agronomy

MDPI

*Article*

# Genome-Wide Analysis of Somatic Embryogenesis-Related Transcription Factors in Cultivated Strawberry (*Fragaria* × *ananassa*) and Evolutionary Relationships among Rosaceae Species

**Adrián Garrido-Bigotes [1,*], Herman Silva [2] and Rodrigo Hasbún [1]**

[1] Laboratorio de Epigenética Vegetal, Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepción, Concepción 4030000, Chile; rodrigohasbun@udec.cl
[2] Laboratorio de Genómica Funcional & Bioinformática, Facultad de Ciencias Agronómicas, Universidad de Chile, La Pintana, Santiago 8820808, Chile; hesilva@uchile.cl
* Correspondence: adrigarrido@udec.cl

**Abstract:** Somatic embryogenesis is a plant regeneration method commonly used in tissue culture. Its molecular mechanisms are well-known in model plants such as *Arabidopsis thaliana* L. *LEAFY COTYLEDON1* (*LEC1*), *LEAFY COTYLEDON2* (*LEC2*), *FUSCA3* (*FUS3*), *ABSCISIC ACID INSENSITIVE3* (*ABI3*), and *BABYBOOM* (*BBM*) genes are considered master regulators in the induction, growth, and maturation of somatic embryos. However, the study of these transcription factors in fruit crops with high agronomic and economic value such as cultivated strawberry (*Fragaria* × *ananassa* Duch.) and other Rosaceae species is scarce. The purpose of this study was the in silico characterization of *LEC1*, *ABI3*, *FUS3*, *LEC2*, and *BBM (LAFL-B)* genes from *F.* × *ananassa* genome and the study of the evolutionary relationships within the Rosaceae family. Synteny analyses and molecular evolutionary rates were performed to analyze the evolution of each transcription factor within the Rosaceae family. Synteny was conserved between *F.* × *ananassa* and other Rosaceae genomes, and paralogous genes were selected through negative selection. Additionally, the exon–intron organization and multiple alignments showed that gene structure and DNA-binding domains were conserved in *F.* × *ananassa* transcription factors. Finally, phylogenetic trees showed close evolutionary relationships between *F.* × *ananassa* and its orthologous proteins in the Rosoideae subfamily. Overall, this research revealed novel insights in the LAFL-B network in *F.* × *ananassa* and other species of the Rosaceae family. These results provide useful in silico information and new resources for the establishment of more efficient propagation systems or the study of ploidy effects on somatic embryogenesis.

**Keywords:** LEAFY COTYLEDON1 (LEC1); LEAFY COTYLEDON2 (LEC2); ABSCISIC ACID INSENSITIVE3 (ABI3); BABY BOOM (BBM); strawberry; Rosaceae family



**Citation:** Garrido-Bigotes, A.; Silva, H.; Hasbún, R. Genome-Wide Analysis of Somatic Embryogenesis-Related Transcription Factors in Cultivated Strawberry (*Fragaria* × *ananassa*) and Evolutionary Relationships among Rosaceae Species. *Agronomy* **2021**, *11*, 356. https://doi.org/10.3390/agronomy11020356

Academic Editor: Rafael A. Cañas and Javier Canales
Received: 19 January 2021
Accepted: 11 February 2021
Published: 17 February 2021



## 1. Introduction

Plants exhibit high plasticity during their developmental stages, and it is a survival strategy to overcome environmental constraints [1]. For instance, zygotic embryogenesis is an essential process for plant reproduction and allows the regeneration of a complete plant from a single cell through the fertilization process [2]. In a similar manner, somatic embryogenesis (SE) is an alternative regeneration process, for which a new individual is generated from a vegetative cell [3,4], occurring naturally in some Kalanchoe species [5]. It can also be manipulated by in vitro tissue culture under appropriate stress signals such as plant growth regulators and wounding, as well as other mechanisms [5,6], offering multiple possibilities to investigate the bases and potential applications of SE in economically important crops. In general, SE is well-known at the morphological, physiological, and molecular level in plant model species such as *Arabidopsis thaliana* [7–9], *Medicago truncatula* [10], and others, such as *Coffea canephora* [11]. However, in fruit crops, such as

strawberry and other Rosaceae species, basic information about the SE process is scarce [12]. For example, to date, only one single study has been conducted examining the molecular aspects of the SE of different *Fragaria* × *ananassa* cultivars [13]. Others have only reported on the developmental and physiological characteristics [14–16].

Molecular mechanisms involved in SE are starting to be well-understood, and a set of transcription factors (TFs) are key regulators during early and late SE events. These participate in the induction and development of embryos from somatic cells, when they are ectopically expressed [7–9,17]. These TFs play specific roles in morphogenesis and tissue identity, accumulation of storage molecules, or acquisition of desiccation tolerance [18], among others. The most important SE-related TFs responsible for embryo formation and development from somatic cells in *A. thaliana* are: *LEAFY COTYLEDON1* (*LEC1*), *LEAFY COTYLEDON2* (*LEC2*), *FUSCA3* (*FUS3*), *ABSCISIC ACID-INSENSITIVE3* (*ABI3*), *BABY BOOM* (*BBM*), *AGAMOUS-LIKE15* (*AGL15*), *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE* (*SERK*), *RWP-RK DOMAIN-CONTAINING 4* (*RKD4*), and *WUSCHEL* (*WUS*) [19]. These genes act through coordinate gene networks, showing high levels of complexity, but only a small group constituted by *LEC1-ABI3-FUS3-LEC2* (*LAFL*) and *BBM* are considered master regulators of the SE process [17,20,21]. First, the *LEC1* gene plays a major role in early embryogenesis, although it also participates in the control of late embryogenesis. Furthermore, its ectopic expression is enough for embryo development from somatic cells [7]. A homologous gene, *LEC1-LIKE* (*L1L*), plays specific roles in the normal embryo development and can also replace *LEC1* function during SE [22]. Both genes are part of HPA3 subunits, belonging to the NF-YB subfamily of TFs. These interact with NF-YA and NF-YC subunits and allow the binding to CCAAT DNA motifs to activate the expression of SE-specific genes in *Arabidopsis thaliana* [22,23]. These TFs are characterized by the presence of CBF-A/NF-YB domain, and 16 amino acidic residues are shared between LEC1 and L1L in Arabidopsis [23]. Second, *LEC2*, *FUS3*, and *ABI3* genes are involved in different stages of induction and development of somatic embryos, participating in acquisition of cell fate, determination of cotyledon identity, and accumulation of reserve compounds in Arabidopsis [8,24,25]. These genes belong to the *LAV* (*LEC2-ABI3-VAL*) family, which is part of the B3 superfamily [26]. LEC2, FUS3, and ABI3 proteins only contain the common B3 domain necessary for DNA binding, but VAL proteins have other functional domains [26]. Third, the *BBM* gene has emerged as a master regulator of somatic embryogenesis, with primary functions in the capacity of totipotency acquisition, cell division, and SE induction [9,20]. The *BBM* gene encoding AP2/ERF TF is characterized by the presence of AP2/ERF domain involved in binding to cis-regulatory elements of DNA [27]. Recently, the *BBM* gene has been considered as a transcriptional activator of the *LAFL* gene network, promoting formation of somatic embryos in *Arabidopsis thaliana* [17]. However, it has been proposed as a putative inhibitor of SE in *F.* × *ananassa* [13].

Based on different publications, *LAFL* and *BBM* (*LAFL-B* network) are critical genes in the acquisition of embryogenic competence of different types of explants, and in the somatic embryo development [17,19,20,28]. The importance of these TFs is highlighted in vascular plants [29,30], and they have been extensively identified as part of their respective families in several crop species, such as *Gossypium hirsutum* [31], *Vitis vinifera* [32,33], *Theobroma cacao* [34], and *Prunus persica* [35]. Most of the evolutionary studies of these TFs are focused on the plant kingdom and, specifically, in vascular plants [29,30]. However, some aspects about genomic characteristics and evolution of these genes in relevant agronomic species with high ploidy, such as *F.* × *ananassa* [36] and other members of the Rosaceae family, are still unknown. The Rosaceae family is composed of Amygdaloideae (*Prunus* spp., *Malus* spp., and *Pyrus* spp.), Rosoideae (*Fragaria* spp., *Rosa* spp., and *Rubus* spp.), and Dryadoideae subfamilies [37–39]. Conventionally, *P. persica*, *Malus* × *domestica*, and *Fragariapersica vesca* are considered model species for Amygdaloideae and Rosoideae subfamilies, respectively [40]. Moreover, genomic resources are available for these species. However, genome-specific properties of each species, including octoploid species such as *F.* × *ananassa*, would provide insights for molecular characteristics of individuals.

To know the master regulator genes underlying the SE process in *F.* × *ananassa* is a key step to elucidating the molecular bases of SE growth and physiology [14–16]. In addition, the evolutionary comparison of these genes with orthologous of other members from the Rosaceae family is essential to establishing bases for understanding the embryogenic ability of this family and the discovery of new biotechnological applications in these crops. Therefore, the goal of this research was the in silico characterization of *LAFL-B* gene networking in *F.* × *ananassa* and the study of their evolutionary relationships with other species of the Rosaceae family. Results suggested that *LEC1, L1L, LEC2, FUS3, ABI3,* and *BBM* genes and proteins were present in variable copy number in *F.* × *ananassa* genome and conserved at the structural level. In addition, evolutionary relationships with *LABL-B* genes in other Rosaceae species suggest that *F.* × *ananassa* genes were closely conserved with genes of Rosoideae subfamily, including *F. vesca*, *Rubus occidentalis,* and *Rosa chinensis*. Taken together, these in silico analyses constitute the first detailed report about SE-related TFs in *F.* × *ananassa* and other Rosaceae species.

## 2. Materials and Methods

### *2.1. Identification of Somatic Embryogenesis-Related Transcription Factors in Fragaria* × *ananassa and Other Rosaceae Species*

Full-length amino acidic sequences of *Arabidopsis thaliana* LEC1, L1L, LEC2, FUS3, ABI3, and BBM were obtained from The Arabidopsis Information Resource (TAIR) database (v10, www.arabidopsis.org) and were used as queries for BLASTP search in *F.* × *ananassa* genome v1.0 (Table S1) [36] using BLAST+ tool available in the Genome Database for Rosaceae (GDR) (www.rosaceae.org) with e-value < $1e^{-3}$. Next, the isoelectric point (pI) and molecular weight (Mw) of putative protein sequences were calculated using the "Compute pI/Mw" tool (https://web.expasy.org/compute_pi/) [41]. Finally, *A. thaliana* protein sequences were also used to identify orthologous proteins by BLASTP (Table S2) in *Fragaria vesca* (v4.0), *Rubus occidentalis* (v3.0), *Rosa chinensis* (v1.0), *Malus* × *domestica* (v1.1), *Prunus persica* (v2.0), and *Pyrus communis* (v2.0) genome (GDR).

### *2.2. Characterization of Gene and Protein Structures for Somatic Embryogenesis-Related Transcription Factors in Fragaria* × *ananassa*

Coding and genomic sequences for *LEC1*, *L1Ls*, *LEC2s*, *FUS3s*, *ABI3s,* and *BBMs* of *F.* × *ananassa* and other Rosaceae genes were obtained from the GDR database. Then, exon–intron structures were analyzed using Gene Structure Display Server 2.0 (GSDS 2.0) software (http://gsds.gao-lab.org/) [42].

Domain and motif structures for LEC1s, L1Ls, LEC2s, FUS3s, ABI3s, and BBMs *F.* × *ananassa* and other Rosaceae species were determined using MEME Suite 5.1.1 (http://meme-suite.org/) [43]. Parameters to determine the protein structure were 10 motifs, with a minimum width of 5 and a maximum width of 50. The CBF-A/NF-YB, B3, and AP2/ERF domains were manually annotated from the motifs obtained. Multiple alignments of LEC1, L1Ls, LEC2s, FUS3s, ABI3s, and BBMs *F.* × *ananassa* and other orthologous Rosaceae proteins were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [44] and visualized with Jalview software [45].

### *2.3. Synteny Analysis and Molecular Evolutionary Rates between Somatic Embryogenesis-Related Transcription Factors in Fragaria* × *ananassa and Other Rosaceae Species*

Chromosomal locations and syntenic relationships between *LEC1s*, *L1Ls*, *LEC2s*, *FUS3s*, *ABI3s,* and *BBMs* of *F.* × *ananassa*, and orthologous genes of *F. vesca*, *R. occidentalis*, *R. chinensis*, *M.* × *domestica*, *P. persica*, and *P. communis* were analyzed using Synteny Viewer tool, available on GDR database, and the diagram was constructed using Basic Circos, available on TBtools software [46]. Molecular evolutionary rates for paralogous gene pairs were calculated using Simple Ka/Ks Calculator, available on TBtools software [46]. FaL1L3 and FvL1L2 coding-sequences were subjected to the identification of site-specific positive selection operating on amino acids using Selecton version 2.4 (http://selecton.tau.ac.il/index.html; [47]), performing the model M8 (positive selection enabled). We

consider that Ka/Ks > 1 and probability < 0.01 indicate amino acid residues under positive selection. FaL1L3 protein was used as reference sequence to show amino acid under positive selection.

