transcript were listed to the right of the corresponding transcript. The conserved motifs of the protein encoded by other alternative transcripts were listed below, and identical protein isoforms were only shown once.

Figure S4. Correlation analysis of gene feature, cis-acting elements and gene expression pattern in poplar, Arabidopsis, grape, and papaya.

Figures

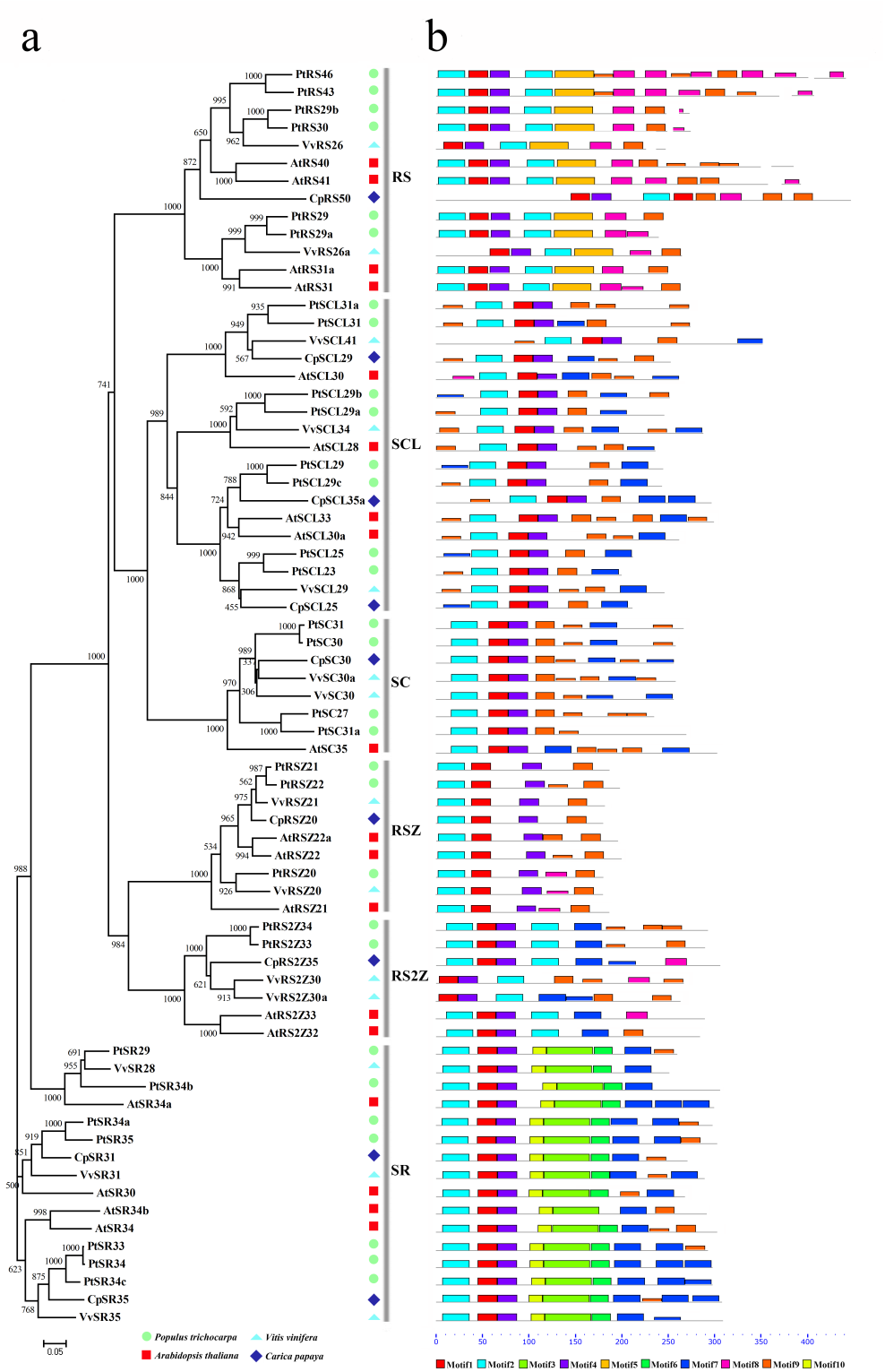

Figure 1

Phylogenetic and motif analysis of poplar, Arabidopsis, grape, and papaya SR genes. a Phylogenetic analysis of 71 SR proteins from poplar, Arabidopsis, grape, and papaya. The unrooted Neighbor-joining tree was constructed using MEGA 7, with 1,000 replications of boot strap values. b The conserved motifs of 71 SR proteins were detected using MEME. The lengths of 10 different motifs were shown in

proportion and accompanied by their phylogenetic relationships. These putative motifs were listed in Table S2.

