Fruit Research

ARTICLE

Open Access

https://doi.org/10.48130/frures-0024-0001

Fruit Research **2024**, 4: e008

Heterodimeric interaction of the C/S1 basic leucine zipper transcription factors in black raspberry: a genome-wide identification and comparative analysis

Ximeng Lin[1,2], Mei Huang[1], Jinwei He[1], Ailing Min[1], Ying Zhou[1], Wendie Ma[1], Xunju Liu[3], Xiaorong Wang[1,4], Haoru Tang[1] and Qing Chen[1,4*]

[1] *College of Horticulture, Sichuan Agricultural University, Chengdu 611130, China*
[2] *College of Horticulture, Nanjing Agricultural University, Nanjing 210014, China*
[3] *University Bordeaux, INRAe, Biologie du Fruit et Pathologie, UMR 1332, av. Edouard Bourlaux, Villenave d'Ornon 33140, France*
[4] *Key Laboratory of Agricultural Bioinformatics, Ministry of Education, Sichuan Agricultural University, Chengdu 611130, China*
* Corresponding author, E-mail: supnovel@sicau.edu.cn

Abstract

The basic leucine zippers (bZIPs) transcription factors play crucial roles in various vital biological processes and have been extensively studied across multiple species. However, their presence in the *Rubus* species has received limited attention. In this study, we identified 49 bZIP transcription factor genes in the genomes of black raspberry (*R. occidentalis*) and red raspberry (*R. idaeus*). Phylogenetic analysis successfully classified the bZIPs into 13 subfamilies. Comparative genomic analysis revealed a conservation of *bZIP* gene numbers among ten Rosaceae species, with gene duplications partially contributing to the expansion of this family in *Rubus*. Microsynteny analysis further supported the evolutionary conservation of this gene family among Rosaceae genomes. Tissue-specific expression of black raspberry *bZIPs* was observed, and gene ontology enrichment analyses revealed their involvement in various biological processes. Those members that respond to nitrogen were identified. Dimerization properties of bZIP proteins were predicted based on the amino acid composition of their coiled-coil domains. Protein interaction network analysis and yeast two-hybrid assays unveiled the formation of heterodimers between the S1/C groups in *R. occidentalis*. This comprehensive characterization of the bZIP transcription factor family provides valuable insights into the evolution and functions of *bZIP* genes in *Rubus*.

Citation: Lin X, Huang M, He J, Min A, Zhou Y, et al. 2024. Heterodimeric interaction of the C/S1 basic leucine zipper transcription factors in black raspberry: a genome-wide identification and comparative analysis. *Fruit Research* 4: e008 https://doi.org/10.48130/frures-0024-0001

Introduction

Transcription factors (TFs) play vital roles in regulating various biological processes in eukaryotic organisms, including plants. One example of TFs that recognize specific DNA elements with ACGT-based motifs is the basic leucine zippers (bZIPs)[1]. The mechanism by which bZIP proteins recognize specific sequences relies on a conserved basic region (BR) located at the N-terminus. This region is approximately 18 amino acids long and has the sequence N-X7-R/K-X9[1,2]. Additionally, the leucine zipper (LZ) domain downstream of the BR is diverse, consisting of heptad repeats of leucine or other hydrophobic amino acids, such as phenylalanine, valine, isoleucine, or methioninende, which are necessary for protein dimerization[3,4].

bZIP TFs have been shown to play a crucial role in diverse biological processes, including seed dormancy, vegetative growth, flowering, fruiting, anthocyanin biosynthesis, lignin regulation, and responses to environmental changes such as nutrient availability, pathogen attacks, and abiotic stresses[5–11]. The number of identified *bZIPs* in plants varies among different species, primarily due to whole-genome duplication and chromosome rearrangement history. In *Arabidopsis*, a total of 78 *bZIP* members have been categorized into 13 related groups[7]. In the Rosaceae family, 114, 50, 47, and 62 members have been identified in apple[12], strawberry[13], peach[14], and pear[15], respectively. The accepted 13-subfamily classification (A to K plus S), which refers to the assignment of *Arabidopsis bZIPs*, is supported by most phylogenetic analysis combined with conserved motif information. Each subfamily can be further divided into smaller clades. Interestingly, a detailed multi-species phylogenetic study of the C and S1 bZIPs, using current plant sequencing data, has revealed the existence of proto-C/S ancestor TFs in early green algae species[16]. However, a comprehensive understanding of the family status of bZIPs in the *Rubus* genus is still lacking.

Among the 13 subgroups of plant bZIPs, the S1 and C groups have gained significant attention in recent years. These two groups play crucial roles in the regulation of energy metabolism and signaling, as well as stress adaptation[17–19]. In particular, the C/S1 bZIP members function extensively in either homo- or heterodimeric forms downstream of the central metabolic kinase SnRK1 signaling[19]. The phosphorylation state of the C group member bZIP63 can influence the preferences of its S1 dimeric partner in responding to varying energy levels[17]. Overexpression of the individual S1 bZIP member *FabZIP1.1* in strawberry fruits or *FabZIP11* in tomato can induce changes in the accumulation of primary metabolites[20,21]. Furthermore,

through CRISPR-A3A-PBE-mediated base editing in the untranslated region of *FvebZIPs1.1*, sugar synthesis in strawberry fruits can be precisely tuned[22], showcasing the potential application of these genes in breeding programs. However, although functional redundancy among the S1 bZIPs has been suggested, questions still remain regarding which homo- or heterodimers are functionally relevant and their specific involvement in signaling pathways.

The Rosaceae family comprises economically significant crop species, such as apple (*Malus* × *domestica* Borkh.), pear (*Pyrus spp.*), peach (*Prunus persica* (L.) Batsch), strawberry (*Fragaria* × *ananassa* Duch.), and raspberry (*Rubus idaeus* L. and *R. occidentalis* L.). While bZIP members have been identified in a few species of this family[12–15], comprehensive research on bZIPs in *Rubus* species, particularly from a multi-species comparison perspective, is lacking. Additionally, our understanding of C/S1 bZIP networks is mainly limited to model plant species[23]. In this study, we systematically identified the *bZIP* family members in the genomes of black raspberry (*R. occidentalis*) and red raspberry (*R. idaeus*) for their available high-quality genomic data. We analyzed gene family expansions and contractions associated with genome duplication, with a particular focus on the C/S1 group members. Conserved motifs and microsynteny relationships were conducted among ten Rosaceae species. Furthermore, we analyzed tissue and stage specificity of the C/S1 group members in *R. occidentalis* using RNA-seq data. We also experimentally curated the C/S1 *bZIP* gene members in the black raspberry genome and validated heterodimeric networks between the two groups using yeast two-hybrid analysis. Our findings not only provide insights into the evolutionary information of bZIP TFs in Rosaceae species but also offer evidence of bZIP protein interactions, guiding future functional genomic studies.

Materials and methods

Identification of the *bZIP* genes

To systematically identify *bZIP* genes in *Rubus* species, including black raspberry (*R. occidentalis*) and red raspberry (*R. idaeus*), we obtained genomic data, such as genome sequences, gene feature files, transcripts, and protein sequences, from the Genome Database for Rosaceae (www.rosaceae.org/). We also included eight other related species for comparative genomic analysis, namely woodland strawberry (*F. vesca*), apple (*Malus* × *domestica*), peach (*P. persica*), pear (*P. pyrifolia*), plum (*P. salicina*), sweet cherry (*P. avium*), apricot (*P. armeniaca*), and rose (*Rose chinensis*) (Supplemental Table S1). To guide our exploration, we downloaded the *Arabidopsis* genome (Araport 11) from the Phytozome database (https://phytozome-next.jgi.doe.gov/). We employed two strategies to identify the bZIP protein coding genes. Firstly, we used the bZIP proteins of *Arabidopsis* from the PlantTFDBv5.0 to construct a plant-specific Hidden Markov Model (HMM) using HMMER (3.1b1). Alternatively, we directly downloaded the seed HMM files for bZIP1 (PF00170) and bZIP2 (PF07716) from the Pfam database (http://pfam.xfam.org/). Subsequently, we individually searched the downloaded protein profiles against these three models using hmmsearch with an e-value of 1e-5. The resulting gene lists were combined, and duplicates were eliminated. Finally, we used the NCBI Batch CD-Search tool (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to search the remaining candidates against the CDD (CDD-58235 PSSMS) database to ensure they possessed bZIP domains.