### *2.4. Phylogenetic Analyses between Somatic Embryogenesis-Related Transcription Factors in Fragaria × ananassa and Other Rosaceae Species*

Multiple alignments of full-length proteins for LEC1s, L1Ls, LEC2s, FUS3s, ABI3s, and BBMs *F. × ananassa*, and their orthologous sequences in *F. vesca*, *R. occidentalis*, *R. chinensis*, *M. × domestica*, *P. persica*, and *P. communis* were performed using Clustal Omega [44]. Phylogenetic analyses were conducted using MEGA X [48] and Neighbor-Joining method (Jones–Taylor–Thornton substitution model and Gamma distribution) with a 1000 bootstrap analysis. Sequences of green algae *Volvox carteri* and *Chlamydomonas reindhartii* were used as outgroups for rooting of phylogenetic tree.

## 3. Results

### *3.1. Identification of LEC1, L1L, LEC2, FUS3, ABI3, and BBM Transcription Factors in Fragaria × ananassa and Six Other Rosaceae Species*

In order to identify *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3*, and *BBM* genes in *F. × ananassa*, a BLASTP search (Table S1) was performed against the first version of the genome [36], available on Genome Database for Rosaceae (GDR). For this analysis, *A. thaliana* orthologous proteins were used as queries. Moreover, the same searches were performed for identifying orthologous genes in *F. vesca* (v4.0) [49,50], *R. occidentalis* (3.0) [51], *R. chinensis* (v1.0) [52], *M. × domestica* (v1.1) [53], *P. persica* (v2.0) [54], and *P. communis* (v2.0) [55] genome (Table S2).

We identified orthologous genes in *F. × ananassa* genome (Table 1): *LEC1* (1), *L1L* (4), *LEC2* (4), *FUS3* (3), *ABI3* (3), and *BBM* (2). These genes were distributed among 12 chromosomes of *F. × ananassa* (Table 1). Loci and alleles were named according to their consecutive chromosome position (Table 1) and considering "Gene Naming Guidelines" proposed for Rosaceae family [56]. The lowest and highest gene lengths corresponded to *FaLEC1* and *FaBBM* (Table 1), respectively. The protein molecular weights varied from 21.16 to 89.80 kDa for FaLEC2.2 and FaBBM2 (Table 1). The isoelectric point (pI) provided biochemical properties for protein functions [57]; FaBBM1, and FaFUS3.3 showed pI values ranged from 6.11 to 5.35 (Table 1), respectively.

Additionally, orthologous genes were also detected in diploid *F. vesca* and other Rosaceae species. In the case of *F. vesca*, the following were identified: *LEC1* (1), *L1L* (2), *LEC2* (1), *FUS3* (1), *ABI3* (1), and *BBM* (1) loci. Two alleles were identified for *F. vesca ABI3* and *BBM* genes (Table 2 and Table S3). In the remaining Rosaceae members, a variable number of SE-related TF was identified (Table 2 and Table S3), although *ABI3* genes could not be found in *P. communis* (Table 2). *LEC1* exhibited a higher number of genes in *R. occidentalis*, *M. × domestica*, and *P. communis* compared to *F. × ananassa* and *F. vesca* (Table 2 and Table S3). *L1L*, *LEC2*, *FUS3*, and *ABI3* genes showed the highest loci number in *F. × ananassa* compared to *F. vesca* and other Rosaceae species (Table 2). In the case of *BBM*, *F. × ananassa* displayed two loci, as well as *R. chinensis*, while other species only showed one locus each (Table 2). In *P. persica*, *L1L*, *LEC2*, and *ABI3* showed two, three, and six alleles, respectively (Table 2). Regarding specific molecular characteristic for *LAFL-B* genes and proteins in other Rosaceae species, LEC1 exhibited the lower length for genes and proteins (Table S3). In contrast, the higher gene and protein lengths were observed for *R. chinensis* and *P. persica* BBM, respectively, according to higher molecular weights in these proteins (Table S3). In relation to pI, RoLEC1.3 and MdLEC2.1 presented the lower values (Table S3).

**Table 1.** Genomic data for *LEC1*, *LEC2*, *FUS3*, *ABI3*, and *BBM* genes in *Fragaria* × *ananassa*.

| Gene | Accession * | Chr | Coordinates | Gene (bp) | CDS (bp) | Protein (aa) | Molecular Weight (kDa) | pI |
|---|---|---|---|---|---|---|---|---|
| *FaLEC1* | augustus_masked-Fvb3-3-processed-gene-175.1 | 3-3 | 17523635-17524327 | 693 | 693 | 230 | 25.46 | 5.42 |
| *FaL1L1* | augustus_masked-Fvb3-1-processed-gene-317.10 | 3-1 | 31777224-31777889 | 666 | 666 | 221 | 24.33 | 5.66 |
| *FaL1L2* | maker-Fvb3-1-snap-gene-259.54 | 3-1 | 25951232-25951844 | 613 | 594 | 197 | 21.29 | 5.88 |
| *FaL1L3* | maker-Fvb3-2-snap-gene-30.79 | 3-2 | 3007308-3009361 | 2054 | 798 | 265 | 29.45 | 6.64 |
| *FaL1L4* | snap_masked-Fvb3-3-processed-gene-44.47 | 3-3 | 4402948-4403610 | 663 | 645 | 214 | 23.69 | 6.05 |
| *FaLEC2.1* | snap_masked-Fvb3-2-processed-gene-48.18 | 3-2 | 4803028-4803975 | 948 | 552 | 183 | 20.98 | 9.44 |
| *FaLEC2.2* | maker-Fvb3-4-snap-gene-246.68 | 3-4 | 24622191-24623184 | 994 | 555 | 184 | 21.16 | 9.03 |
| *FaLEC2.3* | maker-Fvb5-1-snap-gene-133.63 | 5-1 | 13346240-13348435 | 2196 | 1074 | 357 | 40.51 | 6.87 |
| *FaLEC2.4* | augustus_masked-Fvb5-3-processed-gene-156.8 | 5-3 | 15664339-15666680 | 2342 | 1227 | 408 | 45.90 | 5.82 |
| *FaFUS3.1* | maker-Fvb6-1-augustus-gene-7.51 | 6-1 | 758349-761311 | 2963 | 990 | 329 | 37.37 | 5.80 |
| *FaFUS3.2* | maker-Fvb6-2-snap-gene-275.28 | 6-2 | 27584481-27587435 | 2955 | 978 | 325 | 36.78 | 5.45 |
| *FaFUS3.3* | maker-Fvb6-3-augustus-gene-7.37 | 6-3 | 736263-739146 | 2884 | 981 | 326 | 36.96 | 5.35 |
| *FaABI3.1* | maker-Fvb2-1-snap-gene-128.30 | 2-1 | 12810779-12814037 | 3259 | 2205 | 734 | 81.60 | 5.75 |
| *FaABI3.2* | maker-Fvb2-2-augustus-gene-81.38 | 2-2 | 8128310-8131378 | 3069 | 2184 | 727 | 80.61 | 5.81 |
| *FaABI3.3* | maker-Fvb2-4-augustus-gene-151.48 | 2-4 | 15160217-15163448 | 3232 | 2220 | 739 | 82.15 | 6.00 |
| *FaBBM1* | maker-Fvb6-2-augustus-gene-270.50 | 6-2 | 27083071-27087139 | 4069 | 2478 | 825 | 89.76 | 6.11 |
| *FaBBM2* | maker-Fvb6-4-augustus-gene-24.29 | 6-4 | 2470827-2474851 | 4025 | 2478 | 825 | 89.80 | 6.10 |

* Accessions for *F.* × *ananassa* and *A. thaliana* genes were obtained from Genome Database for Rosaceae (GDR) and The Arabidopsis Information Resource (TAIR) database, respectively.

**Table 2.** Loci and allele number of somatic embryogenesis-related transcription factors identified in Rosaceae species.

| | *LEC1* | *L1L* | *LEC2* | *FUS3* | *ABI3* | *BBM* | Protein-Coding Genes | Genome Version ** |
|---|---|---|---|---|---|---|---|---|
| *Fragaria* × *ananassa* | 1 | 4 | 4 | 3 | 3 | 2 | 108,087 | v1.0 [36] |
| *Fragaria vesca* | 1 | 2 | 1 | 1 | 1 (2) * | 1 (2) * | 34,007 | v4.0 [50] |
| *Rubus occidentalis* | 4 | 1 | 1 | 1 | 1 | 1 | 34,545 | v3.0 [51] |
| *Rosa chinensis* | 1 | 1 | 2 | 1 | 1 | 2 | 36,377 | v1.0 [52] |
| *Malus* × *domestica* | 4 | 2 | 2 | 2 | 2 | 1 | 42,140 | v1.1 [53] |
| *Prunus persica* | 1 | 2 | 1 (3) * | 1 | 1 (6) * | 1 | 26,873 | v2.0 [54] |
| *Pyrus communis* | 1 | 2 | 2 | 1 | 0 | 1 | 37,445 | v2.0 [55] |

* Allele numbers are indicated in brackets. ** Genomic data were obtained from GDR database.

### *3.2. Molecular Characterization of LEC1, L1L, LEC2, FUS3, ABI3, and BBM Transcription Factors in Fragaria* × *ananassa*

To gain insights into the conservation of *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3*, and *BBM* TFs in *F.* × *ananassa* and other members of Rosaceae family, structural analyses of genes and proteins were performed (Figures 1–3). Full-length coding sequences were aligned with corresponding genomic sequences to obtain the exon–intron organization. Protein structures were analyzed through discovery of motifs and multiple alignments of DNA binding domains. Finally, *F.* × *ananassa* and other Rosaceae genes and proteins were compared with their orthologous sequences in *A. thaliana*.

The *FaLEC1* gene did not show introns, and information about its UTR regions was missing for *F.* × *ananassa* genome, while one intron and UTR sequences were detected in its *F. vesca* and *A. thaliana* orthologous sequences (Figure 1A). In a similar manner, *RoLEC1.2* showed a single intron, but *RoLEC1.3* displayed four introns (Figure 1A). Similarly, *FaLEC1*, *FaL1L1*, and *FaL1L4* were intronless, and *FaL1L2* and *FaL1L3* displayed one and two introns (Figure 1A), respectively. In the remaining Rosaceae species, almost all *L1L* orthologous genes showed one intron except *RoL1L, RcL1L, and PpL1L2* orthologous genes, all of them being intronless (Figure 1A). Regarding the structure of protein domains, all LEC1s and L1Ls sequences showed the CBF-A/NF-YB domain conserved for its amino acid sequences (Figure 1B). A clear difference in the motif content was observed between LEC1 and L1L protein sequences for the Rosaceae species (Figure 1B). In the case of LEC1 orthologous, two additional motifs were detected for *M.* × *domestica*, *P. persica*, and *P. communis* proteins. They were not present in FaLEC1, FvLEC1, and other orthologous sequences of *R. occidentalis* and *R. chinensis* (Figure 1B). Regard the L1L proteins, a higher number of motifs were detected for *F.* × *ananassa* L1L proteins and the rest of Rosaceae sequences (Figure 1B). Additionally, the CBF-A/NF-YB domain exhibited high conservation of amino acid residues for FaLEC1 and FaL1Ls and orthologous Rosaceae proteins. However, this domain was lower conserved in RoLEC1.1, RoLEC1.3, and RoLEC1.4 sequences (Figure S1). The Asp (D) residue was present for the majority of the proteins except for FaL1L2, FaL1L4, and FvL1L2, showing similar amino acid residue Glu (E) (Figure S1). In the case of RoLEC1.1, RoLEC1.3, and RoLEC1.4, this amino acid was substituted by Lys (K) or Arg (R) (Figure S1).