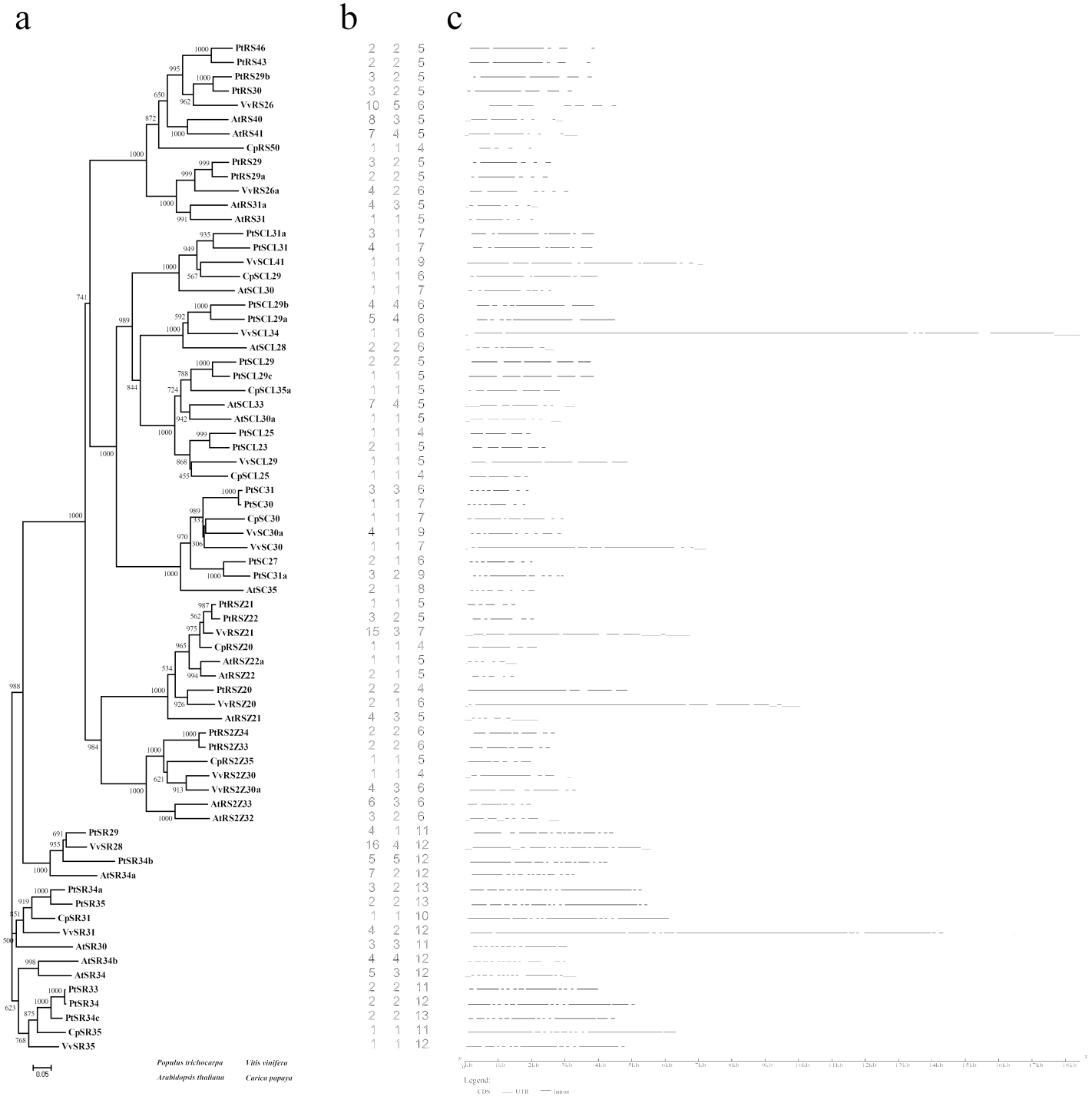

Figure 2

Phylogenetic, transcripts number, protein number, intron number, and gene structures of poplar, Arabidopsis, grape, and papaya SR genes. a Phylogenetic analysis of 71 SR proteins from poplar, Arabidopsis, grape, and papaya. The unrooted Neighbor-joining tree was constructed using MEGA 7, with 1,000 replications of boot strap values. b From left to right, the transcript number, protein number, and

intron number of the corresponding SR gene. c Exons and introns were represented by yellow rectangles and thin lines, respectively. The untranslated area (UTR) was represented by a thick blue line.

