

Research Article

Transcriptome Comparison Reveals Distinct Selection Patterns in Domesticated and Wild Agave Species, the Important CAM Plants

Xing Huang¹, Bo Wang², Jingen Xi¹, Yajie Zhang³, Chunping He¹, Jinlong Zheng¹, Jianming Gao⁴, Helong Chen⁴, Shiqing Zhang⁴, Weihuai Wu¹, Yanqiong Liang¹, and Kexian Yi¹

¹Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

²National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei 430070, China

³Hainan Climate Center, Haikou 570203, China

⁴Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

Correspondence should be addressed to Xing Huang; hxalong@gmail.com and Kexian Yi; yikexian@126.com

Received 27 February 2018; Revised 21 August 2018; Accepted 24 September 2018; Published 22 November 2018

Academic Editor: Giandomenico Corrado

Copyright © 2018 Xing Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Agave species are an important family of crassulacean acid metabolism (CAM) plants with remarkable tolerance to heat and drought stresses (*Agave deserti*) in arid regions and multiple agricultural applications, such as spirit (*Agave tequilana*) and fiber (*Agave sisalana*) production. The agave genomes are commonly too large to sequence, which has significantly restricted our understanding to the molecular basis of stress tolerance and economic traits in agaves. In this study, we collected three transcriptome databases for comparison to reveal the phylogenetic relationships and evolution patterns of the three agave species. The results indicated the close but distinctly domesticated relations between *A. tequilana* and *A. sisalana*. Natural abiotic and biotic selections are very important factors that have contributed to distinct economic traits in agave domestication together with artificial selection. Besides, a series of candidate unigenes regulating fructan, fiber, and stress response-related traits were identified in *A. tequilana*, *A. sisalana*, and *A. deserti*, respectively. This study represents the first transcriptome comparison within domesticated and wild agaves, which would serve as a guidance for further studies on agave evolution, environmental adaptation, and improvement of economically important traits.

1. Introduction

Agave species assembled an important group of crassulacean acid metabolism (CAM) plants with remarkable tolerance to heat and drought stresses in arid regions [1]. CAM plants usually have a higher efficiency in water use than C3 and C4 plants [2]. For this reason, CAM plants brought a great chance to enhance sustainable production of food and bioenergy under the background of limited freshwater resources and global climate change [3]. As a traditional cultivated CAM plant, *Agave tequilana* has been used for the production of distilled spirit tequila for centuries [4]. Further, it also shows a great potential in bioenergy production [5]. Besides, *Agave sisalana* has also been widely cultivated as a cash crop

for fiber production in tropical regions [6]. As a native wild plant in the Sonoran Desert regions of the Southwestern United States and Northwestern Mexico, *Agave deserti* has successfully survived in a severe environment within elevation ranges that experience both hot, dry summers and occasional freezing temperatures in winter [7, 8]. This capacity of high tolerance to multiple stresses has important values to the improvement of the main food crops. These economic and stress-tolerant features have made agave a model CAM crop system for a hot and dry/droughty/xeric environment [9].

However, few reports have revealed the physiology and molecular basis of agaves, especially in *A. sisalana*. To date, *A. tequilana*-related studies mainly focused on fructan, with the obvious purpose of improving fructan production and