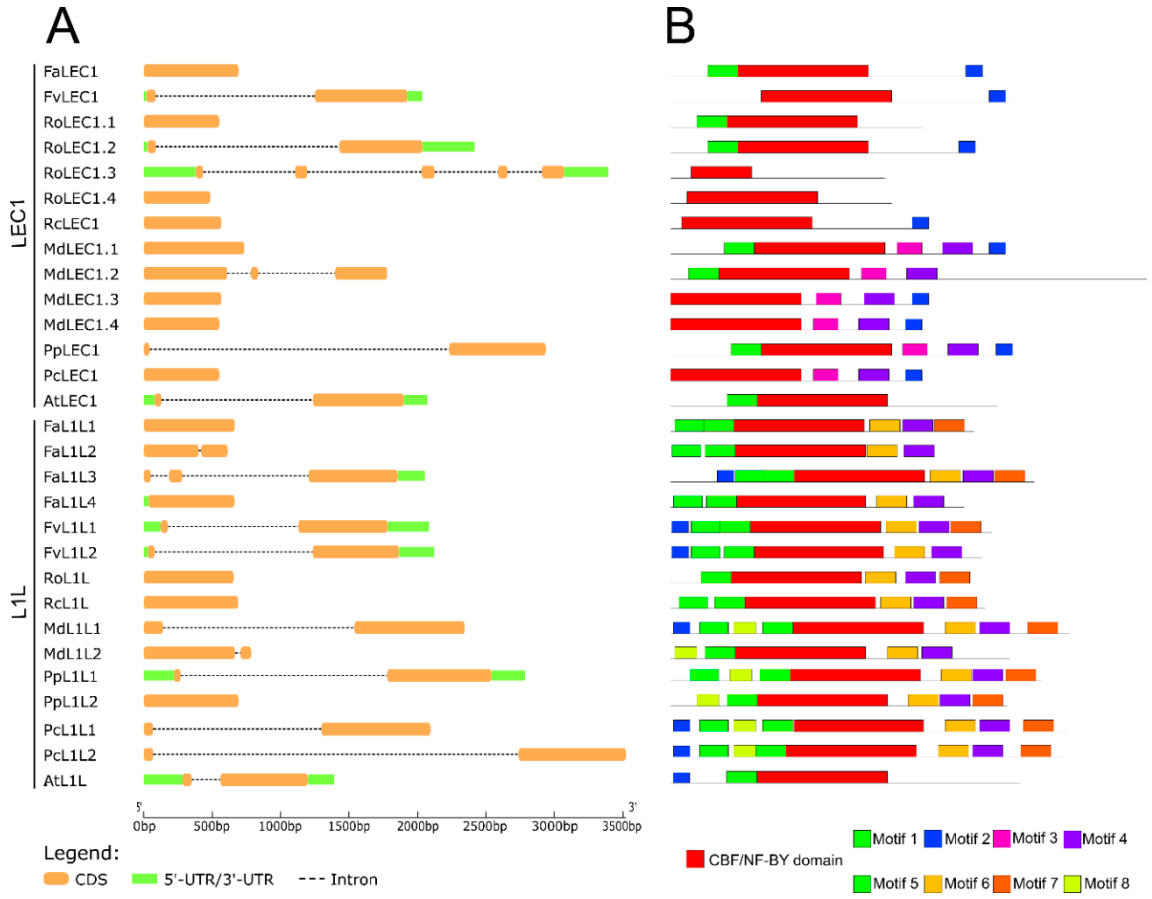


**Figure 1.** LEC1 and LIL1 gene and protein structures in Rosaceae species and *Arabidopsis thaliana*. (**A**) Exon–intron structure of *LEC1* and *L1L* genes. (**B**) CBF-A/NF-YB domain and motifs in LEC1 and L1L proteins. *LEC1*: *LEAFY COTYLEDON1*; *L1L*: *LEAFY COTYLEDON1-LIKE;* Fa: *Fragaria* × *ananassa*; Fv: *Fragaria vesca*; Ro: *Rubus occidentalis*; Rc: *Rosa chinensis*; Md: *Malus* × *domestica*; Pp: *Prunus persica*; Pc: *Pyrus communis;* At: *Arabidopsis thaliana.* A list of accession numbers for protein sequences is available in Table S3.

In the case of *AFL* genes, the exon–intron organization of *FaLEC2s*, *FaFUS3s*, and *FaABI3s* genes shared a similar number of introns (Figure 2A). *FaLEC2.1* and *FaLEC2.2* contained four introns, while *FaLEC2.3*, *FaLEC2.4*, *FaFUS3s*, and *FaABI3s* genes displayed five introns each (Figure 2A). Most of the *LEC2*, *FUS3*, and *ABI3* orthologous genes of other Rosaceae species contained five introns, although some genes, such as *FvLEC2* and *PcFUS3*, only showed two introns (Figure 2A). Others, such as *MdLEC2.1*, displayed six introns (Figure 2A). In the case of *PpLEC2* alleles, a larger 5′-UTR was detected (Figure 2A). Regarding the proteins, orthologous sequences for LEC2s, FUS3, and ABI3 contained the common B3 domain (Figure 2B). FaLEC2.2 and FaLEC2.3 only showed the conserved B3 domain, but FaLEC2.1 and FaLEC2.4 also presented an N-terminal motif (Figure 2B). MdLEC2.2 and PpLEC2 isoforms contained an additional motif (Figure 2B). In the case of FaFUS3s and FvFUS3, the previous motif (motif 2) was conserved and located in the C-terminal region. It was also shared by *R. chinensis* (Figure 2B). The remaining FUS3 sequences of the Rosaceae species showed a similar pattern of B3 domain and motifs (Figure 2B). In general, FaABI3s and orthologous sequences in other Rosaceae species, as well as *A. thaliana*, displayed a high number of motifs and the same profile of domain distribution in the protein sequence (Figure 2B). In general, the B3 domains were conserved, and differences in specific amino acid residues were observed between FaLEC2, FaFUS3, and FaABI3 sequences (Figure S2). Moreover, some Rosaceae sequences, such as RcFUS3, MdFUS3.1, PcFUS3, and MdABI3.1, exhibited additional fragments of sequences compared to the other orthologous (Figure S2).

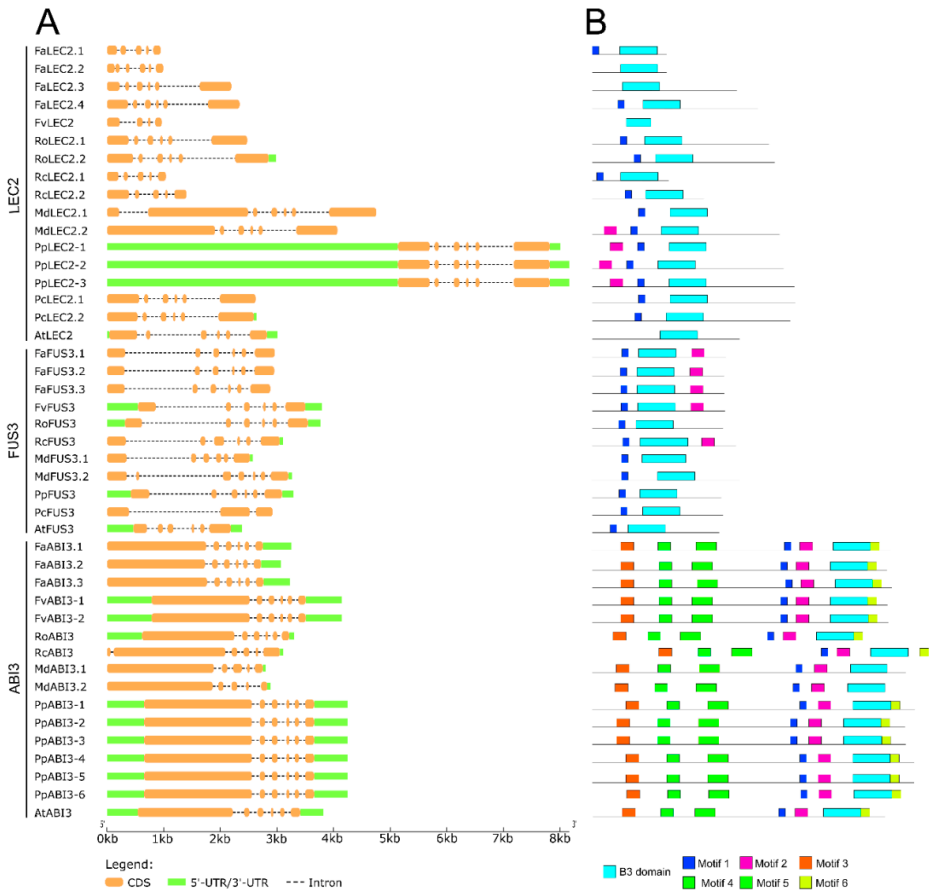


**Figure 2.** LEC2, FUS3, and ABI3 gene and protein structures in Rosaceae species and *Arabidopsis thaliana*. (**A**) Exon–intron structure of *LEC2*, *FUS3*, and *ABI3* genes. (**B**) B3 domain and motifs in LEC2, FUS3, and ABI3 proteins. *LEC2*: *LEAFY COTYLEDON2*; *FUS3*: *FUSCA3*; *ABI3*: *ABSCISIC ACID INSENSITIVE3*; Fa: *Fragaria* × *ananassa*; Fv: *Fragaria vesca*; Ro: *Rubus occidentalis*; Rc: *Rosa chinensis*; Md: *Malus* × *domestica*; Pp: *Prunus persica*; Pc: *Pyrus communis;* At: *Arabidopsis thaliana*. A list of accession numbers for protein sequences is available in Table S3.

The same number and length of introns were observed between *F.* × *ananassa* and *F. vesca BBM* genes (Figure 3A). Both loci *FaBBM1*, *FaBBM2*, and two alleles of *FvBBM* showed six introns (Figure 3A). In a similar manner, *RcBBM1* also showed six introns, but *RcBBM2* and other Rosaceae orthologous displayed seven or eight introns (Figure 3A). The 5′-UTR of *RcBBM2* showed more length than other genes (Figure 3A). In the context of protein structures, FaBMM1 and FaBBM2 proteins exhibited AP2/ERF domain and identical structural motifs for FvBBM isoforms, as well as *R. occidentalis* and *R. chinensis* orthologous sequences (Figure 3B). For *M.* × *domestica*, *P. persica*, and *P. communis*, BBM proteins only retained a single motif compared to the previously mentioned species (Figure 3B). Protein alignments for AP2/ERF domain for BBM proteins displayed conservation nearly to 100% in all Rosaceae species, and slight differences with respect to the AtBBM protein (Figure S3). The results of the molecular characterization of SE-related TFs LEC1, L1L, LEC2, FUS3, ABI3, and BBM (Figures 1–3) indicated that gene and protein structures are highly conserved at structural level in *F.* × *ananassa*, and shared similarity with other species of Rosaceae family.

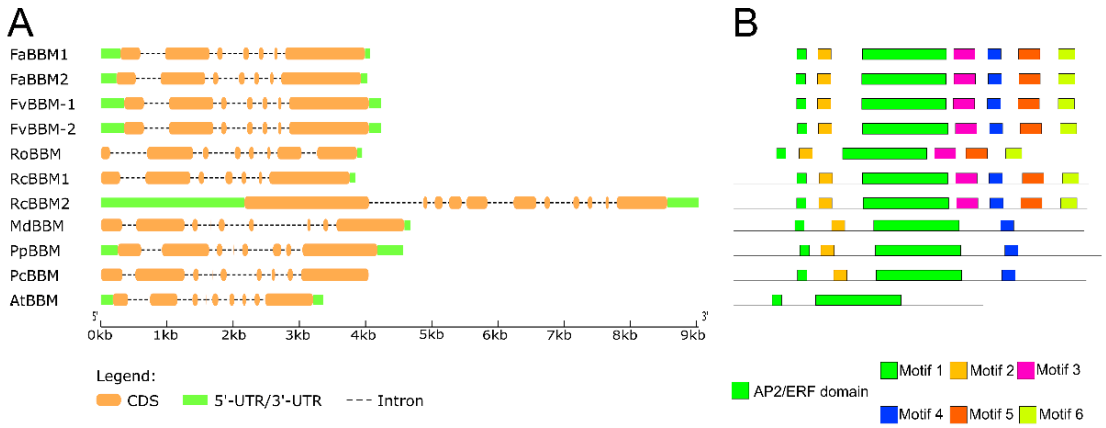


**Figure 3.** BBM gene and protein structures in Rosaceae species and *Arabidopsis thaliana*. (**A**) Exon–intron structure of *BBM* genes. (**B**) AP2/ERF domain and motifs in BBM proteins. *BBM*: *BABYBOOM*; Fa: *Fragaria* × *ananassa*; Fv: *Fragaria vesca*; Ro: *Rubus occidentalis*; Rc: *Rosa chinensis*; Md: *Malus* × *domestica*; Pp: *Prunus persica*; Pc: *Pyrus communis;* At: *Arabidopsis thaliana.* A list of accession numbers for protein sequences is available in Table S3.

### *3.3. Syntenic Relationships and Molecular Evolutionary Analysis of LEC1, L1L, LEC2, FUS3, ABI3, and BBM Genes in Fragaria* × *ananassa and Six Other Rosaceae Species*

To gain a better understanding of the evolution aspect of SE-related TFs in *F.* × *ananassa* and within the Rosaceae family, a comparison of conserved syntenic regions of *F.* × *ananassa* with *F. vesca*, *R. occidentalis*, *R. chinensis*, *M.* × *domestica*, *P. persica,* and *P. communis* chromosomes were performed. We observed that all genes corresponding to *FaLEC1*, *FaL1Ls*, *FaLEC2s*, *FaFUS3s*, *FaABI3s*, and *FaBBMs* were mapped to 12 chromosomes (Table 1 and Figure 4). In other species of the Rosaceae family, all genes except *MdLEC2.1* (unknown position) were also assigned to its chromosome positions (Figure 4). Next, synteny blocks between *F.* × *ananassa* and other genomes were detected (Table S4), and the higher number of conserved syntenic regions was observed with *M.* × *domestica* and *F. vesca* genomes (Figure 4 and Table S4). The lowest synteny was observed between *F.* × *ananassa* and *P. communis* genome (Figure 4 and Table S4). Otherwise, *FaFUS3.2* and *FaBBM1* genes showed a higher number of syntenic relationships with other Rosaceae species (Figure 4 and Table S4), while *FaLEC2.4* were assigned to a single conserved syntenic block with *FvLEC2* gene (Figure 4 and Table S4). In the case of *FaL1L2*, syntenic relationships were not detected with any species. These results indicate that most positions for *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3,* and *BBM* genes are conserved between *F.* × *ananassa* and other Rosaceae family genomes.