Protein properties of the bZIPs

We obtained the protein sequences of all bZIPs from the corresponding genome. Then, using the ProtParam tool (https://web.expasy.org/protparam/), we calculated various protein properties, such as amino acid length, molecular weight (MW), isoelectric point (pI), instability index, and grand average hydropathy.

Phylogenetic tree construction

The MAFFT software (v7.487) was used to perform multiple sequence alignment of all identified bZIP proteins. The resulting alignment matrix was then used as input for tree construction using IQ-TREE2 (v 2.1.4)[24] on a server. The following parameters were used: -s bzips_input -bb 10,000 -nt 60. Finally, the maximum likelihood tree was visualized using the R package ggtree (v3.8.2)[25].

Gene structure, conserved motif and microsynteny analysis of the *bZIP* members

The intron and exon information of the *bZIP*s was extracted from the GFF files. Conserved protein motifs were detected using the Multiple Expectation Maximization for Motif Elicitation (MEME) suite (v5.4.1). Up to 15 motifs were allowed in each protein. The gene structures were viewed using TBtools (v2.008)[26]. Based on the results of Ehlert et al., the members of the S group bZIPs were further divided into three subclasses (S1, S2, and S3)[27].

The LZ domain coiled coil structure consists of a varying number of heptad repeats. Since certain residues, especially those at positions *a*, *d*, *e*, and *g*, could affect dimerization or oligomerization stability and specificity[3], we focused on the categories of amino acids at these positions to predict the dimerization properties of the identified bZIPs, referring to Manzoor et al.[15].

To explore inter- and intra-chromosomal duplication events that may have given rise to the entire *bZIP* family, sequence synteny analysis was conducted. Self-BLASTP was performed using all proteins from the ten selected rosaceae species, with the default parameters, except that the e-value was set to 1e-10. The top five hits were selected as inputs for MCScanX[28]. Default parameters were used to search for traces of recent gene duplications. The duplication types of the bZIP paralogs found by MCScanX were identified as tandem, proximal dispersed, and segmental/whole-genome duplications (WGD) using the 'duplicate_gene_classifier' scripts[28]. The syntenic gene pairs of the C/S1 group bZIPs were plotted using the TBtools software.

Gene family expansion/contraction assay, gene functional annotation, and gene ontology enrichment analysis of the bZIPs

To obtain an overview of the origination of the *bZIP*s among the 11 selected species, we conducted whole-genome gene family expansion/contraction assays. We used OrthoFinder (v2.5.4)[29] to identify orthologs of the selected species based on pairwise BLASTP results. The gene counts in each orthogroup were then inputted into the CAFE (v5.0) software[30] to detect significant changes in gene family members, with three different birth/death rates (gamma) assigned to each three subclades[31]. The changes in the *bZIP* members during ancestral history were visualized using the ggtree package (v3.10)[25].

We also inferred the biofunction of all bZIPs from model plants, specifically *Arabidopsis*, using the obtained orthologous relationships. In this study, we focused solely on the black raspberry bZIPs. The Gene Ontology (GO) information of the matched *Arabidopsis* bZIPs was transferred to the genes from the same orthogroup. To perform gene ontology enrichment analysis of the bZIP members, we followed these steps: First, we functionally annotated all the genome proteins using the TOA pipeline[32]. Briefly, proteins were sequentially searched against the following databases using BLASTP with an e-value of 1e-5: dicots PLAZA 5.0, monocots PLAZA 5.0, NCBI RefSeq Plant (release 20211015), and RefSeq non-redundant proteins (release 20210620). Next, GO enrichment analysis was conducted using the hypergeometric test in the clusterProfiler (v4.8.3)[33], with a Benjamini-Hochberg adjusted *p*-value cutoff of 0.05.

In silco expression analysis of *bZIP*s in black raspberry

The public RNA-seq data derived from different organs (cane, root, young and mature leaf, flower) of black raspberry and fruits of various developmental stages (green, red, and black ripe fruit) were downloaded from the sequence read archive database[34] (Supplemental Table S2). Reads were quality-controlled using the software trim-galore (v 1.18) with a minimal base quality 20. Reads were then directly quantified using the alignment-free kallisto pipeline (v0.46.2)[35]. The expression value was presented as transcripts per million kilobases (TPM). K-means clustering were carried out in the R environment.

To further infer the biofunction of the identified bZIPs, the available RNAseq data from the *Rubus* hybrid cultivar which was treated using different types of nitrogen fertilizers (NH_4^+-N, NO_3^--N, Urea) were analyzed[36]. The raw reads were cleaned and quantified the same as described above. The edgeR package (v4.0.2) was employed to conducted differential gene expression analysis using the Wald test comparing to the zero-N group (CK). Differentially expressed genes were chosen by using the Benjamini-Hochberg adjusted *p* value less than 0.05, and the absolute log2 fold change value greater than one as a threshold.

Construction of PPI network for black raspberry C/S1 bZIP proteins and validation using yeast-two-hybrid assay

The STRING database (v11.5) was employed to predict the protein-protein interactions *via* homology sequence search. The most related species '*F. vesca*' was selected as an organism to search against. To enhance the credibility of the interaction pairs, only the experimentally validated physical subnetwork was mapped. The interaction pair information was used to construct the PPI network in the Cytoscape software (v3.9.1).

To confirm the interaction relationships of the C/S1 group bZIPs of *R. occidentalis*, the high throughput CrY2H-seq yeast two-hybrid system[37] was employed. Firstly, the three C-group and four S1-group *bZIP*s were cloned and sequenced. Total RNA was extracted from fresh young leaves of black raspberry 'Arapoho' using a modified CTAB-based method. The first-strand cDNA was synthesized using the RevertAid H-minus First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA) according to the manufacturer's instructions. The target gene was amplified in a 20 μL PCR mixture containing 1μL cDNA template, 1 μL of forward and reverse gene specific primers (Supplemental Table S3), and 10 μL of PrimeSTAR HS Premix (TaKaRa, Dalian, China). All amplifications were carried out on a PTC100 thermal cycler with 30 cycles of standard three-step PCR reactions annealing at 60 °C. The amplicons were purified using a PCR Clean Up Kit (OMEGA, Madison, WI, USA). To minimize the potential auto-activation of the bZIP proteins in yeast, the S1-group members were cloned into the CrY2H-seq pDBlox vector (ABRC CD3-2421), while the C-group members were cloned into the pADlox vector (ABRC CD3-2420)[37]. All constructs were validated through Sanger sequencing. The confirmed pAD_bZIP (C-group) or pDB_bZIP (S1-group) plasmids were transformed into yeast strains Y8800 or CRY8930, respectively, using a standard lithium acetate and polyethylene glycol method. The two strains with AD and DB were mated. The pDB-bZIP (C-group) yeast strains were mated with Y8800 yeast strains transformed with the pADlox empty plasmid to test autoactivation. Diploid cells were selected on Sc-Leu-Trp media. A series of concentrations of 3-amino-1,2,3-triazole (3-AT) (5 mM, 10 mM, 20 mM, 40 mM, and 100 mM) were used to repress the background expression of the *HIS3* gene. Subsequently, plates of Sc-Leu-Trp-His + 80 mM 3-AT were used to select interacting pairs. Positive clones were picked using a toothpick and tested on selective plates Sc-Leu-Trp + 100 mM 3-AT + 1 mg/L cycloheximide (CHX) to deplete *de novo* self-activations. Only those that exhibited positive activation of the *HIS3* reporter gene and negative growth in the presence of CHX were considered positive interaction pairs.

Results

Genome-wide identification of bZIPs and their protein properties

To maximize the possibility of identifying *bZIP* members in the selected *Rubus* species, two strategies were adopted. The protein signature file from the Pfam database (bZIP1 and bZIP2) and the plant-specific bZIP HMM profile developed in this study were used. As a result, 49 candidate bZIP coding genes were identified in both the black raspberry and red raspberry genomes, respectively. The presence of the BR and LZ domains were confirmed in the corresponding proteins at the CDD database. The properties of all bZIPs were analyzed (Supplemental Table S4). The protein sequences of the bZIPs ranged from 137 (Ro05G03390) to 706 (Ro04G02768) amino acids in length. The molecular weights of the proteins ranged from 15.9 to 76.3 kDa. Proteins in Group B had relatively higher molecular weights compared to other members (Fig. 1a). Group S and Group H members consisted of relatively small proteins. The isoelectric points (pI) of the proteins ranged from 4.6 to 11.43 (Supplemental Table S4). It is worth noting that the majority of the proteins (94.8%) had a pI larger than 8 or less than 7 (Fig. 1b).

