



OPEN ACCESS

EDITED BY

Gavin Huttley,
Australian National University, Australia

REVIEWED BY

Michel-Edwar Mickael,
Polish Academy of Sciences, Poland
Sergio Romero Romero,
National Autonomous University of
Mexico, Mexico

*CORRESPONDENCE

Andreia Carina Turchetto-Zolet,
✉ carina.turchetto@ufrgs.br
Marcia Margis-Pinheiro,
✉ marcia.margis@ufrgs.br

[†]These authors have contributed equally to
this work and share first authorship

RECEIVED 28 June 2024

ACCEPTED 23 May 2025

PUBLISHED 13 June 2025

CITATION

Barros NLF, Filgueiras JPC, Trenz TS, Weber G,
Turchetto-Zolet AC and Margis-Pinheiro M
(2025) ASR gene family: a case of
tandem-drive evolution.
Front. Mol. Biosci. 12:1456645.
doi: 10.3389/fmolb.2025.1456645

COPYRIGHT

© 2025 Barros, Filgueiras, Trenz, Weber,
Turchetto-Zolet and Margis-Pinheiro. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

ASR gene family: a case of tandem-drive evolution

Nicolle Louise Ferreira Barros^{1†}, João Pedro Carmo Filgueiras^{1†},
Thomaz Stumpf Trenz¹, Guilherme Weber²,
Andreia Carina Turchetto-Zolet^{1*} and
Marcia Margis-Pinheiro^{1,2*}

¹Programa de Pós-graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, ²Programa de Pós-graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

Introduction: ABA, Stress, and Ripening (ASR) proteins are characterized by the presence of the ABA/WDS domain and are involved in plant development processes and tolerance to abiotic and biotic stresses. Despite their importance as transcription factors or molecular chaperones, a complete understanding of their biological roles is limited by a lack of information on their mechanisms of action, protein structure, and evolutionary relationships between family members. Our previous molecular dynamics simulation analysis of rice OsASR5 suggested that H⁹¹, R⁹², H⁹³, and K⁹⁴, are the main residues involved in the interaction with DNA, essential for the transcription factor activity of this protein. However, the presence and conservation of the DNA-binding domain among ASR family members remain unknown. Likewise, there is a lack of phylogenetic analyses evaluating the evolutionary history of ASR proteins across major taxonomic groups, outside just the *Solanum* species.

Methods: To address these gaps, we conducted a phylogenetic study and protein sequence analyses to gain insights into the evolution of ASR genes in plants. We performed a genome-wide identification of ASR genes via HMMER, using the ABA/WDS domain, in 163 Archaeplastida genomes.

Results and discussion: Our results reveal that the potential origin of the ASR gene occurred in the common ancestor of Streptophytes (Charophytes and Embryophytes). Moreover, our study identifies ASR genes in seedless plants. The evolutionary relationship between 465 ASR homologs, found in 76 species, was estimated through maximum likelihood analysis. The results reinforce the rapid and dynamic evolution of the ASR gene family, reflected by the low support in the deep nodes of the phylogeny and the great variation in the number of ASRs in the genomes evaluated, and in some cases their complete absence. As for diversification, tandem duplications seem to be the main mechanism involved. Regarding the conservation of residues in the domain, only two of the 78 are widely conserved, such as E⁷⁹ and H⁹³. By analyzing the three-dimensional model, we noticed the interaction between them and we hypothesize that they are essential for the stabilization of the domain during interaction with DNA.

KEYWORDS

gene family evolution, intrinsically disordered proteins, ASR proteins, DNA-binding domain, residues substitution

1 Introduction

Whole-genome duplications and clusters of gene duplication events are evolutionary processes that contribute to the origin and expansion of proteins. The diversity of protein families upholds the complex metabolic pathways that plants activate in response to endogenous and exogenous cues. Considering this, the inference of phylogenetic relationships is crucial for processing genomic data since it offers comparative insights into the roles of these biomolecules through evolutionary distance and guides experimental designs. Besides, these analyses can contribute to the understanding of plant physiology, given that the fate of duplicated genes in association with the ecology of the species determines the common or lineage-specific biological traits (Lynch and Conery, 2000; Carretero-Paulet and Fares, 2012; Li et al., 2016). Reconstructing the evolutionary history of gene families helps identify paralogy and orthology relationships (Jakobson et al., 2024) and, therefore, the classification of nomenclatures and subgroups (Wang et al., 2018), providing valuable information about the enrichment (Castro et al., 2017), redundancy, and functional transition of gene family members through group establishment (Tang et al., 2013). It also contributes to the identification of active residues under different selective pressures.

ASR (ABA, Stress, and Ripening) are so-called plant-exclusive transcription factors (TF) and in addition to regulating gene expression, they are involved with chromatin remodeling and maintenance of the native conformation of proteins under stress (Çakir et al., 2003; Wang et al., 2018; Atassanov et al., 2022). This family is associated with developmental processes and stress response (Chen et al., 2011; Arenhart et al., 2013), although information on the mechanisms of action is limited. ASR proteins are known to localize in different subcellular compartments (Arenhart et al., 2013), but the signals promoting protein translocation, the mode of binding to their targets, and the conformational transition they undergo remain unclear. To date, only five ASR proteins have been characterized as intrinsically disordered (Goldgur et al., 2006; Dai et al., 2011; Hamdi et al., 2017; Barros et al., 2023), of which only one had the topology of the binding complex resolved (Barros et al., 2023).

The conformational plasticity of ASR imposes experimental constraints that can be resolved by *in silico* tools, which can predict the chemical environment of each residue and its interactions to ensure the structural switches between the free and ligand-associated modes. Considering this, our group previously showed that the rice ASR5 protein (OsASR5) is intrinsically disordered since it is enriched with charged residues such as E, H, and K while lacking hydrophobic ones such as L and I, and β and bulky conformation (Barros et al., 2023). The growing interest in intrinsically disordered regions is driven by their abundance in TF protein sequences and the alleged role in stabilizing and guiding the DNA-binding domain in the search for motif-containing sites (for details, see Brodsky et al., 2021). Additionally, these regions might assist the TF transitions between DNA grooves, overall spatial accessibility to co-factors, and post-translational modifications (Már et al., 2023). Besides ASR5 the rice genome encodes for five additional ASR gene members (Frankel et al., 2006), compared to 24 members in wheat (Yoon et al., 2021), one in grape (Çakir et al., 2003), and none in *Arabidopsis thaliana*, for example, whose absence imposes yet another drawback

on the functional analysis of ASR gene family. Although ASR proteins have been identified in many species of economic interest, detailed evolutionary relationships among them remain unexplored.

To address these gaps, our group previously proposed the *in silico* three-dimensional model of the rice ASR5 protein and the zinc-dependent DNA binding (Barros et al., 2023), based on the metal-binding (positions one–16) and DNA-binding (positions 87–94) domains of tomato ASR1 (Kalifa et al., 2004; Rom et al., 2006). Furthermore, we concluded that residues H⁹¹, R⁹², H⁹³, and K⁹⁴ of the DNA-binding site and M¹, K⁵, and K¹⁴ of the putative metal-binding site of OsASR5, along with their flanking sites, are crucial for the stability of the complex with a target gene (Barros et al., 2023). Therefore, the present study aimed to understand the molecular evolution of the ASR gene family across the Archaeplastida species and evaluate the composition of ASR domains and their flanking residues. Our results suggest these proteins are not exclusive to land plants, as the founding gene appears to originate from Charophyta. Also, the phylogenetic analysis indicates that tandem duplications promote the rapid diversification of family members, resulting in two main clades formed by ancient and recent groups and unidentified ASR proteins outside these groups. The results produced here contribute to the identification of active residues among ASR proteins, offering insights into how diffuse the structural disorder between them is, and supporting hypotheses about novel ASR functions.

2 Materials and methods

2.1 Database and sequence retrieval

To explore the evolutionary history of the ASR gene family, we performed a genomic identification of these genes in 163 Archaeplastida species. Among these species, three are Rhodophytes, one Glaucophyta, one Prasinodermophyta, 19 Chlorophyceae, nine Charophytes, seven “Cryptogams”, five Gymnosperms, four “early-diverging Eudicots”, 63 Eudicotyledons, and 51 Monocotyledons (Supplementary Table S1). For this purpose, we used HMMER’s *hmmsearch* (Mistry et al., 2013), employing the PFAM domain present in ASR proteins, the ABA/WDS induced protein (PF02496), and an *e*-value < 1e^{−5} was used as a threshold. All predicted polypeptide sequences from the genomes were retrieved via Ensembl Plants (<https://plants.ensembl.org>), Phytozome v13 (<https://phytozome-next.jgi.doe.gov>), or Phycocosm (<https://phycocosm.jgi.doe.gov>). For Phytozome, we used only the dataset containing the primary transcripts. For Ensembl and Phycocosm, as this dataset is not available, the primary transcripts were filtered after *hmmsearch*. Sequences with less than 65% of the domain coverage were removed from further analysis.

2.2 Sequence alignments and phylogenetic analysis

Two datasets were used to estimate the evolutionary history of the ASR gene family in plants. The first one contained 465

sequences from 76 species (hereby called ‘Streptophyta dataset’). The second one included 490 sequences from 46 species, focusing on the Poaceae family (hereby called the ‘Monocotyledon dataset’). The Streptophyta dataset contains species representing different taxonomic groups, from charophyte algae to angiosperms, while the Monocotyledon dataset only included species of monocotyledons, mainly species from the Poaceae family. Charophyte and Bryophyte ASR were used as roots in the Monocotyledon dataset. Both datasets were aligned using MAFFT with the L-INS-I strategy (Katoh et al., 2017). The resulting alignments were manually curated and trimmed, retaining only the region corresponding to the ABA/WDS protein domain (78 amino acids in length). However, for the Streptophyta dataset, only 74 residues were kept in the final alignment, due to the low quality of the alignment in the last four residues. The curated alignments were then used to estimate the phylogenetic trees. The phylogenetic analysis was carried out with the maximum likelihood method using IQTree v2.2.6 (Minh et al., 2020). The trees were estimated using 10,000 UltraFast bootstrap replicates (Hoang et al., 2017). Additionally, the parameters -pers 0.2 and -nstop 500 were used. Both datasets were analyzed in triplicate, and the final tree was selected based on the best log-likelihood value. The substitution models were determined by ModelFinder (Kalyaanamoorthy et al., 2017) and automatically selected based on the best Bayesian Information Criterion (BIC) value. In the Streptophyta dataset, the selected model was LG + I + R5, while in the Monocotyledon dataset, the Q.plant + R5 was selected. For more confidence in the topology found, we also estimated the phylogeny using PhyML (Guindon et al., 2010), with a support calculated from 200 Transfer bootstrap replications (Lemoine et al., 2018). For the Streptophyta dataset, we also used the LG + I + R5 model, while for the Monocots, we used the second best model proposed by ModelFinder, JTT + R4, due to the absence of “Q.plant” models in PhyML.

