Article

Evidence for Light and Tissue Specific Regulation of Genes Involved in Fructan Metabolism in *Agave tequilana*

Alan D. Gomez-Vargas, Karen M. Hernández-Martínez, Macrina E. López-Rosas, Gerardo Alejo Jacuinde and June Simpson *

Department of Genetic Engineering, CINVESTAV Irapuato, Km. 9.6 Libramiento Norte Carretera Irapuato-León, Irapuato 36821, Guanajuato, Mexico

* Correspondence: june.simpson@cinvestav.mx

Abstract: Plant Glycoside Hydrolase Family 32 (PGHF32) contains the fructosyltransferases and fructan exohydrolase enzymes responsible for fructan metabolism, in addition to closely related vacuolar and cell wall acid invertases. *Agave* species produce complex and dynamic fructan molecules (agavins) requiring 4 different fructosyltransferase activities (1-SST, 1-FFT, 6G-FFT and 6-SFT) for their synthesis. Combined analysis of RNAseq and genome data for *A. tequilana* led to the characterization of the genes encoding 3 fructosyltransferases for this species and support the hypothesis that no separate 6-SFT type enzyme exists in *A. tequilana,* suggesting that at least one of the fructosyltransferases identified may have multiple enzymatic activities. Structures for PGHF32 genes varied for *A. tequilana* and between other plant species but were conserved for different enzyme types within a species. The observed patterns are consistent with the formation of distinct gene structures by intron loss. Promoter analysis of the PGHF32 genes identified abundant putative regulatory motifs for light regulation and tissue-specific expression, and these regulatory mechanisms were confirmed experimentally for leaf tissue. Motifs for phytohormone response, carbohydrate metabolism and dehydration responses were also uncovered. Based on the regulatory motifs, full-length cDNAs for MYB, GATA, DOF and GBF transcription factors were identified and their phylogenetic distribution determined by comparison with other plant species. In silico expression analysis for the selected transcription factors revealed both tissue-specific and developmental patterns of expression, allowing candidates to be identified for detailed analysis of the regulation of fructan metabolism in *A. tequilana* at the molecular level.

Keywords: *A. tequilana*; fructan metabolism; gene structure; promoter motifs; light induction; transcription factors

1. Introduction

Fructans, synthesized by approximately 15% of angiosperms, are a structurally diverse group of molecules made up of multiple fructose units linked to a precursor sucrose molecule [1–3]. The main biological function of fructans is considered to be as a reserve carbohydrate, although it has also been suggested that fructan metabolism in plants arose as an evolutionary adaptation to inhabit environments with prolonged periods of drought or low water availability [1,4–7]. More recently, the roles of fructans in osmoregulation and signaling have also been described [8–11].

The *Agave* genus comprises plants with the ability to accumulate high levels of fructans and these carbohydrates are exploited commercially for the production of tequila, mezcal, pulque and syrups [12–14]. Agave fructans or agavins, are the most complex of all fructans in terms of their highly branched structures [15,16], and their synthesis by fructosyltransferases (FT) requires 4 different enzymatic activities: sucrose: sucrose 1-fructosyltransferase (1-SST), fructan: fructan 1-fructosyltransferase (1-FFT), sucrose: fructan 6 -fructosyltransferase (6-SFT) and fructan: fructan 6G-fructosyltransferase (6G-FFT). In

contrast, fructan exohydrolase enzymes (FEH) are responsible for fructan degradation [17]. Plant FTs and FEHs belong to Plant Glycoside Hydrolase Family 32 (PGHF32), which also includes vacuolar and cell wall invertases [2,18,19]. Enzymes belonging to this family are characterized by highly similar and conserved amino acid sequences. Phylogenetic analyses of these enzymes have suggested that FTs evolved from vacuolar invertases, while FEHs are more closely related to cell wall invertases [18,19].

In *Agave tequilana* Weber var. Azul, extensive transcriptome and gene expression analysis has shown that PGHF32 genes show specific and differential patterns of expression during flowering and in distinct plant tissues [20–22]. In addition, analysis of in vitro grown seedlings revealed that 1-SST and 1-FFT transcripts accumulate in response to the exogenous application of sucrose, abscisic acid (ABA) and other plant hormones (salicylic acid, kinetin and methyl jasmonate) associated with responses to abiotic stress [23,24]. Likewise, it has been shown that fructan synthesis in plants such as wheat and chicory is regulated by metabolic, developmental and environmental stimuli, including light, drought, sucrose and abscisic acid [25–30]. These reports of the diverse stimuli that modify the expression of PGHF32 genes imply complex interactions between individual genes and an array of different transcription factors where potential candidates are MYB, associated with drought responses; GATA and GBF, associated with light responses; and DOF, associated with carbon metabolism, among many others. At the molecular level, however, data is only available for transcription factor (TF) R2R3 type MYB genes studied in wheat (*Triticum aestivum*) and chicory (*Chicorium intybus*). In wheat TaMYB13-1, TaMYB13-2 and TaMYB13-3 transcription factors interact with and induce the expression of the wheat Ta1-SST, Ta6-SFT and Ta1-FFT genes in stem tissue and are induced by sucrose [25]. In chicory, CiMYB-3 and CiMYB-5 induce the expression of the Ci1-FEH1, Ci1-FEH2a, and Ci1-FEH2b genes encoding enzymes for fructan degradation without affecting the expression of genes encoding fructan synthetic enzymes, while CiMYB-17 was identified to be capable of regulating genes for both fructan synthesis and degradation under abiotic stress conditions (drought and freezing) [25,31].

Agave species are some of the few plant species that combine both fructan metabolism and Crassulacean Acid Metabolism (CAM) [14,32]. It is therefore of interest to understand the factors that regulate fructan metabolism in these species and how fructan metabolism is integrated into CAM and general carbohydrate metabolism. Fructan metabolism is strongly linked with sucrose metabolism, since sucrose is crucial for fructan synthesis through PGHF32 enzymes [3]. The tissue-specific expression patterns under biotic and abiotic stimuli suggest that the genes that encode the PGHF32 enzymes involve a complex regulation directed by multiple transcription factors; therefore, determination of the gene structures and putative regulatory motifs in the promoter regions of the members of PGHF32 in *A. tequilana* would allow us to address the evolution of PGHF32 and in particular the enzymes associated with the metabolism of fructans. The identification and characterization of genes encoding transcription factors putatively involved in the regulation of PGHF32 in *A. tequilana* is one of the first steps to understand in detail the regulation of fructan metabolism at the molecular level and to apply this knowledge to future biotechnological applications.

Recently, a draft genome for *A. tequilana* became available (Herrera-Estrella et al. in preparation) and based on this data, we determined the gene structure and identified putative regulatory promoter motifs for each of the members of PGHF32 from *A. tequilana*. Here, we report the results of these analyses and also the in silico identification and characterization of several *A. tequilana* transcription factor cDNAs and predicted proteins. We also confirm light-inducible and tissue-specific regulation of selected PGHF32 encoding genes.

2. Methods

2.1. Characterization of Genes Encoding PGHF32 Members in Agave tequilana

Agave tequilana genomic sequences were obtained using the BLAST algorithm [33] in the *A. tequilana* draft genome (Herrera-Estrella et al. unpublished), permission to ac-

cess the data prior to publication can be arranged by directly contacting Dr. Alfredo Herrera-Estrella: (alfredo.herrera@cinvestav.mx) or Dr. Selene Fernández Valverde: (selene.fernandez@cinvestav.mx) by using as queries the cDNA sequences of PGHF32 previously reported. Genome coordinates were determined for each cDNA to obtain the genomic sequence. An alignment of the genomic and cDNA sequences was carried out using the MUSCLE algorithm [34] in SeaView software version 5.0.2.Available online: http://doua.prabi.fr/software/seaview (accessed on 24 May 2022) in order to determine the exon/intron structure for each genomic sequence. These analyses were also carried out for members of PGHF32 from a selection of fructan and non-fructan accumulating species: asparagus (*Asparagus officinalis*), goatgrass (*Aegilops tauschii*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), *Arabidopsis thaliana*, carrot (*Daucus carota*), beetroot (*Beta vulgaris*), maize (*Zea mays*) and rice (*Oryza sativa*). Genomic sequences were obtained from "Ensembl Plants". Available online: https://plants.ensembl.org/index.html (accessed on 22 August 2020) and "The Arabidopsis Information Resource" (TAIR). Available online: https://www.arabidopsis.org/data-bases (accessed on 28 August 2020). Enzymes and accession numbers are listed in Supplementary Table S1.

2.2. Identification of Regulatory Motifs in the Promoters of PGHF32

Based on the coordinates determined for the PGHF32 genes, a 2 kb region upstream to the transcription start site was analyzed using the PLACE "Plant Cis-Regulatory DNA Elements" database [35]. Available online: https://www.dna.affrc.go.jp/PLACE/?action=newplace (accessed on 6 November 2021) that identifies putative motifs associated with characterized transcription factors. Identified motifs were separated into groups based on regulatory function such as light regulation, response to phytohormones, etc. Detailed comparisons were then carried out for selected motifs related to light regulation, tissue-specific regulation, phytohormone responses, regulation of carbohydrate metabolism and responses to dehydration.

2.3. A. tequilana Transcriptome Database Searches

A search was carried out using the BLAST algorithm in the *A. tequilana* transcriptome reported by [21], using experimentally validated transcription factors as queries for MYB, GATA, DOF and GBF type transcription factors. The best hits with $e < 10^{-5}$ were chosen and the open reading frames (ORF) of the transcripts were determined by using ORF finder. Available online: http://www.bioinformatics.org/sms2/orf_find.html (accessed on 22 June 2021). Sequences encoding the complete predicted proteins were selected as novel transcription factor encoding sequences from *A. tequilana* and were used for comparison with previously characterized amino acid sequences of the same transcription factors from other species. Accession numbers for sequences used in these analyses are listed in Supplementary Table S2 and newly identified *A. tequilana* sequences were deposited in the GenBank database.

2.4. Alignment and Identification of Conserved Motifs

Complete predicted protein sequences were selected for alignment using the MUSCLE algorithm and conserved motifs were compared for each member of the transcription factor family analyzed. Alignments and phylogenetic analyses for MYB, GATA, DOF and GBF transcription factors were performed using Geneious® (version 6.1.8). Available online: www.geneious.com (accessed on 15 February 2022), taking into account all sites including gaps/missing data with the UPGMA algorithm and the Jukes-Cantor genetic distance model with a bootstrap analysis using a total of 1000 repetitions and a support threshold of 50%. PGHF32 phylogenetic analysis was carried out using all sites of the amino acid sequences, including gaps/missing data with the Maximum likelihood algorithm by using the Molecular Evolutionary Genetics Analysis (MEGA 11) software [36]. According to amino acid information, the best substitution model was obtained (WAG + G + I) and used to construct the dendrograms with a bootstrap analysis using a total of 1000 repetitions.

