

## Research Article

# Divergence of the bZIP Gene Family in Strawberry, Peach, and Apple Suggests Multiple Modes of Gene Evolution after Duplication

Xiao-Long Wang, Yan Zhong, Zong-Ming Cheng, and Jin-Song Xiong

College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China

Correspondence should be addressed to Jin-Song Xiong; [jsxiong@njau.edu.cn](mailto:jsxiong@njau.edu.cn)

Received 15 September 2015; Revised 10 November 2015; Accepted 11 November 2015

Academic Editor: Yanbin Yin

Copyright © 2015 Xiao-Long Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The basic leucine zipper (bZIP) transcription factors are the most diverse members of dimerizing transcription factors. In the present study, 50, 116, and 47 *bZIP* genes were identified in *Malus domestica* (apple), *Prunus persica* (peach), and *Fragaria vesca* (strawberry), respectively. Species-specific duplication was the main contributor to the large number of *bZIP*s observed in apple. After WGD in apple genome, orthologous *bZIP* genes corresponding to strawberry on duplicated regions in apple genome were retained. However, in peach ancestor, these syntenic regions were quickly lost or deleted. Maybe the positive selection contributed to the expansion of clade S to adapt to the development and environment stresses. In addition, purifying selection was mainly responsible for *bZIP* sequence-specific DNA binding. The analysis of orthologous pairs between chromosomes indicates that these orthologs derived from one gene duplication located on one of the nine ancient chromosomes in the Rosaceae. The comparative analysis of *bZIP* genes in three species provides information on the evolutionary fate of *bZIP* genes in apple and peach after they diverged from strawberry.

## 1. Introduction

Many of the biological processes in cell or organism, such as responses to the environment and progression through the cell cycle, metabolic and physiological balance are influenced or controlled by regulation of gene expression at the level of transcription. Development is based on the cellular capacity for differential gene expression and is controlled by transcription factors acting as switches of regulatory cascades [1]. Alterations in the expression of genes coding for transcription factors (TFs) are emerging as a major source of the diversity and change that underlie evolution [2]. Presently, at least 64 families of transcription factors have been identified in the plant kingdom [3]. The bZIP proteins represent a large family of TFs with a DNA-binding domain rich in basic amino acid residues, which is adjacent to a leucine zipper dimerization domain (N-x7-R/K-x9) for sequence-specific DNA binding, and a leucine zipper, which is composed of several heptad

repeats of Leu or other bulky hydrophobic amino acids, such as Ile, Val, Phe, or Met, for dimerization specificity [4–7]. In addition, the majority of characterized plant *bZIP* genes to date have been associated with enhancing plant tolerance to diverse types of abiotic stress [8–14].

Recent *bZIP* gene sequence analyses in *Arabidopsis* [5], rice [6], castor bean [15], maize [16], sorghum [17], cucumber [18], and grape [19], further indicated illegitimate recombination (IR) as a major source of duplications and deletions [20]. The evidence obtained from these analyses suggests that gene duplications in a common ancestor of those plants gave rise to *bZIP* genes. Therefore, the very earliest origins of the *bZIP* gene family are associated with a series of gene duplications. A total of 75 and 89 *bZIP* genes have been identified in *Arabidopsis* [5] and rice (*Oryza sativa*) [6], respectively. The *bZIP* genes in these two genera have been classified into 10 groups and 11 groups, respectively, based on DNA binding specificity and sequence similarity.

The Rosaceae is one of the most economical plant families [21] composed by some 90 genera with over 3000 distinct species which have  $x = 7$  to  $x = 17$  chromosomes [22]. According to a phylogenetic treatment based on DNA sequence, data of nuclear and chloroplast genomic regions in Rosaceae reclassified the genus into Dryadoideae, Rosoideae, and Spiraeoideae, each containing a number of distinct supertribes [22]. *Prunus* and *Malus* are included in the Spiraeoideae, supertribe Amygdaleae, and Pyrodae (tribe Pyrinae), respectively, whilst *Fragaria* is included in the Rosoideae, supertribe Rosodae (tribe Fragariinae) [23]. After the rapid evolution of Rosaceae, members of the family display remarkable phenotypic diversity, plant habit, chromosome number, and fruit type which evolved independently on more than one opportunity [24, 25]. A better understanding of how the *bZIP* genes within the Rosaceae arose would provide an insight into how evolution can lead rapidly to diversification. The genomes of three Rosaceous species, woodland strawberry [26], domesticated apple [27], and peach [28], have been recently sequenced, providing an opportunity to conduct a high-resolution comparison of their genomes. In this study, we identified 50, 116, and 47 *bZIP* transcription factors based on the complete genome sequences of strawberry, apple, and peach. Further, through phylogenetic analysis, *Ka/Ks* ratios of genes and *bZIP* domains, and orthologous relationships among chromosomes, we explain the evolutionary history of *bZIP* genes in detail.

## 2. Methods

**2.1. Data Resources and the Identification of *bZIP* Genes.** *Fragaria vesca* (strawberry, v1.1), *Malus domestica* (apple, v1.0), and *Prunus persica* (peach, v1.0) genomic and annotation data were downloaded from the Genome Database for Rosaceae (GDR, <http://www.Rosaceae.org/>) [26–28]. The genome sequences of *Brassica rapa* (v1.3), *Solanum lycopersicum* (iTAG2.3), *Chlamydomonas reinhardtii* (v5.5), *Theobroma cacao* (v1.1), *Selaginella moellendorffii* (v1.0), *Populus trichocarpa* (v3.0), *Medicago truncatula* (Mt4.0v1), *Cucumis sativus* (v1.0), *Carica papaya* (ASGPBv0.4), and *Physcomitrella patens* (v3.0) were downloaded from Phytozome (<http://www.phytozome.net/>) [29]. Genomic data on *M. acuminata* (v1) (<http://banana-genome.cirad.fr/>), *Saccharomyces cerevisiae* (v1) (<http://www.yeastgenome.org/>), and *Cyanidioschyzon merolae* (<http://merolae.biol.s.u-tokyo.ac.jp/>) were also downloaded for inclusion in the analyses. The *bZIP* genes in the genomes of *Vitis vinifera* [19], *Arabidopsis thaliana* [5], and rice (*Oryza sativa*) [6] were previously identified.

The Hidden Markov Model (HMM) profiles of the *bZIP* domain (PF00170) were retrieved from Pfam 27.0 [30] and used for identifying the *bZIP* genes from the downloaded database of genomes using HMMER3.0 [31]. All output genes with a default *E*-value ( $<1.0$ ) were collected and the online software SMART (<http://smart.embl-heidelberg.de/>) was used to confirm the integrity of the *bZIP* domain using an *E*-value of  $<0.1$  [32]. Incorrectly predicted genes were removed. Finally, the sequences of nonredundant genes with high confidence were collected and assigned as *bZIP* genes.

**2.2. Alignment and Phylogenetic Analysis of *bZIP* Genes.** Based on the location (Table S2 in Supplementary Material available online at <http://dx.doi.org/10.1155/2015/536943>) predicted in the Pfam 27.0 [30] of conserved domains in complete predicted *bZIP* protein sequences, the conserved domain sequences of *bZIP* proteins were extracted and aligned using ClustalX (version 1.83) [33]. The phylogenetic trees were generated with MEGA 5.0 [34] using the Neighbor-Joining (NJ) method and number of difference model [35]. 1,000 bootstraps were used to evaluate the significance of the phylogenetic trees.

