

# Evolutionary Significance of RAV Family in 25 Plants from Chlorophyta to Angiospermae

**Wan Li**

Northwest Agriculture and Forestry University

**Minghui Ye**

Northwest Agriculture and Forestry University

**Dongdong Wang**

Northwest Agriculture and Forestry University

**Qin Chen** (✉ [chenpeter2289@nwfau.edu.cn](mailto:chenpeter2289@nwfau.edu.cn))

Northwest Agriculture and Forestry University

---

## Research article

**Keywords:** RAV family, Comparative Genomics, Phylogenetic tree, Evolutionary analysis, Gene duplication

**Posted Date:** August 20th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-58073/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Plants are easily affected by various environmental factors such as temperature, high salinity, and drought which will cause poor plant development and reduce crop yield. Transcription factors (TFs) are related to the stress resistance found in many plant species and play vital roles in the growth and development of plants. Among them, the *RAV* (related to *ABI3/VP1*) gene family, which is a plant-specific type of transcription factor containing *AP2* (*APETALA2*) DNA binding domain and *B3* DNA binding domain, is the focus in this research.

**Results:** There more *RAV* genes/proteins (*RAVs*) were present in Gymnospermae and Angiosperm, a few *RAVs* in Bryophyta and Lycophyte, and no *RAV* gene was predicted in Chlorophyta. *RAVs* could be defined as three main groups, based on our phylogenetic analysis. The members of clade I included all plant stages from Chlorophyta to Angiosperm except Monocots. Clade II only contained eudicots plant species. Clade III, as a branch consisting of monocots plant species and *Marchantia polymorpha*, may help us explore the evolutionary relationship of *RAVs* between Bryophyta and Monocots. According to our results, it is shown that *RAVs* first appeared in bryophytes. While the vasculature continues to develop and evolve, *RAV* gene family also has functional differentiation and has formed new functions due to the existence of gene replication events.

**Conclusions:** Our findings indicate that the *RAVs* originated from Bryophytes, and *RAVs* of these plant species had obvious divisions in the evolutionary tree, which revealed the evolution of *RAVs*. These results have guiding significance for further research.

## Background

As plants continue to evolve, various gene families are constantly evolving. In particular, the numbers of TF families in plants usually grows explosively because of the rapid changes of the environment, which will lead to the formation of family members with new functions or the formation of new transcription factor families to adapt to the new surrounding ecological environment [1–3]. Algae, as the ancestors of plants, are usually aquatic plants, are closely related to the land plants, and play an extremely important role in plant evolution [4]. Affected by the ecological and environmental conditions, the transformation of plants from aquatic to terrestrial is extremely challenging for physiological and genetic adaptation. Studying the evolution of plants will provides an important basis for understanding the growth and reproduction of plants [5].

The research on TFs in the plant kingdom is relatively rich, but is mostly limited to the study of seed plants [6]. In order to advance our knowledge to the evolution of TFs in plants, research on early plants is essential. For example, compared to other land plants, the early plant *Marchantia polymorpha* has different levels of TF diversity [7]. Since the release of the whole-genome sequences of increased numbers of plant species, including algae, moss and other species [6], we can use these data to study and speculate on the evolution and gene function of the transcription factor families in plants.

Plants are continuously affected by various environmental factors such as temperature variation, high salinity, and drought during their growth process. The study of transcription factors (TFs) that can improve the stress resistance of plants by regulating the expression of multiple stress-responsive genes has become a

popular research. The members of the AP2/ERF superfamily which is one of the largest TF families are widely involved in regulating plant growth and development as well as abiotic and biotic stress responses [7, 8], and have one thing in common, that is, they contain at least one AP2 domain. AP2/ERF superfamily can be divided into 4 types: ethylene response factor (*ERF*), *AP2*, *RAV*, and soloists. The *RAV* gene family is a family of TF genes with two characteristic domains, *AP2* and *B3*, where the *AP2* domain is at the C-terminus and the *B3* domain is at the N-terminus [8–10]. The *RAV* family, which is unique to plants, also belongs to the *B3* superfamily because it contains a *B3* domain composed of approximately 110 amino acids [11]. *RAV1* and *RAV2* in *Arabidopsis thaliana* are the earliest found members of the *RAV* family [12].

There are *RAVs* in *Arabidopsis* [13], and this family has been studied in many plant species. *AtRAV1* takes an important part in actively regulating *Arabidopsis* leaf senescence [14]. In soybean (*Glycine max*), *GmRAV* is related to photosynthesis and regulates plant development, especially in terms of regulating senescence of plants [15]. In short-day (SD), soybean *GmRAV* and *Arabidopsis* *RAV1* negatively regulate the flowering and hypocotyl elongation, and the overexpression of soybean *GmRAV* will cause the leaves, roots and stems of soybean grow slowly, implying that they act in controlling plant growth [15–17]. *RAVs* in tomato (*Solanum lycopersicum*) may act as an intermediate TF, somehow linking *AtCBF1* to the genes related to pathogenesis, thereby enhancing tolerance to bacterial wilt (BW) [18]. The results of RT-PCR under *Verticillium dahliae* stress indicated that *RAVs* are related to the stress response in cotton (*Gossypium hirsutum*) [19]. The overexpression of *AtRAV1/2* may enhance the drought resistance in cotton [20]. The overexpression of *CARAV1* of pepper (*Capsicum annuum*) in *Arabidopsis* improved the resistance to high salt stress [21]. In addition, the expression levels of *AtRAV1* and *RAV1L* in *Arabidopsis* were decreased under 24-epibrassinolide (epiBL) and ABA treatments, respectively, and transgenic *Arabidopsis* overexpressing *AtRAV1* is not sensitive to ABA [16, 22].

These research results revealed that *RAVs* may take a part in growth and development of plants and resistance to stress. In view of the importance of *RAVs* in plants, in this study, the *RAVs* in 25 plant species (including each evolutionary stage) will be identified, and the gene structures, protein features, evolution relationships, gene duplications, and Gene Ontology (GO) annotations of *RAVs* will be analyzed to reveal the evolution of *RAVs* in the plant kingdom.

## Results

### Analysis of *RAVs*

We identified the *RAVs* from the genome sequences of the 25 plant species listed below for further phylogenetic analysis. These plant species included Chlorophyta (*Chlamydomonas reinhardtii* (Cr), *Coccomyxa subellipsoidea* (Co), *Dunaliella salina* (Ds), *Micromonas pusilla* (Mi), *Ostreococcus lucimarinus* (Ol), *Volvox carteri* (Vc)), Bryophyta (*Marchantia polymorpha* (Mp), *Physcomitrella patens* (Ph), *Sphagnum fallax* (Sf)), Lycophytes (*Selaginella moellendorffii* (Sm)), Gymnospermae (*Picea abies* (Pa)), and Angiosperms (Amborella: *Amborella trichopoda* (Ar); Eudicots: *Arabidopsis thaliana* (At), *Brassica rapa* (Br), *Cucumis sativus* (Cs, sweet orange), *Capsella rubella* (Cb), *Glycine max* (Gm, soybean), *Gossypium raimondii* (Gr), *Solanum lycopersicum* (Sl, tomato), *Solanum tuberosum* (St, potato); Monocots: *Brachypodium*

*distachyon* (Bd), *Hordeum vulgare* (Hv, barley), *Oryza sativa* (Os, rice), *Triticum aestivum* (Ta, wheat), *Zea mays* (Zm, maize)) (Table 1).