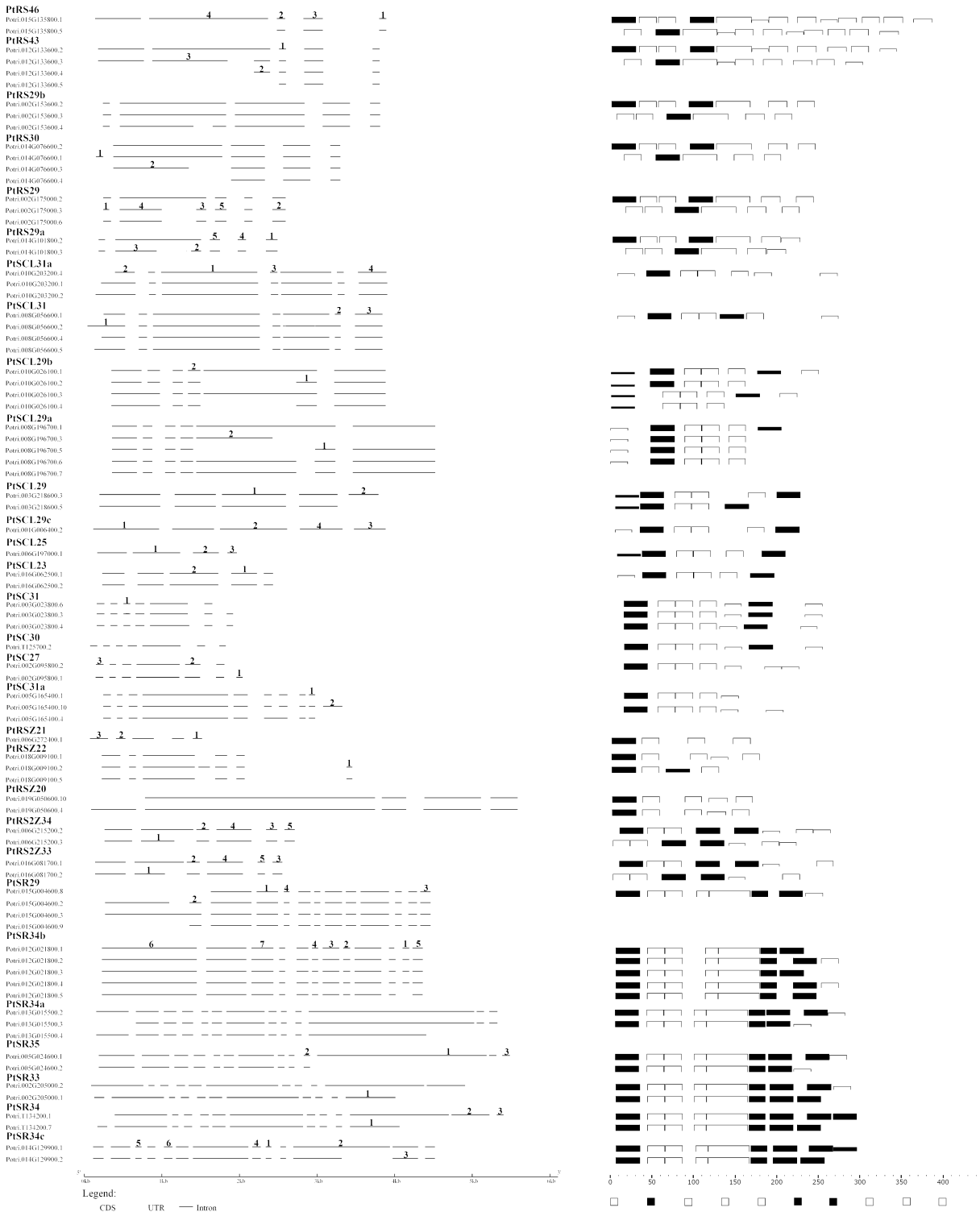

Figure 3

Alternative splicing profile and IR events of poplar SR genes. a Alternative transcripts detected from 30 poplar SR genes were listed in the left panel. For each gene, the isoform designated as the primary transcript by the phytozome database and was placed at the top and the other alternative transcripts

were listed below. Exons and introns were represented by yellow rectangles and thin lines, respectively. The untranslated area (UTR) was represented by a thick blue line. The number indicated that the intron sites undergo IR events under certain conditions. b The conserved motifs of the protein encoded by each primary transcript were listed to the right of the corresponding transcript. The conserved motifs of the protein encoded by other alternative transcripts were listed below, and identical protein isoforms were only shown once.