**2.3. Synteny Analysis of Strawberry, Apple, and Peach Genomes.** For synteny analysis, syntenic genes within the strawberry, apple, and peach genomes, as well as between strawberry and apple, strawberry and peach, peach and apple genomes, were downloaded from the Plant Genome Duplication Database [36] (PGDD, <http://chibba.agtec.uga.edu/duplication/>) and those containing *bZIP* genes were identified and analyzed. We identified the syntenic gene pairs from the same and different species within the same clade from phylogenetic analysis as paralogous and orthologous genes.

**2.4. Estimation of Nonsynonymous Substitutions and Synonymous Substitutions.** The nucleotide sequences of *bZIP* gene and *bZIP* domain in each clade except for UN were aligned by using Clustalw 2.0 [37]. The nonsynonymous substitutions (*Ka*) and synonymous substitutions (*Ks*) and nonsynonymous to synonymous substitution ratios (*Ka/Ks*) were estimated in each gene family according to the alignments in MEGA 5.0 [38]. In order to detect selection pressure of different clades of *bZIPs* in phylogenetic tree (A, B, C, D, E, F, G, H, I, and S), *Ka/Ks* ratio greater than 1, less than 1, and equal to 1 represents positive selection, negative or stabilizing selection, and neutral selection, respectively. The software in SPSS version 19.0 (SPSS, Chicago, IL, USA) was used for statistical analysis. The statistical significance of *Ka/Ks* was defined based on Duncan's multiple range test and *P* value of  $< 0.05$  as statistically significant.

## 3. Results

**3.1. Identification and Comparative Analyses of *bZIP* Genes in Nineteen Species.** The sequences of 1441 *bZIP* sequences in 19 genomes, ranging from fungi to Plantae, including the three Rosaceous species, were used to analyze the evolution of this gene family (Figure 1 and Table S1). In the genome assemblies of strawberry, apple, and peach, 50, 116, and 47 *bZIP* genes were identified, respectively, using the HMM profile from the Pfam database [39] (Table S2). The number of *bZIP* genes varies from 4 (*C. merolae*) to 212 (*P. trichocarpa*) in 19 species with the genome size from 12.2 Mb (*S. cerevisiae*) to 881.3 Mb (*M. domestica*). Furthermore, we found that the number of *bZIP* genes in six higher plant species was more than 100.

The total number of *bZIP* genes in strawberry and peach was very similar. However, it is important to note that the number of *bZIP* genes in these two species was much less than the number in apple (116). The number of *bZIP* genes in strawberry (50) and peach (47) was also much smaller

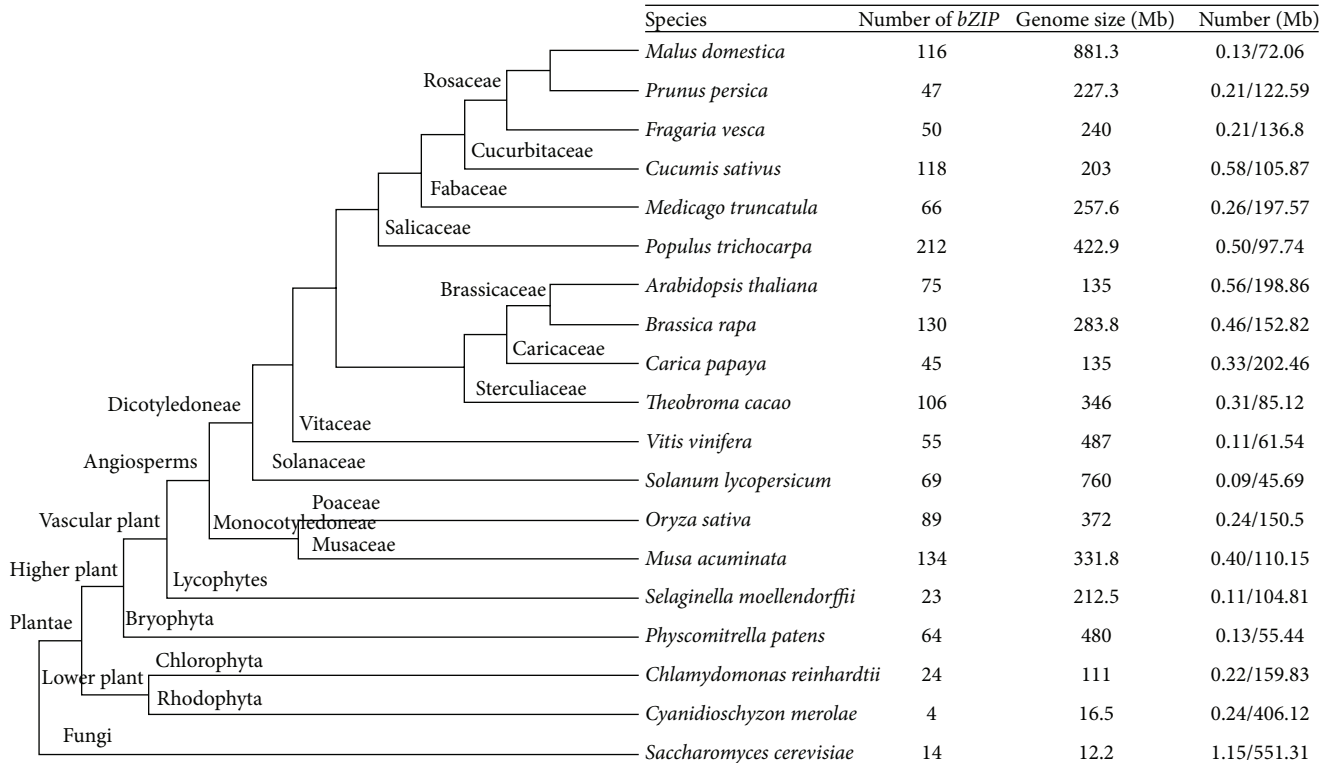


FIGURE 1: Phylogenetic relationships, number of *bZIP* genes, genome size, *bZIP* density, and overall gene density of the nineteen species analyzed. The *bZIP* density and overall gene density of the nineteen species analyzed were separated by parenthesis. The *bZIP* density was followed by overall gene density.

than the number in most of the 19 species (Figure 1), which may infer that only a small number of gene duplication events contributed to the *bZIP* members in these two species. The number of apple *bZIP* genes (116) was similar to the number in *C. sativus* (118) but less than the number found in *P. trichocarpa* (212), *B. rapa* (130), and *M. acuminata* (134) (Figure 1). In contrast, the density of *bZIP* genes in the three Rosaceous species was very distinct and not related to the number of *bZIP* genes present. The *bZIP* density in the apple genome (0.13) was lower than that in strawberry (0.21) and peach (0.21) and only exceeded the density observed in *V. vinifera* (0.11), *S. moellendorffii* (0.11), and *S. lycopersicum* (0.09) (Figure 1). By contrast, the apple genome also had a lower overall gene density (72.06), which is probably the reason for low *bZIP* density in the apple genome.