Phylogenetic and gene family expansion/contraction analysis

Phylogenetic analysis was performed on the bZIP protein sequences to determine their classification and relationships. A total of 78 bZIPs from *Arabidopsis* were included as a reference. The maximum likelihood tree (Fig. 2a) categorized all bZIPs into 13 phylogenetic clades in both raspberry species. The clade *S* had the largest number of members (ten) in the black raspberry genome, followed by clades *A* (seven) and *D* (seven). In contrast, the red raspberry had the same number of bZIPs in groups *S*, *A*, and *I* (eight). Only one member was identified for

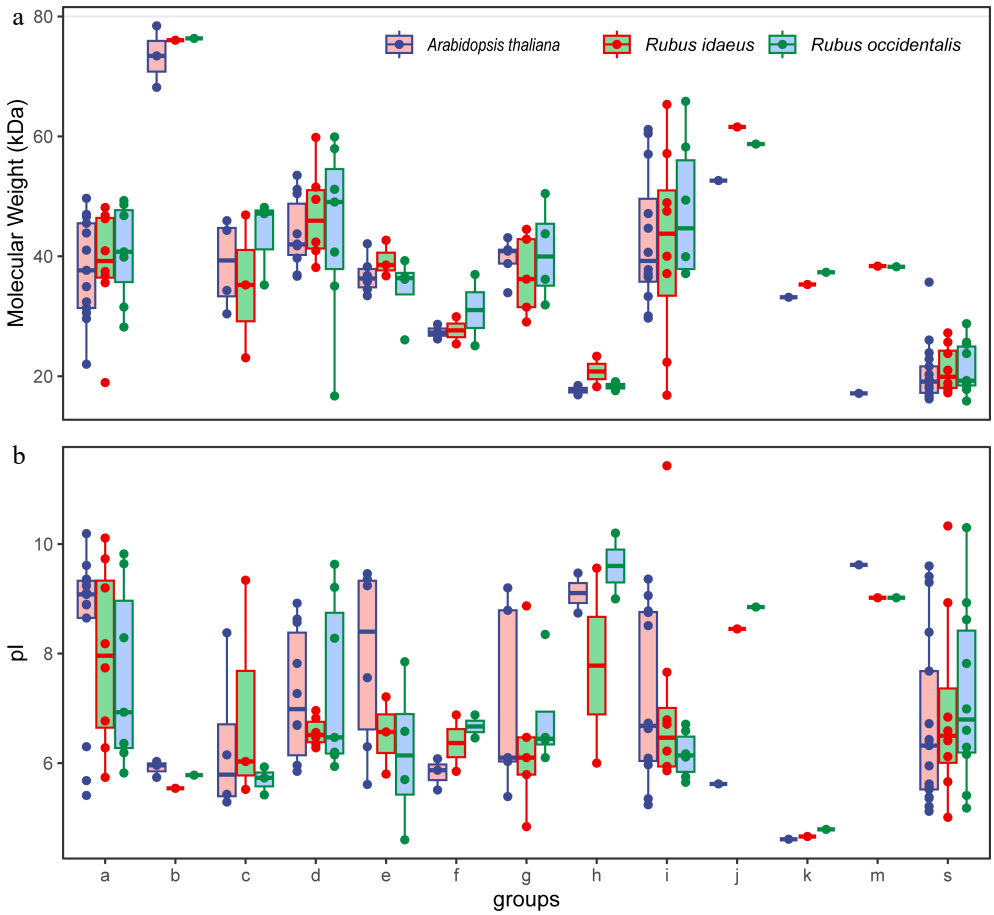

Fig. 1 Molecular weight and protein isoelectric point of the identified bZIPs in each group of *Rubus* species. *Arabidopsis* proteins are included as a reference.

groups *B*, *J*, *K*, and *M* in both species. Interspecies clustering was observed in all clades, indicating that bZIPs originated before the divergence of these two species (Fig. 2a).

To further understand the changes in the *bZIP* gene family members throughout evolutionary history, comparative genomic analysis was conducted among the 10 Rosaceae genomes. The same strategy was applied to identify *bZIP* members in the remaining species (Supplemental Table S4). The number of *bZIP*s in the selected Rosaceae species was significantly smaller than that in *A. thaliana*, considering the ploidy level and genome size, except for the two pome fruits (Fig. 2b). In all species, a relatively large number of members were found in groups *A*, *D*, *I*, and *S*. The identified *bZIP* members were clustered into 57 orthogroups (Supplemental Table S5). No group *M* members were observed in *P. persica* and *R. chinensis*, and no credible group *J* members existed in *P. salicina* and *P. persica*.

Despite the significant number of gene family expansions or contractions observed in the *Rubus* species, no significant changes in birth/death rates were observed for the *bZIP* family members among the Rosaceae species, indicating the conservation of *bZIP*s throughout evolutionary history (Supplemental Fig. S1).

Duplication of *bZIP*s in *Rubus* and micro-synteny of the *bZIP*s among ten Rosaceae species

Gene duplication events have contributed to the amplification and divergence of gene family members during the species' evolution. We analyzed the duplication events that gave rise to the *bZIP* genes in *R. occidentalis* and *R. idaeus* using the MCScanX results. Among the 155 blocks of genome-wide collinearity, 45 *RobZIP*s (91.8%) reside in the duplicated region. Twenty-seven pairs of *RobZIP*s were classified as dispersed duplications, and 18 *bZIP* pairs were categorized as segmental duplications or whole genome duplication (WGD). No tandem repeats were observed. In the red raspberry, 47 *bZIP*s (95.9%) were derived from gene duplication events, with 27 resulting from dispersed gene duplications, 19 from segmental duplications or WGD, and one from a proximal duplication event. These results demonstrate that dispersed and segmental duplication events were the major driving forces behind the evolution of *bZIP* genes in *Rubus*.

The chromosomal level of gene synteny was analyzed among the ten Rosaceae species. A larger number of collinear blocks were detected among rose, raspberry, and strawberry compared to other species (Fig. 3), indicating their closer evolutionary relationships. Consistently, homologous *bZIP* gene pairs were found to reside at almost the same relative locations on chromosomes 2, 4, 5, 6, and 7 in these three species, despite variations in genome size. Segments of chromosome 2 in Rosa, which contained three *RobZIP*s, were found to be syntenic with segments of the same chromosome 6 in strawberry. This observation supports the hypothesis that the *Rosa* genome underwent one fusion and two fission events from ancestral chromosomes[38], contrasting with the single fusion event in the strawberry genome. These results indicate that the *Rubus* genome