2.3 Selection analyses

In the study of coding gene evolution, a widely used measure is the ω value. This parameter estimates the selective pressure acting on a gene, where values less than one indicate purifying selection, with values closer to zero reflecting stronger pressure to conserve residues. Conversely, an ω value greater than one indicates positive selection, and the further it deviates from one, the stronger the selection promoting residue diversification. Values equal to or very close to one are indicative of neutral evolution, meaning that variations in residues do not result in significant changes in protein function (Álvarez-Carretero et al., 2023). To estimate selective pressures in the ASR gene family, a subdataset from monocotyledons was created. Coding sequences were retrieved exclusively from Poaceae species, with *Joinvillea ascendens* as the outgroup. The sequences were aligned at the protein level using MAFFT with the L-INS-I strategy (Katoh et al., 2017). Subsequently, PAL2NAL (Suyama et al., 2006) was used to convert the protein multiple sequence alignment into a codon alignment. The alignment was manually curated to retain only the ABA/WDS domain. At this point, all identical sequences were removed to reduce redundancy and avoid bias due to the high number of polyploids. The phylogeny of this subdataset was estimated using IQTree, and

finally, EasyCodeML (Gao et al., 2019) was employed to detect signals of positive selection in the ASR gene family using the site model strategy (Yang et al., 2000).

2.4 Analysis of the three-dimensional model of the ASR domain

We used MEME (Bailey et al., 2015) for identification of conserved motifs within ASR protein sequences, in both Streptophyta and Monocotyledon dataset, with the maximum number of motifs set to 15, with minimum a width of six and a maximum width of 80 amino acids. WebLogo web software (<https://weblogo.berkeley.edu/logo.cgi>) was used to build the ABA/WDS domain logo from the sequences used. One logo was generated for the alignment of the Streptophyta dataset and another for the Monocotyledon dataset. Furthermore, based on the phylogeny of the Monocotyledon dataset, a sequence logo was constructed for each set of sequences grouped into the clades of ASR1, ASR2, ASR3, ASR4, ASR5, and ASR6 to evaluate whether key amino acids were discriminating these clades. With these data, we analyzed their positioning and interactions in the ASR three-dimensional models. For this purpose, we used PyMOL software and the previously modeled conformation of ASR5 from *Oryza sativa* (Barros et al., 2023). The same software was used to perform the alignment between three-dimensional models by the command “align” in a “one to many” or “one-to-one” mode, with five cycles of iterations and 2 Å as a cutoff. Besides, we used the modeling of free ASR5, which does not contain any ligand, and its DNA interaction form, which also incorporates zinc ions. Additionally, we used the AlphaFold Protein Structure Database (Jumper et al., 2021; Varadi et al., 2021) (<https://alphafold.ebi.ac.uk/>) to retrieve protein structures and models representing examples of each of the other five ASR clades. Applying the structure similarity cluster within the database, we chose the model with the highest average pLDDT (Supplementary Table S2). The chosen model was then verified to ensure it belongs to the appropriate ASR clade according to our phylogeny. Although the work of Barros et al. (2023) showed that AlphaFold modeling had lower scores due to the high number of intrinsically disordered regions present in the ASR, our preliminary analyses, regarding the ABA/WDS domain and mainly in the DNA binding portion, proved that the modeling from AlphaFold was similar to Barros’ work. Therefore, we chose to keep them, since our analysis was focused on the DNA binding portion.

Ancestral sequence reconstruction was performed using GRASP software (Foley et al., 2022), employing the LG substitution model along with the alignment and phylogeny from the Streptophyta dataset. Because accurate alignment is critical for this analysis, reconstruction was restricted to the ABA/WDS domain. Additionally, we used the estimated ancestral sequences of Charophyta to model their three-dimensional conformation. For this purpose, AlphaFold2 (via Google Colab) was utilized with three recycling iterations and minimization applied to all five models (Mirdita et al., 2022). The final model was selected as the top-ranked structure based on the pLDDT value post minimization.

3 Results

3.1 Genome-wide identification of ASR genes

From the 163 genomes analyzed, putative ASR genes were identified in 115. In the remaining 47 genomes, HMMER was unable to recover any ASR genes. The ASR gene possibly originated in the common ancestor of Streptophytes (which comprises Charophytes and Embryophytes) (Figure 1), since ASR genes have only been identified in these two lineages. No other lineage of Archaeplastida evaluated (Rhodophyta, Chlorophyta, Glaucophyta, and Prasinodermophyta) seems to encode ASR genes in their genomes. Notably, within Charophyta, ASR genes are not universally present. They are found in the genera *Klebsormidium*, *Spiroglaea*, *Mesotaenium*, and *Zygnema*, but are absent in *Chara*, *Chlorokybus*, and *Mesostigma*. ASR genes are also not found in 16 angiosperm genomes. Among these, we highlight the Brassicaceae family, which includes the model plant *Arabidopsis thaliana*. We also found no ASR genes in five other Brassicaceae species. The loss of the ASR gene possibly occurred in the common ancestor of the Brassicaceae family, given that ASR genes have been found in other families of the Brassicales order. Even so, the small number of genes found in these species is also noteworthy. A single ASR gene was recovered in *Carica papaya* (Caricaceae) and one in *Cleome violacea* (Cleomaceae). This result suggests a reduction in the number of ASR genes in the Brassicales lineage, which may have facilitated the loss of the ASR gene in Brassicaceae.

As for the diversification of the ASR gene family, there seems to be no correlation between the number of genes and plant diversification (Figure 1; Supplementary Table S1). The high frequency of tandem duplications is noteworthy, possibly serving as the main driver of ASR gene diversification. Among all the species analyzed in this study, the three with the highest number of ASR genes were the monocotyledon *Thinopyrum intermedium* (Poaceae) with 48 putative ASR genes, followed by the bryophyte *Sphagnum magellanicum* (Sphagnaceae) and the lycophyte *Diphasiastrum complanatum* (Lycopodiaceae), both with 35 ASR genes. Among the eudicots, the species with the highest number of ASR genes identified was *Corymbia citriodora* (Myrtaceae) with 27 genes, followed by *Chenopodium quinoa* (Amaranthaceae) with 15 genes and *Eucalyptus grandis* (Myrtaceae) with 13. This result indicates a higher rate of duplications in these species, especially in Myrtaceae. On average, we found 4.17 ASR genes in the evaluated eudicot genomes. Furthermore, from the 47 eudicot species in which we identified ASR genes, 10 of them possibly encode for two ASR genes, and 15 only for one.

Preliminary searches were performed using BLASTp, in the Phytozome database with default options, with the OsASR5 (LOC_Os11g06720) query on the genomes of *Oryza sativa* v7.0 and *Solanum lycopersicum* (ITAG5.0). However, the observed e-values ($\sim 1e-3$) were relatively high, and BLASTp failed to recover all six ASR from the *O. sativa* genome. Because of this, we opted to use HMMER for the comprehensive identification of ASR across the 163 genomes studied. Notably, the BLASTp results revealed an interesting observation: the homologous region identified among the ASR proteins was practically identical across sequences (Supplementary Figure S1). Accordingly, we refer to this

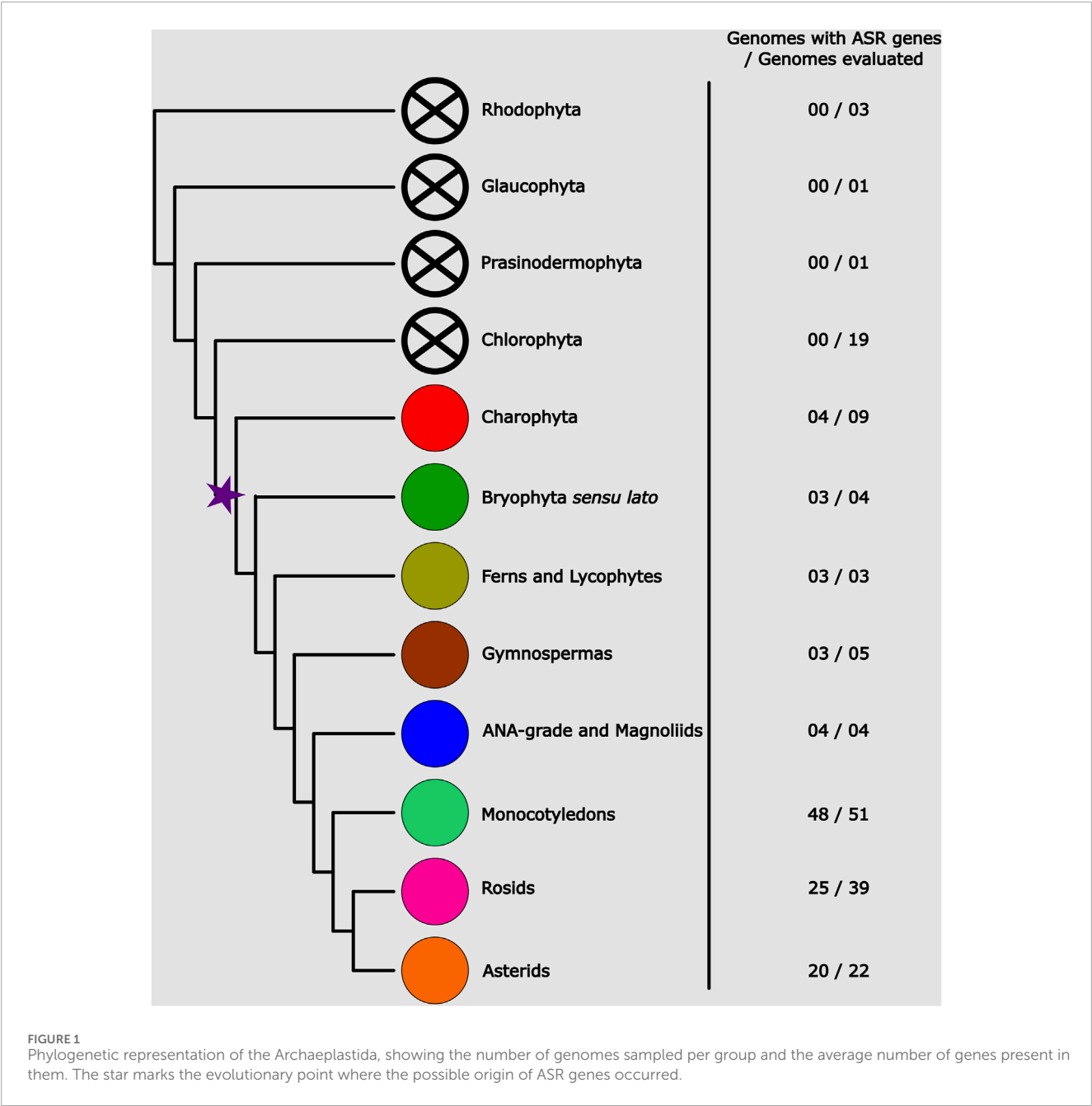
region as the “Core ASR”, as it exhibits the highest identity among the different ASR proteins.