**3.2. Phylogenetic Analysis of *bZIP* Genes in Three Rosaceous Species.** A phylogenetic analysis was performed for the *bZIP* genes in the three Rosaceous species using the *bZIP* domains in strawberry, apple, and peach, as well as *Arabidopsis* [5], in order to further elucidate the evolution of this gene family (Figure S1). Since the *bZIP* genes of *Arabidopsis* have already been clustered, we were able to compare the clustering of the *bZIP* genes of Rosaceous species with the clustering from *Arabidopsis*. Surprisingly, *AtbZIP31*, *AtbZIP33*, and *AtbZIP74* were different from other *bZIP* genes in that they formed individual clades containing only *bZIP* genes of *Arabidopsis*,

suggesting that these individual clades may be specific to *Arabidopsis* (Figure S1).

The results indicated that the ten clades (A, B, C, D, E, F, G, H, I, and S) obtained in our phylogenetic tree were in agreement with the clustering and classification of *bZIP* genes in *Arabidopsis* [5] (Figure 2). However, a few genes formed three small unique clades (UC, Figure 2) in the phylogenetic tree produced from our analyses. This observation supports the hypothesis that these three unique clades may have had independent evolutionary trajectories from the other clades.

All of the clades from Figure 2 include genes from all of the three species. The number of strawberry, apple, and peach *bZIP* genes, respectively, in each of the clades were A (9, 18, 8); B (1, 5, 1); C (3, 6, 4); D (7, 12, 6); E (2, 8, 3); F (2, 6, 2); G (6, 10, 4); H (2, 5, 2); I (6, 17, 6); and S (9, 21, 9). Moreover, the phylogenetic tree of the three Rosaceous species indicated that the *bZIP* genes in strawberry and peach have few paralogs with “one-to-one” topology (two paralogs clustered together in a clade), suggesting that most of them were generated before speciation of strawberry. In contrast, there were many clades with “one-to-one” or “one-to-many” topologies (more than two paralogs clustered together in a clade) in apple, indicating that species-specific duplication events contributed greatly to the large number of apple *bZIP*s.

**3.3. Nonsynonymous and Synonymous Substitution of *bZIP* Genes.** Our result indicates that most clades (A, B, C, D, E, F,

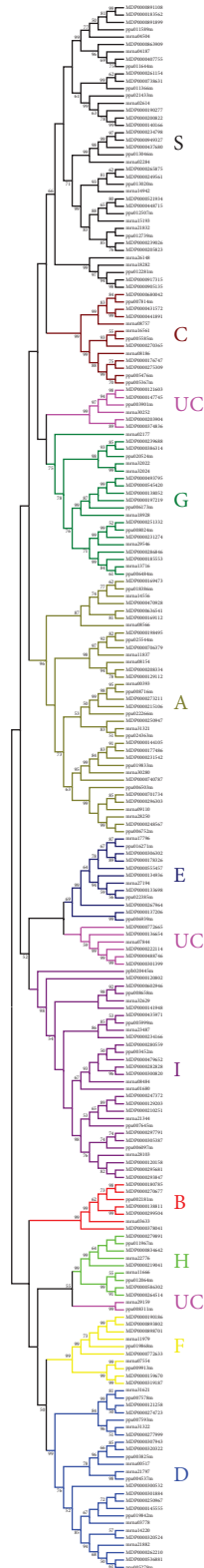


FIGURE 2: Phylogenetic analysis of *bZIP* members in strawberry, apple, and peach. Phylogenetic analysis of bZIP proteins in strawberry (mrna), apple (MDP), and peach (ppa). Only bootstrap values larger than 50% are indicated. Different colors can be used to distinguish the different subgroups. The names of each subgroup are listed on the right. UC represented “unique clades.”

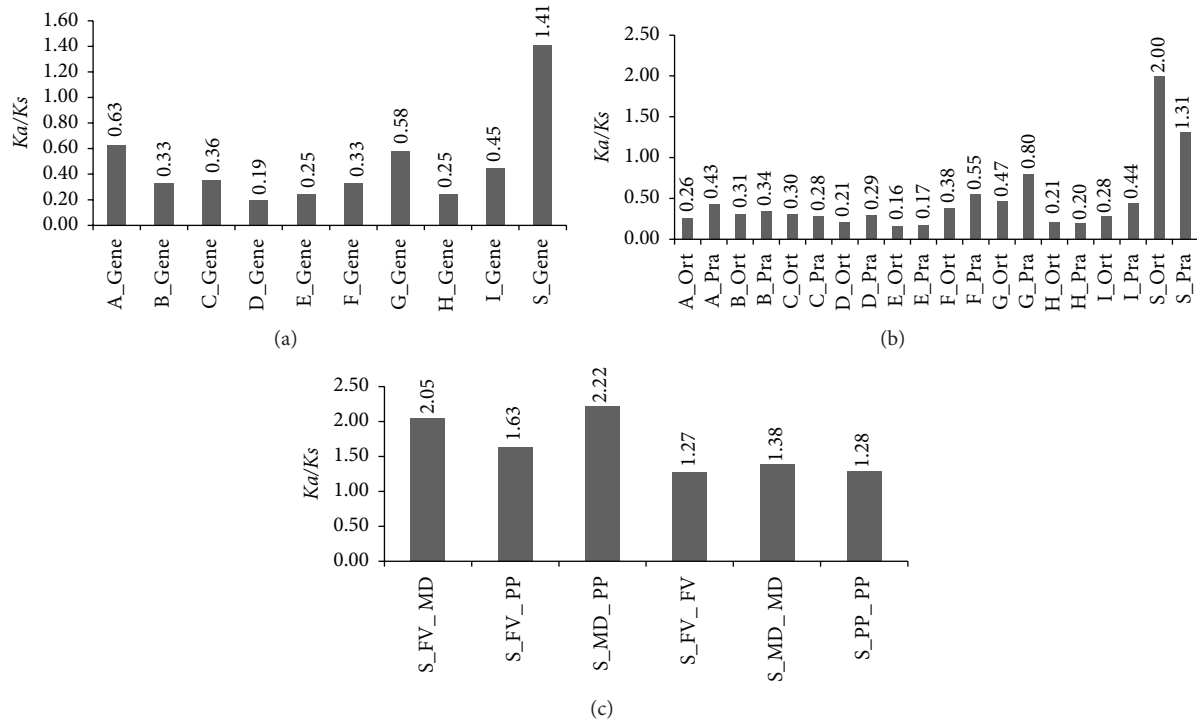


FIGURE 3:  $Ka/Ks$  ratios of *bZIP* genes. (a)  $Ka/Ks$  ratios of genes in clades A–S. (b)  $Ka/Ks$  ratios of paralogous and orthologous gene pairs in clades A–S. (c)  $Ka/Ks$  ratios of paralogs (FV\_FV, MD\_MD, and PP\_PP) and orthologs (FV\_MD, FV\_PP, and MD\_PP) in clade S. The  $Ka/Ks$  ratios are located in the top of the graph.