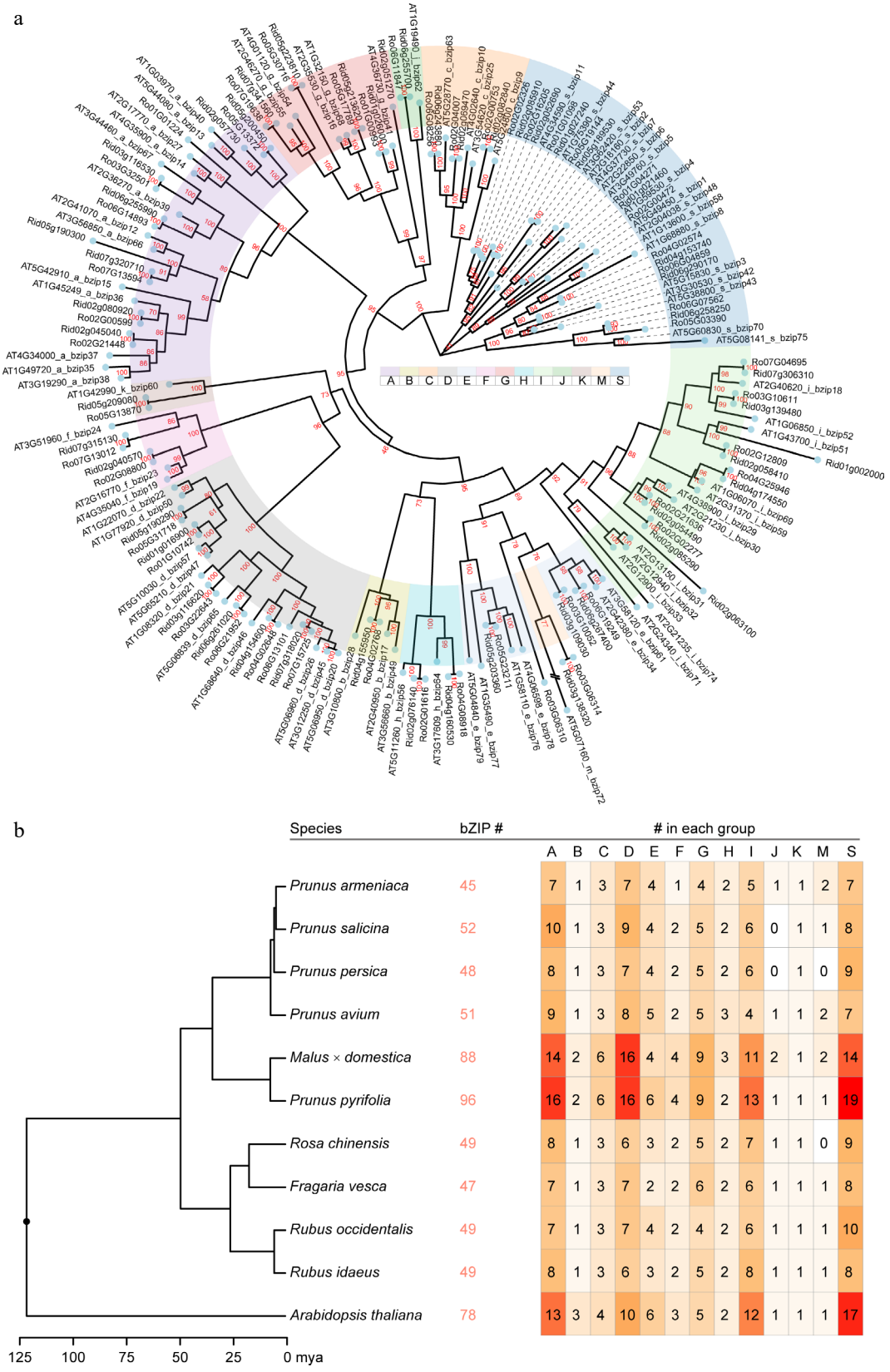

Fig. 2 The phylogenetic tree constructed using the *Rubus* bZIPs, the number of identified *bZIP* members in the selected Rosaceae species, and the assigned subgroup members. (a) The tree was constructed using maximum likelihood method implemented in IQ-TREE with 10,000 fast boostraps to validate the tree topology. Bootstrap values were directed labeled on the branches. (b) The ultrametric tree showing the time of divergence within the selected species scaling in million years ago (mya). The identified total bZIP gene number and the number in each subgroup (a to k, plus m and s) were labelled following the species name (tree tip).

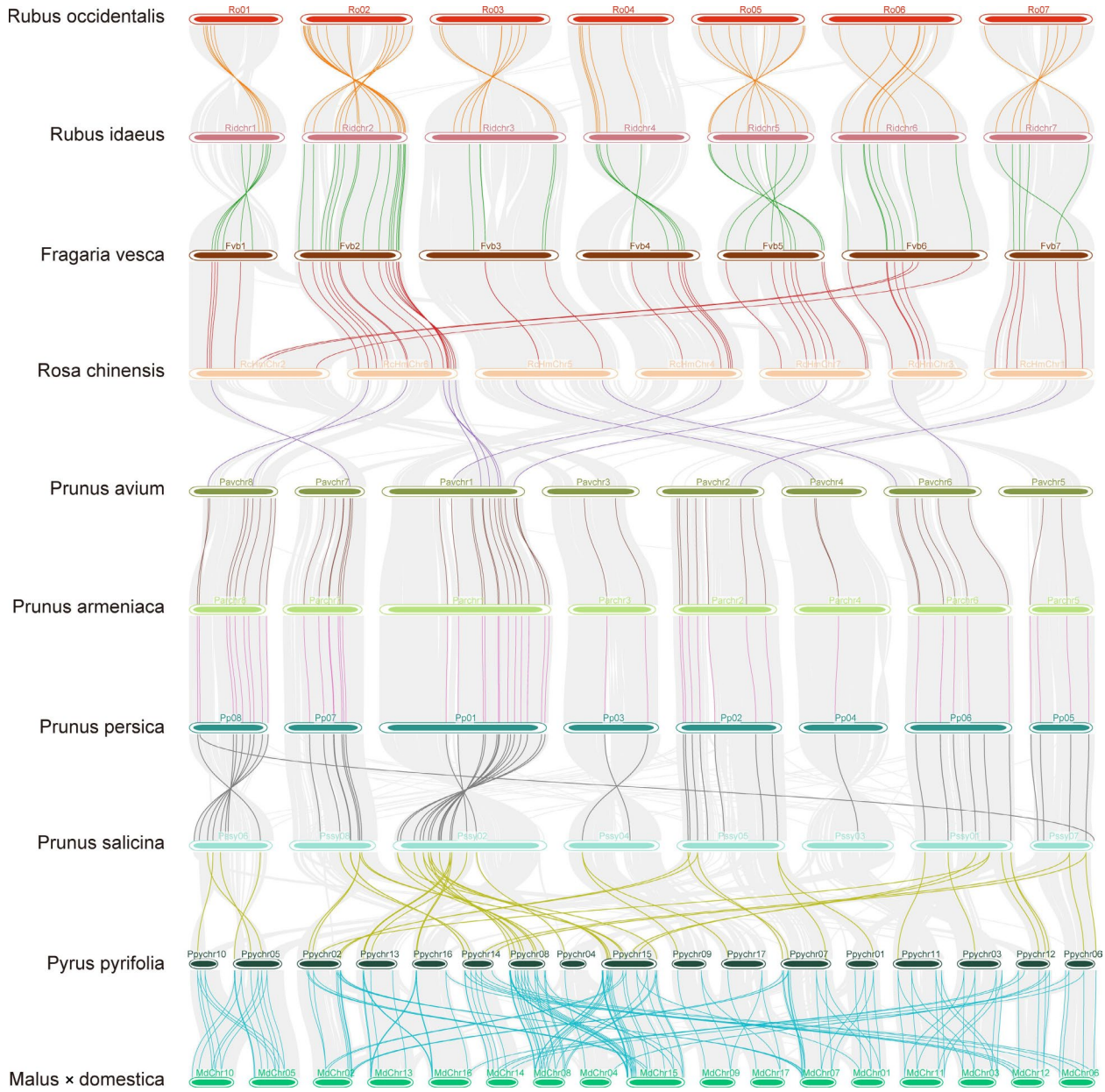

Fig. 3 Micro-synteny of the *bZIP*s among ten Rosaceae species. Genome-wide syntenic blocks detected by using MCScanX between the adjacient species were linked by gray lines defined by their loci on the chromosoms (horizontal bars) in the background. When there was a (several) syntenic pair(s) of bZIPs located in the syntenic region, they were linked by line(s) of a different color other than gray.

has undergone similar chromosomal rearrangements, resembling those in the evolutionary history of wild strawberry. Fewer collinear bZIP pairs were observed between *Rosa* and *Prunus*, while a large number of *bZIP* collinearities were discovered among the *Prunus* species. Thirty *P. salicina bZIP*s have two orthologues in the *P. pyrifolia* genome. Most apple *bZIP*s exhibited two duplicated genomic segments in the pear genome, and *vice versa*, suggesting that the *bZIP* gene family has expanded in the apple and pear genomes, accompanying their unique whole-genome duplication ancestry.

Gene functional annotation and expression analysis of *RobZIP* genes in different tissue types

To understand the biological functions of these *bZIP* genes, gene ontology annotation was performed, and the top two levels of GO terms were obtained. Enrichment analysis revealed that in the molecular function subgroup, DNA-binding, transcription regulatory activity, and protein dimerization activity were significantly enriched (Fig. 4a). These findings align with the main function of these proteins as transcription factors. In terms of biological processes, the RobZIPs were found to participate in various cellular processes, including the regulation of nitrogen compound metabolism, response to carbohydrates, response to stimuli, response to stress, and regulation of cellular macromolecule biosynthesis.