Table 1

The number of RAVs and gene duplication events, and clade of RAVs in 25 different stage plants.

Lineage	Organism (Abbr.)	Numbers of RAV Protein	Segmental Duplication (Pairs)	Tandem Duplication (Pairs)	Clade I	Clade II	Clade III
<b>Chlorophyta</b>							
	<i>Chlamydomonas reinhardtii</i> (Cr)	0	0	0	0	0	0
	<i>Coccomyxa subellipsoidea</i> (Co)	0	0	0	0	0	0
	<i>Dunaliella salina</i> (Ds)	0	0	0	0	0	0
	<i>Micromonas pusilla</i> (Mi)	0	0	0	0	0	0
	<i>Ostreococcus lucimarinus</i> (Ol)	0	0	0	0	0	0
	<i>Volvox carteri</i> (Vc)	0	0	0	0	0	0
<b>Bryophyta</b>							
	<i>Marchantia polymorpha</i> (Mp)	1	0	0	0	0	1
	<i>Physcomitrella patens</i> (Ph)	2	0	0	2	0	0
	<i>Sphagnum fallax</i> (Sf)	2	0	0	2	0	0
<b>Lycophyte</b>							
	<i>Selaginella moellendorffii</i> (Sm)	2	0	0	2	0	0
<b>Gymnospermae</b>							
	<i>Picea abies</i> (Pa)	16	0	0	16	0	0
<b>Angiosperm</b>							
<b>Amborella</b>	<i>Amborella trichopoda</i> (Ar)	1	0	0	1	0	0
<b>Eudicots</b>	<i>Arabidopsis thaliana</i> (At)	6	3	0	2	4	0

Lineage	Organism (Abbr.)	Numbers of RAV Protein	Segmental Duplication (Pairs)	Tandem Duplication (Pairs)	Clade I	Clade II	Clade III
	<i>Brassica rapa</i> (Br)	14	18	1	7	7	0
	<i>Cucumis sativus</i> (Cs)	4	0	0	3	1	0
	<i>Capsella rubella</i> (Cb)	5	2	0	2	3	0
	<i>Glycine max</i> (Gm)	5	6	0	1	4	0
	<i>Gossypium raimondii</i> (Gr)	10	4	2	7	3	0
	<i>Solanum lycopersicum</i> (Sl)	3	1	0	1	2	0
	<i>Solanum tuberosum</i> (St)	2	0	0	0	2	0
<b>Monocots</b>							
	<i>Brachypodium distachyon</i> (Bd)	4	2	1	0	0	4
	<i>Hordeum vulgare</i> (Hv)	8	0	1	0	0	8
	<i>Oryza sativa</i> (Os)	4	3	1	0	0	4
	<i>Triticum aestivum</i> (Ta)	5	0	0	0	0	5
	<i>Zea mays</i> (Zm)	3	2	0	0	0	3
*: The gene mapping information is unclear, it is impossible to determine the tandem duplication pairs; or segmental duplication information cannot be obtained from PGDD.							

In total, 714 RAVs were searched expect *Arabidopsis thaliana*. We analyzed the 714 RAVs based on the conserved domains of RAVs in *Arabidopsis thaliana*, 97 proteins which had *B3* (or *B3\_DNA*, hereinafter indicated by *B3*) and *AP2* (or *AP2 superfamily*, hereinafter indicated by *AP2*) domains were identified (Additional file 4). There was no RAVs were predicted from chlorophyte plant species, the amount of RAVs in *P. abies* were the largest. According to the evolutionary relationship, the RAVs did not evolve until the appearance of Bryophytes. Each bryophyte and tracheophyte plant species also predicted RAVs (Fig. 1A, Table 1).

Some information about *RAVs* was provided, which included the gene information (gene ID, gene location, and numbers of exons and duplication events; Additional file 1a), protein information (protein length, molecular weight, *pl*, etc.; Additional file 1b), protein features (conserved domain location, numbers and diagrams of motifs, predicted subcellular localization, transmembrane domains, and signal peptide; Additional file 1c). These data play an important role in understanding and researching proteins for us.

We analyzed the location of motifs, *B3* and *AP2* domains, and the number of exons in each *RAVs*. While for the most part exons numbered 1–3, and in *C. rubella*, CbRAV4 contained 11 exons, which was obviously different from other genes (Additional file 1a). Generally, the number of exons of the same family member in any plant is relatively stable, and is not related to the length of the sequence [23]. Most proteins had two main domains: *B3* and *AP2*, except for CbRAV4 in *C. rubella*, which had an extra MetAP1 domain (Additional file 1c). Additionally, four to fourth motifs were predicted in each protein. Among them, PaRAV13 contained 4 motifs, AtRAV1, AtRAV2, AtRAV3, BrRAV1, BrRAV9, BrRAV10, BrRAV12, BrRAV13, BrRAV14, CbRAV1, CbRAV5, GmRAV2, GrRAV3, and StRAV2 each contained 14 motifs (Additional file 1c) and we showed all exons, domains, and motifs by Fig. 1.

## Phylogenetic analysis of *RAVs*

97 *RAVs* in 4 different plant stages, from Bryophytes to Angiosperms, were used to construct a phylogenetic tree using the Neighbor-Joining (N-J) method (Fig. 2). According to the unroot tree, we could easily divide 97 *RAVs* into 3 main groups. Different plant stages were distributed in different clades. Clade I had more branching events. Except MpRAV1 (belonged to Clade III) in *M. polymorpha*, other non-seed plants and Gymnospermaes were distributed in Clade I. Seed plants were located in Clade I, Clade II, and Clade III. *A. trichopoda* and 8 eudicot species were located in Clade I and Clade II, and 5 monocot plant species were only distributed in Clade III. We classified these *RAVs* based on the method from Tegeeder and Ward [24]. Early plant *RAVs* appeared in Clade I and Clade III. Clade I mainly contained the *RAV* members of Bryophyta and Lycophytes, Clade III had one *RAV* protein MpRAV1. As the sister group of the remaining flowering plants, *A. trichopoda* also appeared in Clade I. Moreover, there no *RAVs* were distributed in Clade II until the evolution of Eudicots (Table 1).

## Investigation of gene duplication events

The formation of new functional genes and the function differentiation of any gene are inseparable from gene duplication events [25]. We analyzed 97 *RAVs* of each plant species, and the gene duplication events were identified in 10 species (Fig. 3). There were no duplication events from Bryophytes to Gymnospermaes. *A. trichopoda*, *C. sativus*, *S. tuberosum*, and *T. aestivum* exhibited no duplication events, either. Six plant species only had tandem or segmental duplications, and 4 species had both. The highest number of duplication events and segmental duplication events existed in *B. rapa*, and that of tandem duplication events occurred in *G. raimondii* (Table 1, Additional file 1a). The duplication events only occurred in Eudicots and Monocots. The number in Eudicots was higher than in Monocots. The tandem duplications were less than the segmental duplications. Additionally, the duplication events of eudicot species in Clade II were more than that

in Clade I, which suggested that *RAV* protein in Clade II may act a pivotal part in the growth and development in eudicots plant species.