G, H, and I) had  $Ka/Ks$  ratios less than 1 (Figure 3(a)), demonstrating that most genes of those clades were undergoing a purifying selection in the three species. Among all the gene pairs in the clades, 25 (7.99% of clade A), 1 (2.13% of clade E), 16 (9.58% of clade G), and 12 (5.33% of clade I) pairs had  $Ka/Ks$  ratio approximately equal to 1 ( $Ka/Ks$  ratio = 0.8~1.0) for *bZIP* genes in strawberry, apple, and peach (Table S3). However, 52 (16.61% of clade A), 15 (8.98% of clade G), 1 (4.55% of clade H), and 15 (6.67% of clade I) gene pairs had  $Ka/Ks$  ratios greater than 1 for *bZIP* genes (Table S3), which indicates that some of *bZIP* genes were under positive selection or relaxed selection for gene pairs with  $Ka/Ks$  approximately equal to 1. It is worth noting that  $Ka/Ks$  ratio of gene pairs in clade S is significantly greater than other clades ( $P < 0.05$ ), which illustrated that *bZIP* genes were under strongly positive selection (Figure 3(a)).

In order to explain  $Ka/Ks$  ratio distribution of gene pairs in each clade, we compared  $Ka/Ks$  ratio of the orthologous and paralogous gene pairs in strawberry, apple, and peach (Table S4). It is indicated that the  $Ka/Ks$  ratio of paralogs is bigger than orthologs in each clade except for clades C, H, and S (Figure 3(b)). Most of orthologs and paralogs exhibit a low level  $Ka/Ks$  ratio ( $Ka/Ks$  ratio = 0.16~0.80) in different clades (A, B, C, D, E, F, G, H, and I) analyzed (Figure 3(b)). However, the ones of orthologs ( $Ka/Ks$  ratio = 2.00) and paralogs ( $Ka/Ks$  ratio = 1.31) in clade S are obviously greater than 1 and significantly higher than orthologs and paralogs in other clades ( $P < 0.05$ ). Orthologs and paralogs in clade

S could be further divided into three subgroups separately, FV\_PP (between strawberry and peach)/MD\_FV (between apple and strawberry)/MD\_PP (apple and peach) and FV\_FV (within strawberry)/MD\_MD (within apple)/PP\_PP (within peach). Orthologs in the MD\_PP have a highest  $Ka/Ks$  ratio (2.22) and paralogs in the FV\_FV have a lowest  $Ka/Ks$  ratio (1.27) (Figure 3(c)).

**3.4. Nonsynonymous and Synonymous Substitution of *bZIP* Domains.** For getting a more in-depth exploration in selection pressure of *bZIP* genes in different clades during their evolution, we compared the  $Ka/Ks$  ratio of *bZIP* domains in each clade (Table S5). We found that all clades with  $Ka/Ks$  ratios ranging from 0.04 (clade D) to 0.32 (clade G) were less than 0.4 (Figure 4(a)). It is suggested that a strong negative selection plays the leading roles in the evolution of *bZIP* domains.

Basic leucine zipper (*bZIP*) proteins, one of the largest families of transcription factors in plants, are characterized by a basic region (BR) responsible for sequence-specific DNA binding, an adjacent heptad leucine repeat, and the leucine zipper (LZ) [40]. It is concluded that all BR domains ranging from 0.02 (BR of clade C) to 0.24 (BR of clade I) and LZ domains ranging from 0.1 (LZ of clade I) to 0.61 (LZ of clade B) were undergoing negative selection (Figure 4(b), Table S5). Interestingly,  $Ka/Ks$  ratio of BR domain is less than the ones of LZ domain in each clade except for clades H and I (Figure 4(b)).

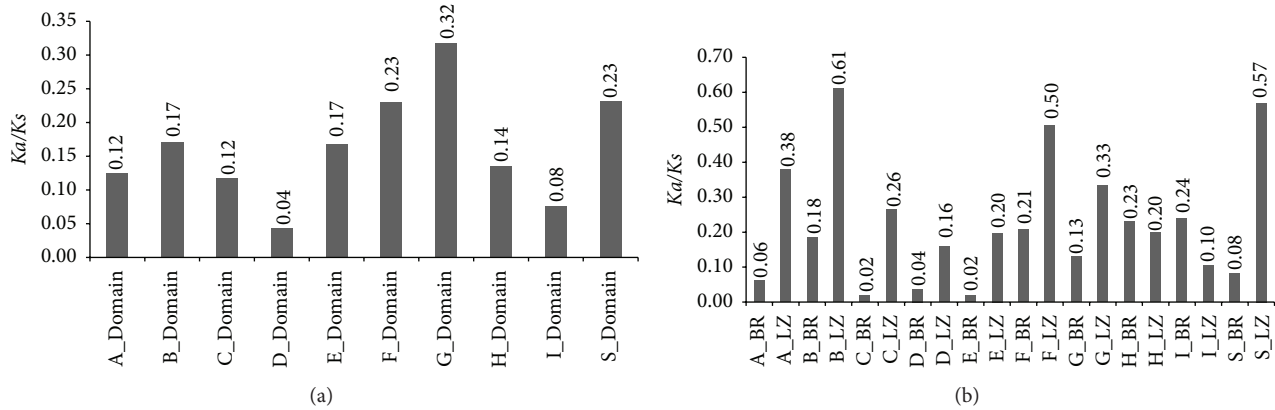


FIGURE 4:  $Ka/Ks$  ratios of *bZIP* domains. (a)  $Ka/Ks$  ratios of domains in clades A–S. (b)  $Ka/Ks$  ratios of BR and LZ domains in clades A–S. The  $Ka/Ks$  ratios are located in the top of the graph.

**3.5. Evaluation of Orthologous *bZIP* Genes between Strawberry, Apple, and Peach.** In order to trace the evolutionary history of *bZIP* genes among the three Rosaceous species, orthologous regions of *bZIP* genes in the three Rosaceous species were subjected to a comparative analysis in order to ascertain the evolutionary history of *bZIP* genes in the Rosaceae. Using Circos software [41], 57 orthologous gene pairs were identified between strawberry and apple (FV\_MD) (Figure 5(a)), 64 between apple and peach (MD\_PP) (Figure 5(b)), and 50 between strawberry and peach (FV\_PP) (Figure 5(c)). Collectively, these data are presented in Table S6 and Figure 5.