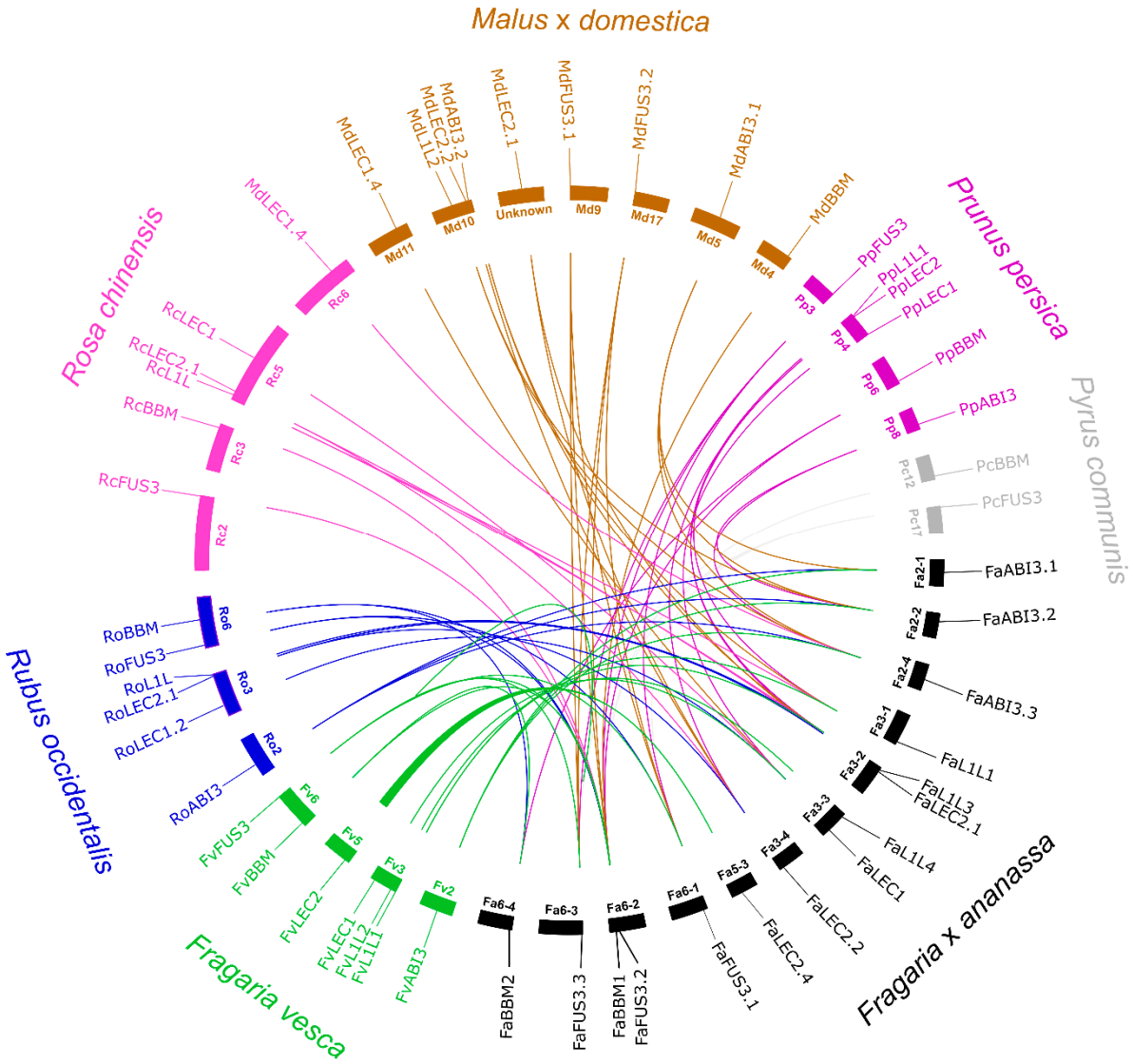


**Figure 4.** Chromosomal location and comparison of syntenic relationships between *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3*, and *BBM* genes for *Fragaria* × *ananassa* and other species of the Rosaceae family. *FaLEC2.3* did not present synteny with any genes of Rosaceae species and it was excluded from the figure. Chromosomes for each species are indicated by color rectangles. Gene names for each species are indicated outside the circle. Colored lines indicate syntenic gene pairs between *F.* × *ananassa* and other species from the Rosaceae family. *LEC1*: *LEAFY COTYLEDON1*; *L1L*: *LEAFY COTYLEDON1-LIKE*; *LEC2*: *LEAFY COTYLEDON2*; *FUS3*: *FUSCA3*; *ABI3*: *ABSCISIC ACID INSENSITIVE3*; *BBM*: *BABYBOOM*; Fa: *Fragaria* × *ananassa*; Fv: *Fragaria vesca*; Ro: *Rubus occidentalis*; Rc: *Rosa chinensis*; Md: *Malus* × *domestica*; Pp: *Prunus persica*; Pc: *Pyrus communis*. Additional information about syntenic blocks is available in Table S4.

To determine the impact of gene duplications of SE-related TFs in the Rosaceae family, molecular evolutionary rates were estimated (Table 3). Firstly, gene duplication patterns were analyzed for each pair of paralogous genes [58,59]. Secondly, paralogous genes pairs were identified considering a coverage of alignment length >70% and an identity for aligned regions >70% [60]. Thirdly, molecular evolutionary rates were calculated for paralogous of *F.* × *ananassa* and between *F.* × *ananassa*—*F. vesca* genes. A total of seven and one paralogous genes were found in *F.* × *ananassa* and *F. vesca* genomes, respectively (Table 3). In the remaining species, only one gene pair was paralogous (Table 3). Most of the intraspecies paralogous genes were generated by dispersed duplications (DSD), and only *MdLEC1.1* and *MdLEC1.2* were a consequence of proximal duplication (PD) (Table 3). The selective pressures were calculated with synonymous site (Ks), non-synonymous site (Ka), and the ratio Ka/Ks, for understanding the gene family expansion. Values of Ka/Ks ~ 0 indicated neutral

selection, while values > or <1 indicated positive and negative selection, respectively [61]. In this study, Ka/Ks values ranged between 0.237 and 1.25 (Table 3). All paralogous genes intraspecies showed Ka/Ks values <1, suggesting that these genes are under negative selection (Table 3). For genes duplicated between *F.* × *ananassa* and *F. vesca*, one gene displayed positive selection (Ka/Ks > 1), but most of the genes were under negative selection (Ka/Ks < 1). These results suggest that the negative selection promoted the removal of deleterious genes during the evolutionary history of *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3,* and *BBM* genes in *Fragaria* × *ananassa* and other species of Rosaceae family.

**Table 3.** Molecular evolutionary rates for *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3,* and *BBM* paralogous genes in *Fragaria* × *ananassa* and other species of the Rosaceae family.

| Gene A | Gene B | Duplication * | Ka | Ks | Ka/Ks | Selection Pressure |
|---|---|---|---|---|---|---|
| *FaL1L1* | *FaL1L3* | DSD | 0.026 | 0.038 | 0.684 | Negative |
| *FaL1L2* | *FaL1L3* | DSD | 0.134 | 0.231 | 0.580 | Negative |
| *FvL1L1* | *FvL1L2* | DSD | 0.043 | 0.173 | 0.248 | Negative |
| *PcL1L1* | *PcL1L2* | DSD | 0.039 | 0.155 | 0.252 | Negative |
| *FaLEC2.1* | *FaLEC2.4* | DSD | 0.130 | 0.236 | 0.551 | Negative |
| *FaLEC2.2* | *FaLEC2.3* | DSD | 0.385 | 0.629 | 0.612 | Negative |
| *RoLEC2.1* | *RoLEC2.2* | DSD | 0.338 | 0.654 | 0.517 | Negative |
| *RcLEC2.1* | *RcLEC2.2* | DSD | 0.327 | 0.802 | 0.408 | Negative |
| *MdLEC1.1* | *MdLEC1.2* | PD | 0.018 | 0.076 | 0.237 | Negative |
| *PcLEC2.1* | *PcLEC2.2* | DSD | 0.097 | 0.203 | 0.478 | Negative |
| *FaFUS3.1* | *FaFUS3.2* | DSD | 0.004 | 0.005 | 0.800 | Negative |
| *FaABI3.1* | *FaABI3.3* | DSD | 0.015 | 0.037 | 0.405 | Negative |
| *FaBBM1* | *FaBBM2* | DSD | 0.009 | 0.035 | 0.257 | Negative |
| *FaLEC1* | *FvLEC1* | | 0.063 | 0.131 | 0.481 | Negative |
| *FaL1L3* | *FvL1L1* | | 0.015 | 0.012 | 1.250 | Positive |
| *FaL1L4* | *FvL1L2* | | 0.033 | 0.063 | 0.524 | Negative |
| *FaLEC2.1* | *FvLEC2* | | 0.218 | 0.294 | 0.741 | Negative |
| *FaFUS3.1* | *FvFUS3* | | 0.003 | 0.009 | 0.333 | Negative |
| *FaABI3.2* | *FvABI3* | | 0.004 | 0.010 | 0.400 | Negative |
| *FaBBM1* | *FvBBM* | | 0.002 | 0.004 | 0.500 | Negative |

* DSD: dispersed duplication; PD: proximal duplication; Fa: *Fragaria* × *ananassa*; Fv: *Fragaria vesca*; Ro: *Rubus occidentalis*; Rc: *Rosa chinensis*; Md: *Malus* × *domestica*; Pc: *Pyrus communis*.

To gain insights into amino acid sites subjected to positive selection, the calculation of Ka/Ks values at each protein position were performed (Figure 5 and Table S5). A total of 59 specific sites showed positive selection (Ka/Ks = 1.3–1.5, Table S5) and were represented in the FaL1L3 sequence (Figure 5). The CBF-A/NF-YB domain presented 18 residues with positive selection, although most of them were part of lower conserved regions (Figure 5).

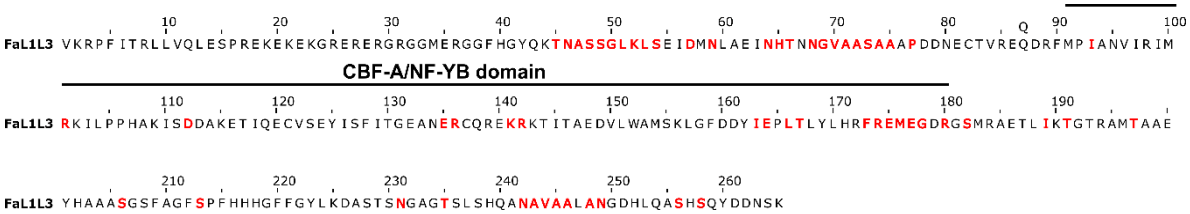


**Figure 5.** Sites under positive selection in *F.* × *ananassa* L1L3 protein used as reference sequence. Red color in letters indicates protein residues subjected to positive selection (Ka/Ks > 1). Ka: non-synonymous site; Ks: synonymous site; L1L: Leafy cotyledon1-like; Fa: *Fragaria* × *ananassa.* Additional information can be found in Table S5.

### *3.4. Phylogenetic Relationships of LEC1, L1L, LEC2, FUS3, ABI3, and BBM Transcription Factors in Fragaria × ananassa and Six Other Rosaceae Species*

To examine evolutionary relationships among LEC1, L1L, LEC2, FUS3, ABI3, and BBM TFs in *F. × ananassa*, *F. vesca*, *R. occidentalis*, *R. chinensis*, *M. × domestica*, *P. persica,* and *P. communis*, phylogenetic trees were constructed using the Neighbor-Joining method [62]. Orthologous proteins from other plants and algae were also sampled (Figure 6). A total of 38 amino acid sequences were considered to perform the LEC1 and L1L evolutionary analysis (Table S3). LEC1 and L1L were divided into two independent groups (Figure 6A), and sequences of *F. × ananassa*, *F. vesca*, *R. occidentalis*, and *R. chinensis* were clustered closely. However, all *F. × ananassa* and *F. vesca* proteins shared a common ancestor (Figure 6A). It is noteworthy that RoLEC1.1 and RoLEC1.4 were clustered closer to *C. canephora* and *C. arabica* orthologs (Figure 6A). In the case of LEC2, FUS3, and ABI3, 57 sequences were used in the construction of the phylogenetic tree (Figure 6B). A clear clustering for LEC2, FUS3, and ABI3 proteins in independent groups was obtained (Figure 6B). *F. × ananassa*, *F. vesca*, *R. occidentalis*, and *R. chinensis* sequences were grouped in the same branch and separated from *M. × domestica*, *P. persica,* and *P. communis* orthologous proteins (Figure 6B). On the other hand, LEC2.2 from *R. chinensis* and *R. occidentalis* were clustered closer to *M. × domestica*, *P. persica,* and *P. communis* species. Similarly, RcFUS3 was grouped closer to FUS3 proteins of *F. × ananassa* orthologous sequences (Figure 6B). Finally, the phylogenetic tree for BBM proteins was constructed using a total of 19 sequences (Figure 6c). FaBBM1, FaBBM2, FvBBM-1, and FvBBM-2 clustered together, close to *R. chinensis* and *R. occidentalis* orthologous proteins. MdBBM, PpBBM, and PcBBM formed an independent group (Figure 6C).