The expression levels of all 49 *RobZIP* genes in different tissue types and at various fruit stages were investigated using RNA-seq data (Fig. 4b). Based on the expression patterns, two major clusters of genes were identified. Genes in the first cluster generally exhibited low expression levels (< 4 TPM) in the investigated tissues, indicating tissue specificity of *bZIP* genes. For

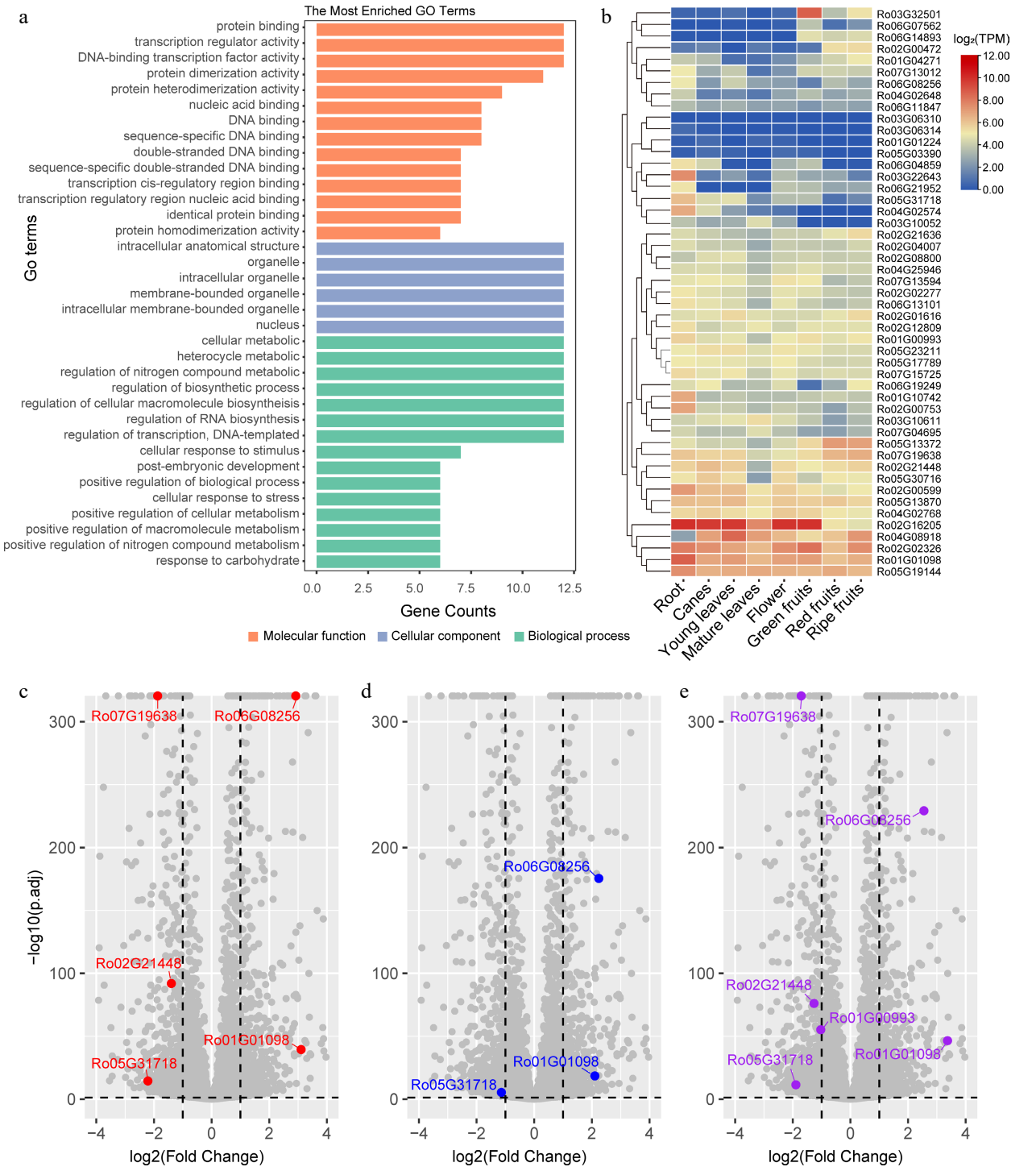

Fig. 4 GO enrichment analysis and expression of the *Rubus occidentalis bZIP*s in different tissues and in leaves upon treated with different types of nitrogen fertilizers. (a) The most significantly enriched GO terms assigned to the *RobZIPs* are presented as bar plots. The GO terms categaried in molecular function, cellular component, and biological process are labeled using orange, light blue, and light green color, respectively. (b) The expression level (log2TPM) of *bZIP* genes in different tissue types are presented as a heatmap. K-means clustering is performed on the rows. Volcano plots showing the differentially expressed genes in the leaves of Rubus spp. when the plants were treated with (c) $(NH_4)_2SO_4$, (d) $Ca(NO_3)_2$, or (e) urea fertilizers. Only differentially expressed *RobZIP*s are colored, with gene names labeled.

instance, the gene *Ro03G32501* showed specific expression in fruits, particularly at the green fruit stage, suggesting its potential role in fruit development regulation. Conversely, *Ro03G22643*, *Ro05G31718*, and *Ro04G02574* were actively transcribed in root tissues. Genes in the second cluster showed high expression levels in the studied tissues. The mRNAs of *Ro02G16205*, *Ro02G02326*, *Ro01G01098*, and *Ro05G19144*, all encoding *S* group bZIP members, exhibited the highest abundance (Fig. 4b). The gene for the H group member *Ro04G08918* also showed high expression, with a similar expression pattern

to that of the *S* group members in most tissues, except for the roots. In summary, these results indicate that *bZIP* members are extensively involved in the regulation of various biological processes in different tissues of black raspberry.

Analysis of the RNAseq data revealed that, out of the 49 identified *RobZIP*s, five exhibited differential expression following the application of NH_4^+-N fertilizer (Fig. 4c), while three showed differential expression in response to NO_3^--N fertilizers (Fig. 4d). Additionally, six *RobZIP*s demonstrated differential transcription, with at least a twofold increase or decrease (Fig. 4e). These findings provide evidence that these RobZIPs are involved in the regulation of nitrogen metabolism.

Prediction of the dimerization properties of *R. occidentalis* bZIPs

Among the 49 identified RobZIPs, the different amino acids at the heptad positions *g*, *a*, *b*, *c*, *d*, *e*, and *f* were summarized (Supplemental Table S6). Based on the criteria that α-helices were terminated by the natural C-terminus or the presence of a proline or two glycines, four bZIPs (Ro03G06310, Ro03G06314, Ro03G10052, Ro06G19249) were predicted to possess only two heptads, while Ro02G02277 and Ro02G21636 possess 11 heptads.

The amino acid composition at the *a*, *d*, *g*, and *e* positions in the heptads was investigated for their involvement in dimerization stabilization and specificity. Aliphatic amino acids (I, V, L, M) comprised approximately 44% of the *a* position, while asparagine occupied 20% of the amino acids (Fig. 5a). This ratio was higher than that observed in *Arabidopsis* (16%) but relatively lower than that in maize (22%)[23,39]. The aliphatic amino acids were mainly positioned at the second and fifth heptads in the LZ domain, accounting for an average of 56.0% and 45.8%, respectively (Fig. 5b). Asparagine was also common in the fourth and eighth heptads, accounting for approximately 10% of the amino acids. However, it was rarely observed at the *a* position in other heptads. Previous research has shown that the interhelical interaction between asparagine (N) at the *a* position and the opposite helical N at the *a'* position of the partner bZIP is the most stable pair compared to other combinations[3]. This suggests that many homodimers could be formed in black raspberry plants. Positive charged amino acids (R, K) or negatively charged amino acids (D, E) were also present, accounting for 8% of all amino acids. These amino acids could drive hetero- or homodimer formation, contributing to dimer stability.

Leucine or other hydrophobic amino acids at the *d* positions were found to be important for dimerization stability. In black raspberry, leucine represented 73% of the amino acids at the *d* position. This frequency was higher than that observed in bZIPs from rice (71%)[40] and much higher than that in *Arabidopsis* (56%)[23]. In addition to the aliphatic amino acids, these hydrophobic amino acids constituted approximately 86% of all amino acids (Fig. 5a), providing supporting evidence that these bZIPs might preferentially function as stable dimers.