2.5. In Silico Expression Analysis

Expression levels of transcripts from each individual transcription factor in different tissues were determined in silico and expressed as "transcripts per million" (TPM), as described by [20,21]. Heatmaps were created based on transcription factor expression data from identified *A. tequilana* transcripts using the ggplot2 function from the tidyverse library in the RStudio statistics package version 1.4.1717. Available online: https://www.rstudio.com/products/rstudio/download/ (accessed on 8 May 2022).

2.6. qRT-PCR Analysis

Total RNA was isolated from green leaf sections of 3 individual 3-year-old *A. tequilana* plants using the TRIzol reagent (Invitrogen; Carlsbad, CA, USA) and the PureLink RNA mini kit (Invitrogen) according to the manufacturer's protocol. For light/dark regimes, green leaf tissue was collected at time 0 before subjecting 2 of the plants to complete darkness for 7 days. A control plant was left under normal light cycle conditions (16 h light/8 h darkness). After 7 days, RNA was again extracted from green leaf tissue and the plants were returned to a normal light cycle for 48 h, after which green leaf RNA was again extracted. For tissue-specific analysis, RNA was extracted from green and white sections of the same leaves (Supplementary Figure S1). qRT-PCR analysis was carried out on 5 members of *A. tequilana* PGHF32: *Atq1SST-1*, *Atq6G-FFT-1* and *Atq6G-FFT-2*, *AtqFEH-4* and *AtqVinv-1* using primers described in [30,37]. cDNA templates were synthesized by incubating 1 μg of total RNA with dT primer and revertaid reverse transcriptase (Thermo Scientific) according to the manufacturer's instructions. qRT-PCR was carried out by using Kapa Sybr Fast qPCR master mix (2X) Universal reagent (Sigma Aldrich, St. Louis, MO, USA) in a StepOne Plus thermocycler (Applied Biosystems, Foster City, CA, USA). Thermal conditions and the relative quantification were determined as described by [22].

3. Results

3.1. Determination of Gene Structures for A. tequilana PGHF32 Encoding Genes

A previous report described the gene structures for 6 of the members of *A. tequilana* PGHF32 [37]. However, more recent RNAseq results revealed several new members of PGHF32 for which gene structures had not been determined [20,21], and in this analysis a further 3 genes encoding vacuolar invertases were identified by analysis of the draft genome (denominated *AtqVinv like 3, 4* and *5*). Genome analysis did not uncover genes encoding 6-SFT type enzymes, supporting the absence of 6-SFT type transcripts in the previous RNAseq analysis and suggesting that in *A. tequilana* (and perhaps more widely in the *Agave* genus), no separate enzymes with 6-SFT activity exist. However, given the complexity of agavin structures, this activity must therefore be carried out by the previously described 1-SST, 6G-FFT or 1-FFT enzymes.

By comparison with transcriptome-derived cDNAs, the structures of each of the *A. tequilana* PGHF32 family genes were determined, confirming the structures previously described by Cortés Romero et al. [37]. Independent genome coordinates could not be determined for *AtqCwinv-3* and *AtqCwinv-2* previously identified from RNAseq data, suggesting that these isoforms actually represent alleles of a single locus. Six different patterns of exon/intron structure were determined, as shown in Figure 1. The 9-nucleotide mini-exon (Exon II), characteristic of PGHF32, was identified in all *A. tequilana* PGHF32 genes, with the exception of those encoding the *A. tequilana* FEH enzymes. The 3 amino acids encoded by Exon II comprise a highly conserved motif characteristic of PGHF32 with the consensus sequence DPN, where the aspartic acid residue (D) is essential for the catalytic activity of the PGHF32 enzymes. The nucleotide ranges of exons and introns are shown in Supplementary Table S3.
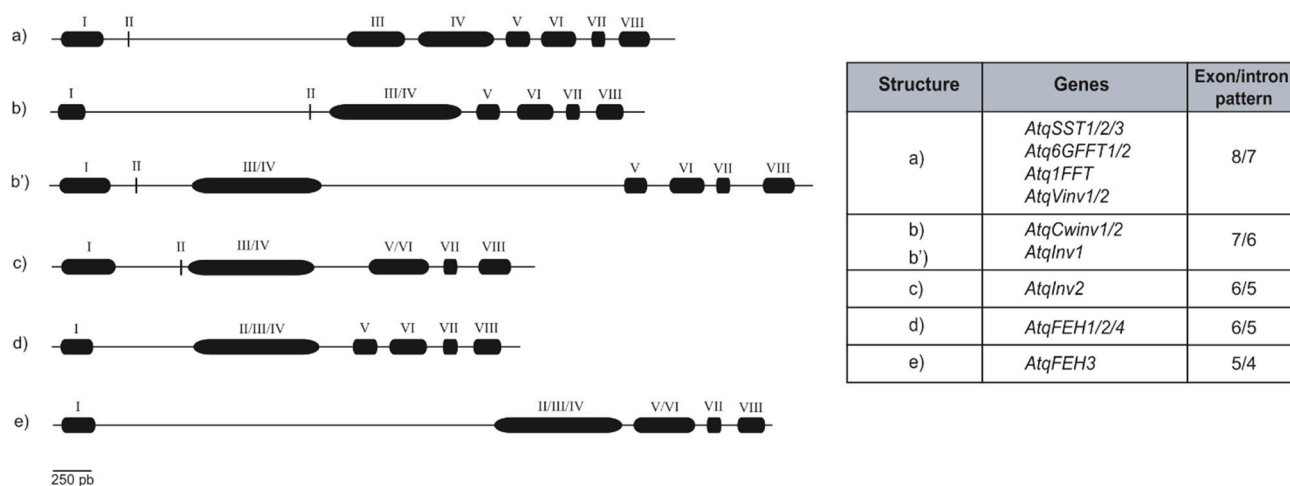
Figure 1. Schematic representation of the genomic structures of *A. tequila* PGHF32 encoding genes. Introns are represented by straight lines and exons with solid boxes. Roman numerals (I-VIII) from left to right identify each exon. Scale bar represents 250 base pairs. The table summarizes the genes and their exon/intron structure.

FT and Vinv genes are the only members of the family with 8 well-defined exons separated by 7 introns (Figure 1a), whereas the rest of the genes present fusions of some exons, specifically exons III/IV in *AtqCwinv1/2* and *AtqInv1* genes (Figure 1b) and exons II/III/IV in the *AtqFEH* genes (Figure 1d). Additionally, *AtqInv2* and *AtqFEH3* present a fusion of exons V/VI (Figure 1b,e). The changes in gene structure produced by the fusion of different exons are consistent with intron loss.

To determine whether the gene structures for PGHF32 were conserved across different species and taxonomic levels, a comparison with PGHF32 gene structures from monocotyledonous fructan producing species: *A. tequilana,* asparagus (*Asparagus officinalis*), goatgrass *(Aegilops tauschii)*, wheat (*Triticum aestivum),* barley (*Hordeum vulgare)* and monocotyledonous and dicotyledonous fructan non-producing species: *Arabidopsis thaliana*, carrot (*Daucus carota*), beetroot (*Beta vulgaris*), maize (*Zea mays*) and rice (*Oryza sativa*) was carried out. The dendrogram shows two main clades: A, containing FTs and vacuolar invertases and B, containing FEH and cell wall invertases. Clade A can further be classified into 4 subclades, all with bootstrap values between 97 and 100. Subclade A/a contains sequences from the gramineous monocotyledons, subclade A/b contains non-gramineous monocotyledons from the Asparagales order (*A. tequilana* and *A. officinalis*), subclade A/c gramineous and non-gramineous monocotyledons and subclade A/d contains dicotyledonous species. Within the subclades, defined clades are also observed that separate enzymes with fructosyltransferase activity from those with vacuolar invertase activity, for example, subclades A/a/1 and A/a/2. In clade B, 3 subclades can also be identified and were denoted as B/e, B/f and B/g. Subclade B/e is specific to dicotyledonous species and supported by a bootstrap value of 100. Subclade B/g is specific to monocotyledonous species, whereas subclade B/f contains both monocotyledonous and dicotyledonous species. However, the node leading to subclades f and g has a very low Boostrap of 52, suggesting that the topology of this section of the dendrogram is less robust. In contrast to clade A in clade B, no clear separation of clades containing enzymes with different activities or specific to gramineous and non-gramineous monocotyledons was observed and although subclade e is specific to dicotyledons, it contains enzymes with different functions (both Cwinv and FEH).

As shown in Supplementary Table S4, the 7 exon/6 intron pattern is the only conserved pattern in all species. Other patterns identified consisted of 4/3, 3/2, 8/7, 6/5 and 9/8 exon/intron conformations, although no consistency was observed in exon/intron patterns regarding taxonomic classification, for example within the monocotyledons or dicotyledons or within the Poales order. However, *A. tequilana* and *A. officinalis*, both of the Asparagales order and non-gramineous monocotyledonous species, share the 8/7, 7/6 and 6/5 exon/intron