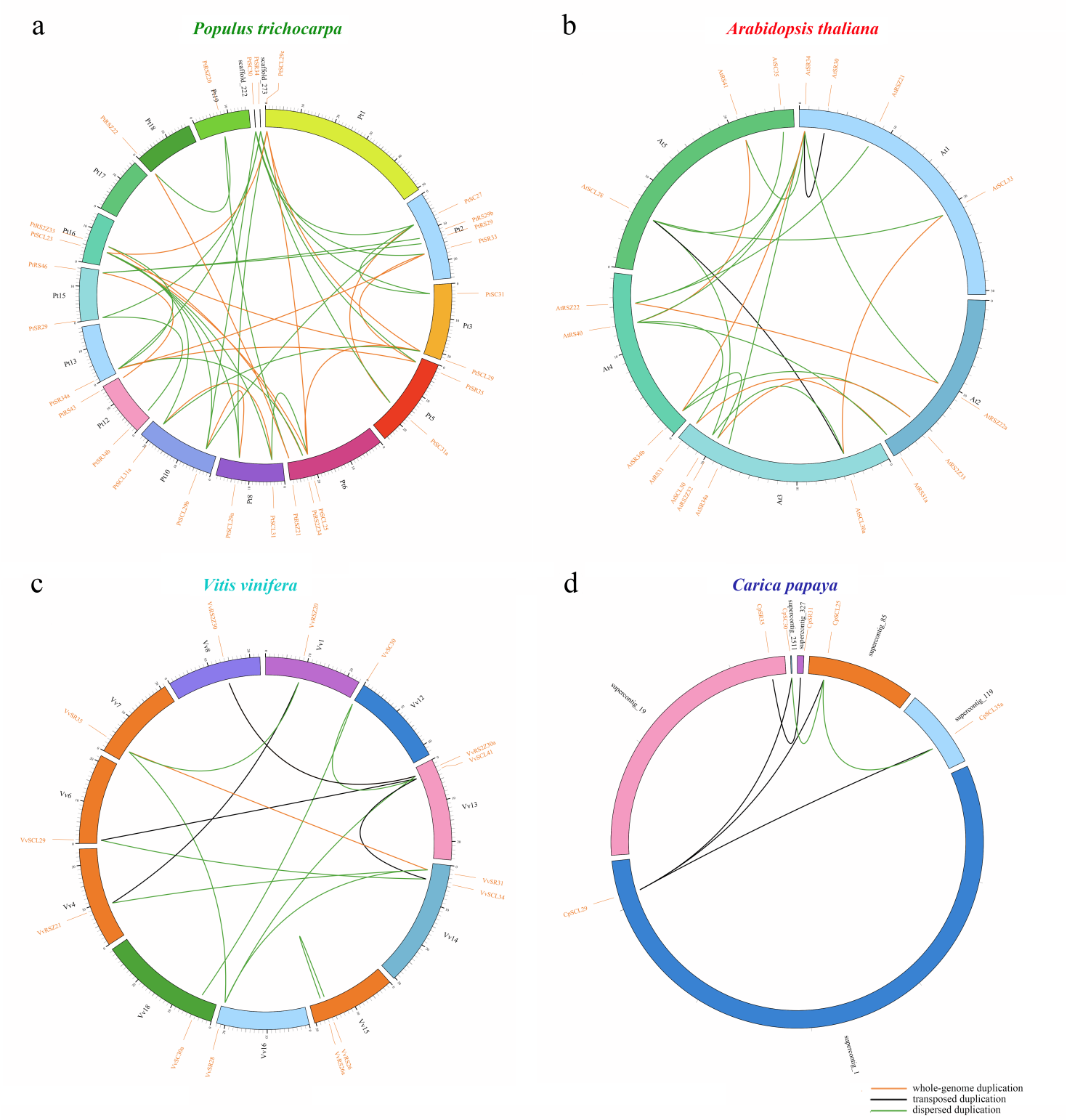

Figure 4

The paralogs analysis of SR genes within poplar, Arabidopsis, grape, and papaya. The whole chromosomes were shown in a circle. The red, black, and green lines represented the whole-genome duplication, transposed duplication, and dispersed duplication of the SR genes, respectively.