It is necessary and useful to calculating the ratio of non-synonymous nucleotides to synonymous nucleotide substitutions ( $Ka/Ks$ ) for us to understand deeply the evolution of gene families. In total, 47 gene pairs were analyzed using DnaSP6 [26].  $Ka/Ks$  ratio values were much below 1.0 in all gene pairs (Additional file 2). We also calculated  $Ka/Ks$  ratio values of all *RAVs* in each plant species (Additional file 3). Overall, in addition to the  $Ka/Ks$  values which could not be calculated, the highest  $Ka/Ks$  value was 2.70 which appeared between PaRAV13 and PaRAV16 in *P. abies* (Additional file 3), and that of the other gene pairs were all less than 1.0.

Ninety-seven proteins were annotated with specific reference, including biological process (BP), molecular function (MF), and cellular component (CC). According to the results of gene annotation, eight aspects of BP were annotated to 97 *RAVs* that were linked to (negative) regulation of transcription and DNA-templated. Six proteins in *Arabidopsis thaliana* participated in ethylene-activated signaling pathway. AtRAV1, BrRAV9, CbRAV1, GrRAV3, MpRAV1, OsRAV3, PhRAV2, SmRAV2, and StRAV1 were related to response to brassinosteroid, negative regulation of flower development, leaf development, and lateral root development. AtRAV3, BdRAV1, BrRAV13, GmRAV1, and SIRAV3 were associated with photoperiodism and flowering. While, CbRAV4 in *C. rubella* was also predicted be related to proteolysis. Six kinds of MF were annotated to these *RAVs*. Among them, all 97 *RAVs* were associated with transcription factor activity (sequence-specific DNA binding) and DNA binding (or transcription regulatory region DNA binding). Further, CbRAV4 were also related to metal ion binding, and metalloexopeptidase and aminopeptidase activity. Only one aspect of CC was annotated to 97 *RAVs*. Forty-nine *RAVs*, which were distributed in various plant stages, might involve in CC, with the location in nucleus (Fig. 4, Additional file 5).

## Discussion

### Analysis of *RAVs*

*RAVs* belonged to the *AP2/ERF* superfamily [27, 28], we took several representative plants in different evolutionary stages as the research objects to analyzed the evolution of *RAVs*. The *RAV* family has been identified in many species, such as sweet orange, *Vitis vinifera*, rice, maize, wheat, soybean, and so on [13, 29–33]. We integrated these results, and we searched and identified *RAV* family again using the methods described in this article. The result we identified was almost the same with that in other articles. However, 13 *RAVs* were identified in soybean [32], several gene which did not contain the AP2 domain should be excluded. In maize, 4 *RAVs* were screened by Zhou et. al using more kinds of database, which was different with the method we used [13].

The *RAVs* are widely associated with the plant growth and development and the regulation against adverse reactions. The overexpression of AtRAV1 and AtRAV2 in cotton can enhance the fiber length differentially under drought stress and delays flowering [34]. The overexpression of CARAV1 which is response to the pathogen infection from pepper increases the tolerance to drought and salt, and enhancing the sensitivity to ABA in transgenic *Arabidopsis* [35]. A *RAV*-like protein from soybean is involved in the process of photosynthesis and senescence. Transgenic *Nicotiana tabacum* (tobacco) plants overexpressing the *RAV*-like

gene will have a slow growth rate, delayed root elongation, delayed flowering time, and a significantly reduced chlorophyll content in the leaves [15]. The *RAV* family has been identified and studied mainly in eudicots plant species and monocots plant species, especially in Eudicots, but rarely in other plant stages. In Bryophytes and Gymnosperms, no research related to *RAVs* is found, while the genome sequences of *S. moellendorffii* in Lycophytes and *A. trichopoda* are reported, respectively [36, 37]. Although *RAVs* have been identified in several other early plants, there has been no further research, either [38]. While we can predict the function of more *RAVs* using the database of the Phytozome website

We predicted 97 *RAVs* (Additional file 6) which had no transmembrane domain and signal peptide (same as other transcription factors) [39] and analyzed *B3* and *AP2* domains in each protein. Each protein had and only had both *B3* and *AP2* domains, except *CbRAV4* which had an extra domain: *MetAP1*. Additionally, A complete gene sequence consists of exons and introns, and exons, as part of the coding sequence, take an crucial part in gene function [23]. The *RAV* members of *P. abies* had 1–3 exons. *PhRAV1* had 3 exons. A few angiosperms' *RAVs* and most *RAVs* of Bryophytes and Lycophytes had two exons, and the most *RAV* members of the angiosperms pant species had only one exon. According to the analysis result, it could be speculated that some introns have been lost from the Bryophytes with the continuous evolution of plants. This is consistent with the previous conclusion [40]. Because the *RAV* protein structures and features in Bryophyta did not vary greatly (Fig. 1, Additional file 1b,c), we inferred that the intron losses that occurred during plant evolution was the reason of the number differences of exons between different species. The phenomenon has been confirmed in previous study [41]. However, *CbRAV4* had 11 exons which is completely incompatible with the plant evolution condition, and we have not been able to find relevant studies to confirm this phenomenon. Combined with the results of conserved domain analysis, we suggest that *CbRAV4* may have evolved into a protein of other families (*MetAP* family) or formed new functions [42], so that it has formed a significant difference from the *RAV* family.

## Evolution of *RAVs*

*RAVs* are divided into 3 main groups, which is consistent with previous research [32]. Phylogenetic results showed that most non-vascular plants (Bryophytes), *A. trichopoda*, *P. abies* and some eudicots plant species were composed of Clade I. Clade II only contained eudicots plant species. The only *RAV* gene in *A. trichopoda* was located in Clade I, which could be explained that because in *A. trichopoda*, which belongs to the Angiosperm and is a sister of flowering plants, six exogenous genomes constructed the mitochondrial genome, one from moss, two from other flowering plants, and three from green algae (Chlorophytes did not contain *RAVs*) [43]. The *RAVs* of eudicots plant species may be derived from the Bryophytes, and have formed the proteins belonging to Clade II with special functions during the evolution process. However, the *RAV* members of the monocots plant species and *M. polymorpha* belong to Clade III, which indicates that the *RAVs* in Monocots may also have evolved from Bryophytes and have common ancestors with Eudicots, but the protein functions are differentiated during the evolution process in Angiosperms. Moreover, there are the most *RAVs* in Gymnosperms, which may mean that there are more obvious evolutions in Gymnosperms than that in Angiosperms. In general, phylogenetic trees suggest that vascular plants may have a common ancestor, but the ancestor of Monocots is inherited from the *MpRAV1* protein of Clade III in *M. polymorpha*, while the ancestor of other seed and non-seed plants may be inherited from the proteins from Bryophytes which are located in Clade I (Fig. 2). With the development of vascular plants, the number of *RAV* members has

increased dramatically (Fig. 3). In addition, due to the use of different protein databases, selection of different species to study, and using a different website/program to analyze, all will affect the analysis results, so the above views need be confirmed by more evidence.