Out of the 57 gene pairs present in the strawberry and apple genomes (Figure 5(a)), 20 strawberry *bZIP* genes correspond to one copy (Type 1), 17 genes correspond to two copies (Type 2), and one gene corresponds to three copies (Type 3) in apple. Therefore, 56 *bZIP* genes in the apple genome have 38 corresponding genes in the strawberry genome. In all three types, some genes have preserved and exhibit the same number of exons (Table S6). Out of 50 gene pairs present in the strawberry and peach genomes (Figure 5(c)), 26 strawberry *bZIP* genes correspond to one copy (Type 1), 9 genes to two copies (Type 2), and 2 genes to three copies (Type 3) in peach. Collectively, 37 strawberry *bZIP* genes corresponded to 38 *bZIP* genes in the peach genome. Genes of all three types in strawberry and peach have preserved similar exon configurations (Table S6). Based on the 30 overlapping *bZIP* strawberry genes, the data collectively indicate that 45 *bZIP* genes, representing 90% of the total number of *bZIP* genes in the strawberry genome, were ancestral and underwent different duplication events after the divergent speciation of apple and peach. Additionally, 56 *bZIP* genes, representing 48.3% of the total number of *bZIP* genes in the apple genome, were retained on duplicated regions. In addition, 38 *bZIP* genes, representing 80.9% of the total number of *bZIP* genes in the peach genome, were retained on syntenic blocks. These data further indicate that most of the *bZIP* genes in strawberry and peach experienced a low level of duplication events compared to the number of duplication events in the apple genome. These findings are consistent with

the results of a previous study which reported that a recent whole genome duplication (WGD) event occurred in apple 60–65 million years ago [27].

**3.6. Orthologous Relationships among Chromosomes.** In order to understand the influence of the WGD in apple on the *bZIP* gene family in the Rosaceae, the major distribution of orthologous chromosomes was identified and compared between paired combinations of strawberry, apple, and peach according to the classification reported by Jung et al. [42] (Table S6, Table S7). The orthologous relationship between chromosomes of peach and strawberry made it evident that the majority of *bZIP* genes on peach chromosomes PC2, PC3, PC5, and PC8 were located on a single homologous FC7, FC6, FC5, and FC2 chromosome in strawberry, respectively. The majority of genes on PC6 and PC7 were also located on strawberry chromosomes, FC1 and FC6. Additionally, 35.71% of the *bZIP* genes on strawberry chromosome FC2 had an orthologous relationship to the PC1. Both ppa016271m and ppa022385m of PC4, however, were located on the FC6 chromosome of strawberry.

The relationship between peach and apple at the chromosome level was more complex than the relationship between peach and strawberry. 66.67%, 66.67%, 50%, 50%, and 50% of *bZIP* genes on five apple chromosome sets, MC2/MC7, MC9/MC17, MC3/MC11, MC14/MC6, and MC2/MC15, respectively, have their orthologous genes corresponding to the chromosomes PC2, PC3, PC4, PC5, and PC7 of peach. Orthologous genes on PC6 corresponded to major genes on four apple chromosomes, MC3, MC4, MC11, and MC12 (Table S6, Table S7).

## 4. Discussion

**4.1. Evolutionary History of *bZIP* Family in Three Species of the Rosaceae.** The *bZIP* transcription factor family is one of the largest and most diverse families of transcriptional regulators in eukaryotic organisms [15]. In the present study, the *bZIP* transcription factor family in 16 species, including 13 higher plants, 2 lower plants, and one fungus, was analyzed, in an

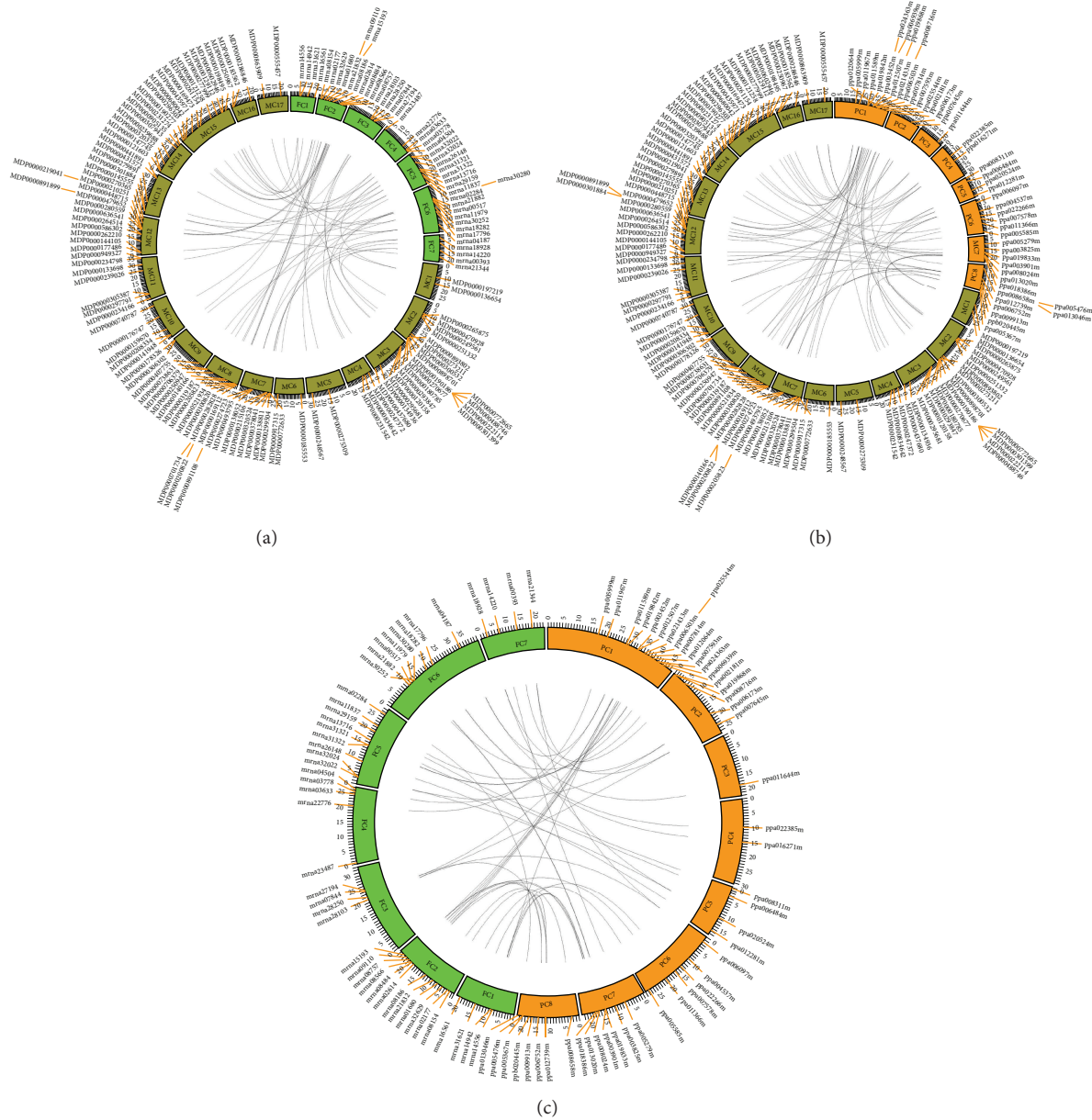


FIGURE 5: Evaluation of orthologous *bZIP* genes between strawberry, apple, and peach. (a) Seven strawberry (FC1 to FC7) and seventeen apple chromosome (MC1 to MC17) maps are based on the orthologous pair position and demonstrate a highly conserved syntenic relationship. (b) Seventeen apple chromosome (MC1 to MC17) and peach (PC1 to PC8) maps are based on the orthologous pair positions and demonstrate highly conserved synteny. (c) Seven strawberry (FC1 to FC7) and peach (PC1 to PC8) maps are based on the orthologous pair positions, and demonstrate highly conserved synteny.

effort to better understand the evolution of this gene family in the Rosaceae. It has been suggested that the *bZIP* gene family existed before the divergence of higher and lower plant species, even in the fungi, which consists with foundation in Wang et al. (2011) research [17]. An uneven distribution of *bZIP* copies among the 19 species was identified, suggesting that the *bZIP* genes within each species had undergone different levels of gene duplication with larger expansion after the divergence of higher and lower plants. For example, the numbers of copies of *bZIP* genes were as follows: *O. sativa* (89), *Cucumis sativus* (118), and *Populus trichocarpa* (212).