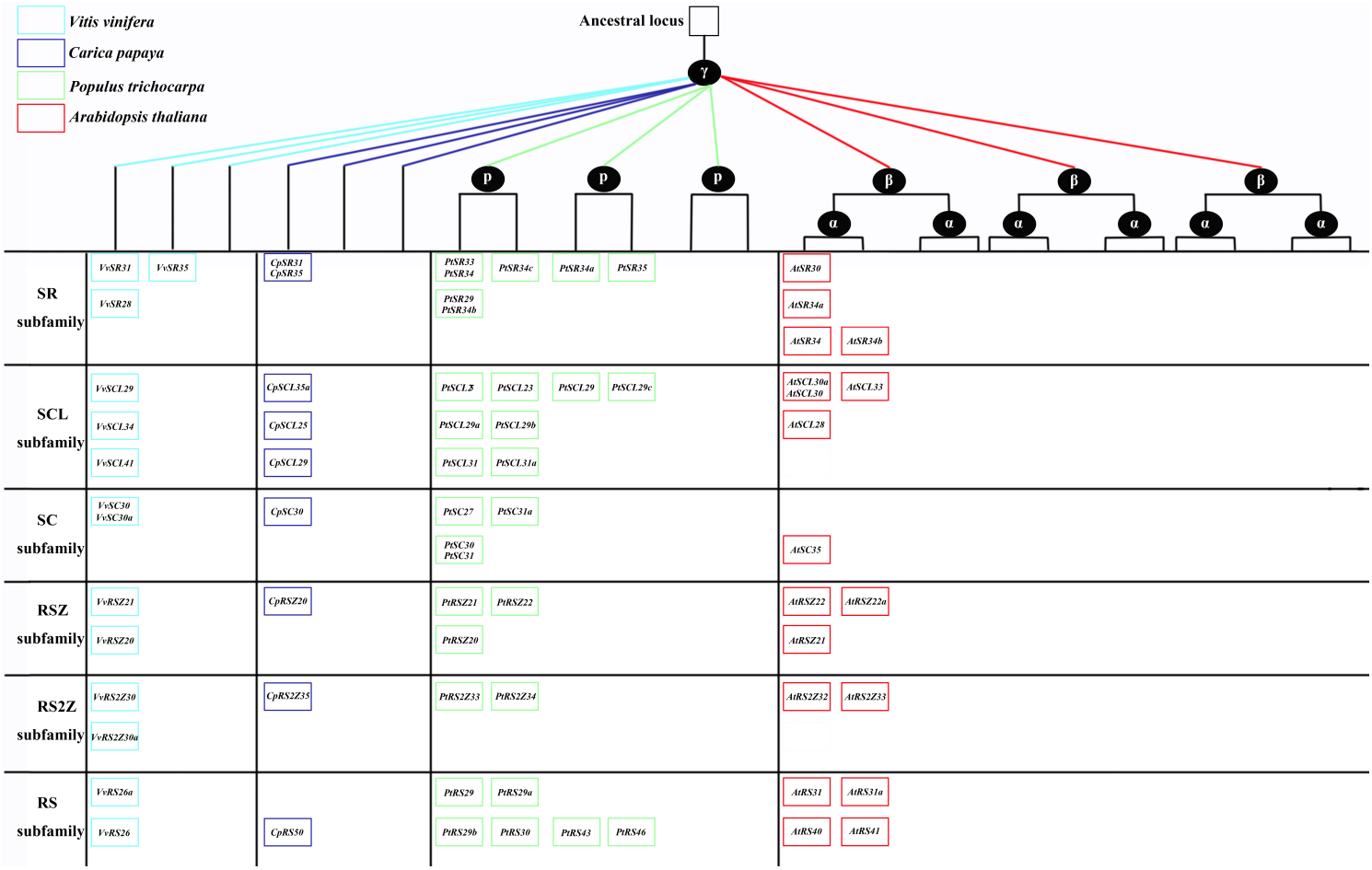

Figure 5

Panoramic picture to visualize the loss and expansion of the ancestral SR genes associated with paleopolyploidy events that had occurred in poplar, Arabidopsis, grape, and papaya. Note: The SR gene, which was not in the same line, indicated that the isolation of these genes was before the γ triplication event. For each row of SR genes of the corresponding species, they were produced by different WGDs. The presence of two SR genes in the box indicated that the two genes were genes produced by other duplicated types after the WGD.