pattern conformations. The mini-exon is also absent in the *A. officinalis* FEH gene, as observed for *A. tequilana*. Interestingly, when the exon/intron patterns are superimposed on a dendrogram showing the relationships between the genes listed in Supplementary Table S3, a good association between clades and exon/intron structures is observed even across different species, as indicated by the exon/intron patterns (4/3, 3/2, 8/7, 6/5, 7/6 and 9/8) in Figure 2.
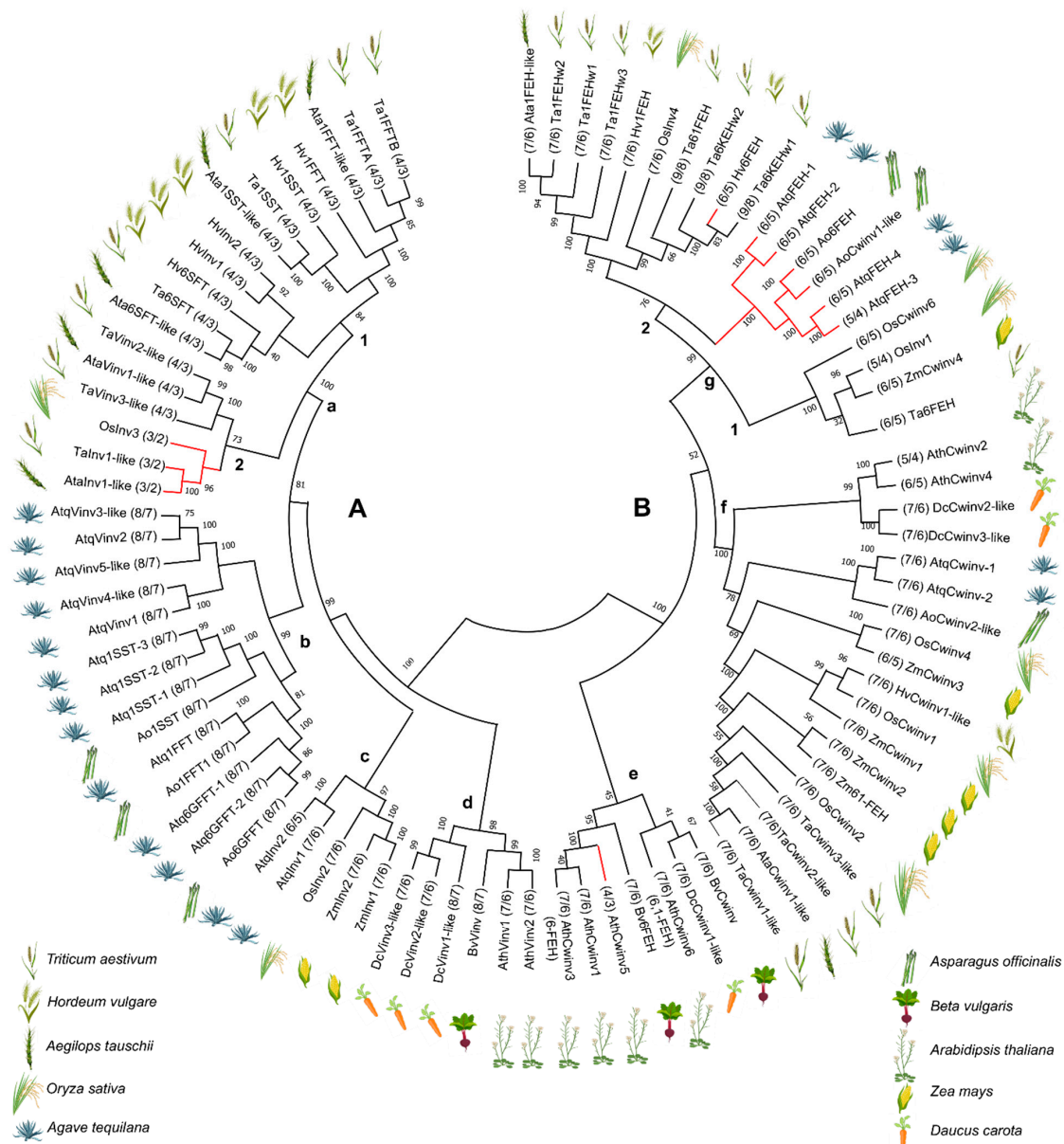
Figure 2. Dendrogram showing relationships between *A. tequilana* PGHF32 enzymes and those of other plant species. Upper case letters indicate the Vacuolar invertase/Fructosyltransferase clade (A) and the Cell wall invertase/FEH clade (B). Lower case letters indicate different subclades within clades A and B. Numbers within the dendrogram represent the Bootstrap analysis values and the numbers followed by the name represent the gene structure (exons/introns). Phylogenetic analysis was carried out by using amino acid sequences and considering all sites including gaps/missing information based on the Maximum likelihood method and the WAG+G+I substitution model. Genes where the mini exon is not present are indicated by red colored branches. Symbology represents the species analyzed.

Some changes in exon/intron structure are directly due to the fusion of the 9-nucleotide mini-exon with the previous or following exon. An example is subclade A/a containing vacuolar invertase and fructosyltransferase enzymes from *A. tauschii*, wheat, rice and barley.

Most of the genes encoding these enzymes have a 4 exon/3 intron structure, whereas 3 members, marked in red, have a 3 exon/2 intron structure. This structural change is due to the fusion of the mini exon with the previous exon. Loss of a discrete mini exon was also observed for the *A. thaliana* Cwinv enzymes in clade B/e and for Cwinv and FEH type enzymes from *A. tequilana*, *A. officinalis* and *H. vulgare* in clade B/g/2 (Figure 2).

In general, the mini-exon was present in all types of PGHF32 members (fructosyltransferases, cell wall and vacuolar invertases and fructan exohydrolases with the exception of the *A. tequilana* and *A. officinalis* FEH genes. In *A. tequilana AtqInv2*, *A. officinalis Ao1-SST* and beetroot *Bv6FEH,* however, the mini-exon encoded 2 rather than 3 amino acids, where the first amino acid of the triad was encoded on the previous exon. Additionally, the third amino acid of the mini-exon triad was found to be variable among the enzymes, encoding either Asparagine (N), Cysteine (C) or Serine (S) (Supplementary Table S3).

3.2. Identification of Putative Regulatory Motifs in the Promotor Regions of Genes Encoding PGHF32 Enzymes in A. tequilana

To locate regulatory motifs denoting putative transcription factor binding sites, 2 Kb regions at the 5' UTR upstream from the coding sequence of 16 *A. tequilana* PGHF32 genes were analyzed. The PLACE program identified 149 putative motifs across all 16 genes. To simplify the comparison across genes, motifs were placed in 24 different groups according to the regulatory process with which they are associated. The numbers of motifs identified for each group showed a similar pattern across all the genes and a typical example of these results is shown for the 4 *A. tequilana* FEH genes in Supplementary Figure S2. As can be seen, the most abundant group of motifs was associated with light regulation, followed by tissue-specific expression, pollen-specific expression, and expression related to phytohormones. In order to analyze motif patterns in more detail, the abundance of selected motifs was compared between the individual genes, as shown in Figure 3.

With the exception of the G-box binding factor (GBF) motif involved in light regulation, the chosen motifs were present in all the *A. tequilana* PGHF32 genes. Differences in abundance may be between the different types of genes; for example, the DOF and GATA motifs are more prevalent in the FT encoding genes than in the other PGHF32 genes. Differences are also observed within the same gene type; for example, the YACT motif is less abundant in *Atq1SST-3* in comparison to *Atq1SST-1* and *2* and the ATATT motif is more abundant in *AtqFEH-3* in comparison to the other FEH genes. Interestingly, although the GBF motif (CACTGT) is known to be necessary for light regulation in many plant species [38–40], only 2 or 4 copies of this element were found and only in *Atq6GFFT-1* and *2*, *AtqCwinv-1* and *2*, *AtqInv1* and *2* and *AtqVinv-2*. The MYB motif (WAACCA), also previously reported to be involved in the regulation of PGHF32 genes [26,28], was also less abundant in comparison to the other motifs.

To confirm the predicted light- and tissue-specific regulation of *A. tequilana* PGHF32 genes, a simple expression analysis was carried out by qRT-PCR using leaf tissue from 3-year-old *A. tequilana* plants. As shown in Figure 4, plants subjected to 7 days in complete darkness showed a decrease in the expression levels of all fructosyltransferase genes (*Atq1SST-1, Atq6GFFT-1* and *2*) as did the vacuolar invertase gene *AtqVinv-1*. The fructan exohydrolase gene *AtqFEH-4* is expressed at low levels in the light but shows a slight induction under darkness. In contrast, when the plants were returned to normal light conditions for 48 h, expression of all genes increased, with the exception of *AtqFEH-4,* which was repressed, confirming a pattern of light-dependent transcriptional regulation. To address the question of tissue-specific patterns of expression, qRT-PCR analysis was carried out on green aerial leaf tissue and white tissue from the leaf base of the same leaf (Supplementary Figure S3). In this case, the fructosyltransferase genes (*Atq1SST-1, Atq6GFFT-1* and *2*) showed increased levels of expression by several thousand fold in white tissue as compared to green tissue and *AtqFEH-4* showed an increase in expression between 2 and 6 fold; however, the vacuolar invertase gene (*AtqVinv-1*) was repressed in white tissue. These results therefore also confirm that the *A. tequilana* PGHF32 genes are transcriptionally regulated in a tissue-specific manner in leaf tissue.
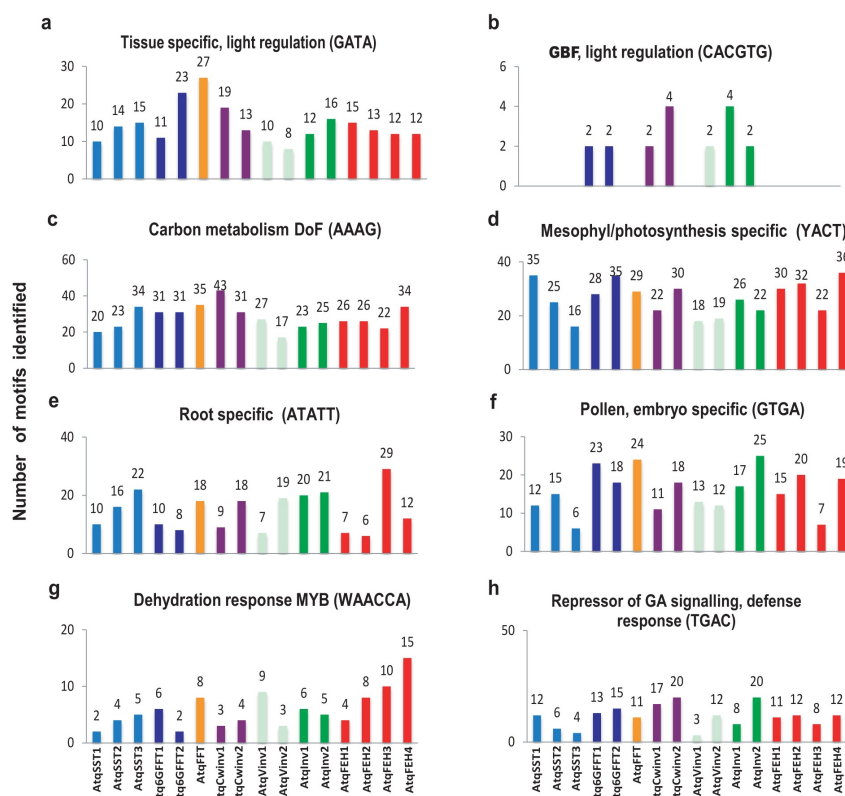
Figure 3. Abundance of transcription factor regulatory motifs identified in PGHF32 promoter regions. Conserved motifs and their associated regulatory process are indicated above each graph. PGHF32 members are indicated at the foot of each column on the x-axis and the y-axis shows the numbers of motifs identified. Each color indicates the different PGHF32 members for present in the *A. tequilana* genome.
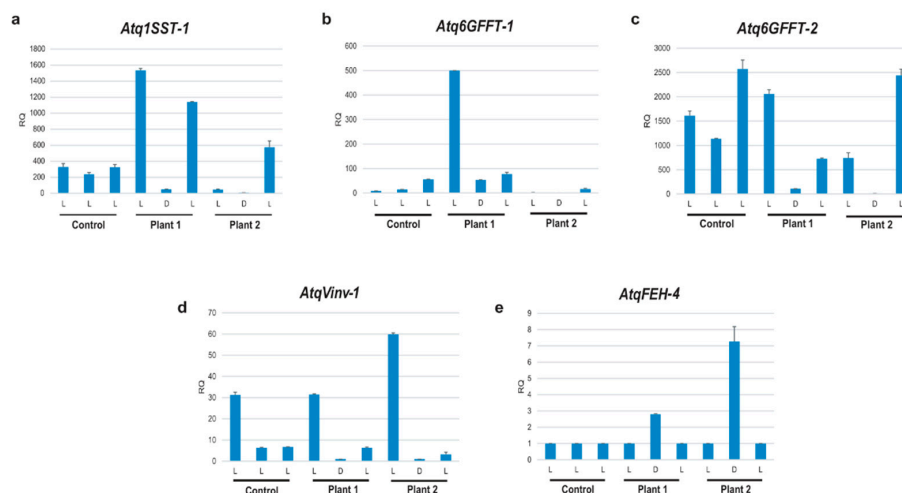

Figure 4. qRT-PCR analysis of PGHF32 genes during light and dark treatments. L, Normal light conditions (16 h light/8 h darkness), D, dark treatment for 7 days. Plants 1 and 2 were placed under normal light conditions (L), then under darkness (D) for 7 days and finally they were moved back to light conditions (L) for 48 h. Error bars indicate the standard deviation. The genes analyzed are indicated above each graph.