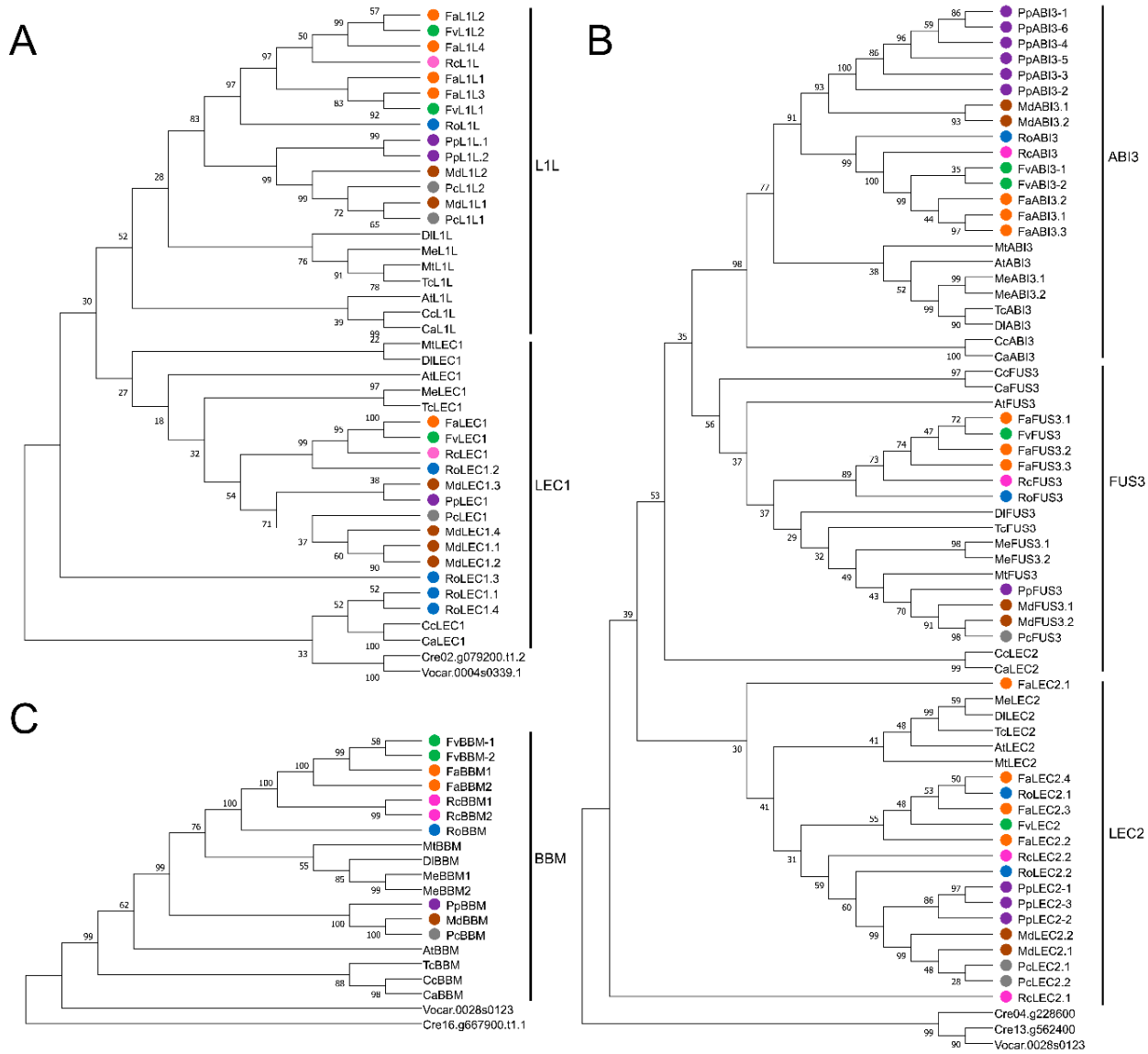


**Figure 6.** Proteins phylogenetic relationships of LAFL-B within the Rosaceae family. (**A**) Phylogenetic tree for LEC1 and, L1L. (**B**) Phylogenetic tree for LEC2, FUS3, and ABI3. (**C**) Phylogenetic tree for Babyboom (BMM) proteins. Numbers in nodes of the phylogenetic trees indicate bootstrap values (1000 replicates). Orthologous sequences of green algae *Volvox carteri* and *Chlamydomonas reindhartii* were used as outgroup. Color circles indicate proteins of different Rosaceae species. LEC1: Leafy cotyledon1; L1L: Leafy cotyledon1-like; LEC2: Leafy cotyledon2; FUS3: Fusca3; ABI3: Abscisic acid insensitive3; BBM: Babyboom; Fa: *Fragaria* × *ananassa*; Fv: *Fragaria vesca*; Ro: *Rubus occidentalis*; Rc: *Rosa chinensis*; Md: *Malus* × *domestica*; Pp: *Prunus persica*; Pc: *Pyrus communis;* At: *Arabidopsis thaliana;* Ca: *Coffea arabica;* Cc: *Coffea canephora;* Dl: *Dimocarpus longan;* Me: *Manihot esculenta;* Mt: *Medicago truncatula;* Tc: *Theobroma cacao*. A list of accession numbers for protein sequences is available in Table S3.

## 4. Discussion

Somatic embryogenesis is a developmental event through which somatic cells experience reprogramming and acquisition of embryogenic competence to form somatic embryos, and then a complete plant [4,5]. This process offers multiple opportunities for the study of molecular bases regulating the zygotic embryogenesis and development pathways in model plants, such as *A. thaliana* and *M. truncatula* [6,10], and even in economically important crops, such as *Coffea canephora* [11,63]. In the Rosaceae family, most studies related to SE have been focused on the establishment of culture conditions for somatic embryo development, based on other studies on *Rubus*, *Rosa*, *Malus*, *Prunus*, and *Pyrus*

genus [64–68] and *F.* × *ananassa* [14–16]. However, the study of molecular bases in these agronomic species has not been addressed in depth [13,69,70].

### *4.1. Fragaria* × *ananassa Genome Contains a Variable Loci Number of SE-related TFs*

Molecular networks of SE are known to be controlled by TFs acting as master regulators of the initiation and development of embryogenic program in *A. thaliana*, *M. truncatula*, and *Coffea canephora* [7–9,20,21], among others. This information will increase and can add to basic knowledge and eventually play a key role in the improvement of cultivated strawberry. Numerous SE-related TFs have been reported, but only six genes, including *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3*, and *BBM*, are major regulators of the cell totipotency, establishment, growth, and maturation of somatic embryos [7–9,17] offering significant biotechnological applications [20,21]. *F.* × *ananassa* is an allopolyploid species (2n = 8x = 56), i.e., its genome is constituted by multiple diploid genomes, including the *F. vesca* subgenome (2n = 2x = 14) [36]. Consequently, *F.* × *ananassa* could contribute with a significant and broader repertoire number of TFs than *F. vesca.* Moreover, it could be useful for the understanding of molecular mechanisms underlying the SE process in different species [36], due to the fact that *F.* × *ananassa* showed the highest loci number for *L1L*, *LEC2*, *FUS3,* and *ABI3* genes (Tables 1 and 2). In a similar way, the doubled-haploid genome of *M.* x *domestica* [53] also contained a higher loci number for SE-related genes (Table 2 and Table S3), due to a higher number of protein-coding genes [36,53]. The polyploidy generated by whole-genome duplications (WGD) is an important driving force in the evolution of plant genomes [71]. These loci numbers could be a result of the polyploidy generated by the *F.* × *ananassa* genome, compared to the diploid genome of *F. vesca* [49]. Surprisingly, some species with diploid genome, such as *R. occidentalis* [51], exhibited a high number of LEC1 loci (Table 2 and Table S3). For instance, this could be the cause of the expansion of loci numbers for SE-related TFs in *F.* × *ananassa*. However, not all diversity within TF families is explained by WGD [72]. It is worthwhile to consider the tandem, transposon-mediated, segmental duplication, and retroduplication as additional mechanisms of gene duplication [73], which could be acting in *F.* × *ananassa* and other Rosaceae species. Otherwise, the loss of duplicated genes is a natural process occurring in plant genomes and can be the result of WGD or loss-of-function mutations [73]. In the case of *LEC1*, four loci were detected in *Rubus* against one locus in the *Rosa* and *Fragaria* genome (Table 2). This could be related with the early *Rubus* divergency from the lineage of *Rosa* and *Fragaria* genus [74]. Interestingly, *ABI3* genes seem to have been subjected both to gain and to loss of duplicated genes in the Amygdaloideae subfamily. For instance, the *ABI3* gene was not found in *P. communis*, while two copies were detected in *M.* × *domestica*, compared to one copy found in *P. persica* (Table 2). This fact suggest that these events have occurred after that *Prunus* diverged from a common ancestor of *Malus* and *Pyrus* [74,75].

### *4.2. Genes and Proteins of LAFL-B Network Are Conserved in Fragaria* × *ananassa and Other Rosaceae Species*

*LEC1* and *L1L* are essential genes responsible for embryo identity in early phases and the development and maturation of embryos [7,22]. Furthermore, these genes are central regulators in the establishment of the developmental program of somatic embryo from vegetative cells [7,22]. In this study, a single *LEC1* locus in octoploid *F.* × *ananassa* and diploid *F. vesca* genome was identified (Table 1). This has been also reported for *A. thaliana*, *M. truncatula*, *Manihot esculenta,* and *V. vinifera* [7,22,33,76], as well as in some Rosaceae species, such as *R. chinensis* and *P. persica* (Table 2, [35]). In the case of *L1L*, four and two loci were found in the octoploid and diploid genome of *F.* × *ananassa* and *F. vesca*, respectively (Tables 1 and 2). Most species contain one or two L1L loci [33,35,76], suggesting that the ploidy of *F.* × *ananassa* [36,77] could be related with the expansion of this gene family. The increase in the number of gene copies in TF families is, in part, a result of WGD events. It is necessary for the emergence of TFs with specific roles in adaptative traits [72]. However, it is noteworthy that the higher ploidy levels promote the efficiency decline of plant in vitro regeneration in some species [78,79]. In general, *LEC1* locus and *L1Ls* loci of *F.* × *ananassa*

displayed similar lengths to their respective *F. vesca* paralogous (Table 2), according to the predominance of the *F. vesca* subgenome as part of the *F.* × *ananassa* genome [36]. Regarding the intron content, some genes, such as *FaLEC1*, *FaL1L1*, and *FaLEC4*, did not show introns in their genomic sequences (Figure 1A). This is similar to reports for *NF-YB4* and *NF-YB8* in *V. vinifera* [33]. Otherwise, the intron content and length in *FaL1L3* locus was different from its *F. vesca* orthologous and paralogous genes (Table 3), respectively. These differences in the absence, presence, or variation of introns is a common characteristic in plant genomes, related to its evolution and the expansion of the functional diversity in proteins [80–82]. Specifically, introns can increase the number of protein isoforms when splicing sites are present [80]. Otherwise, the functionality of LEC1 and L1L TFs is given by the CBF-A/NF-YB domain. This allows the interaction with NF-YA and NF-YC subunits to DNA-binding and regulation of the gene transcription [22,23]. Also, this domain is fully conserved at the amino acid sequence in *F.* × *ananassa* and other Rosaceae species (Figure S1). The Asp (D) residue is considered the key amino acid for the activity of these proteins, in contrast to what has been observed in other proteins of the family NF-YB in Arabidopsis [23]. In general, this residue was conserved, although some proteins, such as FaL1L2, FaL1L4, and FvL1L2, showed the Glu (E) amino acid in the same sequence position (Figure S1). This characteristic was also reported for its orthologous sequences in *V. vinifera* [33]. It is noteworthy that *F.* × *ananassa* LEC1 and L1L and orthologous proteins in other Rosaceae species exhibited similar high distribution of exclusive protein motifs compared to *A. thaliana* sequences (Figure 1B).

*LEC2*, *FUS3*, and *ABI3* genes belong to the LAV family, forming part of the B3 superfamily [26]. These are involved in cell totipotency, embryo identity, and storing reserve compounds during early and late embryogenesis [8,24,25]. *F.* × *ananassa* displayed four *LEC2*, three *FUS3*, and three *ABI3* loci (Table 1), showing the highest loci number compared to *F. vesca* and other Rosaceae species (Table 2). In some species, such as *M.* × *domestica* (Table 2) or *M. esculenta* [76], a maximum number of two loci for each gene was detected at the genome level. In some instances, WGD may be the result of the loss of duplicated genes [83]. In contrast, the higher loci number in *F.* × *ananassa* could be a consequence of a greater chromosome number and the genetic redundancy could be a result of the accumulation of non-deletional mutations through different WGD events [73]. *P. persica LEC2* and *ABI3* were detected with three and six alleles, respectively (Table 2 and Table S3). This indicated that these genes could contribute with different proteins isoforms for fine-tuning of phenotypic responses, similar to observations for some TFs in other species [84]. In general, genes showed conservation of number, length, and exon–intron structure between *F.* × *ananassa* and *A. thaliana* genes. However, slight differences in *FaLECs* and *FvLEC2* with *AtLEC2* were detected (Figure 1A). First, *FaLEC2.3* and *FaLEC2*.4 showed a longer 3′-end intron compared to *A. thaliana* (Figure 1A). Second, *FaLEC2s* contained five introns compared to three introns in *FvLEC2* or *PcFUS3* (Figure 1A). These differences in the organization of genes would have implications for the regulation of gene expression and the generation capacity of splicing protein variants among species [80,81]. Regarding conservation of protein structures, FaLEC2 only showed the conserved B3 domain, similar to those observed in other members of the Rosaceae family (Figure 2B) and *V. vinifera* orthologous sequences [32]. In contrast, FUS3 and ABI3 proteins contained a higher number of conserved motifs in *F.* × *ananassa* and other Rosaceae species. Some were not present in *A. thaliana* proteins (Figure 2B). The acquisition of new genes, including TFs, with conserved sequences could occur by genome duplication from ancestor species [72,73]. On the other hand, amino acid residues defining the function of LEC2 against FUS3 and ABI3 proteins displayed conservation in FaLEC2 proteins and their orthologous in Rosaceae species (Figure S2), according to previous reports regarding *M. esculenta* and *T. cacao* [34,76].