These observations suggest that specific functional expansion may have resulted from environmental selection pressure or specialization in processes of growth and development, including stress responses [14, 43–46] and abscisic acid (ABA) signaling [10, 11, 47]. As a result of evolutionary pressure and/or environmental selection, critical genes or components of genes were retained, whereas others were deleted or lost [48].

We identified 50 and 47 *bZIP* genes in the genomes of strawberry and peach, respectively. This number is similar with those of previous genome-wide studies on some other

species, indicating the presence of 64 *bZIP* homologs in cucumber [18], 55 in grapevine [19], and 49 in castor bean [15]. The *bZIP* homologs in apple (116) were consistent with the numbers in maize (120) [16] and sorghum (92) [17]. *bZIP* genes in strawberry and peach are much lower than that in apple which has a much larger genome size. These observations support the hypothesis that the WGD [27] event which occurred in apple resulted in significant amplification in the number of apple *bZIP* genes. On the other hand, a low level of gene duplication events may have contributed to the number of *bZIP* genes in strawberry and peach.

**4.2. *bZIP* Genes Expansion in the Rosaceae.** The phylogenetic tree of the *bZIP* gene family generated in this study for Rosaceous species is supported by Liu et al. [19], Nijhawan et al. [6], and Wei et al. [16]. Each of the clades included at least 7 *bZIP* genes from the 39 *bZIP* genes identified in the 3 species examined, indicating that many of the *bZIP* genes originated through a process of gene duplication. The widespread existence of paralogs and orthologs with “one-to-one” or “one-to-many” topologies in the Rosaceous species examined suggests that species-specific duplication was the main contributor to the large number of *bZIP*s observed in apple. The number of *bZIP* genes in each of the three species was highly variable, indicating that most of the gene duplication events occurred after evolutionary divergence of each lineage. It is also likely that both WGD and a series of rearrangements occurred during the evolution of certain species.

Extensive genome and EST sequencing of plant species has revealed a substantial history of WGD events [49, 50]. In the Rosaceae, an evolutionary trend toward fruit development and specialization may have been partially based on gene duplication. For example, WGD in apple has resulted in the creation of large families of paralogous genes [27]. In our analysis of the apple genome, 56 (48.3%) *bZIP* genes were retained on duplicated regions. Therefore, the involvement of WGD in the expansion of the *bZIP* gene family in apple is quite evident. Polyploidy provides an excellent genomic resource to study retention and loss of multicopy genes [48, 51]. Following WGD, genes can suffer a variety of fates ranging from massive gene loss to the development of a central role in an essential aspect of the plant [52]. A comparative analysis of *bZIP* genes in strawberry, apple, and peach led us to hypothesize that, after WGD in an apple ancestor, orthologous *bZIP* genes corresponding to strawberry on duplicated regions in apple genome were retained. On the other hand, in the peach ancestor, these syntenic regions were quickly lost or deleted, perhaps due to issues associated with an imbalance in gene dosage [53, 54].

**4.3. Selection Pressure of *bZIP* Genes and *bZIP* Domains in All Clades.** Furthermore, *Ka/Ks* ratios were estimated to detect the diversifying selection pressure on different clades (except for UN clades). The results showed that the *Ka/Ks* ratios for gene pairs in nine clades (A, B, C, D, E, F, G, H, and I) were <1, with most of them being even less than 0.6, suggesting strong purifying selection (Figure 3(a)). However, the other pairs in clade S seemed to be under positive selection, as their *Ka/Ks* ratios were >1. Also, in the phylogenetic tree of Rosaceae, we

found the biggest clade (S) containing 39 genes (21, 9, and 9 for apple, peach, and strawberry, resp.). Much interest focuses on positive selection (adaptive molecular evolution) associated with adaptation and evolution of new forms or functions in that nonsynonymous mutations offer fitness advantages to the protein [55, 56]. Zhao et al. have concluded that functional gain and divergence of transcription factors were driven by distinct positive selection on their transcription activation domains [57]. Based on the derivative data from monocot and dicot species imply that homologues of S *bZIP*s are also transcriptionally activated after stress treatment [58], such as drought, cold, and wounding, or are specifically expressed in defined parts of the flower [59, 60]. The positive selection may have contributed to the expansion of clade S to adapt to the development and environment stresses.

The *bZIP* transcription factors contain a highly conserved *bZIP* domain composed of two structural features: a basic region (N-X<sub>7</sub>-R/K-X<sub>9</sub>) for sequence-specific DNA binding and a leucine zipper composed of several heptad repeats of Leu or other bulky hydrophobic amino acids, such as Ile, Val, Phe, or Met, for dimerization specificity [5–7]. Additionally, *bZIP* domains of all clades also appeared as stronger purifying selection. A purifying selection may aid in the detection of regions or residues of functional importance [55]. These results suggested that functions of genes in major clades did not diverge much along with the genome evolution after the duplication events. Possibly because of the rapid evolution, members of the Rosaceae display remarkable phenotypic diversity, with common morphological synapomorphies not readily identifiable [23]. It is worth noting that paralogs were undergoing stronger purifying selection than orthologs in each clade except for clades C, H, and S (Figure 3(b)), which probably accelerates the process of morphological diversity, plant habit, and fruit type within the Rosaceae. From Figure 4(b), we conclude that BR domains were under stronger purifying selection than LZ domains in each clade except for clades H and I, suggesting that purifying selection was mainly responsible for *bZIP* sequence-specific DNA binding.

**4.4. Orthologous Pairs between Chromosomes.** Peach, at both the macro- and microsyntenic levels, has the most conserved karyotype in relation to the ancestral genome configuration for the Rosaceae [42]. Dirlewanger et al. [61] compared *Malus* and *Prunus* and found strong evidence that single linkage groups in the diploid *Prunus* were homologous to two distinct homologous linkage groups in the amphitetraploid genome of *Malus*. According to orthologous *bZIP* gene pairs analysis, the conserved and syntenic blocks were common to all three genomes analyzed, with a single syntenic block in peach corresponding to one or two syntenic regions in strawberry and two or four syntenic regions in apple. Vilanova et al. [62] compared the diploid reference linkage maps for *Prunus* and *Fragaria* and they identified numerous chromosomal translocations and rearrangements that occurred in the 29 million years since the genera diverged from a common ancestor. Notably, *bZIP* genes on the PC4 peach chromosome corresponded orthologously not to FC6, but rather to FC3. The data indicated that two genes (*ppa016271m* and *ppa022385m*)



located on a nonorthologous chromosome region that had originated from a common ancestor went through some intrachromosomal rearrangements. This interpretation is consistent with the fact that a greater number of small-scale rearrangements occurred in strawberry in comparison to either apple or peach [42]. Whilst an early hypothesis as to the origin of *Malus* implied wide hybridization between ancestral amygdaloid ( $x = 8$ ) and ancestral spiraeoid ( $x = 9$ ) [63], other data suggest that *Malus* may have arisen due to polyploidization of a spiraeoid species [64]. Illa et al. [23] reconstructed a hypothetical ancestral genome for the Rosaceae containing nine chromosomes ( $x = 9$ ), consistent with the report of Vilanova et al. [62]. Based on the analysis of orthologous pairs between chromosomes, we could propose a hypothesis that these orthologs became after one gene duplication located on one of the nine ancient chromosomes in the Rosaceae. An evaluation of the conservation of synteny between *Fragaria*, *Malus*, and *Prunus* based on whole genome sequence data may reveal much about sequence evolution in this closely related family.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Authors' Contribution

Xiao-Long Wang and Yan Zhong contributed equally to this work and should be considered co-first authors.