3.2 Evolution and diversification of the ASR gene family

To understand the origin and evolutionary pattern of ASR genes in plants, we estimated a phylogeny using a dataset including representative species from all the major groups of Streptophyta. This dataset comprises 465 sequences from 76 species (Supplementary Table S1). The rapid diversification and duplication characteristics of ASR genes are evident in the phylogeny, which did not converge with the evolutionary history of the species (Figure 2). The complete phylogeny can be found in the Supplementary Figure S2. We used the Charophyta sequences to root the tree. However, since these sequences did not group together, the root was placed at the node that encompasses all Charophyta sequences, allowing us to establish three major clades: The first clade (C1) is an early-diverging lineage due to the presence of ASR sequences from charophytes. In addition to charophytes, sequences from bryophytes, ferns, and gymnosperms are also grouped in this clade. Notably, there are no angiosperm sequences in this clade. In addition, it is worth noting that the sequences belonging to bryophyte species are exclusive to the C1 clade, while the ASR sequences of gymnosperms and ferns also appear in more derived clades. The second clade (C2) is the smallest in terms of both sequence number and species diversity, being the closest to the C1 clade, and it consists solely of monocots and the fern species *Ceratopteris richardii*. The C2 group also comprises the ASR6 and ASR2 genes from *Oryza sativa*. Finally, the third clade (C3) is the most diverse in terms of species and number of sequences. Eudicots and species representing the earliest lineages to diversify from angiosperms (ANA-grade and Magnoliids) are found exclusively in C3. The most striking feature of this clade is the high degree of polytomies, making it challenging to divide into subgroups (Figure 2). The PhyML results revealed a highly similar topology; however, it identified fewer well-supported groups within C3 compared to IQTree (Supplementary Figure S3).

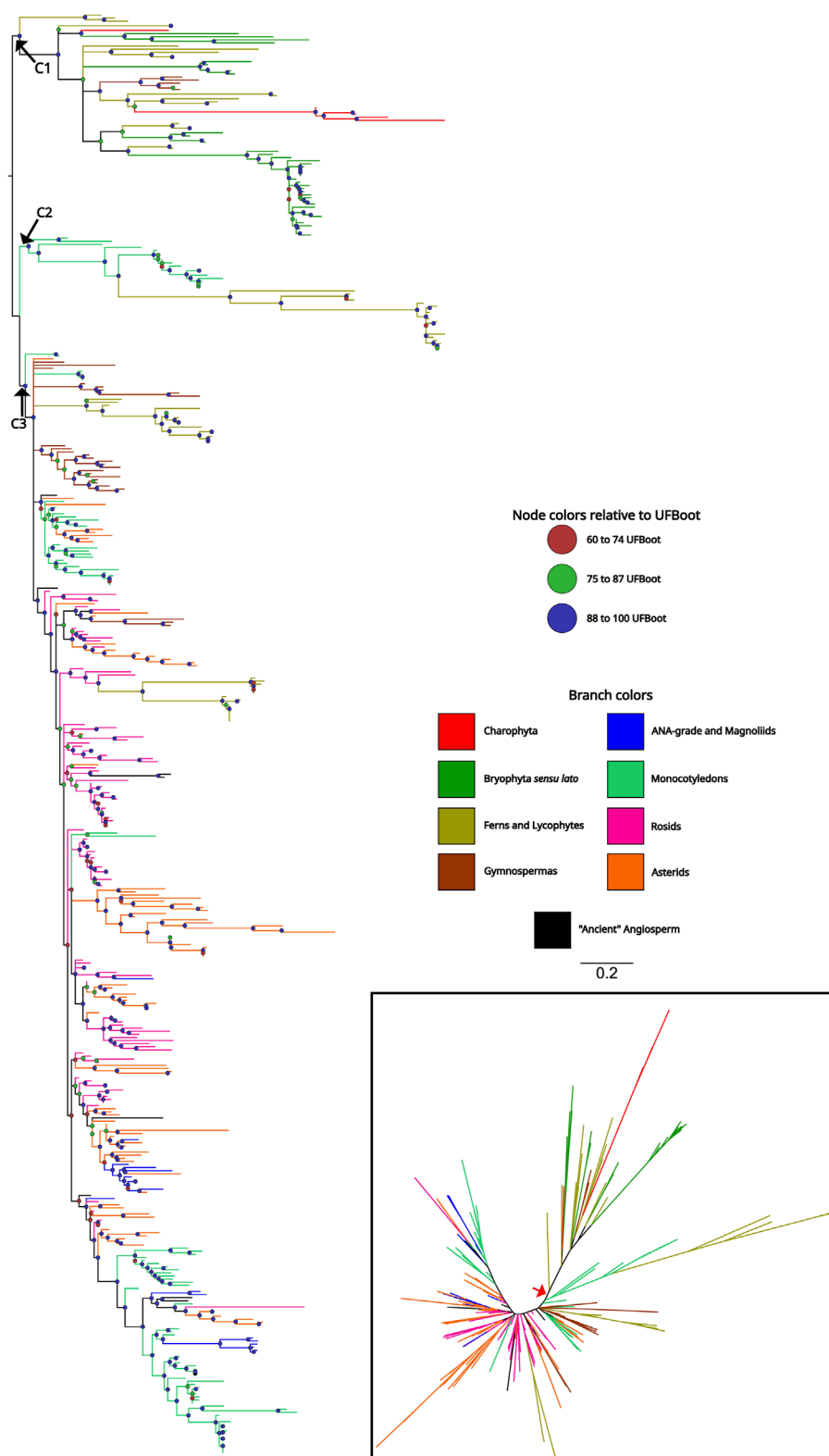
3.3 Phylogeny of ASR genes shows better resolution for recent evolutionary histories

Based on the previous result, we observed that monocot sequences are the first group of angiosperms to appear close to the early-diverging clade and, at the same time, to have sequences that are more distant from the root (Figure 2). This indication of greater differentiation of ASR in monocotyledons led us to carry out a second phylogeny focused on this group, increasing the number of species analyzed. The phylogeny of the monocotyledon dataset provided a clear definition of the evolutionary relationship between ASR genes. The phylogeny can be divided into two large groups, which we named as Group A and Group B (Figure 3). Uncollapsed phylogeny can be found in the Supplementary Figure S4. Group A would be the most ancestral, as it is the closest to the root, and within it, we have the clades comprising the ASR2, ASR5, and ASR6, with



the *Oryza sativa* ASR genes as reference (Frankel et al., 2006). The ASR2 clade is the closest to the root and includes only sequences from Joinvilleaceae and Poaceae (with representation from all five subfamilies evaluated). The same pattern of species is found in the ASR5 clade, which is the most derived within Group A. Finally, the ASR6 clade appears to have diverged later and includes only Joinvilleaceae and three of the Poaceae subfamilies (Chloridoideae, Oryzoideae, Panicoideae). Between the ASR5 and ASR6 clades, there is a clade formed exclusively by sequences from the Pooideae subfamily. These sequences are putative ASR6s, as they are closely related to the rice ASR6 clade. This positioning may indicate that ASR6 of Pooideae has undergone greater differentiation compared to other subfamilies of Poaceae.

Group B consists of more derived ASR genes, distant from the root and forming a monophyletic group that includes ASR1, ASR3, and ASR4 proteins. Both the ASR4 and ASR3 clades contain sequences from all five Poaceae subfamilies and their sister family, the Joinvilleaceae. The ASR1 clade also includes a sequence from *Ananas comosus* (Bromeliaceae); however, sequences from the Pooideae subfamily are absent. Therefore, the Pooideae subfamily is absent from two of the six ASR clades highlighted here. Despite focusing on monocotyledons, the phylogeny's best resolution is restricted to Poaceae and Joinvilleaceae. This result reinforces the rapid evolution of ASR genes, which makes it challenging to recover deeper evolutionary histories with these genes. The other seven monocot species evaluated, belonging to different orders,

**FIGURE 2**

Phylogenetic tree of ASR proteins from Streptophyta. The topology was recovered after applying the maximum likelihood method, in IQTree, on the amino acid sequences of the ABA/WDS domain (Pfam ID: PF02496) that characterize ASR proteins. The colored circles on the nodes represent the Ultrafast Bootstrap obtained from 10,000 replicates. Branches with Ultrafast Bootstrap less than or equal to 59 were deleted and represented as polytomies. The phylogeny included 76 species representing the main taxa of the Streptophyta clade. Branches colors indicate which clade the sequence belongs to. The unrooted tree is represented in the lower right corner and the red arrow indicates the point where the tree was rooted.