Different groups of *RAVs* may have different functions in different plant species. The overexpressing of *AtRAV1* in *Arabidopsis* results in a retardation of lateral root and rosette leaf development, and the expression inhibition will lead to an earlier flowering phenotype, indicating that *AtRAV1* may function as a negative regulatory component of growth and development [16]. The mRNA level of *AtRAV1* from *Arabidopsis* showed an increase, the highest level, and a decrease in the three leaf stages of late maturity, early senescence and late senescence, which suggested that *AtRAV1* has an significant role in regulating leaf senescence [14]. In transgenic tomato plants, with the overexpression and silencing *SIRAV2*, the expression levels of *SIERF5* and *PR5* increased and decreased, which enhanced or weakened the BW tolerance, respectively [18]. The *BnaRAV-1-HY15* TF in *B. napus* is highly identical to *AtRAV2* in *Arabidopsis*. The expression of *BnaRAV-1-HY15* will be induced by cold, salt and PEG treatments, and is insensitivity to ABA [44]. The more similar the distribution of genes in the same family on the phylogenetic tree, the more similar their functions in different plants [45]. The same class probably means similar characteristic of *RAV* family. Therefore, we can speculate the function of a *RAV* gene in any species based on the existing research results, thus providing effective ideas and directions for further research [39].

### Gene duplication events, Ka/Ks ratio, and GO annotation

In all living organisms, gene duplication events are a very common matter, which provides a basis for the formation of new functional genes and the function differentiation of genes in organisms. This also leads to species differences and species specificity [46, 47]. Our finding indicated that gene duplication events appeared only present in Eudicots and Monocots with the evolution of plants. There was no gene duplication in other species because of a small number or none of *RAVs*. But in *P. abies*, no gene duplications were identified from 16 *RAVs*, either. The evolution of *RAVs* was also related to gene duplications. The increasing of the duplications of *RAVs* following the evolution of plants, providing evidence for confirming the roles of this family in plant evolution. In addition, this phenomenon that gene duplication events only exist in Eudicots and Monocots also is consistent with the appearance of the function differentiation of *RAVs*.

The *Ka/Ks* values calculated by DnaSP software indicated all gene duplication events were suffering purifying selection, which meaning that these plants eliminate deleterious mutations during evolution process, and keeping the protein as it is [48]. We calculated *Ka/Ks* values in each of the plant species' *RAVs*, and found that except for *PaRAV13* and *PaRAV16* in *P. abies* with a *Ka/Ks* value of 2.70, the other genes were all stable and were all under purify selection (Additional file 3). Overall, *RAV* gene family is relatively stable through the process of plant evolution.

GO annotation is one way to predict the function of genes in terms of BP, MF and CC [49]. Based on these results of GO annotations, about a half of *RAVs* clustered in the nucleus, the main molecular functions of *RAVs* were in transcription factor activity and DNA binding, and the main biology process was regulation of transcription and DNA-template (Fig. 4, Additional file 5). While, *CbRAV4* in *C. rubella* was also predicted be related to proteolysis, metal ion binding, metalloexopeptidase activity, and aminopeptidase activity, which was

the same as the result of conserved domain and exons analysis, and showed the specificity of the gene CbRAV4.

## Conclusion

In recent years, with the completion of genome sequencing of various species, there have been many data and support for the study of basic lineages and the evolution of gene families. In this study, we first identified the *RAVs* and predicted the structure and functions of *RAVs* from 25 plant species using these existing databases. These results showed that *RAV* protein did not exist in Chlorophytes, which implied *RAV* protein was not formed until vascular plants appeared. Secondly, the phylogenetic relationships among the members of the whole *RAV* family showed that 97 identified *RAVs* were clearly divided into three clades. Moreover, we found that the *RAV* members of Clade II only belonged to Eudicots, while the *RAVs* of monocots plant species existed only in Clade III. This indicated that the *RAVs* might have a functional differentiation in Monocots and Eudicots during the evolution of plants. Then, analysis of gene duplication events and the calculation of *Ka/Ks* values showed that *RAVs* were almost only in purification selection, indicating that it was a stable gene family during evolution of plants. In addition, analysis of domains and motifs showed that *RAV* protein was relatively conservative during evolution process, and the number of exons and introns changed with the evolution of plants. It should be noted that CbRAV4 in *C. rubella*, which was different with other *RAVs* in gene structure and gene function prediction, suggested that CbRAV4 may have evolved new functions.

## Methods

### Analysis of *RAVs* in 25 plant species

Twenty-three plants protein/genome/CDS sequences were download from the Phytozome website (<https://phytozome.jgi.doe.gov/pz/portal.html>). The relevant sequences of *Arabidopsis thaliana* were obtained from TAIR website (<https://www.arabidopsis.org/>). The PGDD website (<http://chibba.agtec.uga.edu/duplication/>) provided the sequences of *Picea abies*. We used *AtRAV* protein sequences as a query to blast against other plants ( $e\text{-value} < 1^{-10}$ ). To ensure each protein belonged to the *RAV* family, all target proteins were analyzed by NCBI-CDD (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) and MEGA7 (<https://www.megasoftware.net/>) software to check that each protein had completed *B3* and *AP2* domain. After excluding the incorrect and redundant proteins, the remaining sequences were considered as putative proteins.

The gene structure was analyzed by GSDS website (<http://gsds.cbi.pku.edu.cn/index.php>). The motifs were predicted through the MEME website (<http://meme-suite.org/tools/meme>) using the default parameters. The protein features were analyzed by using the ProtParam tool (<https://web.expasy.org/protparam/>). A combined transmembrane domain and signal peptide was predicted by the TMHMM website (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>).

### Phylogenetic analysis of *RAVs*

The phylogenetic inference was carried out using the MEGA7 software. Twenty-five plant species were included in the tree. The Neighbor-Joining (N-J) method was used to calculate genetic distance with the parameters of the Jones-Taylor-Thornton (JTT) model and pairwise deletion option to handle missing data, and 1000 replicates for bootstrap analysis [50]. The N-J method was especially well-suited for datasets comprising lineages with largely varying rates of evolution. It can be used in combination with methods that allow for correction of superimposed substitutions [51].

### **Investigation of gene duplication events, Ka/Ks values and annotation information**

Gene tandem duplication events of *RAVs* were analyzed following the methods of Gu et al. (2002) [60], the major criteria were that the length of alignable sequence covers > 75% of longer gene, and similarity of aligned regions > 75%. Two genes were regarded as tandem pairs if they were located on the same chromosome and were separated by no more than 10 unrelated genes [61]. The segmental duplication pairs were analyzed by using the PGDD website (<http://chibba.agtec.uga.edu/duplication>) and the MCScanX program (<http://chibba.pgml.uga.edu/mcscan2/>). Ka and Ks values were calculated by DnaSP6 software.

Gene annotation was carried out by searching Gene Ontology (GO) through the InterPro website (<http://www.ebi.ac.uk/interpro/>) and PlantTFDB website (<http://planttfdb.cbi.pku.edu.cn/>).