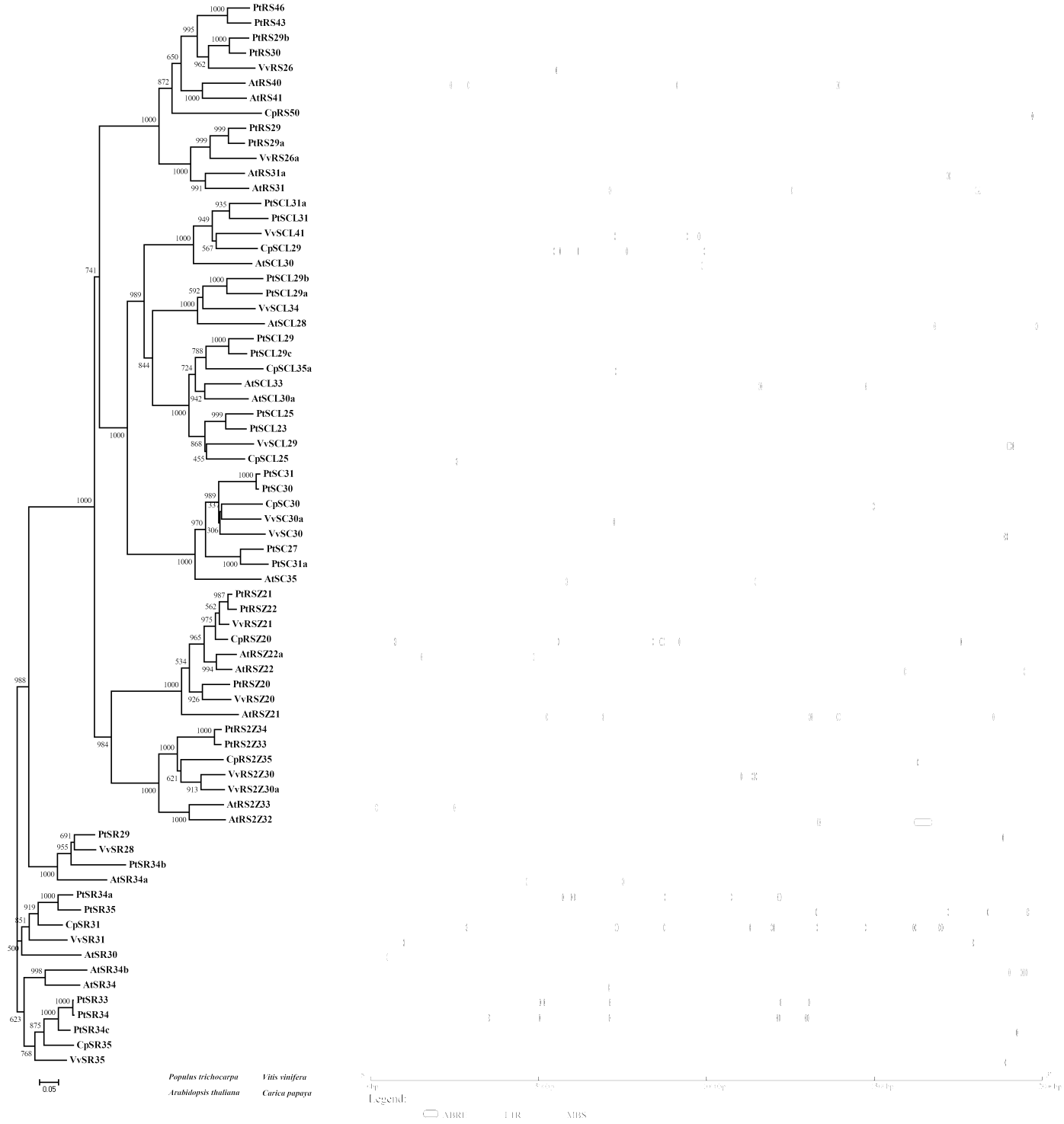

Figure 6

Distribution of major stress-related cis-elements in the promoter sequences of the 71 SR genes. Putative ABRE, LTR, MBS and TC-rich repeats core sequences were represented by drop-down arrows of different color as indicated. ABRE: cis-acting element involved in the abscisic acid responsiveness; LTR: cis-acting element involved in low-temperature responsiveness; MBS: MYB binding site involved in drought-inducibility; TC-rich repeats: cis-acting element involved in defense and stress responsiveness.

Figure 7

Expression of SR genes under stress in poplar, Arabidopsis, grape, and papaya. The expression level data of the poplar and papaya SR genes was obtained from RNA-seq data [34, 35], and Arabidopsis and grape chip data GSE5620, GSE5621, GSE5623, GSE5624, GSE5628 and GSE31594 were obtained from the NCBI GEO database. The values in the figure represented the log2-transformed foldchange values. Note: Gene expression value was the total expression value (the sum of all transcript expression readings) of the gene.