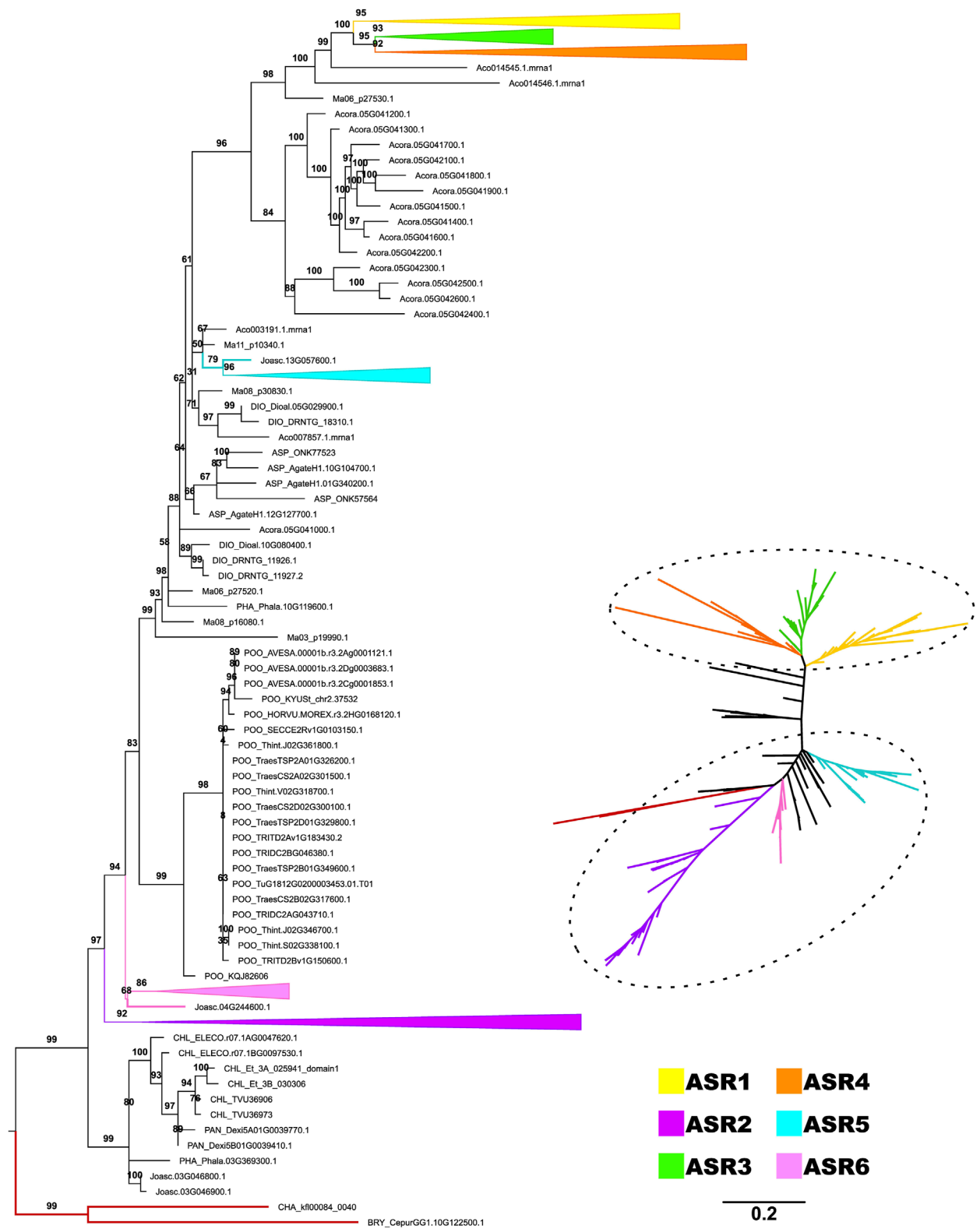


FIGURE 3
Diversification of ASR protein family from monocots. The topology was recovered after applying the maximum likelihood method, in IQTree, on the amino acid sequences of the ABA/WDS domain (Pfam ID: PF02496) that characterize ASR proteins. The phylogeny included 46 species belonging to the monocotyledon clade. The numbers on each branch represent the Ultrafast Bootstrap obtained from 10,000 replicates. The Charophyta and Bryophyta species established the root (red branches). The colors highlight the groups composed of putative proteins ASR1 (yellow), ASR2 (purple), ASR3 (green), ASR4 (gray), ASR5 (blue), and ASR6 (pastel pink), as seen in the trees rooted (left) and unrooted (right). The six ASR clades are collapsed in the rooted phylogeny.

did not group with any defined clade. In some cases, they did not fall within the two highlighted groups (Group A and Group B), being positioned between them. Notably, most sequences from *Acorus americanus* (Acoraceae) (14 of 15) occupy this intermediate position. Additionally, two sequences from *A. comosus* (out of seven) and one from *Musa acuminata* (out of six) are also in this position. All other sequences are found exclusively within the large group of ancestral ASR, except for one single sequence from *A. comosus* grouped within the ASR1 clade.

It is worth mentioning that Group B found in monocots (ASR1, ASR3, and ASR4) remains the most distant and derived even in the phylogeny using the Streptophyta dataset. These findings suggest further differentiation and, potentially, the acquisition of new functions within this ASR group. This result reinforces the probable diversification and importance of this gene family in monocotyledons, especially in Poaceae, compared to other angiosperms. For the monocotyledon dataset, PhyML estimated a phylogeny highly similar to that obtained with IQTree, allowing the identification of the six ASR clades and their classification into Groups A and B, along with some intermediate sequences (Supplementary Figure S5).

3.4 Composition of motifs follows the ancestry recovered by phylogeny

Motif analysis of the Monocotyledon dataset revealed specific patterns associated with different ASR clades (Supplementary Figure S6). Motifs 0, 1, and 2 correspond to the ABA/WDS domain, which is conserved across all clades. Motif 6 is exclusive to group A (early-divergent clade), although it is not universally present. Similarly, Motif 10 is exclusive to group A, except for three sequences from group B. It is located in the C-terminal portion and it is absent in the ASR2 clade. In contrast, Motif 8, while also present in the C-terminal portion, is exclusive to group B (derived clade), with only three sequences from group A exhibiting this motif. Motif 3 corresponds to the putative zinc-binding domain and shows a scattered distribution, albeit with enrichment in the ASR5 and ASR1 clades. Motif 11 occurs multiple times within the same protein and is enriched in the ASR6 clade, being shared between the ASR2 and ASR6 clades. Motifs 9, 7, 4, and 13 are restricted to the Pooideae family. Motif 9 is enriched in ASR3 genes and in some members of the ASR2 clade, whereas the remaining motifs are generally exclusive to the ASR2 clade, with few exceptions. Motif 12 is also enriched in Pooideae, being found in the ASR3 and ASR2 clades, though not universally. The last two motifs are clade-specific, with fewer than three exceptions: Motif 14 is exclusive to the ASR6 clade, while Motif 5 is exclusive to the ASR2 clade.

Considering the Streptophyta dataset, the C1 early-divergent lineage encompasses sequences with aggregated motives compared to the other two lineages, and an overall Motif 11-5-4-1-0-9 organization (Supplementary Figure S7). It is noteworthy the replacement of Motif 3 for Motif 11 in the N-terminal portion of two ancestral sequences (BRY_Sphmag05G052700.1 and BRY_Sphmag01G112900.1) to reach the above mentioned pattern, and the acquisition of Motif 4. Besides, the sequences without Motifs 2 or 9 in the C-terminal portion are made of Motifs 6 or 7 in exchange. On the other hand, the C3 clade encompasses less-structural diverse

sequences composed mainly of core Motifs (ABA/WDS) and Motifs 3 and 6. Other members mark the expansion of Motif 10 and a progressive introduction and expansion of Motif 13 in the C3 clade, where one of four sequences bear two copies of Motif 13. The change is noticed in the C2 clade, whose sequences have up to three copies of Motif 13 and are the only ones with Motifs 12 and 14, besides the predominance of Motif 8 (except by three sequences in C3).

3.5 ASR clades are formed by proteins with active sites in the DNA-binding domain

The composition of the recovered clades' DNA-binding domain showed a predominance of Q residues at position 91 in the early diverged clades of ASR proteins, formed by ASR2 and ASR6 sequences (Figure 4A). On the other hand, the intermediate ASR5 clade marks the transition to the emergence of positively charged residues at the same position, as noticed among the more recent ASR1, ASR3, and ASR4 clades. Moreover, the divergent clades also highlight the transition in the pattern of positive residues found in the first-divergent proteins (ASR2 and ASR6) at position 92, to a more diverse class of amino acids, represented by residues such as S, G, and E, in the other clades. In contrast, the amino acid at position 93 is exclusively a H across all recovered groups. Finally, a tendency towards positive residues was observed in most clades, where K and R residues are found at position 94, except by the ASR2 clade, where Q and M residues are conserved.

Regarding the residues flanking the DNA-binding domain, which are also crucial for the stability of the complex with the target to be regulated (Barros et al., 2023), the sequence logos highlight prominent patterns: the derived clades ASR1, ASR3, and ASR4 have acquired a D at position 50, while K is conserved at positions 59 and 85 across all clades. In the first diverged clade, ASR2, residue position 89 is mainly occupied by G, a non-polar residue. This glycine is replaced by R, a positively charged amino acid residue, in ASR6. This class of amino acids became predominant in the other clades, which are exclusively formed by K in this position.

As expected, the residues found in the ABA/WDS domain within the Streptophyta dataset are more diverse than those in the Monocotyledon dataset. Besides H⁹³, found in the DNA-binding domain, only two E residues, at positions 79 and 114, and an A, at position 76, are conserved in most of the evaluated sequences (Figure 4B). After setting the ASR5 from *Panicum hallii* as a reference, the alignment between this three-dimensional model and other representatives of the remaining five ASR clades of Monocots, modeled by AlphaFold, showed RMSD values of 0.23 Å (ASR1, 288 atoms), 0.38 Å (ASR3, 342 atoms), 0.41 Å (ASR4, 338 atoms), and 0.41 Å (ASR6, 316 atoms). These results corroborate the conservation of the ABA/WDS domain among these paralogs despite the disparity of the OsASR2 model (RMSD 8.76 Å, 433 atoms) (Supplementary Figure S8). Consider checking out Supplementary Table S2 for clarification about the above-mentioned representatives of each ASR clade. Additionally, for a better understanding of ASR protein domain organization, consider looking for the distribution and residue composition of each domain in light of a rice ASR5 protein as a reference in Supplementary Figure S9. Interestingly, even the reconstructed ancestral sequence of the ABA/WDS domain