The *BBM* gene was initially considered as an auxiliary gene in the SE process that increased the number of somatic embryos when overexpressed [9]. However, recently, Hortsman et al. [17] reported that the *BBM* TF is responsible for the expression activation of *LEC1*, *LEC2*, *FUS3*, and *ABI3* genes in *A. thaliana*. A total of two *BBM* loci were

identified in *F.* × *ananassa* (Table 1), similar to *R. chinensis* (Table 2) and *Rosa canina* [85]. In the case of *F.* × *ananassa* genes, the chromosome number was not related to the *BBM* loci number (Table 1). It could be related to losses of duplicated genes during genome evolution [83]. Regarding the gene structure, *FaBBM1* and *FaBBM2* genes exhibited similar length of introns compared to *F. vesca* and other Rosaceae species. However, the number of introns was lower than the *A. thaliana* orthologous gene (Figure 3A). Although the *BBM* genes conserved intron structures in the Rosaceae species, its impaired molecular function proposed for the SE process in *F.* × *ananassa* [13] or its function in the shoot regeneration in *R. canina* [85] appeared to be contrary to that traditionally reported in other species, such as *A. thaliana* [19] or *C. canephora* [11]. Otherwise, clear differences were observed in *FaBBMs* and other *BBM* orthologous proteins (Figure 3B) of Rosoideae against Amygdaloideae subfamily [37,38]. For instance, *BBM* proteins in *Fragaria*, *Rubus*, and *Rosa* genus contained additional conserved motifs compared to their orthologous in *Malus*, *Prunus*, and *Pyrus* genus, suggesting that these proteins evolved independently in these subgroups of the Rosaceae family [40,86]. The AP2/ERF domain is used for DNA-binding [27] and displayed small differences in the amino acid sequence with *A. thaliana* (Figure S3), according to those previously reported in other species such as *R. canina, M. truncatula*, *C. arabica*, and *Glycine max* [85,87]. The *FaBBM1* gene reported in this study is the same as the *BBM* gene previously identified in the cultivar "Benihopp" by Gao et al. [13]. Furthermore, it could act as a putative inhibitor of SE because of its lower expression levels in embryogenic callus than in non-embryogenic callus and somatic embryos [13]. In contrast, the expression of the *BBM* orthologous gene promotes the development of somatic embryos in other species such as *Saccharum officinarum*, *C. canephora*, and *C. arabica* [11,87,88]. Therefore, although *BBM* genes of *F.* × *ananassa* and other Rosaceae species were similar to its orthologous gene in other species, such as the model plant *A. thaliana*, the molecular function of this SE-related TF could depend on the species or even a particular plant lineage.

### *4.3. LAFL-B Genes and Proteins of Rosoideae Evolved Independently of Amygdaloideae Subfamily*

Most of the evolutionary approaches consider the study of loci number, gene duplications, synteny analyses, calculation of molecular evolutionary rates of genes, and phylogenetic relationships between proteins. Firstly, all Rosaceae species showed *LAFL* and *BBM* genes in their genomes, except for the *ABI3* gene that was not found in *P. communis*, and loci, and allele number were different between species (Table 2). *F.* × *ananassa* holds a polyploid genome constituted by 56 chromosomes [36] and in general, contains a higher number of SE-related genes than diploid *F. vesca* and other Rosaceae genomes (Tables 1 and 2). Similarly, the hybrid *M.* × *domestica* also have a high number of chromosomes [53] and showed a high number of *LEC1* genes (Table 2). Moreover, a number of *LAFL-B* genes appeared to be based on the number of protein-coding genes compared to what was observed among *F.* × *ananassa* and *M.* × *domestica* versus *F. vesca* and *P. persica* (Table 2). Secondly, gene duplications contribute to expansion of gene families, and they are grouped into four types [58]. These include the tandem duplications (TD), where two genes are adjacent in the same chromosome, and the proximal duplications (PD), where two genes are in the same chromosome and separated by a few genes. Transposed duplications (TRD), DNA, or RNA-based molecular mechanisms generate two gene copies distantly; and dispersed duplications (DSD) generate two copies of genes, which are not close nor colinear. In the case of *LAFL-B* genes belonging to the Rosaceae family species, the majority have been generated by DSD through the genome (Table 3). These are a result of the polyploidization and are the type of duplication more prevalent in the Rosaceae family [58]. Thirdly, regarding the interspecies synteny of the *LAFL-B* genes in Rosaceae family, a greater number of syntenic genes were observed between *F.* × *ananassa* genome compared to *M.* × *domestica* and *F. vesca* (Figure 4). This could be related to the ploidy level [89] and the existence of a common ancestor [36,49], respectively. A weak synteny was detected between *F.* × *ananassa* and *P. communis* (Figure 4), reflecting greater evolutionary distances [38,40]. Fourthly, the estimation of the relationship between the

number of non-synonymous and synonymous substitutions (Ka/Ks) inform the type of selective pressures on gene sequences [61]. In the case of the *LAFL-B* genes, this selection pressure was negative for paralogous pairs of *F.* × *ananassa* and other genomes (Table 3). Otherwise, the duplication of paralogous genes from *F.* × *ananassa* and *F. vesca* ancestor was similar, and these genes were selected by negative selection. The predominance of negative selection for SE-related genes is a decreasing mechanism of genetic diversity [90], which could have an adverse impact on the efficiency of SE in these species. However, *FaL1L3* and *FvL1L1* were under positive selection (Table 3), showing 59 amino acid residues subjected to a positive selective pressure, mostly out of the CBF-A/NF-YB domain (Figure 5). These facts indicate that *FaL1L3* and its paralogous sequence in *F. vesca* contain beneficial mutations [61], suggesting diversification and functional adaptation of these genes from a common ancestor [91]. Fifthly, phylogenetic trees display evolutionary relationships between protein sequences for different species (Figure 6). On the other hand, LEC1, L1L, FUS3, ABI3, and *BBM* proteins of *F.* × *ananassa* were clustered in the same group of *F. vesca*, *R. chinensis*, and *R. occidentalis.* However, RcLEC2.2 and RoLEC2.2 were grouped closer to Amygdaloideae species. These are members of Rosoideae subfamily, while *P. persica, M.* × *domestica,* and *P. communis* belong to Amygdaloideae, and showed lower similarity (Figure 6). Overall, the *LAFL-B* gene network in *F.* × *ananassa* contained the higher loci number within the Rosaceae family; is more closely related to *F. vesca* and other Rosoideae subfamily species; was generated by dispersed duplications; and was under negative selection.

The Rosaceae family is composed of Dryadoideae, Rosoideae, and Amygdaloideae subfamilies [39]. Rosoideae includes *Fragaria*, *Rosa*, and *Rubus* genus, *F.* × *ananassa* and *F. vesca* being evolutionarily closer to *R. chinensis* than R. *occidentalis* [74]. On the other hand, *Prunus*, *Malus* and *Pyrus* are members of the Amygdaloideae subfamily [38,75]. *M.* × *domestica* and *Pyrus communis* share a common ancestor, while *P. persica* belongs to a different clade within this subfamily [38,75]. In this sense, our results about *LAFL-B* gene families exhibited clear relationships according to the evolutionary history of the Rosaceae family [38,74,75]. Taking into account the diversity of loci number in SE-related gene families (Table 2), multiple duplication events occurring in the two families and in specific genus [38,75] could have triggered the gain or loss of gene number in a specific lineage. Additionally, *F.* × *ananassa*, *F. vesca,* and R. *chinensis* showed closer relationships at structural level of genes and proteins than *R. occidentalis*. Moreover, Rosoideae species presented higher differences with orthologous genes of Amygdaloideae species (Figures 1–4 and Figure 6) according to higher evolutionary distance [38,74,75]. In a similar manner, genes and proteins shared similar structural properties within Amygdaloideae lineage. In the case of *P. persica,* the divergency from the clade of *M.* × *domestica* and *P. communis* [38,75] could be the event that determined higher differences between genes and proteins of LAFL-B (Figures 1–4 and Figure 6). Regarding syntenic relationships, the results suggest that most of the genes related to the SE process conserve their position in each genome (Figure 4). Lower phylogenetic distances determine a high number of syntenic regions, according to that observed in Fragaria genomes [36,49]. Furthermore, syntenic genes of Rosaceae may be present in a common ancestor of two larger families (Figure 4). Finally, the positive selection of genes suggested that paralogous genes *FaL1L3* and *FvL1L1* contain beneficial mutations (Figure 5), contributing to some functional advantages in the early SE [22]. However, although many genes duplicated are retained in the genome, not all genes are functionals [73], making necessary other additional epigenomic, transcriptomic or proteomic studies.

## 5. Conclusions

*F.* × *ananassa* is an important crop belonging to the Rosaceae family. Genomic information is available to address the molecular basis for the SE process in a polyploid species that has not been explored in depth. In this study, the presence of *LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3,* and *BBM* genes of *F.* × *ananassa* and other Rosaceae genomes provides insights into

the TFs that would act as regulators of the SE process. In general, genes and proteins of the LAFL-B network showed conservation at a structural level, through gene structure and DNA-binding domains of proteins in each TF family. In addition, evolutionary analyses indicated that *F.* × *ananassa* contained the highest loci number for *L1L*, *LEC2*, *FUS3*, *ABI3*, and *BBM* with respect to other Rosaceae species. Furthermore, it may contribute to a wider range of targets for the establishment of in vitro regeneration systems of *F.* × *ananassa* than *F. vesca*. Regarding the evolutionary history of these TFs, interspecies synteny analyses displayed a greater number of synteny blocks for *LAFL-B* genes among *F.* × *ananassa*, *F. vesca*, and *M.* × *domestica* genomes. The lower molecular evolutionary rates indicated that the negative selection was predominant in genes for *F.* × *ananassa* and other Rosaceae species. Finally, phylogenetic analyses showed that LAFL-B TFs were most closely related to its orthologous proteins of Rosoideae, compared to the Amygdaloideae subfamily.

Globally, the knowledge about SE-related TFs involved in the induction, development, and maturation of somatic embryos offers genomic targets for obtaining new *F.* × *ananassa* varieties with better and more efficient characteristics for clonal propagation. Moreover, new biotechnological approaches incorporating omics techniques could be used for the subsequent breeding of this species, considering that SE is a versatile regeneration system allowing the transformation and generation of new plants for cultivated strawberry with interesting and improved agronomic traits. However, future studies need to be directed to reveal how the dynamics of these TFs work during the SE and how the ploidy affects this process in *F.* × *ananassa.* These studies will be important for improved strawberry propagation.

**Supplementary Materials:** The following are available online at https://www.mdpi.com/2073-4395/11/2/356/s1. **Table S1.** Results of BLASTP search in *Fragaria* × *ananassa* genome, using *Arabidopsis thaliana* proteins as queries. **Table S2.** Results of BLASTP search in Fragaria vesca, Rubus occidentalis, Rosa chinensis, Malus × domestica, Prunus persica, and Pyrus communis genome, using A. thaliana proteins as queries. **Table S3.** Genomic data for LEC1, LEC2, FUS3, and ABI3 genes in Fragaria vesca, Rubus occidentalis, Rosa chinensis, Malus × domestica, Prunus persica, Pyrus communis, and Arabidopsis thaliana. **Table S4.** Synteny blocks between *Fragaria* × *ananassa LEC1*, *L1L*, *LEC2*, *FUS3*, *ABI3*, and *BBM* genes with their orthologous genes in the Rosaceae family. **Table S5.** Codon Ka/Ks values for *Fragaria* × *ananassa* L1L3 protein sequence obtained using the Selecton software. **Figure S1.** Multiple alignment of LEC1 and L1L protein sequences of *Fragaria* × *ananassa*, *F. vesca*, and *Arabidopsis thaliana*. Inverted triangles in the multiple alignments indicate amino acid residues described for the functionality of LEC1 and L1L proteins. *LEC1*, *LEAFY COTYLEDON1*; *L1L*, *LEAFY COTYLEDON1-LIKE*. Fa, *Fragaria* × *ananassa*; Fv, *Fragaria vesca*; *At, Arabidopsis thaliana.* A list of Accession numbers for protein sequences is available in Table S3. **Figure S2.** Multiple alignment of LEC2, FUS3, and ABI3 protein sequences of *Fragaria* × *ananassa*, *F. vesca*, and *Arabidopsis thaliana*. Inverted triangles in the multiple alignments indicate amino acid residues described for the functionality of LEC2 functionality compared to FUS3 and ABI3 proteins. LEC2, LEAFY COTYLEDON2; FUS3, FUSCA3; ABI3, ABSCISIC ACID INSENSITIVE3. Fa, *Fragaria* × *ananassa*; Fv, *Fragaria vesca*; At, *Arabidopsis thaliana*. A list of Accession numbers for protein sequences is available in Table S3. **Figure S3.** Multiple alignment of BBM protein sequences in *Fragaria* × *ananassa*, *F. vesca*, and *Arabidopsis thaliana*. AP2/ERF, APETALA2/ETHYLENE RESPONSIVE FACTOR; BBM, BABYBOOM. Fa, *Fragaria* × *ananassa*; Fv, *Fragaria vesca*; At, *Arabidopsis thaliana*. A list of Accession numbers for protein sequences is available in Table S3.