3.3. Identification and Analysis of Coding Sequences and Predicted Proteins for Transcription Factors Putatively Involved in the Regulation of PGHF32 Genes in A. tequilana

Based on the motifs identified in the promoter regions of the *A. tequilana* PGHF32 genes and previous reports of transcription factors that regulate PGHF32 in other plant species, a

search of the *A. tequilana* transcriptome database was carried out in order to obtain coding sequences for MYB, GATA, DOF and GBF type transcription factors and complete cDNAs could be retrieved for 16 MYB, 13 GATA, 6 DOF and 6 GBF type proteins. Complete cDNA sequences were translated in silico to obtain predicted amino acid sequences for each gene and analyzed to confirm that the conserved domains expected for each type of transcription factor were present (Figure 5). As can be observed, the expected domains were identified for all predicted proteins. To further classify the identified transcription factors within each gene family, dendrograms were produced by aligning the *A. tequilana* transcription factor amino acid sequences with representative members of the same transcription factor gene families from other plant species (Figure 6).
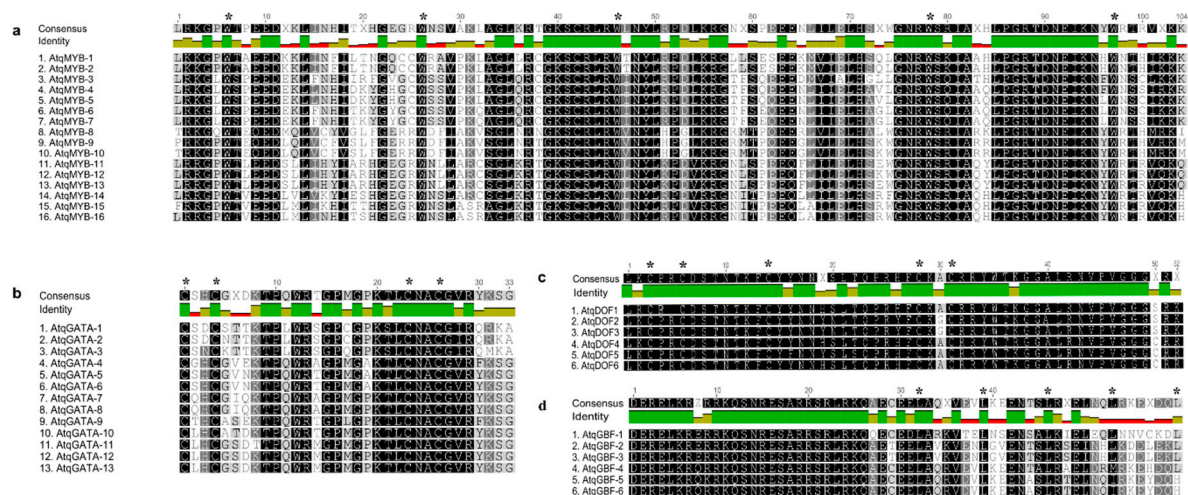
Figure 5. Comparison of amino acid motifs of the MYB, GATA, DOF and GBF transcription factors in *A. tequilana*. (**a**) R2R3 motif of the MYB family, (**b**,**c**) zinc finger motif of the GATA and DOF family and (**d**) the basic leucine zipper (bZIP) of the GBF family. Asterisks (*) indicate tryptophan (W), cysteine (C) and leucine (L) conserved residues that are crucial for the MYB, GATA/DOF and GBF function, respectively. Graphs above alignments indicate highly conserved residues shown as dark green.

For MYB-type transcription factors, the R1R2R3 type MYB proteins (mainly involved in cell cycle regulation) [41] and the R2R3 type MYB proteins are clearly separated in clades A and B, respectively. Clade A is divided into 3 subclades (A-1, A-2, and A-3) (Figure 6a). Clade A-1 comprises only MYB-type transcription factors from *A. thaliana.* In clade A-2, 7 *A. tequilana* MYB are grouped together with MYB from *C. intybus* and rice. In subclade A-3a, 6 putative MYB proteins from *A. tequilana* were grouped with the MYB from *A. thaliana* and *C. intybus.* In subclade A-3b, 3 MYB proteins from *A. tequilana* are grouped together with TaMYB-13 from *T. aestivum.* OsMYB-3 and 5 did not group within subclade A-1, 2 or 3, although they fall within clade A.

For GATA-like transcription factors, 3 major clades can also be distinguished (Figure 6b). In clade A, AthGATA21 from *A. thaliana* proteins is grouped with three *A. tequilana* GATA proteins (AtqGATA-1, AtqGATA-2 and AtqGATA-3). Clade B contains several subclades, B-1a, B-1b and B-1c. Subgroup B-1 contains AtqGATA-7 and AtqGATA-8 from *A. tequilana* and OsGATA-11 from rice. In subgroup B-1b, 4 *A. tequilana* proteins group most closely with the *A. thaliana* protein AthGATA-9, whereas AtqGATA-9 is grouped most closely with *A. thaliana* proteins AthGATA-2 and 4 shown to respond to abiotic stress and light and is involved in the regulation of the circadian cycle, respectively [42,43]. In subclade B-1c, 3 *A. tequilana* proteins are grouped with proteins from *A. thaliana*. Clade C contains only GATA proteins from *A. thaliana*.

The 6 DOF type proteins (Figure 6c) identified for *A. tequilana* were compared to DOF proteins from wheat, *A. thaliana* and maize and two main clades, A and B, can be distinguished. In clade A, *A. tequilana* DOF proteins form a discrete group. In clade B, *A. tequilana* proteins also form a discrete group but are associated with DOF proteins from

A. thaliana, which correspond to transcription factors involved in regulation of sucrose metabolism and light and /or show temperature dependent regulation [44,45]. *A. tequilana* proteins did not group with the maize and wheat proteins included in the dendrogram, also known to be involved in regulation of sucrose metabolism and light induction [27,46] or with AthDOF6.

For members of the GBF family, three main clades were suggested (Figure 6d). Clade A contains only AthGBF-4, which differs from the rest of the Arabidopsis GBFs and is unable to form homodimers [47]. In clade B, AthGBF-2 and AthGBF-3 from *A. thaliana* are grouped together with the AtqGBF-2 and AtqGBF-3 sequences from *A. tequilana* and in clade C, AthGBF-1 from *A. thaliana* is grouped together with four proteins from *A. tequilana*.
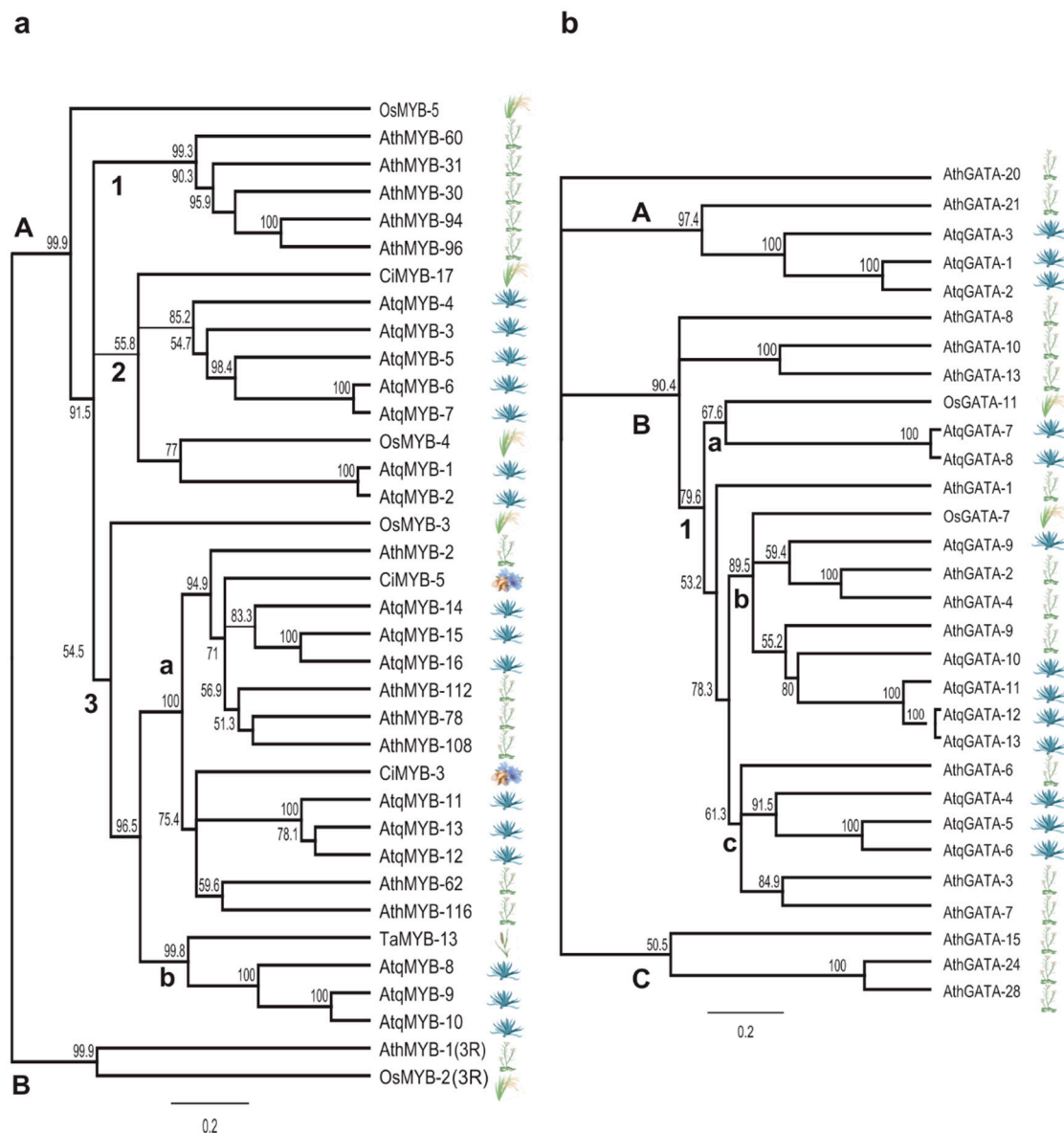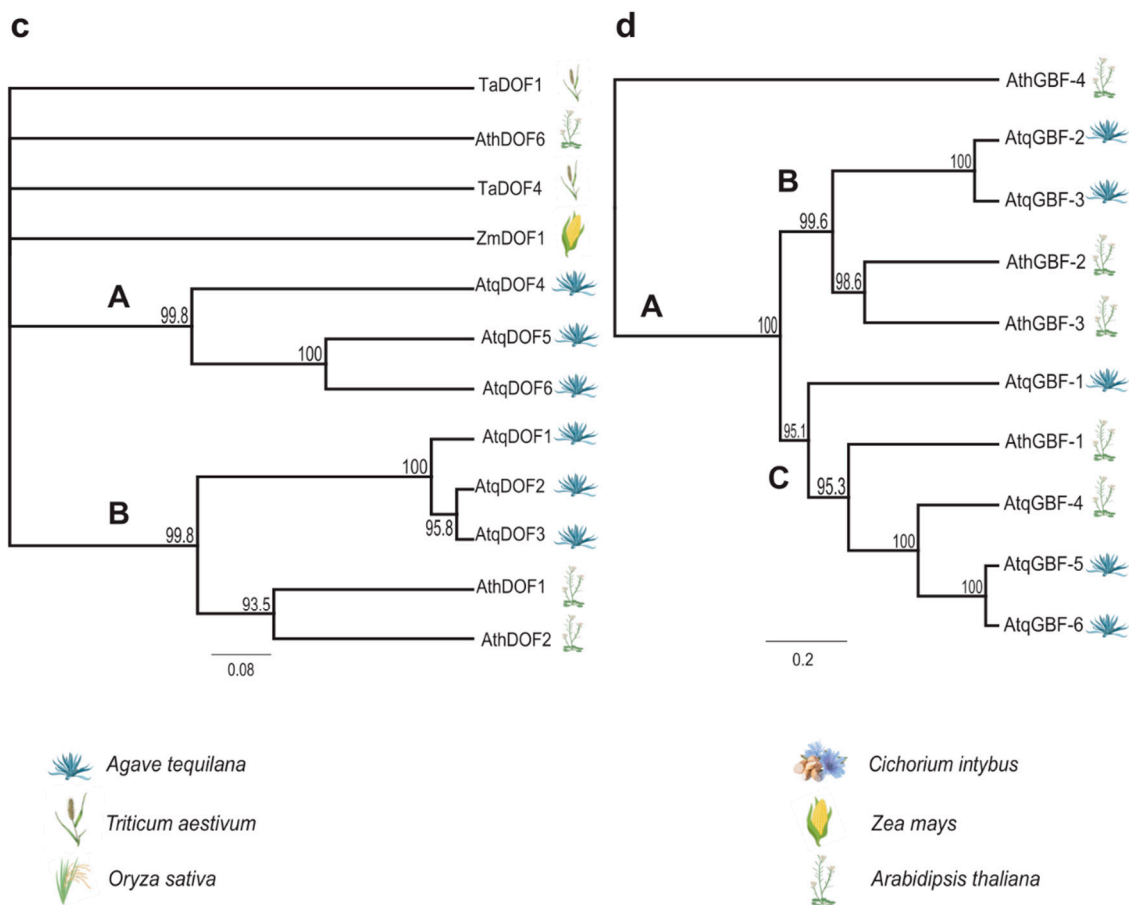