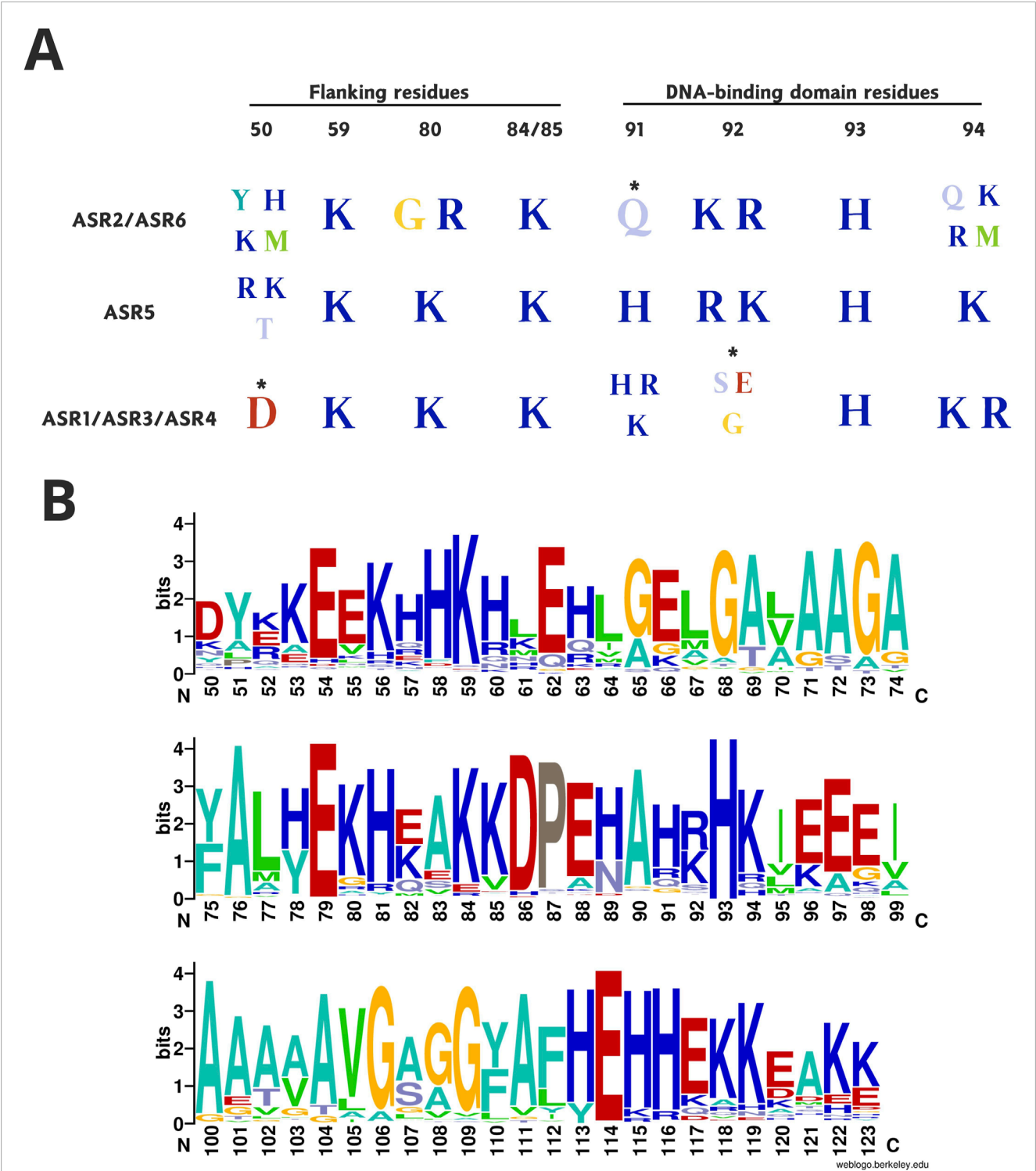


FIGURE 4
Comparative residue composition between ASR proteins. The illustration in (A) depicts the dynamics of residue substitutions between ASR groups of monocots according to the diversification of each, that is, from the most ancestral (ASR2, ASR6, and ASR5) to the most derived (ASR1, ASR3, and ASR4) ones. The highlighted positions belong to flanking sites (50, 59, 80, 84, and 85) and the DNA-binding domain (positions 91–94) predicted by [Barros et al. \(2023\)](#) as crucial for the OsASR5 transcription factor activity. Asterisks (*) mark positions not occupied exclusively by the residues presented. The sequence logos (B) depicts the composition of the ABA/WDS (Pfam ID: PF02496) of all ASR proteins belonging to the Streptophyta dataset. In the plot, the x-axis points out the positions in the sequences, and the y-axis represents the amplitude of each residue, which is proportional to the conservation in the alignment between the sequences.

(KAKKEKKHKRNMAGVGALAAAGGFAAWEAHEAFVDPGH AKKHMEAGVAGAVGAGGYALHEHHEKKLEK) exhibited a fold highly similar to those observed in the six ASR clades mentioned above. The RMSD values ranged from 0.35 to 0.79 Å, with the exception of ASR2, which showed a significantly higher value of 8.6 Å. The reconstructed ancestral domain had an average pLDDT value of 85.

3.6 The ABA/WDS domain is under purifying selection in Poaceae ASR genes

Our analyses using our dataset did not reveal any sites under positive selection in the Poaceae ASR genes. The phylogeny used to perform the site model test is found in [Supplementary Figure S10](#). The estimated ω values based on the M8 model indicates a predominance of purifying selection, with values ranging from 0.050 to 0.357 ([Supplementary Figure S11](#)). We observed that the highest ω values occurred at the extremities of the ABA/WDS domain (both C and N terminal). These peaks suggest that these regions experience lower selective pressure relative to other sites, thereby permitting greater residue variability, albeit still limited (indicated by the low ω values). Particularly interesting are the peaks in the central region of the domain, specifically at sites 91 and 92, which are located within the DNA binding site. These residues may contribute to binding specificity, facilitating the recognition of different DNA motifs.

4 Discussion

Evolutionary studies of ASR genes are limited, with the first and most comprehensive one conducted by [Frankel et al. \(2006\)](#). Our work applies broad sampling and robust phylogenetic methods, such as maximum likelihood, to investigate the evolution of ASR genes. Prior to the review by [González and Iusem \(2014\)](#), ASR genes were only described in Embryophytes, a clade that includes Gymnosperms and Angiosperms. Our study identifies ASR genes in early groups of land plants, such as mosses and ferns.

The likely origin of the ASR genes lies in the common ancestor of Charophytes and Embryophytes, as we did not detect their presence in Chlorophyta genomes ([Figure 1](#); [Supplementary Table S1](#)). Since ASR genes are involved in response to desiccation and are ABA-responsive ([Yang et al., 2005](#); [Sachdeva et al., 2020](#)), they might have played a role in the adaptation of plants to terrestrial environments ([Delwiche and Cooper, 2015](#); [Kapoor et al., 2022](#)). One of the points highlighted by [Frankel et al. \(2006\)](#) was the greater similarity among ASR paralogues when compared to their orthologues. These observations are in line with the hypothesis of the functional redundancy between OsASR1 and OsASR5 proposed by [Arenhart et al. \(2012\)](#), which are co-localized in the nucleus and cytoplasm, and whose conformational pattern is highly similar ([Supplementary Figure S12](#)). Our results corroborate this observation, as the phylogeny showed low support at deep nodes ([Figure 2](#)). The C3 clade is a clear example, exhibiting numerous polytomies and well-supported minor subclades. These smaller subclades are generally composed exclusively of sequences

derived from species within the same botanical family and, in certain cases, from the same order.

The distribution of genes suggests that the process of tandem duplications is the main responsible for the expansion of the ASR family. From the 465 sequences analyzed in the Streptophyta dataset, 340 came from tandem duplications [they are found on the same contig/chromosome separated by less than 100 Kb ([Liu et al., 2021](#))]. Since 21 are unique sequences (one ASR in the genome), the other 104 came from other duplication mechanisms.

[Frankel et al. \(2006\)](#) pointed out that ASR2 and ASR6 from *Oryza sativa* were the most divergent compared to other ASR genes analyzed. This statement is corroborated and reinforced by our analyses. In the Streptophyta dataset, the C2 clade is the closest to the ancestral ASR genes ([Figure 2](#)), consisting mostly of monocot sequences, which include ASR2 and ASR6 from rice ([Supplementary Figure S2](#)). The other four ASR sequences from rice are grouped in C3. For ASR6, the *O. sativa* ssp. japonica sequence (LOC_Os04g34600) is the only one in the clade, within the Oryzoideae subfamily, that contains the ABA/WDS domain incomplete. Even though the *O. sativa* ssp. indica genome contains ASR6 (BGIOGA015105) with the complete domain. This finding indicates a fast accumulation of mutations in ASR6 genes during the differentiation of the *O. sativa* ssp. japonica ([Londo et al., 2006](#); [Campbell et al., 2020](#)). However, the possibility that this could be a sequencing/assembly error can not be ruled out, due to the presence of seven uncertain nucleotides (N) in the first exon ([O'Rawe et al., 2015](#)). For tomato ASR, the five genes were recovered, and as expected, four of them are grouped in the same subclade in C3, along with other sequences from Solanaceae ([Supplementary Figure S2](#)).

The pLDDT values serve as a measure of confidence in the models generated by AlphaFold2. This scale ranges from 0 to 100, where values exceeding 90 indicate high precision, values between 70 and 90 suggest good precision, and values below 70 denote lower confidence, with those below 50 often indicating intrinsically disordered regions ([Middendorf and Eicholt, 2024](#)). The overall average pLDDT of representatives from each selected clade shows that only ASR1 has a confident prediction, reflecting the intrinsically disordered nature of the ASR proteins. Nevertheless, our study focuses on the “ABA/WDS induced protein” domain. When considering the pLDDT values solely for the domains, only ASR2 and ASR6 exhibit values below 70; furthermore, when evaluating just the core of the domain, only ASR2 demonstrates low confidence ([Supplementary Table S2](#)).

Some ASR proteins have already been characterized as intrinsically disordered, including the “ABA/WDS induced protein” domain. Experiments with two representatives of the ASR5 clade revealed that they undergo conditional folding, meaning that under certain conditions, they adopt tertiary structures ([Hamdi et al., 2017](#)). The authors tested glycerol (which mimics dehydration), presence of zinc ions, and 2,2,2-trifluoroethanol (TFE), which mimics the effects of a protein's hydrophobic interactions with its target ([Hamdi et al., 2017](#)). Conditional folding was similarly reported for tomato ASR1 under dehydration conditions and in presence of zinc ions ([Goldgur et al., 2006](#)). Furthermore, a previous *in silico* study by our group demonstrated that the addition of zinc ions stabilizes ASR and promotes the formation of secondary structures ([Barros et al., 2023](#)). Moreover, we observed a strong similarity between this model and those

predicted by AlphaFold, particularly in the core of the domain. Proteins are dynamic, constantly changing conformations, and the AlphaFold prediction represents a single snapshot of the protein's conformational landscape (Guo et al., 2022). A study shows that AlphaFold2 can systematically identify disordered regions that present conditional folding (Alderson et al., 2023). Based on these results, we propose that the core of the domain adopts a functional helix–turn–helix fold, a structural motif well documented in the literature and associated with DNA binding.