**Author Contributions:** Conceptualization: A.G.-B.; methodology: A.G.-B.; validation: A.G.-B., H.S., and R.H.; formal analysis: A.G.-B., H.S., and R.H.; investigation: A.G.-B., H.S., and R.H.; resources: A.G.-B. and R.H.; data curation: A.G.-B.; writing—original draft preparation: A.G.-B.; writing—review and editing: A.G.-B., H.S., and R.H.; visualization: A.G.-B., H.S., and R.H.; supervision: A.G.-B. and R.H.; project administration: A.G.-B.; funding acquisition: A.G.-B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Research and Development Agency (ANID, Chile) grant FONDECYT/Postdoctorado 3190894 to A.G.-B.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References


1. Ahanger, M.A.; Akram, N.A.; Ashraf, M.; Alyemeni, M.N.; Wijaya, L.; Ahmad, P. Plant Responses to Environmental Stresses—from Gene to Biotechnology. *AoB Plants* **2017**, *9*. [CrossRef]
2. Radoeva, T.; Vaddepalli, P.; Zhang, Z.; Weijers, D. Evolution, Initiation, and Diversity in Early Plant Embryogenesis. *Dev. Cell* **2019**, *50*, 533–543. [CrossRef] [PubMed]
3. Bidabadi, S.S.; Jain, S.M. Cellular, Molecular, and Physiological Aspects of In Vitro Plant Regeneration. *Plants* **2020**, *9*, 702. [CrossRef] [PubMed]
4. Fehér, A. Callus, Dedifferentiation, Totipotency, Somatic Embryogenesis: What These Terms Mean in the Era of Molecular Plant Biology? *Front. Plant Sci.* **2019**, *10*. [CrossRef] [PubMed]
5. Pasternak, T.; Dudits, D. Epigenetic Clues to Better Understanding of the Asexual Embryogenesis in Planta and in Vitro. *Front. Plant Sci.* **2019**, *10*. [CrossRef]
6. Fehér, A. Somatic Embryogenesis—Stress-Induced Remodeling of Plant Cell Fate. *Biochimica et Biophysica Acta* **2015**. [CrossRef]
7. Lotan, T.; Ohto, M.; Yee, K.M.; West, M.A.L.; Lo, R.; Kwong, R.W.; Yamagishi, K.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Arabidopsis *LEAFY COTYLEDON1* Is Sufficient to Induce Embryo Development in Vegetative Cells. *Cell* **1998**, *93*, 1195–1205. [CrossRef]
8. Stone, S.L.; Braybrook, S.A.; Paula, S.L.; Kwong, L.W.; Meuser, J.; Pelletier, J.; Hsieh, T.-F.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Arabidopsis *LEAFY COTYLEDON2* Induces Maturation Traits and Auxin Activity: Implications for Somatic Embryogenesis. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3151–3156. [CrossRef]
9. Boutilier, K.; Offringa, R.; Sharma, V.K.; Kieft, H.; Ouellet, T.; Zhang, L.; Hattori, J.; Liu, C.-M.; van Lammeren, A.A.M.; Miki, B.L.A.; et al. Ectopic Expression of *BABY BOOM* Triggers a Conversion from Vegetative to Embryonic Growth. *Plant Cell* **2002**, *14*, 1737–1749. [CrossRef] [PubMed]
10. Rose, R.J. Somatic Embryogenesis in the Medicago Truncatula Model: Cellular and Molecular Mechanisms. *Front. Plant Sci.* **2019**, *10*. [CrossRef] [PubMed]
11. Nic-Can, G.I.; López-Torres, A.; Barredo-Pool, F.; Wrobel, K.; Loyola-Vargas, V.M.; Rojas-Herrera, R.; De-la-Peña, C. New Insights into Somatic Embryogenesis: *LEAFY COTYLEDON1, BABY BOOM1* and *WUSCHEL-RELATED HOMEOBOX4* Are Epigenetically Regulated in *Coffea Canephora*. *PLoS ONE* **2013**, *8*, e72160. [CrossRef] [PubMed]
12. Carrasco, B.; Meisel, L.; Gebauer, M.; Garcia-Gonzales, R.; Silva, H. Breeding in Peach, Cherry and Plum: From a Tissue Culture, Genetic, Transcriptomic and Genomic Perspective. *Biol. Res.* **2013**, *46*, 219–230. [CrossRef]
13. Gao, L.M.; Zhang, J.; Hou, Y.; Yao, Y.C.; Ji, Q.L. RNA-Seq Screening of Differentially-Expressed Genes during Somatic Embryogenesis in *Fragaria* × *ananassa* Duch. 'Benihopp'. *J. Hort. Sci. Biotech.* **2015**, *90*, 671–681. [CrossRef]
14. Biswas, M.K.; Islam, R.; Hossain, M. Somatic Embryogenesis in Strawberry (Fragaria Sp.) through Callus Culture. *Plant Cell Tiss. Org.* **2007**, *90*, 49–54. [CrossRef]
15. Husaini, A.M.; Abdin, M.Z. Interactive Effect of Light, Temperature and TDZ on the Regeneration Potential of Leaf Discs of *Fragaria* × *Ananassa* Duch. *In Vitro Cell. Dev. Biol. Plant* **2007**, *43*, 576–584. [CrossRef]
16. Gerdakaneh, M.; Mozafari, A.-A.; Sioseh-mardah, A.; Sarabi, B. Effects of Different Amino Acids on Somatic Embryogenesis of Strawberry (*Fragaria* × *ananassa* Duch.). *Acta Physiologiae Plantarum* **2011**, *33*, 1847–1852. [CrossRef]
17. Horstman, A.; Li, M.; Heidmann, I.; Weemen, M.; Chen, B.; Muino, J.M.; Angenent, G.C.; Boutilier, K. The BABY BOOM Transcription Factor Activates the *LEC1-ABI3-FUS3-LEC2* Network to Induce Somatic Embryogenesis. *Plant Physiol.* **2017**, *175*, 848–857. [CrossRef]
18. Jia, H.; McCarty, D.R.; Suzuki, M. Distinct Roles of LAFL Network Genes in Promoting the Embryonic Seedling Fate in the Absence of VAL Repression. *Plant Physiol.* **2013**, *163*, 1293–1305. [CrossRef]
19. Horstman, A.; Bemer, M.; Boutilier, K. A Transcriptional View on Somatic Embryogenesis. *Regeneration* **2017**, *4*, 201–216. [CrossRef] [PubMed]
20. Jha, P.; Kumar, V. *BABY BOOM (BBM):* A Candidate Transcription Factor Gene in Plant Biotechnology. *Biotech. Lett.* **2018**, *40*, 1467–1475. [CrossRef]
21. Kumar, V.; Jha, P.; Van Staden, J. *LEAFY COTYLEDONs (LECs):* Master Regulators in Plant Embryo Development. *Plant Cell Tiss. Org.* **2020**, *140*, 475–487. [CrossRef]
22. Kwong, R.W.; Bui, A.Q.; Lee, H.; Kwong, L.W.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. *LEAFY COTYLEDON1-LIKE* Defines a Class of Regulators Essential for Embryo Development. *Plant Cell* **2003**, *15*, 5–18. [CrossRef]
23. Lee, H.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Arabidopsis LEAFY COTYLEDON1 Represents a Functionally Specialized Subunit of the CCAAT Binding Transcription Factor. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2152–2156. [CrossRef] [PubMed]
24. Ledwoń, A.; Gaj, M.D. *LEAFY COTYLEDON1, FUSCA3* Expression and Auxin Treatment in Relation to Somatic Embryogenesis Induction in Arabidopsis. *Plant Growth Regul.* **2011**, *65*, 157–167. [CrossRef]

25. Parcy, F.; Valon, C.; Raynal, M.; Gaubier-Comella, P.; Delseny, M.; Giraudat, J. Regulation of Gene Expression Programs during Arabidopsis Seed Development: Roles of the *ABI3* Locus and of Endogenous Abscisic Acid. *Plant Cell* **1994**, *6*, 1567–1582. [CrossRef] [PubMed]
26. Swaminathan, K.; Peterson, K.; Jack, T. The Plant B3 Superfamily. *Trends Plant Sci.* **2008**, *13*, 647–655. [CrossRef] [PubMed]
27. El Ouakfaoui, S.; Schnell, J.; Abdeen, A.; Colville, A.; Labbé, H.; Han, S.; Baum, B.; Laberge, S.; Miki, B. Control of Somatic Embryogenesis and Embryo Development by AP2 Transcription Factors. *Plant Mol. Biol.* **2010**, *74*, 313–326. [CrossRef]
28. Kumar, R.; Khurana, A.; Sharma, A.K. Role of Plant Hormones and Their Interplay in Development and Ripening of Fleshy Fruits. *J. Exp. Bot.* **2014**, *65*, eru277. [CrossRef]
29. Han, J.-D.; Li, X.; Jiang, C.-K.; Wong, G.K.-S.; Rothfels, C.J.; Rao, G.-Y. Evolutionary Analysis of the LAFL Genes Involved in the Land Plant Seed Maturation Program. *Front. Plant Sci.* **2017**, *8*. [CrossRef]
30. Cagliari, A.; Turchetto-Zolet, A.C.; Korbes, A.P.; Maraschin, F.d.S.; Margis, R.; Margis-Pinheiro, M. New Insights on the Evolution of Leafy Cotyledon1 (LEC1) Type Genes in Vascular Plants. *Genomics* **2014**, *103*, 380–387. [CrossRef]
31. Chen, Y.; Yang, Z.; Xiao, Y.; Wang, P.; Wang, Y.; Ge, X.; Zhang, C.; Zhang, X.; Li, F. Genome-Wide Analysis of the *NF-YB* Gene Family in *Gossypium hirsutum* L. and Characterization of the Role of *GhDNF-YB22* in Embryogenesis. *Int. J. Mol. Sci.* **2018**, *19*, 483. [CrossRef] [PubMed]
32. Ahmad, B.; Zhang, S.; Yao, J.; Rahman, M.U.; Hanif, M.; Zhu, Y.; Wang, X. Genomic Organization of the B3-Domain Transcription Factor Family in Grapevine (*Vitis vinifera* L.) and Expression during Seed Development in Seedless and Seeded Cultivars. *Int. J. Mol. Sci.* **2019**, *20*, 4553. [CrossRef] [PubMed]
33. Ren, C.; Zhang, Z.; Wang, Y.; Li, S.; Liang, Z. Genome-Wide Identification and Characterization of the *NF-Y* Gene Family in Grape (*Vitis vinifera* L.). *BMC Genomics* **2016**, *17*, 605. [CrossRef]
34. Zhang, Y.; Clemens, A.; Maximova, S.N.; Guiltinan, M.J. The Theobroma Cacao B3 Domain Transcription Factor *TcLEC2* Plays a Duel Role in Control of Embryo Development and Maturation. *BMC Plant Biol.* **2014**, *14*, 106. [CrossRef] [PubMed]
35. Li, M.; Li, G.; Liu, W.; Dong, X.; Zhang, A. Genome-Wide Analysis of the *NF-Y* Gene Family in Peach (*Prunus persica* L.). *BMC Genomics* **2019**, *20*, 612. [CrossRef]
36. Edger, P.P.; Poorten, T.J.; VanBuren, R.; Hardigan, M.A.; Colle, M.; McKain, M.R.; Smith, R.D.; Teresi, S.J.; Nelson, A.D.L.; Wai, C.M.; et al. Origin and Evolution of the Octoploid Strawberry Genome. *Nat. Genet.* **2019**, *51*, 541–547. [CrossRef]
37. Stevens, P.F. Onwards. Angiosperm Phylogeny Website. Version 12, July 2012 [and More or Less Continuously Updated since]. 2001. Available online: Http://www.Mobot.Org/MOBOT/Research/APweb/ (accessed on 3 February 2021).
38. Xiang, Y.; Huang, C.-H.; Hu, Y.; Wen, J.; Li, S.; Yi, T.; Chen, H.; Xiang, J.; Ma, H. Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. *Mol. Biol. Evol.* **2017**, *34*, 262–281. [CrossRef]
39. Potter, D.; Gao, F.; Bortiri, P.E.; Oh, S.-H.; Baggett, S. Phylogenetic Relationships in Rosaceae Inferred from Chloroplast Mat K and Trn L-Trn F Nucleotide Sequence Data. *Plant Syst. Evol.* **2002**, *231*, 77–89. [CrossRef]
40. Shulaev, V.; Korban, S.S.; Sosinski, B.; Abbott, A.G.; Aldwinckle, H.S.; Folta, K.M.; Iezzoni, A.; Main, D.; Arús, P.; Dandekar, A.M.; et al. Multiple Models for Rosaceae Genomics. *Plant Physiol.* **2008**, *147*, 985–1003. [CrossRef] [PubMed]
41. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Springer Protocols Handbooks; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607. ISBN 978-1-59259-890-8.
42. Hu, B.; Jin, J.; Guo, A.-Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An Upgraded Gene Feature Visualization Server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]
43. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef] [PubMed]
44. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]
45. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2-a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef]
46. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef] [PubMed]
47. Stern, A.; Doron-Faigenboim, A.; Erez, E.; Martz, E.; Bacharach, E.; Pupko, T. Selecton 2007: Advanced Models for Detecting Positive and Purifying Selection Using a Bayesian Inference Approach. *Nucleic Acids Res.* **2007**, *35*, W506–W511. [CrossRef]
48. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
49. Edger, P.P.; VanBuren, R.; Colle, M.; Poorten, T.J.; Wai, C.M.; Niederhuth, C.E.; Alger, E.I.; Ou, S.; Acharya, C.B.; Wang, J.; et al. Single-Molecule Sequencing and Optical Mapping Yields an Improved Genome of Woodland Strawberry (*Fragaria* vesca) with Chromosome-Scale Contiguity. *Gigascience* **2018**, *7*. [CrossRef]
50. Li, Y.; Pi, M.; Gao, Q.; Liu, Z.; Kang, C. Updated Annotation of the Wild Strawberry Fragaria Vesca V4 Genome. *Hort. Res.* **2019**, *6*, 1–9. [CrossRef]
51. VanBuren, R.; Wai, C.M.; Colle, M.; Wang, J.; Sullivan, S.; Bushakra, J.M.; Liachko, I.; Vining, K.J.; Dossett, M.; Finn, C.E.; et al. A near Complete, Chromosome-Scale Assembly of the Black Raspberry (*Rubus occidentalis*) Genome. *Gigascience* **2018**, *7*. [CrossRef]