Figure 6. *Cont.*

Figure 6. (**a**–**d**) Relationships between amino acid sequences of the (**a**) MYB, (**b**) GATA, (**c**) DOF and (**d**) GBF transcription factors from *A. tequilana* and other plant species. Letters indicate different clades and numbers indicate Bootstrap analysis values. Phylogenetic analyses were carried out using complete amino acid sequences and considering all sites including gaps/missing information based on the UPGMA method and the Jukes-Cantor genetic distance model. Symbology represents the species analyzed.

3.4. In Silico Expression Patterns of MYB, GATA GBF and DOF Transcription Factors Identified for A. tequilana

In order to identify candidate transcription factors that could be responsible for the light- or tissue-specific modes of regulation of PGHF32 suggested by the promoter motif analysis and confirmed for leaf tissue by the qRT-PCR analysis, transcriptome databases were analyzed to determine the in silico expression patterns for each of the transcription factors described above (Figure 7). As can be observed, individual transcription factor-encoding genes showed distinct tissue expression patterns (Figure 7A,B). Specific genes for each selected transcription factor showed specific or very strong expression in anthers, leaves or ovaries, as in the case of AtqMYB-15, AtqGATA-1, Atq-GBF-2 and 3 and AtqDOF-4, whereas in contrast other genes showed null or very low expression in all tissues tested, such as AtqMYB-1, 10, 11 and 12, AtqGATA-4 and 9, AtqGBF-5 and AtqDOF-5 (Figure 7A). Figure 7B shows heatmaps of differential expression patterns in leaves and shoot apical meristem (SAM) tissue of the selected transcription factors during the vegetative to reproductive transition in *A. tequilana* and patterns specific to leaf (AtqGATA-1) or SAM (AtqGATA-9) tissue can be observed and for a specific tissue at a specific stage of the transition, for example, leaf tissue in the vegetative and sunken stages (AtqMYB-4) and SAM tissue at the vegetative stage (AtqMYB-9). In the case of the DOF genes, a pattern specific to meristem tissue at all developmental stages during the reproductive transition was observed for AtqDOF-4 and 5, whereas AtqDOF-1 was constitutively expressed at all

developmental stages and AtqDOF-2, 3 and 6 showed null or very low levels of expression. AtqGBF-2 and 3 show expression across all tissues tested but higher expression in floral and root tissue and AtqGBF-1, 2 and 5 are expressed at all stages and in both leaf and meristem tissue during the reproductive transition.
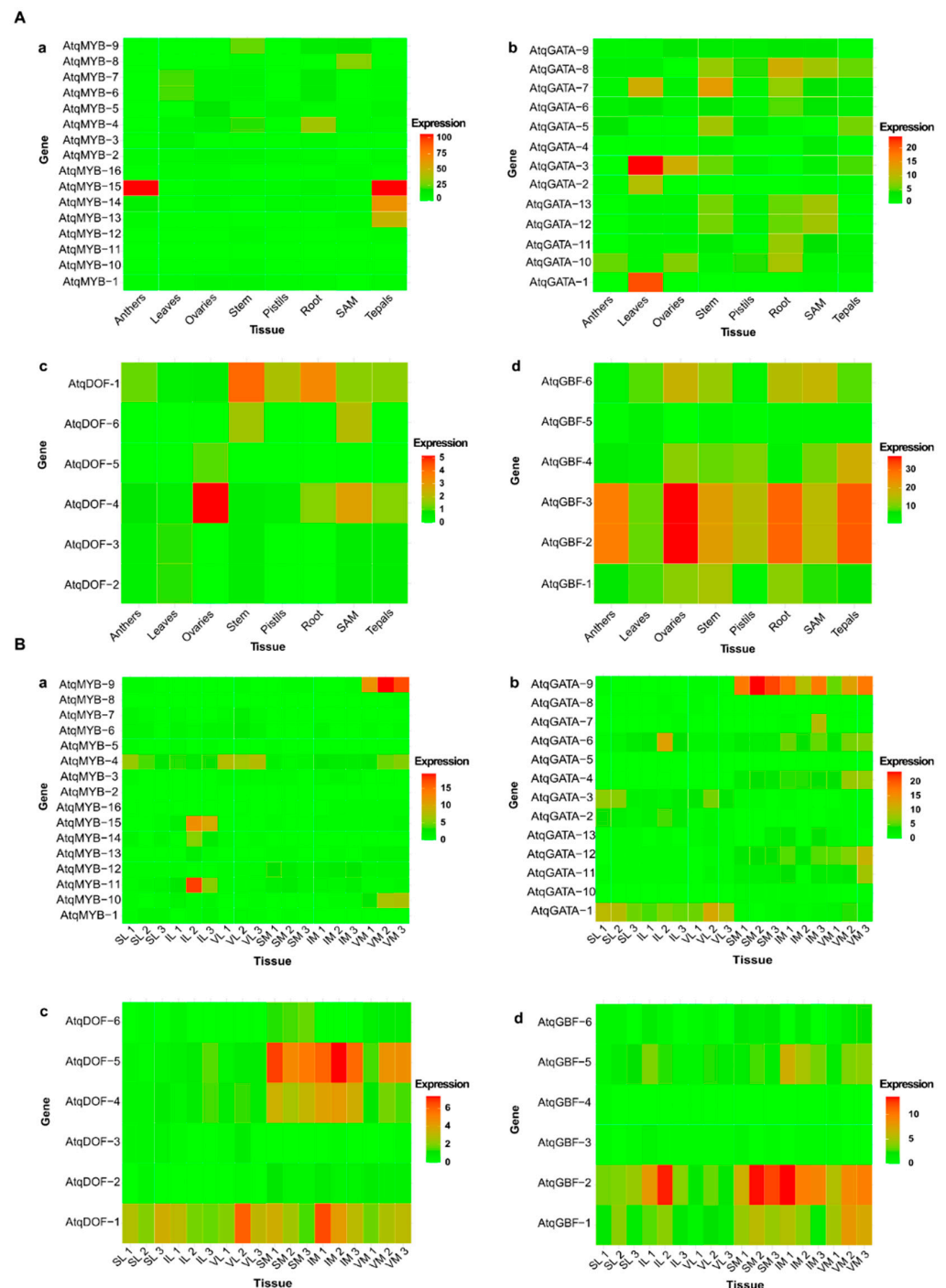
Figure 7. Heat diagrams of in silico gene expression patterns of (**a**) MYB genes, (**b**) GATA genes, (**c**) DOF genes and (**d**) GBF genes. (**A**) vegetative and floral tissues from *A. tequilana* plants as indicated on the x-axis, (2 different plants sampled for each tissue) and (**B**) during the vegetive to reproductive transition as indicated on the x-axis, (3 different plants sampled for each tissue) in *A. tequilana* plants. Red indicates the highest number of transcripts (expression) whereas green indicates the lowest number of transcripts (expression). VL, vegetative leaf, SL, sunken leaf, IL, inflorescence leaf, VM, vegetative shoot apical meristem, SM, sunken shoot apical meristem and IM, inflorescence meristem.