Despite the reasonable level of amino acid conservation in ASR sequences, only four residues are highly conserved (A⁷⁶, E⁷⁹, H⁹³, and E¹¹⁴) throughout all clades (Figure 4B). Glutamic acid (E) commonly increases the accessibility of proteins to the solvent (Uversky, 2013). This parameter contributes to the determination of binding hotspots (Barik et al., 2015), as it is associated with the formation of complexes (Chakravarty et al., 2013). The H⁹³ residue is critical for OsASR5 binding to the *STAR1* target (Barros et al., 2023). From the alignment between the free and bound OsASR5 models (Figures 5A,B), we observed that the α -helix involved in DNA binding, and containing the H⁹³, is missing in the free OsASR5 conformation (Figure 5C, yellow line) compared to the bound one (Figure 5D). We also noticed that E⁷⁹ interacts with H⁹³ and are closer to each other (1.8 Å) in the bound OsASR5 model (Figure 5D). Therefore, E⁷⁹ residue can potentially contribute to ASR stability when binding to DNA, allowing the α -helix establishment. Nevertheless, site-directed mutagenesis experiments should be performed to confirm such hypotheses.

In terms of the monocots ASR groups, the main amino acids involved in DNA-binding display different sequence consensus (Figure 4A; Supplementary Figure S13). The first-divergent ASR proteins in Group A (ASR2 and ASR6) are mostly composed of E at position 91, and basic amino acids at position 92. Conversely, Group B contains basic amino acids at position 91, and a greater diversity at 92, containing polar and non-polar residues (Figure 4A). This variety in the DNA-binding domain can imply an increased diversification in which *cis*-elements the ASR proteins can recognize and regulate, as the different residues interact distinctly with DNA bases (Hossain et al., 2023). Functional studies using the ChIP-seq approach would pave the way to understanding how ASR proteins differentiate themselves, regarding the promoters they bind and which set of genes they regulate (Ricardi et al., 2014).

Regarding the composition of the sequences, there is a conservation of residues that promote structural disorder, such as E, H, K, A, and G (Romero et al., 2000), and a low frequency of hydrophobic residues among ASR proteins of the Streptophyta dataset (Figure 4B). In addition to this evidence, the structural disorder is also revealed by the difficulties in building multiple alignments (Riley et al., 2023), such as what occurred during the assembly of our Streptophyta dataset. It is important to highlight that the predominance of charged residues, as observed among ASR proteins, favors the activity of molecular chaperones (Volkin et al., 1993; Narberhaus, 2002), and the enrichment of E promotes interaction with histones (Uversky, 2013), as noticed in grape ASR (Atassanov et al., 2022).

We hypothesized that ASR proteins are versatile and that the ancestral ones worked as molecular chaperones. According to

Rebeaud et al. (2021), the increase in quality control required by the expansion of eukaryotic proteomes did not occur through the diversification of new families of core chaperones but rather through the duplication of existing members. Moreover, the tendency towards purifying selection of mutations that destabilize protein conformations (Smock et al., 2016), the lack of investment in sequence variability for interaction with new substrates due to the reduced binding specificity of chaperones, the uniformity between types of substrates (Bogumil et al., 2013), and the compensatory effects between copies may have contributed to the neofunctionalization and diversification of regulatory ASR proteins.

Since the duplication of transcription factors can be followed by evolutionary novelties in regulatory networks (Voordeckers et al., 2015), mutations in these proteins produce adverse pleiotropic effects (Hsia and McGinnis, 2003), in contrast to those that affect only *cis*-regulatory elements (Wray, 2007). Because of this, the conservation of the ABA/WDS domain, in addition to functional evidence (Arenhart et al., 2013; Dominguez and Carrari, 2015; Arenhart et al., 2012), corroborates the role of ASR proteins as transcriptional regulators, whose family expansion is essentially based on duplications so that the original copy maintains the stability of the regulatory network, while the other is subject to greater evolutionary flexibility, resulting in subfunctionalization, such as responses to different environmental cues, or neofunctionalization (Lynch and Force, 2000; Conant and Wolfe, 2008).

Additionally, the existence of few ASR genes among unicellular eukaryotes and the noticeable abundance of them in higher plants suggest involvement with the occupation of the terrestrial environment (Kapoor et al., 2022), and it corroborates their role as TFs assigned to some members of this family (Arenhart et al., 2013; Dominguez and Carrari, 2015; Arenhart et al., 2012) since this class of proteins is more abundant in plants compared to other eukaryotes (Shiu et al., 2005).

5 Conclusion

ASR genes seem to have emerged before the rise of Charophytes and evolved dynamically in different groups, or even in a species-specific manner, which encompasses several tandem duplications and gene losses. Regarding Angiosperms, Poaceae is the only family possessing ASR genes that are evolutionarily closer to their ancestors, in addition to their presence in the most derived clades, indicating a greater importance of ASR genes in Poaceae. Regarding ASR proteins from monocots, we highlight the DNA-binding site, particularly residues 91 and 92. These residues exhibit greater variability compared to residues 93 and 94, as indicated by the estimated ω values. This suggests that sites 91 and 92 may play a role in recognizing different DNA motifs, which has yet to be tested. Gene functional analyses are still necessary to unveil which set of genes the divergent ASR can regulate, or whether they are actually functionally redundant. Furthermore, we proposed the structural importance of the E(X)₁₄H motif within the ABA/WDS domain, based on the conservation of these residues and the three-dimensional model of ASR.

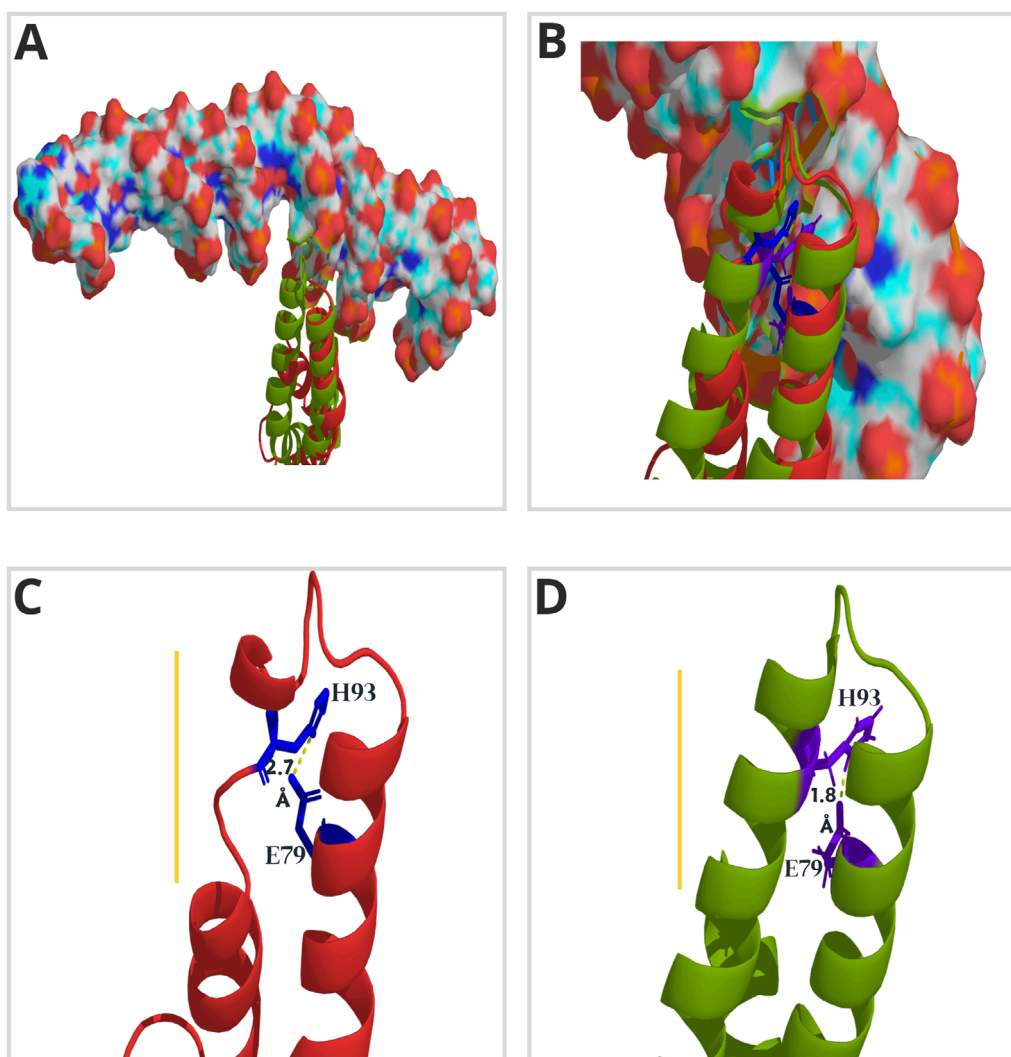


FIGURE 5

Three-dimensional (3D) model of ASR5 protein from rice (*Oryza sativa* ssp. *japonica*). The lateral view highlights the α -helices and loops that form the DNA-binding domain, as proposed by Barros et al. (2023). The model in red shows the free OsASR5 protein, while the one in green represents the protein bound to the *cis*-element (illustrated in surface mode) of the promoter region of the *STAR1* (Barros et al., 2023). Both conformations are aligned in (A) and (B) to emphasize the structural gain during OsASR5-*STAR1* complex formation. In (C) and (D), the distances in Angstrom ($\text{\AA}$) between residues H93 and E79 (illustrated in stick mode) in the free (red) and bound (green) versions of OsASR5, respectively, are illustrated. The yellow line points out the relevant structural differences.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

NLFB: Conceptualization, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing. JF: Conceptualization, Data curation, Formal Analysis, Methodology, Visualization, Writing – original draft. TT: Conceptualization, Methodology, Visualization, Writing – original

draft, Writing – review and editing. GW: Data curation, Writing – review and editing. AT-Z: Methodology, Resources, Supervision, Writing – review and editing. MM-P: Conceptualization, Resources, Supervision, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). NLFB is supported by CNPq

Grant number 140586/2021-0. TST is supported by FAPERGS (PDT12 - Proc. 23/2551-0001893-4). ACT-Z is supported by CNPq Grant number 313949/2023-9.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2025.1456645/full#supplementary-material>

SUPPLEMENTARY TABLE S1

Species subjected to genome-wide identification of ASR genes. This table includes the taxonomic classification of each species, the prefixes used to identify the sequences in their respective datasets, the number of genes initially recovered, and the number retained after refinement and preliminary analyses. Genomes with no recovered ASR genes are highlighted in purple. Genomes that lacked ASR genes after filtering are highlighted in yellow. Monocotyledon genomes included in the Streptophyta dataset are highlighted in light green.