## **Abbreviations**

**TFs:** Transcription factors

**AP2:** APETALA2

**RAV:** Related to ABI3/VP1

**SD:** Short-day

**BW:** Bacterial wilt

**GO:** Gene Ontology

**BP:** Biological process

**MF:** Molecular function

**CC:** Cellular component

**N-J:** Neighbor-Joining

**JTT:** Jones-Taylor-Thornton

## **Declarations**

**Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

All data generated or analyzed during this study are included in this article.

### **Competing interests**

The authors declare that they have no competing interests.

### **Funding**

This work was funded mainly by the National Key Research and Development Program of China (2018YFD0200805); the Project of Science and Technology from Shaanxi Province (No. 2017ZDXM-NY-004); and major collaborative innovation projects for production, education and research in Yangling demonstration zone (No. 2016CXY-05). In addition, this work was partially supported by State Key Laboratory of Crop Stress Biology in Arid Areas, China. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### **Author Contributions**

All authors have read and approved the manuscript, and ensure that this is the case. W.L., D.W. and Q.C. designed the experiments; W.L., Q.C., D.W., and M.Y. carried out the experiments; W.L. and M.Y. analysed the experimental results; W.L., M.Y., and D.W. analysed the data and developed the analysis tools; W.L. wrote the manuscript, and Q.C. revised the manuscript.

### **Acknowledgements**

We thank to Prof. Chen for suggesting the funding. We thank to Northwest A&F University for suggesting the laboratory.

## **References**

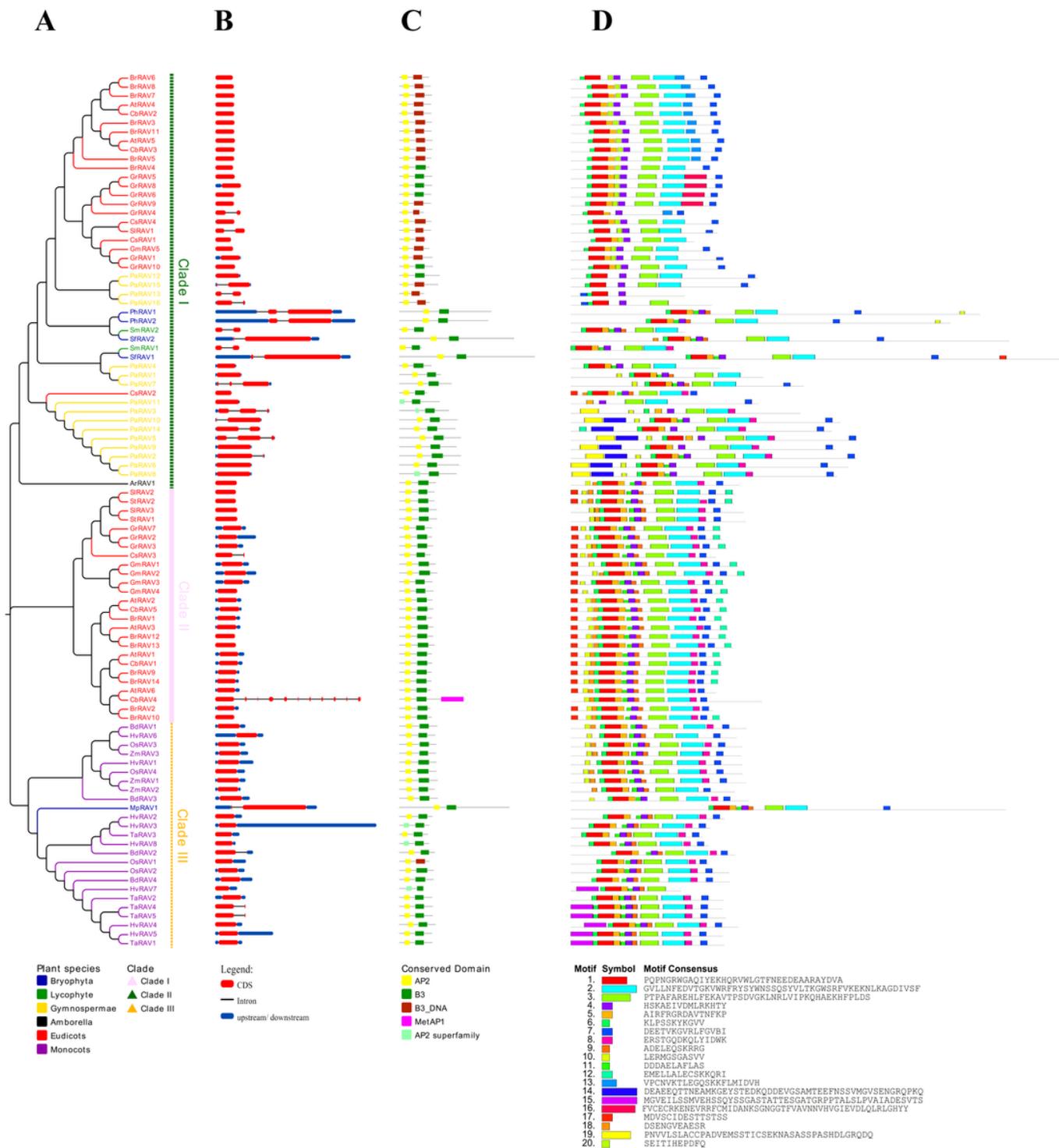
1. Catarino B, Hetherington AJ, Emms DM, Kelly S, Dolan L. The Stepwise Increase in the Number of Transcription Factor Families in the Precambrian Predated the Diversification of Plants On Land. *Mol Biol Evol.* 2016; **33**(11):2815.
2. Rensing SA. Gene duplication as a driver of plant morphogenetic evolution. *Curr Opin Plant Biol.* 2014; **17**(1):43-48.
3. Moghe G, Last R. Something Old, Something New: Conserved Enzymes and the Evolution of Novelty in Plant Specialized Metabolism. *Plant Physiol.* 2015; **169**(3):1512-1523.