4. Discussion

Searches of the *A. tequilana* genome database (currently under development at the Advanced Genome Unit at Cinvestav, Mexico) allowed us to identify the genes belonging to PGHF32 in this species. In addition to genes previously identified using RNAseq data [20,21], 3 new genes closely related to genes determined to encode vacuolar invertases were found and tentatively named *AtqVinv-3, 4* and *5*, although confirmation of invertase activity has not yet been carried out. Genome analysis failed to identify candidates for 6-SFT type fructosyltransferase encoding genes, supporting the results from extensive RNAseq analysis in *A. tequilana*, *A. striata*, and *A. victoria-reginae*, where no transcripts encoding putative 6-SFT type enzymes have been found [20]. The complex agavin structures of *A. tequilana* [15] indicate that at least 4 enzyme activities are necessary for their synthesis: 1-SST, 1-FFT, 6G-FFT and 6-SFT. The lack of a separate 6-SFT encoding gene suggests that one of the other enzymes may carry out multiple activities, as has been reported for onion and asparagus, where a bifunctional 6G-FFT/1-FFT enzyme has been described [48,49]. The gene structures for *A. tequilana* PGHF32 are consistent with those described previously by Cortés-Romero et al. 2012 [37] for *Atq1SST-1* and *2*, *Atq6GFFT-1* and *2*, *AtqCwinv-1* and *AtqVinv-1*. For the first time, the gene structures for *Atq1SST-3, AtqVinv-2, 3, 4* and *5*, *Atqinv1* and *2*, *AtqCwinv-2* and *AtqFEH-1, 2, 3* and *4* are described. A third Cwinv gene (*AtqCwinv-3)* had been identified based on transcriptome data; however, *AtqCwinv-2* and *AtqCwinv-3* mapped to the same genome location, suggesting that the distinct transcripts identified previously are encoded by different alleles at the same locus.

All *A. tequilana* FT and Vinv encoding genes conserve the 8 exon/7 intron structure, whereas the Cwinv encoding genes and *Atqinv1* show a 7 exon/6 intron pattern where exons III and IV are fused. *Atqinv2* has a unique structure with a fusion of exons III and IV and V and VI. *Atqinv1* and *Atqinv2* have not yet been characterized in detail, and although they group with vacuolar invertases from other species (*A. deserti*, rice and maize), their position on the dendrogram and in a previous report indicates that although they are members of PGHF32, their putative function is unclear [20]. They may enter the category of defective invertases that have altered enzyme activities and putative regulatory functions [50]. All the *A. tequilana* FEH encoding genes lack the 9 nucleotide mini-exon characteristic of PGHF32 due to the fusion of exons II and III. In addition, *AtqFEH-3* lacks the mini-exon but also carries an extended first intron and fusion of exons III and IV to produce a 6 exon/5 intron pattern. Introns 2 and 3 are shown to be most variable in size and the relocation of the mini-exon closer to exon III in structures b′ and c may suggest a possible inversion of the region covering introns 1 and 2.

Comparison of PGHF32 gene structures from a selection of different plant species uncovered 7 different exon/intron patterns, which, with the exception of the 9/8 pattern, were shared across the taxonomic groups. However, superimposing the exon/intron patterns on a dendrogram showing the relationships between gene types revealed a very good correlation. The exon/intron structures observed for fructan-producing species also support the hypothesis that fructosyltransferases evolved from vacuolar invertases and that fructan exohydrolases evolved from cell wall invertases. For example, fructosyltransferases and vacuolar invertases from *A. tauschii*, wheat and barley all share either 4/3 or 3/2 exon/intron structures. In comparison, all cell wall invertases from these species have a 7/6 pattern and fructan exohydrolases have a mixture of 9/8 and 7/6 patterns. Similarly for *A. tequilana* and *A. officinalis*, all fructosyltransferase and vacuolar invertase genes have and 8/7 structures whereas cell wall invertase genes in common with *A. tauschii*, wheat and barley have the 7/6 structure and FEH genes have mainly a 6/5 structure.

Gene function, however, is not directly related to the evolution of the exon/intron patterns since, for example, all *A. tequilana* FT and vacuolar invertases share the same pattern but have clearly different enzyme activities. The only case where a good correlation between function and structure was observed in the case of the *A. tequilana* FEH genes where all have lost the second intron and the mini-exon is fused to exon III, whereas the Cwinv genes maintain intron 2 and the mini-exon. The presence or absence of the

mini-exon is also relevant for the regulation of PGHF32 genes since it has been shown that this 9 nucleotide exon, which encodes a highly conserved amino acid motif, can undergo differential splicing under specific conditions leading to regulation at the post transcriptional level [51]. In general, the modifications in exon/intron patterns can be explained by progressive loss of introns and in some cases (shown in red in Figure 2) can be directly attributed to loss of intron 2 and mini-exon fusion. A possible scenario for the progression of intron loss for *A. tequilana* PGHF32 is shown in Figure 8. As genome data become available, it will be interesting to extend comparisons of PGHF32 gene structure in both fructan synthesizing and non-synthesizing species and determine whether structural changes are related to gene function and regulation.
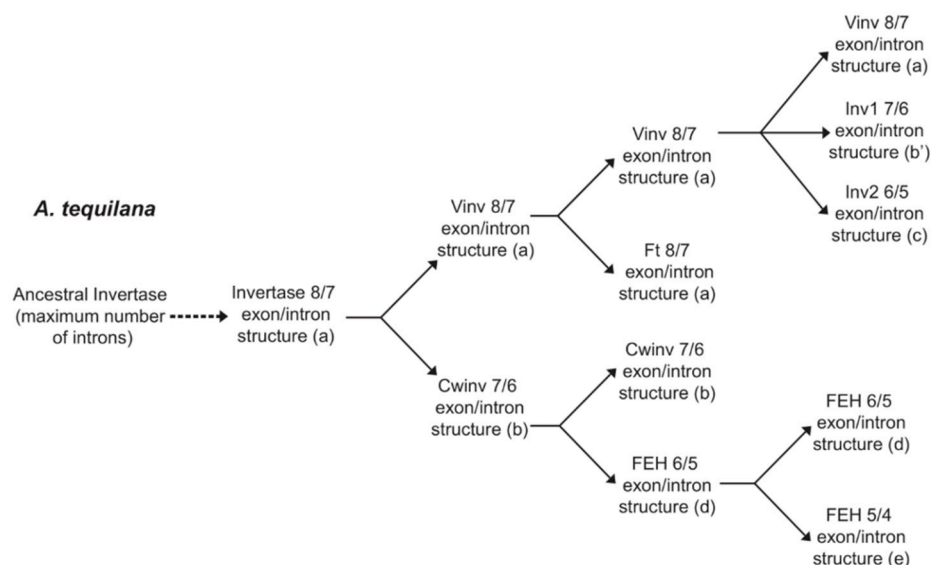
Figure 8. Diagram showing a possible scenario to describe how gene duplication and intron loss have produced the exon/intron patterns observed for the *A. tequilana* PGHF32 genes. Exon/intron patterns and structures a, b, b′, c, d and e are those described in Figure 1.

RNAseq data has shown that PGHF32 genes are differentially regulated in a tissue-specific manner and during different stages of development [20,21]. Individual genes are also regulated by stress and hormone treatments [23,24]. These results are consistent with reports for PGHF32 members in many other species [26,28,31]. The availability of genome data allowed investigation for the first time of the promoter regions of the *A. tequilana* PGHF32 genes in order to identify putative regulatory motifs responsible for their transcriptional regulation under distinct conditions and tissue types. Perhaps surprisingly, the most abundant motifs were found to be associated with light regulation; however, this may reflect the importance of both fructan and Crassulacean Acid Metabolism (CAM) for *Agave* species. In *Agaves*, only very low levels of starch, located mainly in the guard cells, are found in photosynthetically active tissues, whereas fructooligosaccharides are abundant [22,52] and may be transported to sink tissues [5]. This implies a need for close coordination between CAM photosynthesis cycles and fructan metabolism. Since few plant species combine both fructan and CAM metabolism [14,32], *Agaves* represent an important model for determining the coordinated regulation of these processes that also have significance for the capacity of *Agave* species to tolerate arid environments. One surprising aspect of the light regulatory motifs identified was the low number or absence of GBF motifs in the *A. tequilana* PGHF32 promoters since GBF has been shown in other species to be essential components of light regulation in association with the GATA motif [53]. It will therefore be of interest to explore in detail the light regulatory mechanisms in *A. tequilana*.

Other abundant motifs were consistent with previous reports of factors affecting the expression of genes encoding enzymes involved in fructan metabolism in both *Agaves* and other plant species [25–30]. The DOF and YACT motifs associated with carbohydrate

metabolism and photosynthetically active tissues, respectively, also support a strong coordination between fructan and CAM metabolism [54–56]. Regulation by phytohormones, such as cytokinins and especially gibberellins, is also consistent with previous reports on differential gene expression profiles during the vegetative to reproductive transition [29], and significant numbers of motifs associated with floral tissue (pollen and embryogenesis) and root tissue all confirm previous reports for PGHF32 genes in *A. tequilana* [28]. MYB transcription factors have been shown to play essential roles in the regulation of fructan metabolism genes in *C. intybus* and in response to light, biotic stress and sucrose or phytohormone signaling in *A. thaliana* [26,31,57]. However, although MYB motifs associated with dehydration were identified in all *A. tequilana* PGHF32 genes, they were less abundant than those associated with the other transcription factors described above. The dendrograms comparing *A. tequilana* transcription factor proteins with those of other species in some cases showed very low bootstrap values, indicating that for these clades or branches the topology is less robust. This may reflect the observation that although the characteristic domains for each transcription factor are highly conserved, as shown in Figure 5, these proteins are quite heterogeneous in the surrounding sequences. The inclusion of greater numbers of proteins from a wider variety of species in the comparisons should lead to dendrograms with more robust topologies.

The qRT-PCR analyses experimentally confirm that the selected group of PGHF32 genes are regulated by light and show tissue-specific patterns of regulation in leaves. Although all the PGHF32 genes shared elements for light- and tissue-specific regulation, they did not all show the same regulatory responses. Whereas the FT encoding genes (*Atq1SST-1, Atq6GFFT1* and *2*) and the vacuolar invertase gene *AtqVinv-1* all showed clear effects of light regulation, the exohydrolase gene *AtqFEH-4* was expressed at a low level and showed a slight induction under darkness. When green and white leaf tissues were compared, the 3 FT encoding genes shared the same pattern of high expression in white leaf tissue in comparison to green leaf tissue, whereas *AtqVinv-1* was repressed in white tissue in comparison to green tissue. On the other hand, *AtqFEH-4* was slightly induced in white tissue compared to green tissue. These results reflect the complexity and precision of transcriptional gene regulation accomplished by the interactions of a battery of transcription factors that respond to localized cellular conditions or metabolites, as well as environmental factors, such as light or water stress.