SUPPLEMENTARY TABLE S2

Access code for models retrieved from the AlphaFold Protein Structure Database, which ASR clade in monocotyledons it represents, species name, and average quality value of the modeled protein (pLDDT), for the domain, and for the "core portion".

SUPPLEMENTARY TABLE S3

Sequence acronyms used in the phylogenies estimated by PhyML and their corresponding acronyms in IQTree. The modifications were made to comply with the ten-character header limit imposed by PhyML.

SUPPLEMENTARY FIGURE S1

Results of BLAST using OsASR5 as the query. Schematic representation of the regions where BLASTp returned hits in the genome of *Oryza sativa* and *Solanum lycopersicum*, along with the corresponding e-values.

SUPPLEMENTARY FIGURE S2

Phylogram of the ASR gene family in Streptophytes. Maximum-likelihood phylogeny, estimated by IQTree, containing 76 species, with representatives of the main clades of Streptophytes. The phylogeny was estimated from proteins, using the portion of the ABA/WDS induced domain and the model LG+I+R5. Branches colors indicate which taxonomic group the sequence belongs to. UltraFast Bootstrap values are indicated in their respective branches.

SUPPLEMENTARY FIGURE S3

Phylogram of the ASR gene family in Streptophyte, estimated by PhyML using the model LG+I+R5. Phylogeny was estimated from proteins, using the

ABA/WDS-induced domain portion. Branch colors indicate to which taxonomic group the sequence belongs. Bootstrap values are indicated in their respective branches. Correspondence codes are found in [Supplementary Table S3](#).

SUPPLEMENTARY FIGURE S4

Phylogram of the ASR gene family in Monocotyledons. Maximum-likelihood phylogeny of monocotyledons, estimated by IQTree, based on the peptide sequence of the ABA/WDS induced domain and using the model Q.plant+R5. UltraFast Bootstrap values are indicated in their respective branches. A sequence of Charophyta and a sequence of Bryophyta were used to root the tree, and its branches are highlighted in red. The ASR1 clade is highlighted in yellow, ASR2 in purple, ASR3 in green, ASR4 in orange, ASR5 in blue, and ASR6 in light pink.

SUPPLEMENTARY FIGURE S5

Phylogram of the ASR gene family in Monocotyledons, estimated by PhyML using the model JTT+R4. Phylogeny was estimated from proteins, using the ABA/WDS-induced domain portion. A sequence of Charophyta and a sequence of Bryophyta were used to root the tree, and its branches are highlighted in red. The ASR1 clade is highlighted in yellow, ASR2 in purple, ASR3 in green, ASR4 in orange, ASR5 in blue, and ASR6 in light pink. Correspondence codes are found in [Supplementary Table S3](#).

SUPPLEMENTARY FIGURE S6

Protein motifs of the Monocotyledons ASR gene family. The Pooideae sequences, although not grouped into the ASR6 clade, show similarity in the composition of the motifs with it. Therefore, we consider them as putative ASR6.

SUPPLEMENTARY FIGURE S7

Protein motifs of the Streptophytes ASR gene family.

SUPPLEMENTARY FIGURE S8

Three-dimensional (3D) models of ASR proteins representatives of each Monocot ASR clade. The lateral view shows the alignment between the models of an ASR1 (cyan), ASR2 (yellow), ASR3 (purple), ASR4 (orange), ASR5 (magenta), and ASR6 (black) modeled by AlphaFold, emphasizing the conservation of the ABA/WDS domain (Pfam ID: PF02496).

SUPPLEMENTARY FIGURE S9

Domain organization and regions of interest of a rice ASR5 protein. The OsASR5 three-dimensional (3D) model in (A) highlights, by a color code, the putative metal-binding site (green), the ABA/WDS domain (black) scrutinized by this study, which encompasses the core peptide (light blue) conserved among ASR proteins, including the DNA-binding domain (DBD) (dark blue). The color red represents the remaining regions of OsASR5. The supporting schematic illustration depicted in (B) labels each domain following the color code, along with the (C) protein sequence of OsASR5.

SUPPLEMENTARY FIGURE S10

Phylogram representing the ASR genes sub-dataset used in CodeML. The phylogeny was estimated by IQTree using the model TIM3+F+R5 and the nucleotide sequences. The phylogram was rooted only for visualization. The unrooted tree was used for CodeML. The arrow indicates the branch where the tree was rooted.

SUPPLEMENTARY FIGURE S11

Posterior mean ω for each amino acid site within the ABA/WDS domain. These values were estimated using the M8 model in EasyCodeML, based on ASR gene sequences from Poaceae and Joinvilleaceae. Site numbering corresponds to the position of the ABA/WDS domain in ASR5 of *Oryza sativa* (LOC_Os11g06720).

SUPPLEMENTARY FIGURE S12

Three-dimensional (3D) models of ASR1 and ASR5 proteins from rice (*Oryza sativa* ssp. japonica). The lateral view shows the alignment between the OsASR1 (cyan) and OsASR5 (magenta) models provided by AlphaFold.

SUPPLEMENTARY FIGURE S13

Comparative residue composition between monocot ASR protein groups. The sequence logos depict the composition of the ABA/WDS (Pfam ID: PF02496) of the (A) ASR1, (B) ASR2, (C) ASR3, (D) ASR4, (E) ASR5, and (F) ASR6 clades. In the plot, the x-axis points out the positions in the sequences, and the y-axis represents the amplitude of each residue, which is proportional to the conservation in the alignment between the sequences.