52. Raymond, O.; Gouzy, J.; Just, J.; Badouin, H.; Verdenaud, M.; Lemainque, A.; Vergne, P.; Moja, S.; Choisne, N.; Pont, C.; et al. The Rosa Genome Provides New Insights into the Domestication of Modern Roses. *Nat. Genet.* **2018**, *50*, 772–777. [CrossRef]
53. Daccord, N.; Celton, J.-M.; Linsmith, G.; Becker, C.; Choisne, N.; Schijlen, E.; van de Geest, H.; Bianco, L.; Micheletti, D.; Velasco, R.; et al. High-Quality de Novo Assembly of the Apple Genome and Methylome Dynamics of Early Fruit Development. *Nat. Genet.* **2017**, *49*, 1099–1106. [CrossRef]
54. Verde, I.; Jenkins, J.; Dondini, L.; Micali, S.; Pagliarani, G.; Vendramin, E.; Paris, R.; Aramini, V.; Gazza, L.; Rossini, L.; et al. The Peach v2.0 Release: High-Resolution Linkage Mapping and Deep Resequencing Improve Chromosome-Scale Assembly and Contiguity. *BMC Genomics* **2017**, *18*, 225. [CrossRef]
55. Chagné, D.; Crowhurst, R.N.; Pindo, M.; Thrimawithana, A.; Deng, C.; Ireland, H.; Fiers, M.; Dzierzon, H.; Cestaro, A.; Fontana, P.; et al. The Draft Genome Sequence of European Pear (*Pyrus communis* L. 'Bartlett'). *PLoS ONE* **2014**, *9*, e92644. [CrossRef] [PubMed]
56. Jung, S.; Bassett, C.; Bielenberg, D.G.; Cheng, C.-H.; Dardick, C.; Main, D.; Meisel, L.; Slovin, J.; Troggio, M.; Schaffer, R.J. A Standard Nomenclature for Gene Designation in the Rosaceae. *Tree Genet. Genomes* **2015**, *11*, 108. [CrossRef]
57. Mohanta, T.K.; Khan, A.; Hashem, A.; Abd Allah, E.F.; Al-Harrasi, A. The Molecular Mass and Isoelectric Point of Plant Proteomes. *BMC Genomics* **2019**, *20*, 631. [CrossRef]
58. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene Duplication and Evolution in Recurring Polyploidization–Diploidization Cycles in Plants. *Genome Biol.* **2019**, *20*. [CrossRef] [PubMed]
59. Maher, C.; Stein, L.; Ware, D. Evolution of Arabidopsis MicroRNA Families through Duplication Events. *Genome Res.* **2006**, *16*, 510–519. [CrossRef] [PubMed]
60. Lopez-Ortiz, C.; Peña-Garcia, Y.; Natarajan, P.; Bhandari, M.; Abburi, V.; Dutta, S.K.; Yadav, L.; Stommel, J.; Nimmakayala, P.; Reddy, U.K. The Ankyrin Repeat Gene Family in Capsicum Spp: Genome-Wide Survey, Characterization and Gene Expression Profile. *Sci. Rep.* **2020**, *10*, 4044. [CrossRef]
61. Hurst, L.D. The Ka/Ks Ratio: Diagnosing the Form of Sequence Evolution. *Trends Genet.* **2002**, *18*, 486–487. [CrossRef]
62. Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [CrossRef]
63. Etienne, H.; Breton, D.; Breitler, J.-C.; Bertrand, B.; Déchamp, E.; Awada, R.; Marraccini, P.; Léran, S.; Alpizar, E.; Campa, C.; et al. Coffee Somatic Embryogenesis: How Did Research, Experience Gained and Innovations Promote the Commercial Propagation of Elite Clones From the Two Cultivated Species? *Front. Plant Sci.* **2018**, *9*. [CrossRef] [PubMed]
64. Sabooni, N.; Shekafandeh, A. Somatic Embryogenesis and Plant Regeneration of Blackberry Using the Thin Cell Layer Technique. *Plant Cell Tiss. Org.* **2017**, *130*, 313–321. [CrossRef]
65. Ameri, A.; Davarynejad, G.H.; Moshtaghi, N.; Tehranifar, A. The Role of Carbohydrates on The Induction of Somatic Embryogenesis and The Biochemical State of The Embryogenic Callus in Pyrus *communis* L. Cv. 'Dar Gazi'. *Erwerbs-Obstbau* **2020**, *62*. [CrossRef]
66. Bao, Y.; Liu, G.; Shi, X.; Xing, W.; Ning, G.; Liu, J.; Bao, M. Primary and Repetitive Secondary Somatic Embryogenesis in *Rosa hybrida* 'Samantha'. *Plant Cell Tiss. Org.* **2012**, *109*, 411–418. [CrossRef]
67. Paul, H.; Belaizi, M.; Sangwan-Norreel, B.S. Somatic Embryogenesis in Apple. *J. Plant Physiol.* **1994**, *143*, 78–86. [CrossRef]
68. Pesce, P.G.; Rugini, E. Influence of Plant Growth Regulators, Carbon Sources and Iron on the Cyclic Secondary Somatic Embryogenesis and Plant Regeneration of Transgenic Cherry Rootstock 'Colt' (*Prunus avium* × *P. pseudocerasus*). *Plant Cell Tiss. Org.* **2004**, *79*, 223–232. [CrossRef]
69. Zakizadeh, H.; Stummann, B.M.; Lütken, H.; Müller, R. Isolation and Characterization of Four Somatic Embryogenesis Receptor-like Kinase (*RhSERK*) Genes from Miniature Potted Rose (*Rosa hybrida* Cv. Linda). *Plant Cell Tiss. Org.* **2010**, *101*, 331–338. [CrossRef]
70. Zheng, L.; Ma, J.; Mao, J.; Fan, S.; Zhang, D.; Zhao, C.; An, N.; Han, M. Genome-Wide Identification of *SERK* Genes in Apple and Analyses of Their Role in Stress Responses and Growth. *BMC Genomics* **2018**, *19*, 962. [CrossRef]
71. Adams, K.L.; Wendel, J.F. Polyploidy and Genome Evolution in Plants. *Curr. Op. Plant Biol.* **2005**, *8*, 135–141. [CrossRef]
72. Lehti-Shiu, M.D.; Panchy, N.; Wang, P.; Uygun, S.; Shiu, S.-H. Diversity, Expansion, and Evolutionary Novelty of Plant DNA-Binding Transcription Factor Families. *Biochimica et Biophysica Acta Gene Regul. Mech.* **2017**, *1860*, 3–20. [CrossRef] [PubMed]
73. Panchy, N.; Lehti-Shiu, M.; Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [CrossRef]
74. Chen, X.; Li, J.; Cheng, T.; Zhang, W.; Liu, Y.; Wu, P.; Yang, X.; Wang, L.; Zhou, S. Molecular Systematics of Rosoideae (Rosaceae). *Plant Syst. Evol.* **2020**, *306*, 9. [CrossRef]
75. Chin, S.-W.; Shaw, J.; Haberle, R.; Wen, J.; Potter, D. Diversification of Almonds, Peaches, Plums and Cherries—Molecular Systematics and Biogeographic History of Prunus (Rosaceae). *Mol. Phylogenet. Evol.* **2014**, *76*, 34–48. [CrossRef]
76. Brand, A.; Quimbaya, M.; Tohme, J.; Chavarriaga-Aguirre, P. Arabidopsis *LEC1* and *LEC2* Orthologous Genes Are Key Regulators of Somatic Embryogenesis in Cassava. *Front. Plant Sci.* **2019**, *10*. [CrossRef] [PubMed]
77. Yang, Y.; Davis, T.M. A New Perspective on Polyploid Fragaria (Strawberry) Genome Composition Based on Large-Scale, Multi-Locus Phylogenetic Analysis. *Genome Biol. Evol.* **2017**, *9*, 3433–3448. [CrossRef] [PubMed]
78. Gmitter, F.G.; Ling, X.; Cai, C.; Grosser, J.W. Colchicine-Induced Polyploidy in Citrus Embryogenic Cultures, Somatic Embryos, and Regenerated Plantlets. *Plant Sci.* **1991**, *74*, 135–141. [CrossRef]

79. Ghotbi Ravandi, E.; Rezanejad, F.; Dehghan, E. In Vitro Regeneration Ability of Diploid and Autotetraploid Plants of *Cichorium intybus* L. *Cytol. Genet.* **2014**, *48*, 166–170. [CrossRef]
80. Jo, B.-S.; Choi, S.S. Introns: The Functional Benefits of Introns in Genomes. *Genom. Inform.* **2015**, *13*, 112–118. [CrossRef] [PubMed]
81. Zhu, L.; Zhang, Y.; Zhang, W.; Yang, S.; Chen, J.-Q.; Tian, D. Patterns of Exon-Intron Architecture Variation of Genes in Eukaryotic Genomes. *BMC Genom.* **2009**, *10*, 47. [CrossRef]
82. Chaudhary, S.; Khokhar, W.; Jabre, I.; Reddy, A.S.N.; Byrne, L.J.; Wilson, C.M.; Syed, N.H. Alternative Splicing and Protein Diversity: Plants Versus Animals. *Front. Plant Sci.* **2019**, *10*. [CrossRef]
83. Paterson, A.H.; Chapman, B.A.; Kissinger, J.C.; Bowers, J.E.; Feltus, F.A.; Estill, J.C. Many Gene and Domain Families Have Convergent Fates Following Independent Whole-Genome Duplication Events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. *Trends Genet.* **2006**, *22*, 597–602. [CrossRef] [PubMed]
84. Zong, Y.; Xi, X.; Li, S.; Chen, W.; Zhang, B.; Liu, D.; Liu, B.; Wang, D.; Zhang, H. Allelic Variation and Transcriptional Isoforms of Wheat *TaMYC1* Gene Regulating Anthocyanin Synthesis in Pericarp. *Front. Plant Sci.* **2017**, *8*. [CrossRef] [PubMed]
85. Yang, H.F.; Kou, Y.P.; Gao, B.; Soliman, T.M.A.; Xu, K.D.; Ma, N.; Cao, X.; Zhao, L.J. Identification and Functional Analysis of *BABY BOOM* Genes from *Rosa canina*. *Biologia Plantarum* **2014**, *58*, 427–435. [CrossRef]
86. Hummer, K.E.; Janick, J. Rosaceae: Taxonomy, Economic Importance, Genomics. In *Genetics and Genomics of Rosaceae*, 1st ed.; Folta, K.M., Gardiner, S.E., Eds.; Springer: New York, NY, USA, 2009; pp. 1–17. ISBN 978-0-387-77490-9.
87. Silva, A.T.; Barduche, D.; do Livramento, K.G.; Paiva, L.V. A Putative BABY BOOM-like Gene (CaBBM) Is Expressed in Embryogenic Calli and Embryogenic Cell Suspension Culture of *Coffea arabica* L. *In Vitro Cell. Dev. Biol. Plant* **2015**, *51*, 93–101. [CrossRef]
88. Maulidiya, A.U.K.; Sugiharto, B.; Dewanti, P.; Handoyo, T. Expression of Somatic Embryogenesis-Related Genes in Sugarcane (*Saccharum officinarum* L.). *J. Crop Sci. Biotech.* **2020**, *23*, 207–214. [CrossRef]
89. Considine, M.J.; Wan, Y.; D'Antuono, M.F.; Zhou, Q.; Han, M.; Gao, H.; Wang, M. Molecular Genetic Features of Polyploidization and Aneuploidization Reveal Unique Patterns for Genome Duplication in Diploid Malus. *PLoS ONE* **2012**, *7*. [CrossRef]
90. Yamasaki, M.; Wright, S.I.; McMullen, M.D. Genomic Screening for Artificial Selection during Domestication and Improvement in Maize. *Ann. Bot.* **2007**, *100*, 967–973. [CrossRef] [PubMed]
91. Roth, C.; Liberles, D.A. A Systematic Search for Positive Selection in Higher Plants (Embryophytes). *BMC Plant Biol.* **2006**, *6*, 12. [CrossRef]