Although the qRT-PCR results do not include the whole set of PGHF32 genes, the expression patterns observed are consistent with previous reports on fructan metabolism in *Agave* species, as summarized in Figure 9. Sucrose is synthesized as the product of photosynthesis and can be exploited for the synthesis of fructans in green leaf tissue. Both sucrose and fructans can be mobilized to sink tissues, such as the stem. Sucrose can be used to synthesize more complex fructans in the leaf base and stem tissue or converted to starch in the peripheral growth meristem (to be utilized for growth and expansion of the stem). It is consistent that invertase activity is lower in white tissue so that sucrose can be used by fructosyltransferases and starch synthesizing enzymes. It is possible that a low level of fructo exohydrolase activity is needed to ensure that mainly fructooligosaccharides are present in leaf base tissue and can be transported to the stem where they are then converted into more complex fructan polymers for long-term storage.

The in silico expression patterns determined for the transcription factor genes are also consistent with putative roles in the regulation of PGHF32 genes in *A. tequilana.* Specific GATA encoding genes are highly expressed in leaf tissue in general and in SAM tissue at different stages of the vegetative to reproductive transition. Specific DOF genes are strongly expressed in floral and stem tissue and in SAM tissue throughout the vegetative to reproductive transition, perhaps reflecting changes in carbohydrate metabolism in these actively growing tissues. Specific MYB genes are also strongly expressed in floral tissues, leaves and in vegetative SAM tissue. GBF transcription factors showed strong expression in floral and root tissues, consistent with patterns of PGHF32 expression.

　　　In order to isolate and characterize in detail the FT encoding genes involved in the regulation of *A. tequilana* PGHF32, it is essential to identify the best candidates of these multigene families based on their phylogenetic relations with previously characterized genes in other plant species and on their expression patterns in *A. tequilana*. For example, AtqGATA-1 and 3 are strongly expressed in leaf tissue and are found in a clade containing genes from *A. thaliana* known to be involved in light and circadian regulation. AtqGATA-1 and 3 are therefore strong candidates for participating in the light regulation of the PGHF32 genes. Knowledge of in silico transcription levels in combination with the full-length cDNAs used to produce the dendrograms are useful tools for the cloning of the candidate TF genes for further analysis.
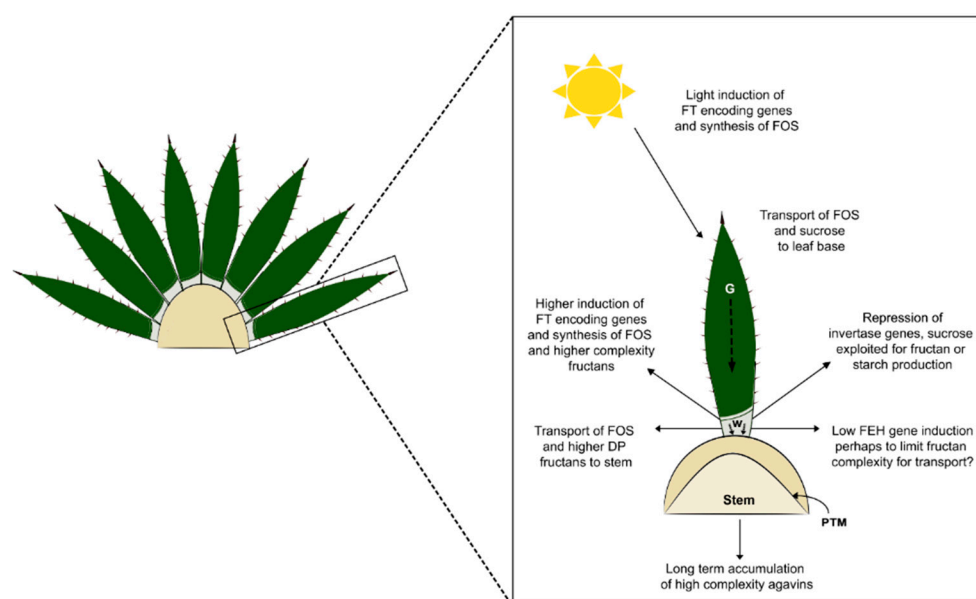
Figure 9. Model for the integration of light and leaf specific regulation of PGHF32 members into overall carbohydrate metabolism in *A. tequilana*. FT, fructosyltransferases, FOS, fructooligosaccharide, DP, degree of polymerization, G, green leaf tissue, W, white leaf tissue, PTM, primary thickening meristem.

5. Conclusions

　　　Based on genome analysis, members of *A. tequilana* PGHF32 previously identified by RNAseq were confirmed with 3 newly identified Vinv genes completing the family. The lack of a separate 6-SFT encoding gene was also confirmed. Analysis of gene structures for all *A. tequilana* PGHF32 genes revealed exon/intron patterns consistent with other plant species, although specific for *Agaves* and supports the hypothesis of evolution of FT from vacuolar invertases and FEH from cell wall invertases. In addition, the presence of distinct exon/intron patterns can be attributed to the loss of specific introns and fusion of surrounding exons. Analysis of promoter motifs suggests that PGHF32 genes in *A. tequilana* are strongly regulated by light and are tissue specific. In silico expression analysis, in combination with phylogenic analysis comparing selected *A. tequilana* transcription factor genes with those of other plant species, identified candidate genes for future detailed characterization of the regulatory mechanisms associated with PGHF32.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants11162153/s1, Figure S1: *Agave tequilana* leaf showing the photosynthetically active aerial green tissue and the non-photosynthetically active basal white tissue; Figure S2: Example of the pattern of abundance of regulatory motifs found in *Agave tequilana* PGHF32 promoter regions based on *A. tequilana* FEH genes; Figure S3: qRT-PCR analysis of PGHF32 genes in green (G) and white (W) leaf regions; Table S1: List of species and accession numbers for PGHF32 sequences used in alignments and phylogenetic analyses; Table S2: List of species and accession numbers for all sequences of transcription factors MYB, GATA, DOF and GBF used in alignments and

phylogenetic analyses; Table S3: Gene structure conformation of PGHF32 genes in monocotyledonous and dicotyledonous species; Table S4: Nucleotide conformation in exons and introns of genes encoding the PGHF32 enzymes in *Agave tequilana*.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: *A. tequilana* draft genome (Herrera-Estrella et al. unpublished), permission to access the data prior to publication can be arranged by directly contacting Dr. Alfredo Herrera-Estrella: (alfredo.herrera@cinvestav.mx) or Dr. Selene Fernández Valverde: (selene.fernandez@cinvestav.mx).

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

References

1. Hendry, G. Evolutionary origins and natural functions of fructans—A climatological, biogeographic and mechanistic appraisal. *New Phytol.* **1993**, *123*, 3–14. [CrossRef]
2. Banguela, A.; Hernández, L. Fructans: From natural sources to transgenic plants. *Biotecnol. Apl.* **2006**, *23*, 202–210.
3. Versluys, M.; Kirtel, O.; Toksoy Öner, E.; Van den Ende, W. The fructan syndrome: Evolutionary aspects and common themes among plants and microbes. *Plant Cell Environ.* **2017**, *41*, 16–38. [CrossRef]
4. Pilon-Smits, E.; Ebskamp, M.; Paul, M.J.; Jeuken, M.; Weisbeek, P.J.; Smeekens, S. Improved Performance of Transgenic Fructan-Accumulating Tobacco under Drought Stress. *Plant Physiol.* **1995**, *107*, 125–130. [CrossRef] [PubMed]
5. Wang, N.; Nobel, P.S. Phloem Transport of Fructans in the Crassulacean Acid Metabolism Species *Agave deserti*. *Plant Physiol.* **1998**, *116*, 709–714. [CrossRef] [PubMed]
6. Kawakami, A.; Yoshida, M. Molecular characterization of sucrose:sucrose 1-fructosyltransferase and sucrose: Fructan 6-fructosyltransferase associated with fructan accumulation in winter wheat during cold hardening. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 2297–2305. [CrossRef] [PubMed]
7. Hisano, H.; Kanazawa, A.; Yoshida, M.; Humphreys, M.O.; Iizuka, M.; Kitamura, K.; Yamada, T. Coordinated expression of functionally diverse fructosyltransferase genes is associated with fructan accumulation in response to low temperature in perennial ryegrass. *New Phytol.* **2008**, *178*, 766–780. [CrossRef]
8. Bieleski, R.L. Fructan hydrolysis drives petal expansion in the ephemeral daylily flower. *Plant Physiol.* **1993**, *103*, 213–219. [CrossRef]
9. Vergauwen, R.; Van den Ende, W.; Van Laere, A. The role of fructan in flowering of *Campanula rapunculoides*. *J. Exp. Bot.* **2000**, *51*, 1261–1266. [CrossRef]
10. Valluru, R. Fructan and hormone connections. *Front. Plant Sci.* **2015**, *6*, 180. [CrossRef]
11. Versluys, M.; Tarkowski, Ł.P.; Van den Ende, W. Fructans as DAMPs or MAMPs: Evolutionary prospects, cross-tolerance, and multistress resistance potential. *Front. Plant Sci.* **2017**, *7*, 2061. [CrossRef] [PubMed]
12. Narváez-Zapata, J.A.; Sánchez-Teyer, L.F. Agaves as a raw material: Recent technologies and applications. *Recent Pat. Biotechnol.* **2009**, *3*, 185–191. [CrossRef] [PubMed]
13. Cushman, J.C.; Davis, S.C.; Yang, X.; Borland, A.M. Development and use of bioenergy feedstocks for semi-arid and arid lands. *J. Exp. Bot.* **2015**, *66*, 4177–4193. [CrossRef] [PubMed]
14. Davis, S.C.; Simpson, J.; Gil Vega, K.C.; Niechayev, N.A.; van Tongerlo, E.; Hurtado Castano, N.; Dever, L.V.; Búrquez, A. Undervalued potential of crassulacean acid metabolism for current and future agricultural production. *J. Exp. Bot.* **2019**, *70*, 6521–6537. [CrossRef]
15. Mancilla-Margalli, N.A.; Lopez, M.G. Water-soluble carbohydrates and fructan structure patterns from *Agave* and *Dasylirion* species. *J. Agric. Food Chem.* **2006**, *54*, 7832–7839. [CrossRef]