References

- Alderson, T. R., Pritišanac, I., Kolaric, D., and Forman-Kay, J. D. (2023). Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Biophys. Comput. Biol.* 120 (44), e2304302120. doi:10.1073/pnas.2304302120
- Álvarez-Carretero, S., Paschalia, K., and Yang, Z. (2023). Beginner's guide on the use of PAML to detect positive selection. *Mol. Biol. Evol.* 40 (4), msad041. doi:10.1093/molbev/msad041
- Arenhart, R. A., de Lima, J. C., Pedron, M., Carvalho, F. E. L., da Silveira, J. A. G., Rosa, S. B., et al. (2013). Involvement of ASR genes in aluminium tolerance mechanisms in rice. *Plant Cell Environ.* 36 (1), 52–67. doi:10.1111/j.1365-3040.2012.02553.x
- Arenhart, R. A., Margis, R., and Margis-Pinheiro, M. (2012). The rice ASR5 protein: a putative role in the response to aluminum photosynthesis disturbance. *Plant Signal. & Behav.* 7 (10), 1263–1266. doi:10.4161/psb.21662
- Atanassov, H., Parrilla, J., Artault, C., Verbeke, J., Schneider, T., Grossmann, J., et al. (2022). Grape ASR-silencing sways nuclear proteome, histone marks and interplay of intrinsically disordered proteins. *Int. J. Mol. Sci.* 23 (3), 1537. doi:10.3390/ijms23031537
- Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43 (W1), W39–W49. doi:10.1093/nar/gkv416
- Barik, A., Nithin, C., Karampudi, N., Mukherjee, S., and Bahadur, R. P. (2015). Probing binding hot spots at protein–RNA recognition sites. *Nucleic Acids Res.* 44 (2), e9. doi:10.1093/nar/gkv876
- Barros, N. L. F., Siqueira, A. S., Arenhart, R. A., and Margis-Pinheiro, M. (2023). Modeling the zinc effect on OsASR5-STAR1 promoter interaction by molecular dynamics. *Proteins* 91 (7), 944–955. doi:10.1002/prot.26481
- Bogumil, D., Alvarez-Ponce, D., Landan, G., McInerney, J. O., and Dagan, T. (2013). Integration of two ancestral chaperone systems into one: the evolution of eukaryotic molecular chaperones in light of eukaryogenesis. *Mol. Biol. Evol.* 31 (2), 410–418. doi:10.1093/molbev/mst212
- Brodsky, S., Jana, T., and Barkai, N. (2021). Order through disorder: the role of intrinsically disordered regions in transcription factor binding specificity. *Curr. Opin. Struct. Biol.* 71, 110–115. doi:10.1016/j.sbi.2021.06.011
- Çakir, B., Agasse, A., Gaillard, C., Saumonneau, A., Delrot, S., and Atanassova, R. (2003). A grape ASR protein involved in sugar and abscisic acid signaling. *Plant Cell* 15 (9), 2165–2180. doi:10.1105/tpc.013854
- Campbell, M. T., Du, Q., Liu, K., Sharma, S., Zhang, C., and Walia, H. (2020). Characterization of the transcriptional divergence between the subspecies of cultivated rice (*Oryza sativa*). *BMC Genomics* 21 (1), 394. doi:10.1186/s12864-020-06786-6
- Carretero-Paulet, L., and Fares, M. A. (2012). Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Mol. Biol. Evol.* 29 (11), 3541–3551. doi:10.1093/molbev/mss162
- Castro, P. H., Lilay, G. H., Muñoz-Mérida, A., Schjøerring, J. K., Azevedo, H., and Assunção, A. G. L. (2017). Phylogenetic analysis of F-bZIP transcription factors indicates conservation of the zinc deficiency response across land plants. *Sci. Rep.* 7 (1), 3806. doi:10.1038/s41598-017-03903-6
- Chakravarty, D., Guharoy, M., Robert, C. H., Chakrabarti, P., and Janin, J. (2013). Reassessing buried surface areas in protein–protein complexes. *Protein Sci. A Publ. Protein Soc.* 22 (10), 1453–1457. doi:10.1002/pro.2330
- Chen, J., Liu, D., Jiang, Y., Zhao, M., Shan, W., Kuang, J., et al. (2011). Molecular characterization of a strawberry FaASR gene in relation to fruit ripening. *PLoS ONE* 6 (9), e24649. doi:10.1371/journal.pone.0024649
- Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9 (12), 938–950. doi:10.1038/nrg2482
- Dai, J.-R., Liu, B., Feng, D.-R., Liu, H., He, Y., Qi, K., et al. (2011). MpAsr encodes an intrinsically unstructured protein and enhances osmotic tolerance in transgenic *Arabidopsis*. *Plant Cell Rep.* 30 (7), 1219–1230. doi:10.1007/s00299-011-1030-1
- Delwiche, C., and Cooper, E. (2015). The evolutionary origin of a terrestrial flora. *Curr. Biol.* 25 (19), R899–R910. doi:10.1016/j.cub.2015.08.029
- Dominguez, P. G., and Carrari, F. (2015). ASR1 transcription factor and its role in metabolism. *Plant Signal. & Behav.* 10 (4), e992751. doi:10.4161/15592324.2014.992751
- Foley, G., Mora, A., Ross, C. M., Bottoms, S., Sützl, L., Lamprecht, M. L., et al. (2022). Engineering indel and substitution variants of diverse and ancient enzymes using graphical representation of ancestral sequence predictions (GRASP). *PLOS Comput. Biol.* 18 (10), e1010633. doi:10.1371/journal.pcbi.1010633
- Frankel, N., Carrari, F., Hasson, E., and Iusem, N. D. (2006). Evolutionary history of the Asr gene family. *Gene* 378, 74–83. doi:10.1016/j.gene.2006.05.010
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., and Ho, S. Y. W. (2019). EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol. Evol.* 9 (7), 3891–3898. doi:10.1002/ece3.5015
- Goldgur, Y., Rom, S., Ghirlando, R., Shkolnik, D., Shadrin, N., Konrad, Z., et al. (2006). Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. *Plant Physiol.* 143 (2), 617–628. doi:10.1104/pp.106.092965
- González, R. M., and Iusem, N. D. (2014). Twenty years of research on Asr (ABA-stress-ripening) genes and proteins. *Planta* 239 (5), 941–949. doi:10.1007/s00425-014-2039-9
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59 (3), 307–321. doi:10.1093/sysbio/syq010
- Guo, H. B., Perminov, A., Bekele, S., Kedziora, G., Farajollahi, S., Varaljay, V., et al. (2022). AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci. Rep.* 12, 10696. doi:10.1038/s41598-022-14382-9
- Hamdi, K., Salladini, E., O'Brien, D., Brier, S., Chenal, A., Yacoubi, I., et al. (2017). Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins. *Sci. Rep.* 7 (1), 15544. doi:10.1038/s41598-017-15299-4
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2017). UFBoot2: improving the Ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35 (2), 518–522. doi:10.1093/molbev/msx281
- Hossain, K. A., Kogut, M., Slabonska, J., Sappati, S., Wiczcór, M., and Czuj, J. (2023). How acidic amino acid residues facilitate DNA target site selection. *Proc. Natl. Acad. Sci. U. S. A.* 120 (3), e2212501120. doi:10.1073/pnas.2212501120
- Hsia, C. C., and McGinnis, W. (2003). Evolution of transcription factor function. *Curr. Opin. Genet. & Dev.* 13 (2), 199–206. doi:10.1016/s0959-437x(03)00017-0
- Jakobson, L., Möttus, J., Suurväli, J., Sömera, M., Tarassova, J., Nigul, L., et al. (2024). Phylogenetic insight into ABCE gene subfamily in plants. *Front. Genet.* 15, 1408665. doi:10.3389/fgene.2024.1408665
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Kalifa, Y., Gilad, A., Konrad, Z., Zaccari, M., Scolnik, P. A., and Bar-Zvi, D. (2004). The water- and salt-stress-regulated Asr1 (abscisic acid stress ripening) gene encodes a zinc-dependent DNA-binding protein. *Biochem. J.* 381 (2), 373–378. doi:10.1042/bj20031800
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14 (6), 587–589. doi:10.1038/nmeth.4285
- Kapoor, B., Kumar, P., Verma, V., Irfan, M., Sharma, R., and Bhargava, B. (2022). How plants conquered land: evolution of terrestrial adaptation. *J. Evol. Biol.* 35 (5), 5–14. doi:10.1111/jeb.14062
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinforma.* 20 (4), 1160–1166. doi:10.1093/bib/bbx108
- Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556 (7702), 452–456. doi:10.1038/s41586-018-0043-0
- Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., and De Smet, R. (2016). Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* 28 (2), 326–344. doi:10.1105/tpc.15.00877
- Liu, C., Wu, Y., Liu, Y., Yang, L., Dong, R., Jiang, L., et al. (2021). Genome-wide analysis of tandem duplicated genes and their contribution to stress resistance in pigeonpea (*Cajanus cajan*). *Genomics* 113 (1), 728–735. doi:10.1016/j.ygeno.2020.10.003
- Londo, J. P., Chiang, Y.-C., Hung, K.-H., Chiang, T.-Y., and Schaal, B. A. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci.* 103 (25), 9578–9583. doi:10.1073/pnas.0603152103
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290 (5494), 1151–1155. doi:10.1126/science.290.5494.1151
- Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154 (1), 459–473. doi:10.1093/genetics/154.1.459
- Már, M., Nitsenko, K., and Heidarrson, P. O. (2023). Multifunctional intrinsically disordered regions in transcription factors. *Chem. - A Eur. J.* 29, e202203369. doi:10.1002/chem.202203369
- Middendorf, L., and Eicholt, L. A. (2024). Random, de novo, and conserved proteins: How structure and disorder predictors perform differently. *Proteins*, 92 (6), 757–767. doi:10.1002/prot.26652
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37 (5), 1530–1534. doi:10.1093/molbev/msaa015

- Mirdita, M., Schütze, K., Moriaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi:10.1038/s41592-022-01488-1
- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41 (12), e121. doi:10.1093/nar/gkt263
- Narberhaus, F. (2002). Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol. Mol. Biol. Rev.* 66 (1), 64–93. doi:10.1128/mmbr.66.1.64-93.2002
- O’Rawe, J. A., Ferson, S., and Lyon, G. J. (2015). Accounting for uncertainty in DNA sequencing data. *Trends Genet.* 31 (2), 61–66. doi:10.1016/j.tig.2014.12.002
- Rebeaud, M. E., Mallik, S., Goloubinoff, P., and Tawfik, D. S. (2021). On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life. *Proc. Natl. Acad. Sci.* 118 (21), e2020885118. doi:10.1073/pnas.2020885118
- Ricardi, M. M., González, R., Zhong, S., Dominguez, P. G., Duffy, T., Turjanski, P., et al. (2014). Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought stress-responsive transcription factor. *BMC Plant Biol.* 14 (1), 29. doi:10.1186/1471-2229-14-29
- Riley, A. C., Ashlock, D. A., and Graether, S. P. (2023). The difficulty of aligning intrinsically disordered protein sequences as assessed by conservation and phylogeny. *PLoS one* 18 (7), e0288388. doi:10.1371/journal.pone.0288388
- Rom, S., Gilad, A., Kalifa, Y., Konrad, Z., Karpasas, M. M., Goldgur, Y., et al. (2006). Mapping the DNA- and zinc-binding domains of ASR1 (abscisic acid stress ripening), an abiotic-stress regulated plant specific protein. *Biochimie* 88 (6), 621–628. doi:10.1016/j.biochi.2005.11.008
- Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2000). Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* 42 (1), 38–48. doi:10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3
- Sachdeva, S., Bharadwaj, C., Singh, R. K., Jain, P. K., Patil, B. S., Roorkiwal, M., et al. (2020). Characterization of ASR gene and its role in drought tolerance in chickpea (*Cicer arietinum* L.). *PLOS ONE* 15 (7), e0234550. doi:10.1371/journal.pone.0234550
- Shiu, S.-H., Shih, M.-C., and Li, W.-H. (2005). Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol.* 139 (1), 18–26. doi:10.1104/pp.105.065110
- Smock, R. G., Yadid, I., Dym, O., Clarke, J., and Tawfik, D. S. (2016). *De novo* evolutionary emergence of a symmetrical protein is shaped by folding constraints. *Cell* 164 (3), 476–486. doi:10.1016/j.cell.2015.12.024
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi:10.1093/nar/gkl315
- Tang, Y., Wei, Y., He, W., Wang, Y., Zhong, J., and Qin, C. (2013). GATA transcription factors in vertebrates: evolutionary, structural and functional interplay. *Mol. Genet. Genomics* 289 (2), 203–214. doi:10.1007/s00438-013-0802-4
- Uversky, V. N. (2013). The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord. Proteins* 1 (1), e24684. doi:10.4161/idp.24684
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50 (D1), D439–D444. doi:10.1093/nar/gkab1061
- Volkin, D. B., Tsai, P.-K., Dabora, J. M., Gress, J., Burke, C., Linhardt, R. J., et al. (1993). Physical stabilization of acidic fibroblast growth factor by polyanions. *Archives Biochem. Biophysics* 300 (1), 30–41. doi:10.1006/abbi.1993.1005
- Voordeckers, K., Pougach, K., and Verstrepen, K. J. (2015). How do regulatory networks evolve and expand throughout evolution? *Curr. Opin. Biotechnol.* 34, 180–188. doi:10.1016/j.copbio.2015.02.001
- Wang, Z., Zhao, K., Pan, Y., Wang, J., Song, X., Ge, W., et al. (2018). Genomic, expressional, protein-protein interactional analysis of Trihelix transcription factor genes in *Setaria italica* and inference of their evolutionary trajectory. *BMC genomics* 19 (1), 665. doi:10.1186/s12864-018-5051-9
- Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8 (3), 206–216. doi:10.1038/nrg2063
- Yang, C.-Y., Chen, Y.-C., Jauh, G. Y., and Wang, C.-S. (2005). A lily ASR protein involves abscisic acid signaling and confers drought and salt resistance in *Arabidopsis*. *Plant Physiol.* 139 (2), 836–846. doi:10.1104/pp.105.065458
- Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A.-M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155 (1), 431–449. doi:10.1093/genetics/155.1.431
- Yoon, J. S., Kim, J. Y., Kim, D. Y., and Seo, Y. W. (2021). A novel wheat ASR gene, TaASR2D, enhances drought tolerance in *Brachypodium distachyon*. *Plant Physiology Biochem.* 159, 400–414. doi:10.1016/j.plaphy.2020.11.014