4. Harholt J, Moestrup Ø, Ulvskov P. Why plants were terrestrial from the beginning. *Trends Plant Sci.* 2016; **21**(2):96-101.
5. Delwiche CF, Cooper ED. The Evolutionary Origin of a Terrestrial Flora. *Curr Biol.* 2015; **25**(19):R899-910.
6. Zhang C, Zhang L, Wang D, Ma H, Liu B, Shi Z, Ma X, Chen Y, Chen Q. Evolutionary history of the glycoside hydrolase 3 (GH3) family based on the sequenced genomes of 48 plants and identification of Jasmonic acid-related GH3 proteins in *Solanum tuberosum*. *Int J Mol Sci.* 2018; **19**(7):1850.
7. Bowman J, Kohchi T, Yamato K, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F. Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell.* 2017; **171**(2):287-304.
8. Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM. Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell.* 1992; **4**(10):1251-1261.
9. Riechmann JL, Lain S, Garcia JA. Highlights and prospects of potyvirus molecular biology. *Journal of General Virology.* 1992; **73**(1):1-16.
10. Suzuki M, Kao C, McCarty DR. The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *Plant Cell.* 1997; **9**(5):799-807.
11. Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. *Trends Plant Sci.* 2008; **13**:647-655.
12. Kagaya Y, Ohmiya K, Hattori T. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* 1999; **27**:470-478.
13. Zhou ML, Tang YX, Wu YM. Genome-Wide Analysis of AP2/ERF Transcription Factor Family in *Zea Mays*. *Curr Bioinform.* 2012; **7**(3):-.
14. Woo HR, Jin HK, Kim J, Kim J, Lee U, Song IJ, Kim JH, Lee HY, Hong GN, Lim PO. The RAV1 transcription factor positively regulates leaf senescence in Arabidopsis. *J Exp Bot.* 2010; **61**(14):3947-3957.
15. Zhao I, Li W, Luo Q, Han Y, Li W. A RAV-like transcription factor controls photosynthesis and senescence in soybean. *Planta.* 2008; **227**:1389-1399.
16. Hu YX, WANG YH, LIU XF, LI JY. Arabidopsis RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. *Cell Res.* 2004; **14**:8-15.
17. Lu QY, Zhao L, Li DM, Hao DQ, Zhan Y, Li WB. A GMRAV ortholog is involved in photoperiod and sucrose control of flowering time in soybean. *PLoS One.* 2014; **9**:e89145.
18. Li CW, Su RC, Cheng CP, Sanjaya, You SJ, Hsieh TH, Chao TC, Chan MT. Tomato RAV Transcription Factor Is a Pivotal Modulator Involved in the AP2/EREBP-Mediated Defense Pathway. *Plant Physiol.* 2011; **156**(1):213-227.
19. Lu H, Zhou Z, Chen H, Ling J, Liu F, Cai X, Wang X, Wang C, Wang Y, Wang K. Genome-Wide Analysis of the RAV Gene Family in Cotton. *Cotton Science.* 2014; **26**(6):471-482.
20. Matías-Hernández L, Aguilar-Jaramillo AE, Marin-Gonzalez E, SuarezLopez P, Pelaz S. RAV genes: regulation of floral induction and beyond. *Ann Bot.* 2014; **114**:1459-1470.
21. Lee SC, Choi DS, Hwang IS, Hwang BK. The pepper oxidoreductase CaOXR1 interacts with the transcription factor CaRAV1 and is required for salt and osmotic stress tolerance. *Plant Mol Biol.* 2010; **73**:409-424.

22. Fu M, Kang HK, Son SH, Kim SK, Nam KH. A subset of Arabidopsis RAV transcription factors modulates drought and salt stress responses independent of ABA. *Plant and Cell Physiology*. 2014; **55**:1892-1904.
23. Zhang C, Kong N, Cao M, Wang D, Chen Y, Chen Q. Evolutionary significance of amino acid permease transporters in 17 plants from Chlorophyta to Angiospermae. *BMC Genomics* 2020; **21**:391-401.
24. Tegeder M, Ward JM. Molecular Evolution of Plant AAP and LHT Amino Acid Transporters. *Front Plant Sci*. 2012; **3**:21.
25. Lee TH, Kim J, Robertson JS, Paterson AH. Plant Genome Duplication Database. *Methods Mol Biol*. 2017; **1533**:267-277.
26. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-GA. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Molecular Biology & Evolution*. 2017; **34**(12).
27. Wessler SR. Homing into the origin of the AP2 DNA binding domain. *Trends Plant Sci*. 2005; **10**(2):54-56.
28. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim Biophys Acta*. 2011; **1819**(2):86-96.
29. Ito TM, Polido PB, Rampim MC, Kaschuk G, Souza SGH. Genome-wide identification and phylogenetic analysis of the AP2/ERF gene superfamily in sweet orange (*Citrus sinensis*). *Genetics & Molecular Research Gmr*. 2014; **13**(3):7839-7851.
30. Rashid M, He G, Yang G, Hussain J, Yan X. AP2/ERF Transcription Factor in Rice: Genome-Wide Canvas and Syntenic Relationships between Monocots and Eudicots. *Evol Bioinform Online*. 2012; **8**(8):321-355.
31. Thamilarasan SK, Park JI, Jung HJ, Nou IS. Genome-wide analysis of the distribution of AP2/ERF transcription factors reveals duplication and CBFs genes elucidate their potential function in *Brassica oleracea*. *BMC Genomics*. 2014; **15**(1):422.
32. Zhao SP, Xu ZS, Zheng WJ, Wan Z, Wang YX, Yu TF, Ming C, Zhou YB, Min DH, Ma YZ. Genome-Wide Analysis of the RAV Family in Soybean and Functional Identification of GmRAV-03 Involvement in Salt and Drought Stresses and Exogenous ABA Treatment. 2017.
33. Zhuang J, Peng RH, Cheng ZM, Zhang J, Cai B, Zhang Z, Gao F, Zhu B, Fu XY, Jin XF. Genome-wide analysis of the putative AP2/ERF family genes in *Vitis vinifera*. *Sci Hortic (Amsterdam)*. 2009; **123**(1):0-81.
34. Mittal A, Jiang Y, Ritchie GL, Burke JJ, Rock CD. AtRAV1 and AtRAV2 overexpression in cotton increases fiber length differentially under drought stress and delays flowering. *Plant Science An International Journal of Experimental Plant Biology*. 2015; **241**:78-95.
35. Sohn KH, Lee SC, Jung HW, Hong JK, Hwang BK. Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. *Plant Mol Biol*. 2006; **61**:897-915.
36. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA *et al*. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science*. 2011; **332**(6032):960-963.
37. DePamphilis C, Palmer J, Rounsley S, Sankoff D, Schuster S, Ammiraju J, Barbazuk W, Chamala S, Chanderbali A, Determann R *et al*. The Amborella genome and the evolution of flowering plants. *Science*.

- 2013; **342**(6165):1467.
38. Jin J, He Z, Lei K, Ge G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 2014; **42**:1182-1187.
39. Li W, Dong J, Cao M, Gao X, Wang D, Liua B, Chen Q. Genome-wide identification and characterization of HD-ZIP genes in potato. *Gene.* 2019; **697**:103-117.
40. Rogozin I. Origin and evolution of spliceosomal introns. *Biol Direct.* 2012; **7**(1):11.
41. Roy S, Penny D. Patterns of intron loss and gain in plants: intron loss dominated evolution and genome-wide comparison of *O. sativa* and *A. thaliana*. *Mol Biol Evol.* 2007; **24**(1):171-181.
42. Kang J-M, Ju H-L, Sohn W-M, Na B-K. Characterization of the biochemical properties of two methionine aminopeptidases of *Cryptosporidium parvum*. *Parasitol Int.* 2012; **61**(4):707-710.
43. Rice D, Alverson A, Richardson A, Young G, Sanchezpuerta M, Munzinger J, Barry K, Boore J, Zhang Y, Depamphilis C. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. 2013. *Science*; **342**(6165):1468-1473.
44. Zhuang J, Sun C-C, Zhou X-R, Xiong A-S, Zhang J. Isolation and characterization of an AP2/ERF-RAV transcription factor BnaRAV-1-HY15 in *Brassica napus* L. HuYou15. *Mol Biol Rep.* 2011; **38**:3921-3928.
45. Blanc G, Barakat A, Guyot R, Cooke R, Delseny I. Extensive duplication and reshuffling in the arabidopsis genome. *Plant Cell.* 2000; **12**:1093-1101.
46. Zheng W, Zhengkui Z, Yunfeng L, Tengfei L, Qing L, Yuanyuan J, Congcong L, Chao F, Min W, Mian W *et al.* Functional evolution of phosphatidylethanolamine binding proteins in soybean and Arabidopsis. *Plant Cell.* 2015; **27**(2):323.
47. Zhang J. Evolution by gene duplication: an update. *Trends Ecol Evol.* 2003; **18**(6):292-298.
48. Hurst L. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 2002; **18**(9):486-487.
49. Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Production.* 2008; **2008**:619832.
50. Singh VK, Jain M, Garg R. Genome-wide analysis and expression profiling suggest diverse roles of GH3 genes during development and abiotic stress responses in legumes. *Front Plant Sci.* 2014; **5**:789.
51. Zhang W, Sun Z. Random local neighbor joining: A new method for reconstructing phylogenetic trees. *Molecular Phylogenetics & Evolution.* 2008; **47**(1):117-128.