Charged amino acids at the *g* and *e'* positions are involved in determining dimer specificity, as attractive or repulsive electrostatic interactions, together with the amino acids at the *a* (*a'*) position, could promote hetero- or homo-dimer formation, respectively. Charged amino acids accounted for more than 49% of the *g* and *e* positions in black raspberry. Specifically, in the *g* position, positive charged residues were more abundant than negatively charged ones (30% vs 19%). In contrast, these charged amino acids were almost equally distributed in the *e* position (Fig. 5a). The frequency of complete charged *g*→*e'* pairs (with both *g* and the corresponding *e'* being charged) was further analyzed across the first nine heptads (Supplemental Table S6). Four groups of interactions were defined as proposed by Vinson et al.[3]. Attractive pairs were predominant in the first, second, fifth, and sixth heptads (Fig. 5c). A small proportion of repulsive pairs was observed in the first and fifth repeats. Moreover, the second, seventh, and ninth heptads exclusively contained attractive interaction pairs when charged amino acids were present (Fig. 5c & Supplemental Table S6). In the analysis, incomplete pairs at these two positions were not included, but they were labeled in the Supplemental Table S6 as they might still participate in heterodimerization.

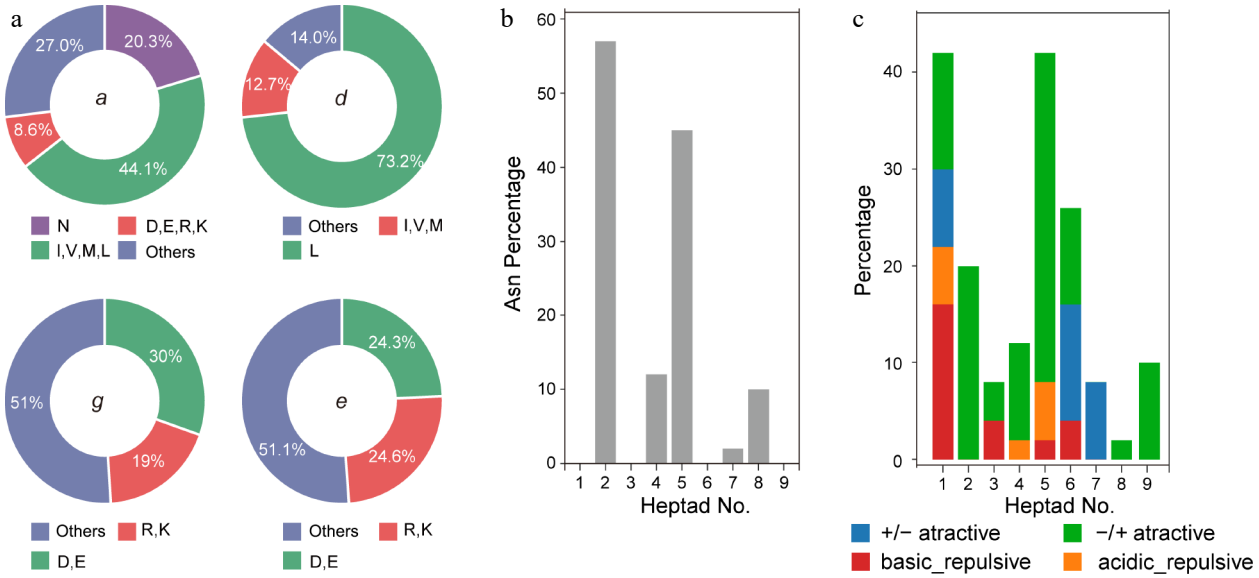

Fig. 5 (a) Pie chart presenting the frequency of amino acids in all the *g*, *e*, *a* and *d* positions in bZIPs, (b) histogram of the frequency of Asn residual at the *a* position, (c) attractive or repulsive g and e' pairs per heptad for *R. occidentalis*.

The identified bZIPs were classified into 13 subfamilies based on the interhelical electrostatic interactions between the *g*→*e'* residuals and the residual preference in the *a* position described above (Supplemental Table S6). Majority of the subfamilies exhibited complex dimer preferences because both attractive and repulsive interactions were found in the heptads. It has been proposed that the attractive electrostatic forces would favor either homo- or hetero- dimer formation. Only those proteins exclusively possess repulsive pairs would encourage heterodimerization. According to this principle, a very small number of proteins were predicted to have heterodimer specificity. For example, Ro03G06314 and Ro03G06310 had apparent heterodimer preference due to the presence of a repulsive pair in *g* and *e'* position and lacking any attractive pairs. Asparagine was also absent in the *a* position of this protein. Moreover, changed amino acids in the *a* position would also favor heterodimer formation[41]. Although we indicate that the LZ stops after the second heptad due to the presence of a proline residual in the third heptad, it appears very canonical from the third to the tenth heptads. For this reason, we propose that the Ro03G10052 and Ro06G19249 might tend to inhibit homodimer formation.

Protein-protein network prediction and validation of the protein interaction using yeast two-hybrid assays

We utilized the STRING PPI database to predict potential protein interactions based on homologous sequence searches. Out of the 49 identified bZIPs, 20 were predicted to interact with at least one other bZIP member (Fig. 6a). For example, the *S* group factor Ro05G19144 was found to interact with eight other bZIPs from the *C* group (Ro02G00753, Ro06G08256, and Ro02G04007), the *A* group (Ro02G00599, Ro02G21448, Ro05G 13372), as well as the S group (Ro02G16205, Ro02G02326).

To validate these predicted interactions, we selected seven genes, including the three members from the *C* group and the four members of the *S1* subgroup in the *S* clade. We used specific primers to amplify the expected products, but two of the genes produced multiple bands (Fig. 6b). We purified the PCR amplicons by slicing them from the gel and then ligated them into a cloning vector. We sequenced at least five clones from each sample. The sequencing results revealed that the larger bands observed on the gel were derived from non-specific amplification and were therefore excluded from further experiments. Interestingly, we obtained two different transcript isoforms from the gene *Ro02G04007*. The longer isoform (t2) retained two short introns (Fig. 6c), which resulted in the introduction of a premature stop codon. Consequently, we did not include the longer isoform in subsequent assays.

To further validate the protein interactions, we employed the CrY2H-seq yeast two-hybrid system. The results confirmed that Ro1G01098 and Ro5G19144 were able to interact with all members of the *C* group of RobZIPs. Additionally, Ro2G02326 interacted with Ro2G04007 and Ro6G08256, while Ro2G16205 could only interact with one member of the *C* group, Ro2G 04007.

Discussion

The *bZIP* gene family is widely present in eukaryotes and participates in various biological processes. In this study, we identified 49 *bZIP* genes in the genomes of two *Rubus* species (*R. idaeus* and *R. occidentalis*) and conducted a comprehensive comparative analysis with several other important fruit tree species in the Rosaceae family. Previous studies have identified 114, 50, 47, and 62 members in apple, strawberry, peach, and pear, respectively[12–15]. The number of bZIPs identified in our study was slightly lower than in previous studies. This difference can be attributed, in part, to our strict selection criteria for the presence of both the BR and LZ domains, as well as the number of heptad repeats in the C-termini. For example, we excluded Ro03G06962 from our list because it matched the bZIP motif pattern but lacked the canonical LZ heptads. The updated genome information may also contribute to the differences in results. Wang et al. first identified bZIPs in diploid strawberry when the initial genome annotation was released[13]. As omics-related techniques have progressed, genome data has been updated and renewed in recent years.

Various number of subgroup members were identified in the related Rosaceae species, with the *A*, *D*, *I*, *S* groups being the most dominant (Fig. 2). When compared with *Arabidopsis*, the number was generally smaller in these Rosaceae plants, except the cultivated apple and pear, in which similar amount of bZIPs were found. Two recent WGDs events (named α and β) have been proposed to be occured in the Brassicaceae lineage (where *Arabidopsis* resides)[42], which was absent in most Roside linage. Instead, a recent WGD has shaped the genome of the domesitcated apple and 'Cuiguan' pear in the history[43,44]. This indicated that the WGD envents might have contributed to the epansion of the bZIP members in the selected fruit trees. In accordance, intrachromosomal gene duplication analysis demonstrated that about 1/3 of the *bZIP*s were derived from segmental duplication in the *Rubus*. However, more *bZIP*s were raised from dispersed gene duplication, which might continue occurring after WGD, because we did not detect significant expansion/contraction for the *bZIP* gene family using the birth/death modeling. These results were in contrast to the results observed in the Chinese white pear genome, in which WGD contributed equally to that of the dispersed duplication events[15].