16. Mellado-Mojica, E.; López, M.G. Fructan metabolism in *A. tequilana* Weber Blue variety along its developmental cycle in the field. *J. Agric. Food Chem.* **2012**, *60*, 11704–11713. [CrossRef]
17. Van den Ende, W. Multifunctional fructans and raffinose family oligosaccharides. *Front. Plant Sci.* **2013**, *4*, 247. [CrossRef]
18. Wei, J.Z.; Chatterton, N.J. Fructan biosynthesis and fructosyltransferase evolution: Expression of the 6-SFT (sucrose: Fructan 6-fructosyltransferase) gene in crested wheatgrass (*Agropyron cristatum*). *J. Plant Physiol.* **2001**, *158*, 1203–1213. [CrossRef]
19. Van Den Ende, W.; De Coninck, B.; Clerens, S.; Vergauwen, R.; Van Laere, A. Unexpected presence of fructan 6-exohydrolases (6-FEHs) in non-fructan plants: Characterization, cloning, mass mapping and functional analysis of a novel "cell-wall invertase-like" specific 6-FEH from sugar beet (*Beta vulgaris* L.). *Plant J.* **2003**, *36*, 697–710. [CrossRef]
20. Avila de Dios, E.; Gomez Vargas, A.D.; Damián Santos, M.L.; Simpson, J. New insights into plant glycoside hydrolase family 32 in Agave species. *Front. Plant Sci.* **2015**, *6*, 594. [CrossRef]
21. Avila de Dios, E.; Delaye, L.; Simpson, J. Transcriptome analysis of bolting in *A. tequilana* reveals roles for florigen, MADS, fructans and gibberellins. *BMC Genom.* **2019**, *20*, 473. [CrossRef] [PubMed]
22. Pérez López, A.V.; Simpson, J.; Clench, M.R.; Gomez Vargas, A.D.; Ordaz Ortiz, J.J. Localization and Composition of Fructans in Stem and Rhizome of *Agave tequilana* Weber var. azul. *Front. Plant Sci.* **2021**, *11*, 608850. [CrossRef] [PubMed]
23. Suárez-González, E.M.; López, M.G.; Délano-Frier, J.P.; Gómez-Leyva, J.F. Expression of the 1-SST and 1-FFT genes and consequent fructan accumulation in *Agave tequilana* and *A. inaequidens* is differentially induced by diverse abiotic stress related elicitors. *J. Plant Physiol.* **2014**, *171*, 359–372. [CrossRef] [PubMed]
24. Suárez-González, E.M.; Palmeros Suárez, P.A.; Cruz-Rubio, J.M.; Martínez-Gallardo, N.A.; Cisneros Hernández, I.; Délano-Frier, J.P.; Gómez-Leyva, J.F. Differential fructan accumulation and expression of fructan biosynthesis, invertase and defense genes is induced in *Agave tequilana* plantlets by sucrose or stress-related elicitors. *Agri Gene* **2016**, *2*, 17–28. [CrossRef]
25. Kooiker, M.; Drenth, J.; Glassop, D.; McIntyre, C.L.; Xue, G.P. TaMYB13-1, a R2R3 MYB transcription factor, regulates the fructan synthetic pathway and contributes to enhanced fructan accumulation in bread wheat. *J. Exp. Bot.* **2013**, *64*, 3681–3696. [CrossRef]
26. Wei, H.; Zhao, H.; Su, T.; Bausewein, A.; Greiner, S.; Harms, K.; Rausch, T. *Chicory* R2R3-MYB transcription factors CiMYB5 and CiMYB3 regulate fructan 1-exohydrolase expression in response to abiotic stress and hormonal cues. *J. Exp. Bot.* **2017**, *68*, 4323–4338. [CrossRef]
27. Shaw, L.M.; McIntyre, C.L.; Gresshoff, P.M.; Xue, G.P. Members of the Dof transcription factor family in *Triticum aestivum* are associated with light-mediated gene regulation. *Funct. Integr. Genom.* **2009**, *9*, 485–498. [CrossRef]
28. Xue, G.P.; Kooiker, M.; Drenth, J.; McIntyre, C.L. TaMYB13 is a transcriptional activator of fructosyltransferase genes involved in β-2,6-linked fructan synthesis in wheat. *Plant J.* **2011**, *68*, 857–870. [CrossRef]
29. Chiang, Y.H.; Zubo, Y.O.; Tapken, W.; Kim, H.J.; Lavanway, A.M.; Howard, L.; Pilon, M.; Kieber, J.J.; Schaller, G.E. Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. *Plant Physiol.* **2012**, *160*, 332–348. [CrossRef]
30. Zhao, H.; Greiner, S.; Scheffzek, K.; Rausch, T.; Wang, G.A. 6&1-FEH encodes an enzyme for fructan degradation and interact with invertase inhibitor protein in maize (*Zea mays* L.). *Int. J. Mol. Sci.* **2019**, *20*, 3807. [CrossRef]
31. Wei, H.; Bausewein, A.; Greiner, S.; Dauchot, N.; Harms, K.; Rausch, T. CiMYB17, a stress-induced chicory R2R3-MYB transcription factor, activates promoters of genes involved in fructan synthesis and degradation. *New Phytol.* **2017**, *215*, 281–298. [CrossRef] [PubMed]
32. Escamilla-Treviño, L. Potential of Plants from the Genus Agave as Bioenergy Crops. *Bioenergy Res.* **2011**, *5*, 1–9. [CrossRef]
33. Altschul, S.F.; Gertz, E.M.; Agarwala, R.; Schäffer, A.A.; Yu, Y.K. PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res.* **2009**, *37*, 815–824. [CrossRef] [PubMed]
34. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113. [CrossRef]
35. Higo, K.; Ugawa, Y.; Iwamoto, M.; Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* **1999**, *27*, 297–300. [CrossRef]
36. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]
37. Cortés-Romero, C.; Martínez-Hernández, A.; Mellado-Mojica, E.; López, M.G.; Simpson, J. Molecular and Functional Characterization of Novel Fructosyltransferases and Invertases from *Agave tequilana*. *PLoS ONE* **2012**, *7*, e35878. [CrossRef]
38. Sibéril, Y.; Doireau, P.; Gantet, P. Plant bZIP G-box binding factors. *Eur. J. Biochem.* **2001**, *268*, 5655–5666. [CrossRef]
39. Jakoby, M.; Weisshaar, B.; Dröge-Laser, W.; Vicente-Carbajosa, J.; Tiedemann, J.; Kroj, T.; Parcy, F. bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci.* **2002**, *7*, 106–111. [CrossRef]
40. Ezer, D.; Shepherd, S.J.K.; Brestovitsky, A.; Dickinson, P.; Cortijo, S.; Charoensawan, V.; Box, M.S.; Biswas, S.; Jaeger, K.E.; Wigge, P.A. The G-box transcriptional regulatory code in arabidopsis. *Plant Physiol.* **2017**, *175*, 628–640. [CrossRef]
41. Wang, W.; Sijacic, P.; Xu, P.; Lian, H.; Liu, Z. *Arabidopsis* TSO1 and MYB3R1 form a regulatory module to coordinate cell proliferation with differentiation in shoot and root. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3045–E3054. [CrossRef] [PubMed]
42. Gupta, P.; Nutan, K.K.; Singla-Pareek, S.L.; Pareek, A. Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice. Front. *Front. Plant Sci.* **2017**, *8*, 1944. [CrossRef] [PubMed]

43. Manfield, I.W.; Devlin, P.F.; Jen, C.H.; Westhead, D.R.; Gilmartin, P.M. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the *Arabidopsis* GATA gene family. *Plant Physiol.* **2007**, *143*, 941–958. [CrossRef]

44. Soitamo, A.J.; Piippo, M.; Allahverdiyeva, Y.; Battchikova, N.; Aro, E.M. Light has a specific role in modulating *Arabidopsis* gene expression at low temperature. *BMC Plant Biol.* **2008**, *8*, 13. [CrossRef]

45. Tsujimoto-Inui, Y.; Naito, Y.; Sakurai, N.; Suzuki, H.; Sasaki, R.; Takahashi, H.; Ohtsuki, N.; Nakano, T.; Yanagisawa, S.; Shibata, D.; et al. Functional genomics of the Dof transcription factor family genes in suspension-cultured cells of *Arabidopsis thaliana*. *Plant Biotechnol.* **2009**, *26*, 15–28. [CrossRef]

46. Noguero, M.; Atif, R.M.; Ochatt, S.; Thompson, R.D. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Sci.* **2013**, *209*, 32–45. [CrossRef]

47. Ramegowda, V.; Gill, U.S.; Sivalingam, P.N.; Gupta, A.; Gupta, C.; Govind, G.; Nataraja, K.N.; Pereira, A.; Udayakumar, M.; Mysore, K.S.; et al. GBF3 transcription factor imparts drought tolerance in *Arabidopsis thaliana*. *Sci. Rep.* **2017**, *7*, 9148. [CrossRef]

48. Ritsema, T.; Smeekens, S.C.M. Engineering fructan metabolism in plants. *J. Plant Physiol.* **2003**, *160*, 811–820. [CrossRef]

49. Ueno, K.; Onodera, S.; Kawakami, A.; Yoshida, M.; Shiomi, N. Molecular characterization and expression of a cDNA encoding fructan:fructan 6G-fructosyltransferase from asparagus (*Asparagus officinalis*). *New Phytol.* **2005**, *165*, 813–824. [CrossRef]

50. Le Roy, K.; Vergauwen, R.; Struyf, T.; Yuan, S.; Lammens, W.; Mátrai, J.; De Maeyer, M.; Van den Ende, W. Understanding the role of defective invertases in plants: Tobacco Nin88 fails to degrade sucrose. *Plant Physiol.* **2013**, *161*, 1670–1681. [CrossRef]

51. Simpson, C.G.; Hedley, P.E.; Watters, J.A.; Clark, G.P.; McQuade, C.; Machray, G.C.; Brown, W.S. Requirements for mini-exon inclusion in potato invertase mRNAs provide evidence for exon-scanning interactions in plants. *RNA* **2000**, *6*, 422–433. [CrossRef] [PubMed]

52. Zavala-García, L.E.; Sánchez-Segura, L.; Avila de Dios, E.; Pérez-López, A.; Simpson, J. Starch accumulation is associated with active growth in *A. tequilana*. *Plant Physiol. Biochem.* **2018**, *130*, 623–632. [CrossRef] [PubMed]

53. Jiao, Y.; Lau, O.; Deng, X. Light-regulated transcriptional networks in higher plants. *Nat. Rev. Genet.* **2007**, *8*, 217–2300. [CrossRef]

54. Yanagisawa, S. Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *Plant J.* **2000**, *21*, 281–288. [CrossRef] [PubMed]

55. Yang, G.; Gao, X.; Ma, K.; Li, D.; Jia, C.; Zhai, M.; Xu, Z. The walnut transcription factor JrGRAS2 contributes to high temperature stress tolerance involving in Dof transcriptional regulation and HSP protein expression. *BMC Plant Biol.* **2018**, *18*, 367. [CrossRef] [PubMed]

56. Banu, S.; Huda, K.; Tuteja, N. Isolation and functional characterization of the promoter of a DEAD-box helicase Psp68 using Agrobacterium-mediated transient assay. *Plant Signal. Behav.* **2014**, *9*, e28992. [CrossRef] [PubMed]

57. Kelemen, Z.; Sebastian, A.; Xu, W.; Grain, D.; Salsac, F.; Avon, A.; Berger, N.; Tran, J.; Dubreucq, B.; Lurin, C.; et al. Analysis of the DNA-Binding Activities of the Arabidopsis R2R3-MYB Transcription Factor Family by One-Hybrid Experiments in Yeast. *PLoS ONE* **2015**, *10*, e0141044. [CrossRef]