### Acknowledgments

This research was financially supported in part by Natural Science Foundation of China (Grant no. 31401849) and the Priority Academic of Jiangsu Province.

### References

- [1] M. P. Scott, "Development: the natural history of genes," *Cell*, vol. 100, no. 1, pp. 27–40, 2000.
- [2] S. B. Carroll, "Endless forms: the evolution of gene regulation and morphological diversity," *Cell*, vol. 101, no. 6, pp. 577–580, 2000.
- [3] P. Pérez-Rodríguez, D. M. Riaño-Pachón, L. G. G. Corrêa, S. A. Rensing, B. Kersten, and B. Mueller-Roeber, "PlnTFDB: updated content and new features of the plant transcription factor database," *Nucleic Acids Research*, vol. 38, supplement 1, Article ID gkp805, pp. D822–D827, 2009.
- [4] H. C. Hurst, "Transcription factors. 1: bZIP proteins," *Protein Profile*, vol. 1, no. 2, pp. 123–168, 1993.
- [5] M. Jakoby, B. Weisshaar, W. Dröge-Laser et al., "bZIP transcription factors in *Arabidopsis*," *Trends in Plant Science*, vol. 7, no. 3, pp. 106–111, 2002.
- [6] A. Nijhawan, M. Jain, A. K. Tyagi, and J. P. Khurana, "Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice," *Plant Physiology*, vol. 146, no. 2, pp. 333–350, 2008.
- [7] S. C. Lee, H. W. Choi, I. S. Hwang, D. S. Choi, and B. K. Hwang, "Functional roles of the pepper pathogen-induced bZIP transcription factor, CABZIP1, in enhanced resistance to pathogen infection and environmental stresses," *Planta*, vol. 224, no. 5, pp. 1209–1225, 2006.
- [8] O. Yang, O. V. Popova, U. Süthoff, I. Lüking, K.-J. Dietz, and D. Goldack, "The *Arabidopsis* basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance," *Gene*, vol. 436, no. 1-2, pp. 45–55, 2009.
- [9] F. Weltmeier, F. Rahmani, A. Ehlert et al., "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*, vol. 69, no. 1-2, pp. 107–119, 2009.
- [10] Y. Xiang, N. Tang, H. Du, H. Ye, and L. Xiong, "Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice," *Plant Physiology*, vol. 148, no. 4, pp. 1938–1952, 2008.
- [11] G. Lu, C. Gao, X. Zheng, and B. Han, "Identification of OsbZIP72 as a positive regulator of ABA response and drought tolerance in rice," *Planta*, vol. 229, no. 3, pp. 605–615, 2009.
- [12] H. Shimizu, K. Sato, T. Berberich et al., "LIP19, a basic region leucine zipper protein, is a Fos-like molecular switch in the cold signaling of rice plants," *Plant and Cell Physiology*, vol. 46, no. 10, pp. 1623–1634, 2005.
- [13] M. Zou, Y. Guan, H. Ren, F. Zhang, and F. Chen, "A bZIP transcription factor, OsABI5, is involved in rice fertility and stress tolerance," *Plant Molecular Biology*, vol. 66, no. 6, pp. 675–683, 2008.
- [14] M. A. Hossain, J.-I. Cho, M. Han et al., "The ABRE-binding bZIP transcription factor OsABF2 is a positive regulator of abiotic stress and ABA signaling in rice," *Journal of Plant Physiology*, vol. 167, no. 17, pp. 1512–1520, 2010.
- [15] Z. Jin, W. Xu, and A. Liu, "Genomic surveys and expression analysis of bZIP gene family in castor bean (*Ricinus communis* L.)," *Planta*, vol. 239, no. 2, pp. 299–312, 2014.
- [16] K. Wei, J. Chen, Y. Wang et al., "Genome-wide analysis of bZIP-encoding genes in maize," *DNA Research*, vol. 19, no. 6, pp. 463–476, 2012.
- [17] J. Wang, J. Zhou, B. Zhang, J. Vanitha, S. Ramachandran, and S.-Y. Jiang, "Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on sorghum," *Journal of Integrative Plant Biology*, vol. 53, no. 3, pp. 212–231, 2011.
- [18] M. C. Baloglu, V. Eldem, M. Hajyzadeh, and T. Unver, "Genome-wide analysis of the bZIP transcription factors in cucumber," *PLoS ONE*, vol. 9, no. 4, Article ID e96014, 2014.
- [19] J. Liu, N. Chen, F. Chen et al., "Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*)," *BMC Genomics*, vol. 15, no. 1, article 281, 2014.
- [20] T. Wicker, N. Yahiaoui, and B. Keller, "Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes," *The Plant Journal*, vol. 51, no. 4, pp. 631–641, 2007.
- [21] E. Dirlwanger, P. Cosson, M. Tavaud et al., "Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.)," *Theoretical and Applied Genetics*, vol. 105, no. 1, pp. 127–138, 2002.