To further investigate the function of *bZIP*s in black raspberry, we annotated the entire genome and conducted GO enrichment analysis on the identified *bZIP* genes (Fig. 4). The enrichment analysis indicated that the molecular function of *bZIP* genes is mainly related to DNA-binding transcription factor activity, their cellular localization is mainly in the nucleus, and their biological processes mainly include nitrogen compound metabolism, response to carbohydrate, response to stimulus, response to stress, and regulation of cellular macromolecule biosynthesis. These findings are consistent with the functions of *bZIP* gene family members in other species. For instance, *AtbZIP1* have been reported to be involved in plant resistance[45]; *AtbZIP11* regulates both carbon and nitrogen metabolism[46]. The homologous gene of *AtbZIP53* in strawberry, *FabZIPs1.1*, not only enhances plant resistance but also plays a role in regulating sugar accumulation[20]. Therefore, it is speculated that the collinear gene *Ro05G19144* in black raspberry has the potential to regulate sugar accumulation in fruit, making it a candidate gene for improving the quality of black raspberry fruits. As found in our results, the RobZIPs (Ro06G08256, Ro05G31718, Ro01G01098) clearly participated in the process of nitrogen compound metabolism since they respond to all types of nitrogen fertilizers (Fig. 4c, d). Further experimental validation is expected.

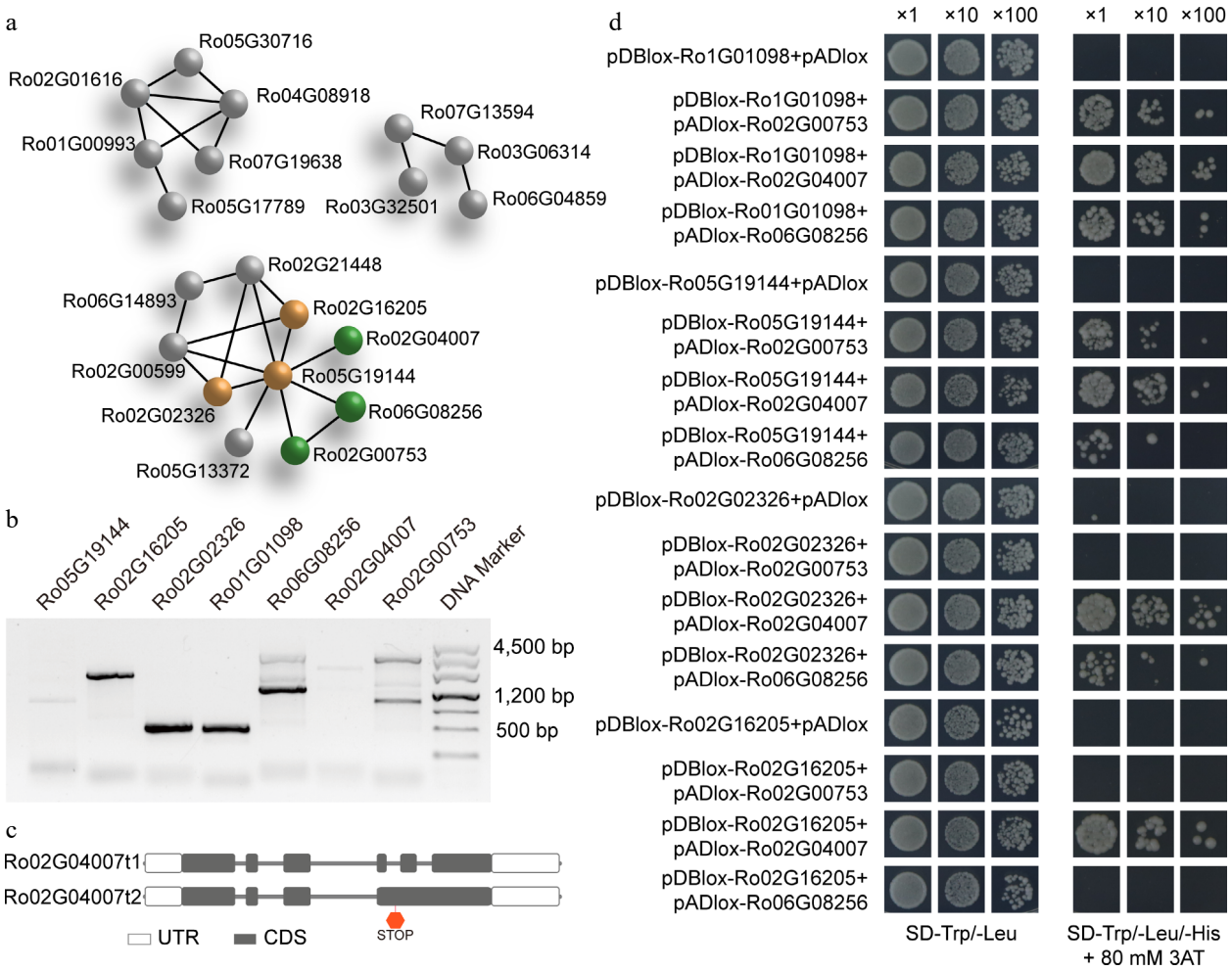

Fig. 6 Protein-protein network prediction, PCR amplification, and interaction validation for the C/S1 group *R. occidentalis* bZIP members using yeast two-hybrid analysis. PPI network prediction of all RobZIPs was conducted using the STRING PPI database. The *S* group members were colored in brown, and the C group members were colored in green. (a) Members of other groups are in gray. (b) PCR amplification of the four S1 and three C group *RobZIP* members and (c) two isoforms for the *Ro02G04007* gene due to alternative splicing. Validation of the heterodimer interaction of the C/S1 RobZIPs using yeast two-hybrid system.

Previous studies have shown that the C and S1 group of bZIP can form specific dimers named bZIP C/S1 network. The bZIP C/S1 network has been established as a signaling hub coordinating plant development and stress responses[47]. The Arabidopsis bZIP C/S1 interaction network mediates downstream signal transduction of SnRK1 during low energy stress and is also known as the SnRK1-C/S1 signaling model[19]. Using amino acid sequence analysis, we determined that the dimerization ability and stability of RobZIP were stronger than those of *Arabidopsis* and rice. Furthermore, we predicted its interaction network using the STRING database and verified it with the yeast two-hybrid assay (Fig. 6). However, there are fewer interacting proteins than that in *Arabidopsis*[19,45]. This results can partly ascribe to the fact that we did not investigate the interaction between homo-RobZIPs, as we know little about the functions and mechanisms of homologous interactions in RobZIPs. In the recent results depicted through double DAP-seq, the interactions occurring in either C/S1 bZIP heterodimer or the S1/S1 homodimer form could significantly expand the DNA binding preferences of these TFs[48]. Additionally, this dimer combination pattern would change as phosphorylation modifications of the protein might occur under energy-sufficient or -deficient environmental conditions[17]. We have identified nine hetero interaction pairs of the C/S1 RobZIPs using yeast two-hybrid analysis (Fig. 6), which represents the first bZIP interaction map in the *Rubus* species. Furthermore, bZIP members from other groups were also able to function as hetero- or homodimers. For example, the famous FD and FDP (both A group bZIPs) interaction module is indispensable in inflorescence meristem development[49]. Much more work is needed for future studies.

In conclusion, we have conducted a genome-wide analysis of the *bZIP* family in two *Rubus* species and have demonstrated their ability to respond to both internal and external energy signals, which enhances plant adaptability and survival. Our investigation may lead to further insight into the relationship between the function and structure of bZIP family members.

Author contributions

The authors confirm contribution to the paper as follows: conceptualization: Chen Q, Lin X; data curation, formal analysis: Lin X; funding acquisition, project administration: Chen Q; methodology: Lin X, Huang M, He J; software, visualization: Min

A, Zhou Y, Ma W; original draft, review & editing: Lin X, Liu X; review and editing: Wang X, Tang H. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The datasets used during and/or analyzed in the current study are publicly available in the Sequence Read Archive database. The code used are available from the corresponding author on reasonable request.