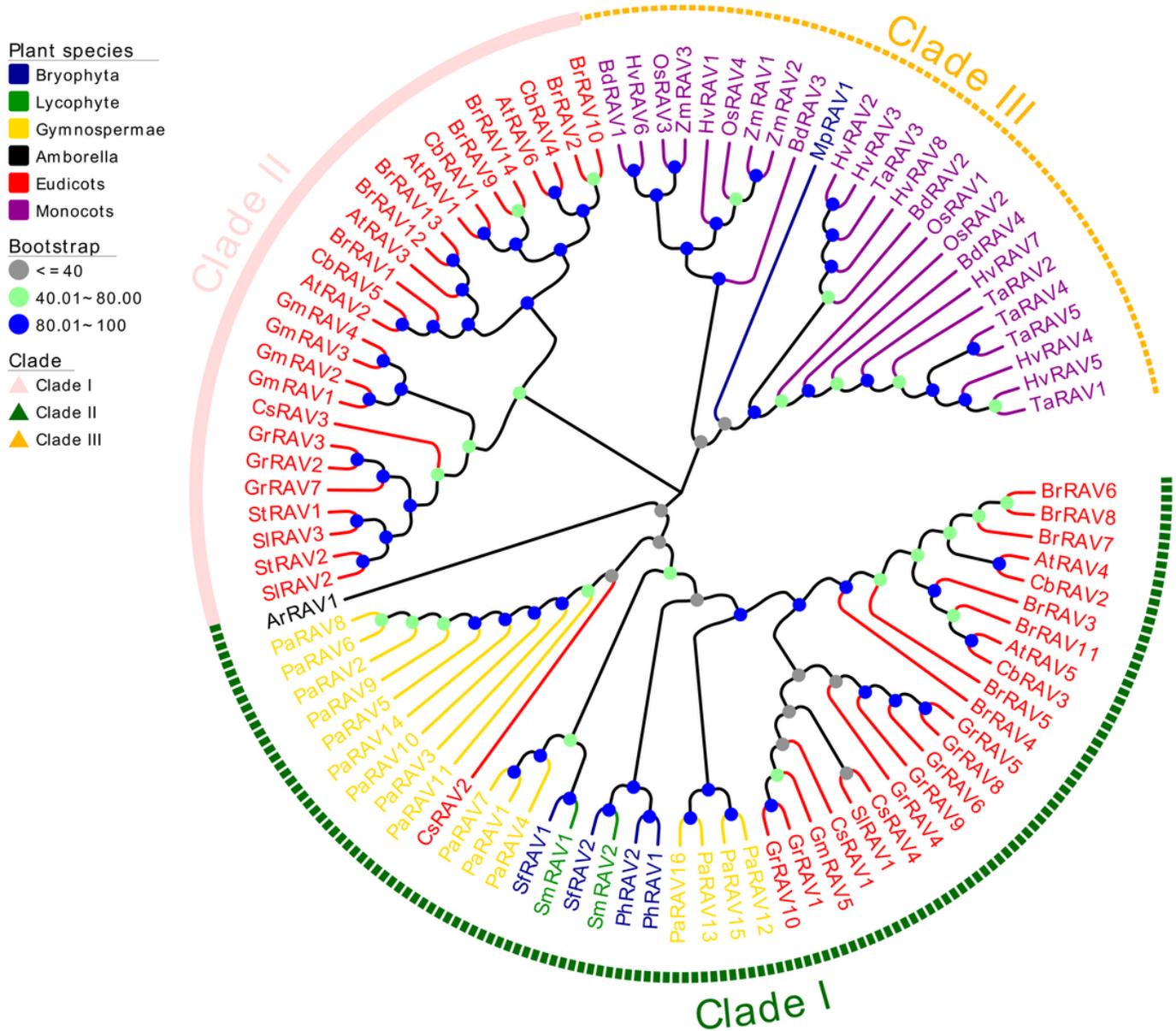
## Figures



**Figure 1**

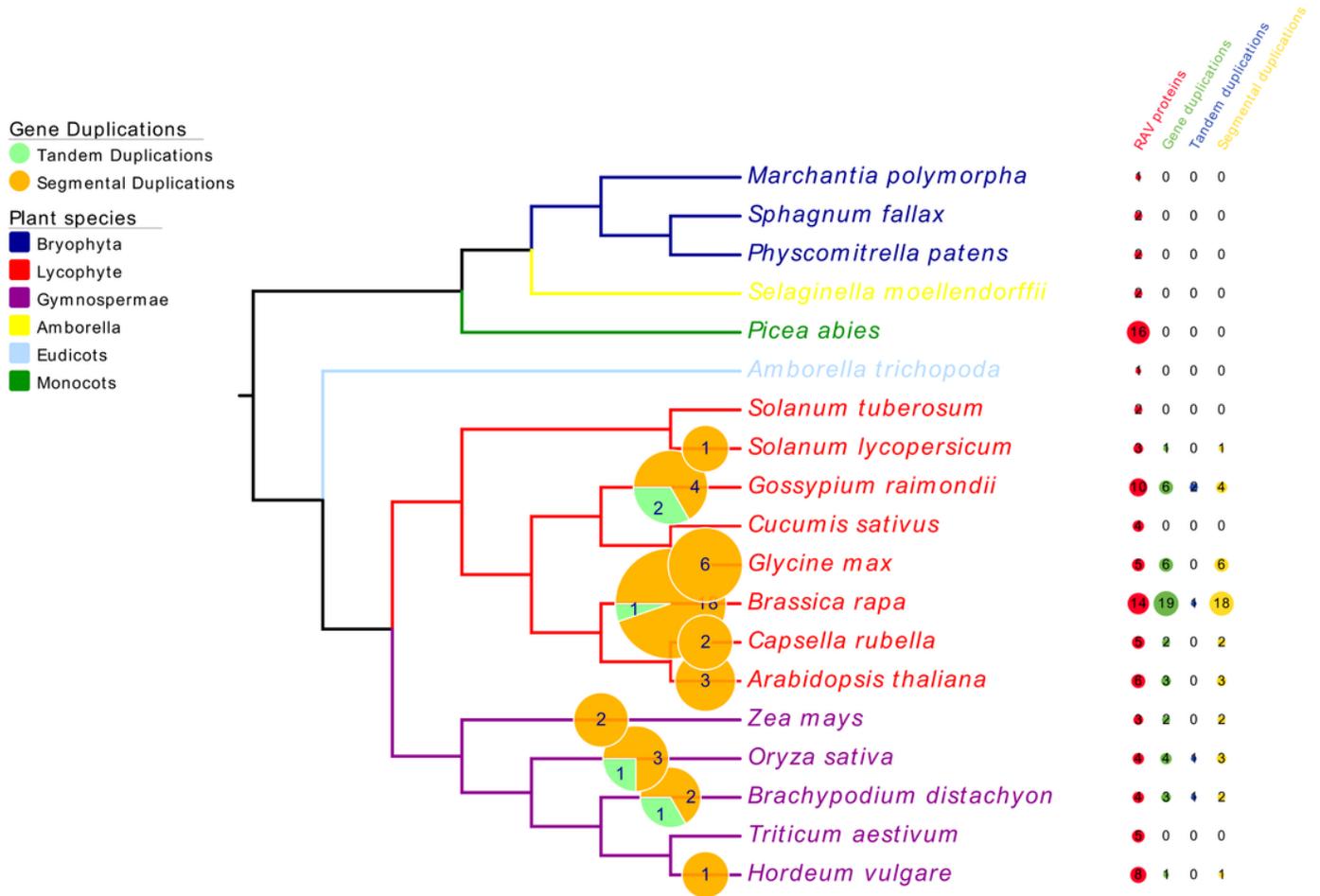
The division of whole RAVs. A: Phylogenetic tree. The evolutionary history was inferred using the Neighbor-Joining method in MEGA7. Bootstrap values of 1000 replications were executed. The 3 main groups and the different plant species are distinguished by different colors in the tree. B: The gene structures. The exon/intron distribution of corresponding 97 RAVs was detected by comparing these predicted coding sequences (CDS) with their corresponding genomic sequences by using the GSDS online tool (<http://gsds.cbi.pku.edu.cn>); the red boxes represent CDS; the discontinuous lines refer to the introns of these genes; the blue boxes show the