- [22] D. Potter, T. Eriksson, R. C. Evans et al., "Phylogeny and classification of Rosaceae," *Plant Systematics and Evolution*, vol. 266, no. 1-2, pp. 5-43, 2007.
- [23] E. Illa, D. J. Sargent, E. L. Girona et al., "Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family," *BMC Evolutionary Biology*, vol. 11, article 9, 2011.
- [24] D. R. Morgan, D. E. Soltis, and K. R. Robertson, "Systematic and evolutionary implications of RBCL sequence variation in Rosaceae," *American Journal of Botany*, vol. 81, no. 7, pp. 890-903, 1994.
- [25] D. Potter, F. Gao, P. E. Bortiri, S.-H. Oh, and S. Baggett, "Phylogenetic relationships in Rosaceae inferred from chloroplast matK and trnL-trnF nucleotide sequence data," *Plant Systematics and Evolution*, vol. 231, no. 1-4, pp. 77-89, 2002.
- [26] V. Shulaev, D. J. Sargent, R. N. Crowhurst et al., "The genome of woodland strawberry (*Fragaria vesca*)," *Nature Genetics*, vol. 43, no. 2, pp. 109-116, 2011.
- [27] R. Velasco, A. Zharkikh, J. Affourtit et al., "The genome of the domesticated apple (*Malus domestica* Borkh.)," *Nature Genetics*, vol. 42, no. 10, pp. 833-839, 2010.
- [28] I. Verde, A. G. Abbott, S. Scalabrin et al., "The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution," *Nature Genetics*, vol. 45, no. 5, pp. 487-494, 2013.
- [29] D. M. Goodstein, S. Shu, R. Howson et al., "Phytozome: a comparative platform for green plant genomics," *Nucleic Acids Research*, vol. 40, no. 1, pp. D1178-D1186, 2012.
- [30] M. Punta, P. C. Coghill, R. Y. Eberhardt et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 40, no. 1, pp. D290-D301, 2012.
- [31] S. R. Eddy, "Profile hidden Markov models," *Bioinformatics*, vol. 14, no. 9, pp. 755-763, 1998.
- [32] I. Letunic, T. Doerks, and P. Bork, "SMART 7: recent updates to the protein domain annotation resource," *Nucleic Acids Research*, vol. 40, no. 1, pp. D302-D305, 2012.
- [33] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins, "The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools," *Nucleic Acids Research*, vol. 25, no. 24, pp. 4876-4882, 1997.
- [34] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596-1599, 2007.
- [35] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 406-425, 1987.
- [36] T.-H. Lee, H. Tang, X. Wang, and A. H. Paterson, "PGDD: a database of gene and genome duplication in plants," *Nucleic Acids Research*, vol. 41, no. 1, pp. D1152-D1158, 2013.
- [37] M. A. Larkin, G. Blackshields, N. P. Brown et al., "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947-2948, 2007.
- [38] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731-2739, 2011.
- [39] R. D. Finn, A. Bateman, J. Clements et al., "Pfam: the protein families database," *Nucleic Acids Research*, vol. 42, no. 1, pp. D222-D230, 2013.
- [40] W. H. Landschulz, P. F. Johnson, and S. L. McKnight, "The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins," *Science*, vol. 240, no. 4860, pp. 1759-1764, 1988.
- [41] M. Krzywinski, J. Schein, I. Birol et al., "Circos: an information aesthetic for comparative genomics," *Genome Research*, vol. 19, no. 9, pp. 1639-1645, 2009.
- [42] S. Jung, A. Cestaro, M. Troggio et al., "Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceous subfamilies," *BMC Genomics*, vol. 13, article 129, 2012.
- [43] M. A. Hossain, Y. Lee, J.-I. Cho et al., "The bZIP transcription factor OsABF1 is an ABA responsive element binding factor that enhances abiotic stress signaling in rice," *Plant Molecular Biology*, vol. 72, no. 4, pp. 557-566, 2010.
- [44] J.-Y. Kang, H.-I. Choi, M.-Y. Im, and Y. K. Soo, "Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling," *The Plant Cell*, vol. 14, no. 2, pp. 343-357, 2002.
- [45] S.-J. Oh, I. S. Sang, S. K. Youn et al., "Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth," *Plant Physiology*, vol. 138, no. 1, pp. 341-351, 2005.
- [46] H. Tak and M. Mhatre, "Cloning and molecular characterization of a putative bZIP transcription factor VvbZIP23 from *Vitis vinifera*," *Protoplasma*, vol. 250, no. 1, pp. 333-345, 2013.
- [47] Y. Uno, T. Furihata, H. Abe, R. Yoshida, K. Shinozaki, and K. Yamaguchi-Shinozaki, "Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 21, pp. 11632-11637, 2000.
- [48] J. Yu, S. Tehrim, F. Zhang et al., "Genome-wide comparative analysis of NBS-encoding genes between *Brassica* species and *Arabidopsis thaliana*," *BMC Genomics*, vol. 15, no. 1, article 3, 2014.
- [49] Y. Jiao, N. J. Wickett, S. Ayyampalayam et al., "Ancestral polyploidy in seed plants and angiosperms," *Nature*, vol. 473, no. 7345, pp. 97-100, 2011.
- [50] S. Proost, P. Pattyn, T. Gerats, and Y. Van De Peer, "Journey through the past: 150 million years of plant genome evolution," *Plant Journal*, vol. 66, no. 1, pp. 58-65, 2011.
- [51] F. Cheng, J. Wu, and X. Wang, "Genome triplication drove the diversification of *Brassica* plants," *Horticulture Research*, vol. 1, Article ID 14024, 2014.
- [52] M. Kellis, B. W. Birren, and E. S. Lander, "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*," *Nature*, vol. 428, no. 6983, pp. 617-624, 2004.
- [53] M. Freeling, "The evolutionary position of subfunctionalization, downgraded," *Genome Dynamics*, vol. 4, pp. 25-40, 2008.
- [54] M. Freeling, "Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition," *Annual Review of Plant Biology*, vol. 60, pp. 433-453, 2009.
- [55] J. Yang, Z. L. Wang, X. Q. Zhao et al., "Natural selection and adaptive evolution of leptin in the *Ochotona* family driven by the cold environmental stress," *PLoS ONE*, vol. 3, no. 1, Article ID e1472, 2008.

- [56] Z. Yang, R. Nielsen, N. Goldman, and A.-M. K. Pedersen, "Codon-substitution models for heterogeneous selection pressure at amino acid sites," *Genetics*, vol. 155, no. 1, pp. 431–449, 2000.
- [57] X. Zhao, Q. Yu, L. Huang, Q. Liu, and A. Asakura, "Patterns of positive selection of the myogenic regulatory factor gene family in vertebrates," *PLoS ONE*, vol. 9, no. 3, Article ID e92873, 2014.
- [58] T. Kusano, T. Berberich, M. Harada, N. Suzuki, and K. Sugawara, "A maize DNA-binding factor with a bZIP motif is induced by low temperature," *Molecular & General Genetics*, vol. 248, no. 5, pp. 507–517, 1995.
- [59] J. F. Martínez-García, E. Moyano, M. J. C. Alcocer, and C. Martín, "Two bZIP proteins from *Antirrhinum* flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors," *Plant Journal*, vol. 13, no. 4, pp. 489–505, 1998.
- [60] A. Strathmann, M. Kuhlmann, T. Heinekamp, and W. Dröge-Laser, "BZI-1 specifically heterodimerises with the tobacco bZIP transcription factors BZI-2, BZI-3/TBZF and BZI-4, and is functionally involved in flower development," *The Plant Journal*, vol. 28, no. 4, pp. 397–408, 2001.
- [61] E. Dirlwanger, E. Graziano, T. Joobeur et al., "Comparative mapping and marker-assisted selection in Rosaceae fruit crops," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9891–9896, 2004.
- [62] S. Vilanova, D. J. Sargent, P. Arús, and A. Monfort, "Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry)," *BMC Plant Biology*, vol. 8, no. 1, article 67, 2008.
- [63] K. Sax, "The origin of the Pomoideae," *Proceedings of the American Society for Horticultural Science*, vol. 30, pp. 147–150, 1933.
- [64] C. Sterling, "Comparative morphology of the carpel in the Rosaceae. I. Prunoideae: *Prunus*," *American Journal of Botany*, vol. 51, no. 1, pp. 36–44, 1964.