Acknowledgments

This work was financially supported by the National Natural Foundation of China (No. 31972387) and Sichuan Agricultural University's education reform project No. X202143.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary Information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/frures-0024-0001)

Dates

Received 16 October 2023; Accepted 26 December 2023; Published online 1 February 2024

References

1. Foster R, Izawa T, Chua NH. 1994. Plant bZIP proteins gather at ACGT elements. *The FASEB Journal* 8:192–200
2. Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J. 2002. bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science* 7:106–11
3. Vinson CR, Hai T, Boyd SM. 1993. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes and Development* 7:1047–58
4. Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, et al. 2002. Classification of human B-ZIP proteins based on dimerization properties. *Molecular and Cellular Biology* 22:6321–35
5. Iglesias-Fernández R, Barrero-Sicilia C, Carrillo-Barral N, Oñate-Sánchez L, Carbonero P. 2013. *Arabidopsis thaliana* bZIP44: a transcription factor affecting seed germination and expression of the mannanase-encoding gene *AtMAN7*. *The Plant Journal* 74:767–80
6. Ye L, Wu Y, Zhang J, Zhang J, Zhou H, et al. 2023. A bZIP transcription factor (*CiFD*) regulates drought- and low- temperature-induced flowering by alternative splicing in citrus. *Journal of Integrative Plant Biology* 65:674–91
7. Dröge-Laser W, Snoek BL, Snel B, Weiste C. 2018. The *Arabidopsis* bZIP transcription factor family — an update. *Current Opinion in Plant Biology* 45:36–49
8. Zhong L, Chen D, Min D, Li W, Xu Z, et al. 2015. AtTGA4, a bZIP transcription factor, confers drought resistance by enhancing nitrate transport and assimilation in *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* 457:433–39
9. Tang C, Li T, Klosterman SJ, Tian C, Wang Y. 2020. The bZIP transcription factor VdAtf1 regulates virulence by mediating nitrogen metabolism in *Verticillium dahliae*. *New Phytologist* 226:1461–79
10. Wang H, Xu K, Li X, Blanco-Ulate B, Yang Q, et al. 2023. A pear S1-bZIP transcription factor PpbZIP44 modulates carbohydrate metabolism, amino acid, and flavonoid accumulation in fruits. *Horticulture Research* 10:uhad140
11. Zhang X, Wu Y, Zhang Y, Yin X, van Nocker, et al. 2022. Identification of potential key genes in resveratrol biosynthesis via transcriptional analyses of berry development in grapevine (*Vitis* spp.) genotypes varying in *trans*-resveratrol content. *Fruit Research* 2:6
12. Li Y, Meng D, Li M, Cheng L. 2016. Genome-wide identification and expression analysis of the bZIP gene family in apple (*Malus domestica*). *Tree Genetics & Genomes* 12:82
13. Wang X, Chen X, Yang T, Cheng Q, Cheng Z. 2017. Genome-wide identification of *bZIP* family genes involved in drought and heat stresses in strawberry (*Fragaria vesca*). *International Journal of Genomics* 2017:3981031
14. Sun M, Fu X, Tan Q, Liu L, Chen M, et al. 2016. Analysis of basic leucine zipper genes and their expression during bud dormancy in peach (*Prunus persica*). *Plant Physiology and Biochemistry* 104:54–70
15. Manzoor MA, Manzoor MM, Li G, Abdullah M, Han W, et al. 2021. Genome-wide identification and characterization of bZIP transcription factors and their expression profile under abiotic stresses in Chinese pear (*Pyrus bretschneideri*). *BMC Plant Biology* 21:413
16. Peviani A, Lastdrager J, Hanson J, Snel B. 2016. The phylogeny of C/S1 bZIP transcription factors reveals a shared algal ancestry and the pre-angiosperm translational regulation of S1 transcripts. *Scientific Reports* 6:30444
17. Mair A, Pedrotti L, Wurzinger B, Anrather D, Simeunovic A, et al. 2015. SnRK1-triggered switch of bZIP63 dimerization mediates the low-energy response in plants. *eLife* 4:e05828
18. Weltmeier F, Ehlert A, Mayer CS, Dietrich K, Wang X, et al. 2006. Combinatorial control of *Arabidopsis* proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. *The EMBO Journal* 25:3133–43
19. Dröge-Laser W, Weiste C. 2018. The C/S_1 bZIP network: a regulatory hub orchestrating plant energy homeostasis. *Trends in Plant Science* 23:422–33
20. Chen Q, Tang Y, Wang Y, Sun B, Chen T, et al. 2020. Enhance sucrose accumulation in strawberry fruits by eliminating the translational repression of FabZIPs1.1. *Scientia Horticulturae* 259:108850
21. Zhang Y, Li S, Chen Y, Liu Y, Lin Y, et al. 2022. Heterologous overexpression of strawberry *bZIP11* induces sugar accumulation and inhibits plant growth of tomato. *Scientia Horticulturae* 292:110634
22. Xing S, Chen K, Zhu H, Zhang R, Zhang H, et al. 2020. Fine-tuning sugar content in strawberry. *Genome Biology* 21:230
23. Deppmann CD, Acharya A, Rishi V, Wobbes B, Smeekens S, et al. 2004. Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to *Homo sapiens* B-ZIP motifs. *Nucleic Acids Research* 32:3435–45
24. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–34
25. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8:28–36
26. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194–202
27. Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, et al. 2006. Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *The Plant Journal* 46:890–900
28. Wang Y, Tang H, Debarry JD, Tan X, Li J, et al. 2012. *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40:e49
29. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20:238

30. Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36:5516–18
31. Xiang Y, Huang C, Hu Y, Wen J, Li S, et al. 2017. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular Biology and Evolution* 34:262–81
32. Mora-Márquez F, Chano V, Vázquez-Poletti JL, de Heredia UL. 2021. TOA: a software package for automated functional annotation in non-model plant species. *Molecular Ecology Resources* 21:621–36
33. Yu G, Wang L, Han Y, He Q. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology* 16:284–87
34. VanBuren R, Wai C, Colle M, Wang J, Sullivan S, et al. 2018. A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. *GigaScience* 7:giy094
35. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34:525–27
36. Duan Y, Yang H, Yang H, Wu Y, Fan S, et al. 2023. Integrative physiological, metabolomic and transcriptomic analysis reveals nitrogen preference and carbon and nitrogen metabolism in blackberry plants. *Journal of Plant Physiology* 280:153888
37. Wanamaker SA, Garza RM, MacWilliams A, Nery JR, Bartlett A, et al. 2017. CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nature Methods* 14:819–25
38. Raymond O, Gouzy J, Just J, Badouin H, Verdenaud M, et al. 2018. The *Rosa* genome provides new insights into the domestication of modern roses. *Nature Genetics* 50:772–77
39. Wei K, Chen J, Wang Y, Chen Y, Chen S, et al. 2012. Genome-wide analysis of bZIP-encoding genes in maize. *DNA Research* 19:463–76
40. Nijhawan A, Jain M, Tyagi AK, Khurana JP. 2008. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiology* 146:333–50
41. Acharya A, Ruvinov SB, Gal J, Moll JR, Vinson C. 2002. A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. *Biochemistry* 41:14122–31
42. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100
43. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, et al. 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* 42:833–39
44. Gao Y, Yang Q, Yan X, Wu X, Yang F, et al. 2021. High-quality genome assembly of 'Cuiguan' pear (*Pyrus pyrifolia*) as a reference genome for identifying regulatory genes and epigenetic modifications responsible for bud dormancy. *Horticulture Research* 8:197
45. Kang SG, Price J, Lin PC, Hong JC, Jang JC. 2010. The *Arabidopsis* bZIP1 transcription factor is involved in sugar signaling, protein networking, and DNA binding. *Molecular Plant* 3:361–73
46. Hanson J, Hanssen M, Wiese A, Hendriks MMWB, Smeekens S. 2008. The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of *ASPARAGINE SYNTHETASE1* and *PROLINE DEHYDROGENASE2*. *The Plant Journal* 53:935–49
47. Weltmeier F, Rahmani F, Ehlert A, Dietrich K, Schütze K, et al. 2009. Expression patterns within the *Arabidopsis* C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Molecular Biology* 69:107–19
48. Li M, Yao T, Lin W, Hinckley WE, Galli M, et al. 2023. Double DAP-seq uncovered synergistic DNA binding of interacting bZIP transcription factors. *Nature Communications* 14:2600
49. Romera-Branchat M, Severing E, Pocard C, Ohr H, Vincent C, et al. 2020. Functional divergence of the *Arabidopsis* florigen-interacting bZIP transcription factors FD and FDP. *Cell Reports* 31:107717