non-coding region of the upstream/downstream. C: The main domain, B3 and AP2 location. The different domains are showed by different colors. D: The motif composition. The motifs, numbered 1–20, are displayed in different colored boxes.



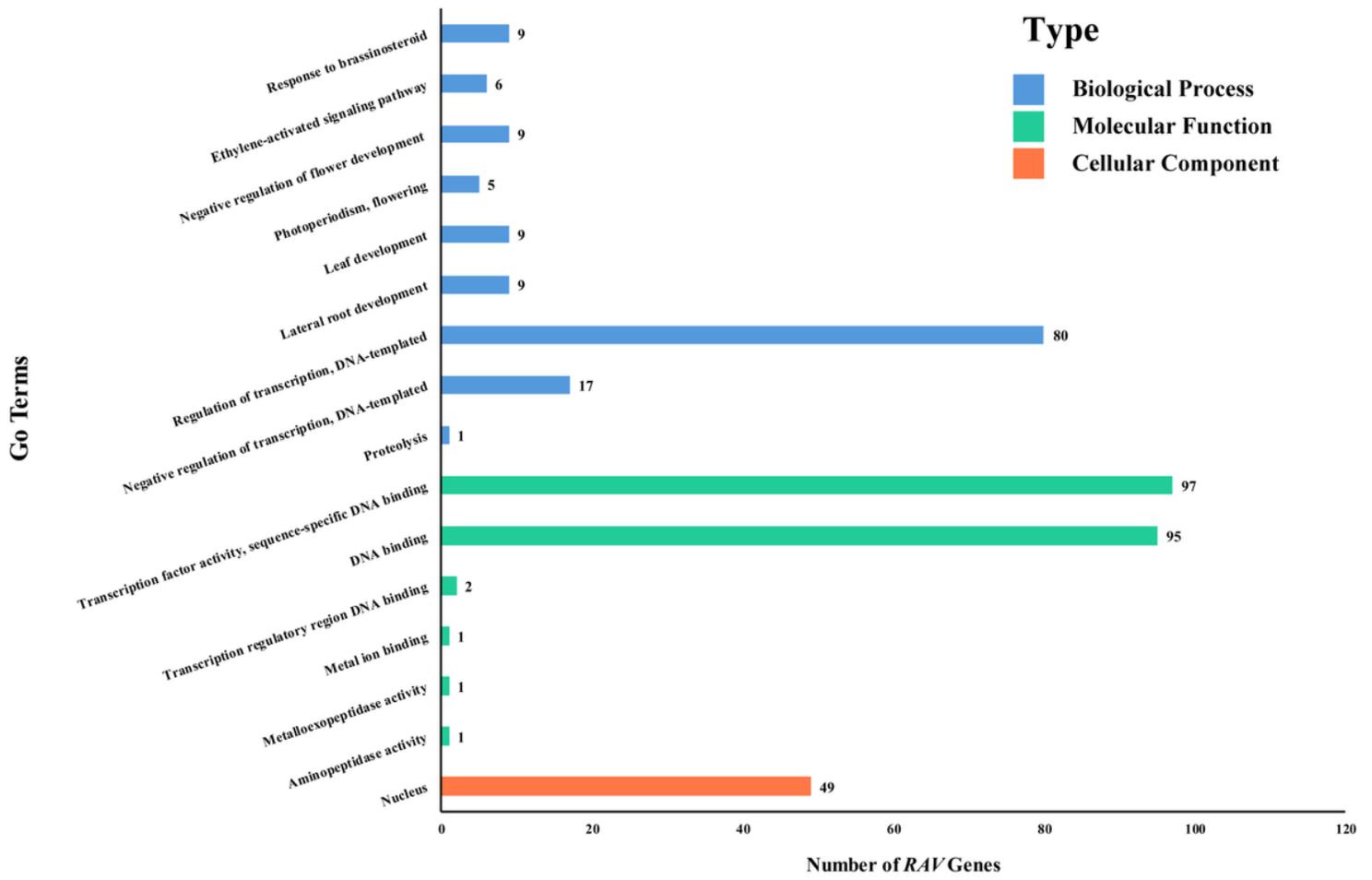
**Figure 2**

Phylogenetic tree of RAVs. Ninety-seven RAVs from Chlorophyta to Angiosperms were distributed in the unroot tree. The RAVs from different stages was shown in different colors. The three main parts were showed by pink, orange and green colors' curve lines, respectively.



**Figure 3**

Hypothetical evolutionary models for RAVs from plants. The branch tips are marked with the representative species of different plant stages. Branches are colored based on the taxonomy classification. The circles on the clades represent gene duplication events with different colors. The light-green and orange colors display the numbers of tandem duplications and segmental duplications, respectively. The doplots show the number of RAV members, gene duplication events, tandem duplication, and segmental duplication with numbers and different colors.



**Figure 4**

The GO annotation information of whole RAVs. The X-axis shows the numbers of RAVs, and the Y-axis indicates the types of GO terms. Different types of gene annotation are shown in different colors.

## Supplementary Files

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

- [Additionalfile6.docx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile1.xlsx](#)