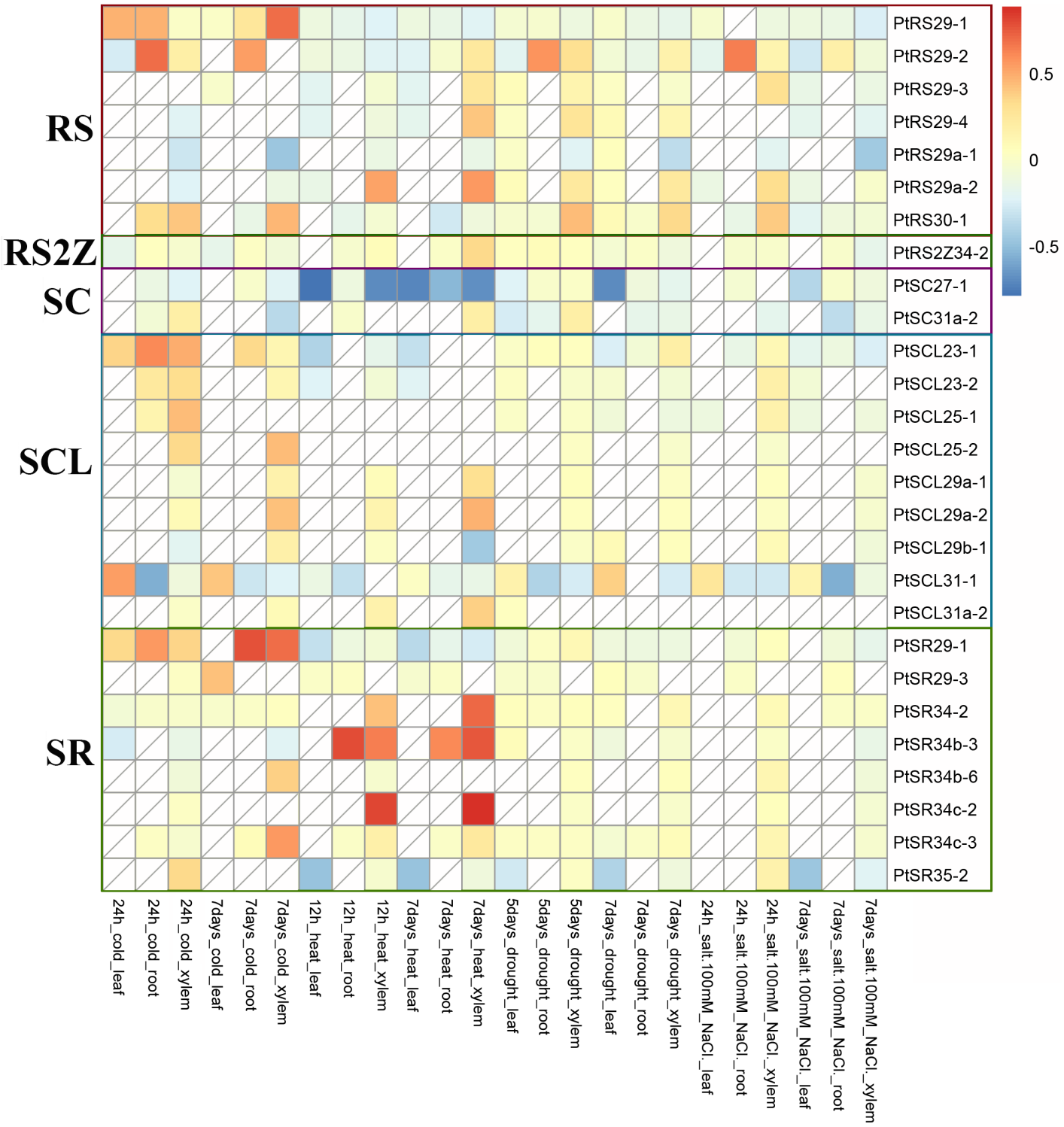

Figure 8

The relative IR rates (stress treated samples vs. untreated control samples) (IRR_ratio_diff) of the SR genes changed significantly under cold, heat, drought, and salt in leaf, root and xylem. Significant change indicated that relative IR rates changed more than 30% under a certain stress. The blank indicated no difference in IR rates. The number to the right of the gene name indicated the intron site of the corresponding gene in Figure 3.

Supplementary Files

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

- TableS5.xls
- FigureS1.tif
- TableS4.xls
- FigureS2.tif
- TableS2.docx
- FigureS4.tif
- TableS1.docx
- TableS3.docx
- FigureS3.tif