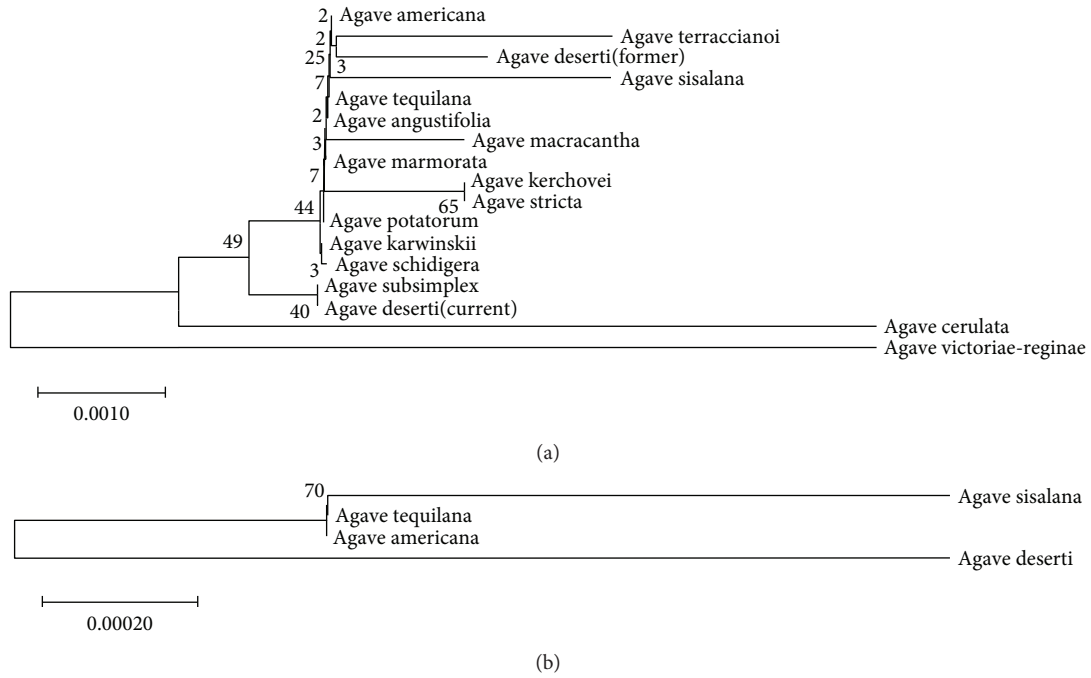


FIGURE 1: Phylogenetic analysis of chloroplast sequences in *Agave* species. (a) The *trnL*+*trnL*-*trnF* sequences (about 900 bp) were used to construct a phylogenetic tree. (b) The sequences of *A. tequilana* (GAHU01110124), *A. deserti* (GAHT01022741), and *A. americana* (KX519714) were from NCBI. *A. sisalana* (CL7065.Contig2) sequence and the other 13 sequences were from previous studies [15, 24].

TABLE 1: Summary of the transcriptomes for the 3 *Agave* species.

Species	<i>A. tequilana</i>	<i>A. deserti</i>	<i>A. sisalana</i>
Total sequenced high-quality data	293.5 Gbp	184.7 Gbp	11.3 Gbp
Number of <i>Agave</i> unigenes	204,530	128,869	131,422
Sum length of <i>Agave</i> unigenes	204.9 Mbp	125.0 Mbp	77.6 Mbp
N50 length	1387 bp	1323 bp	861 bp

application to generate bioenergy [5, 10, 11]. *A. deserti*-related studies were mainly conducted for its ecological and physiological adaptation to a severe environment, which is highly valuable for the improvement of the main food crops [12]. Besides, a series of saponin-related researches have been reported, while few reports were related to fiber in *A. sisalana* [13, 14]. These three agave species are closely related species but with different remarkable biological features. A previous study has revealed the phylogenetic relations according to their *trnL* sequences [15]. The result showed that they were closely related species in spite of different origins for *A. tequilana* (Jalisco), *A. deserti* (the Sonoran Desert), and *A. sisalana* (Chiapas) [4, 6, 7]. Agaves have very large genomes, which has significantly restricted their genome assembly and limited our understanding to their evolution pattern [16]. In other crops, accessible genomes and genome-wide association analysis have revealed many economically important traits for crop improvement [17–19]. Recently, the development of next-generation sequencing has brought a new direction for gene-related studies without the restriction of genome data [20, 21]. Furthermore, transcriptome

comparison has also been conducted for evolution analysis and searching economically important traits in some genome unavailable crops [22]. In this study, we selected three transcriptome databases for comparison to reveal the phylogeny and evolution pattern of the three agave species [23, 24]. Those genes related to species-specific traits would be also identified and evaluated for their importance in agronomy production and environmental adaptation of agaves. This study represents the first transcriptome comparison within domesticated and wild agaves, which would serve as a guidance for further studies on agave evolution, environmental adaptation, and improvement of economically important traits.

2. Materials and Methods

2.1. Phylogenetic Analysis. Phylogenetic analysis was conducted by MEGA 5.0 software with the minimum-evolution method [25]. The methods and parameters were according to the previous study [15]. The bootstrap method was employed for confidence in nodes with 1000 replicates. Partial chloroplast sequences for *A. tequilana* and *A. deserti* were

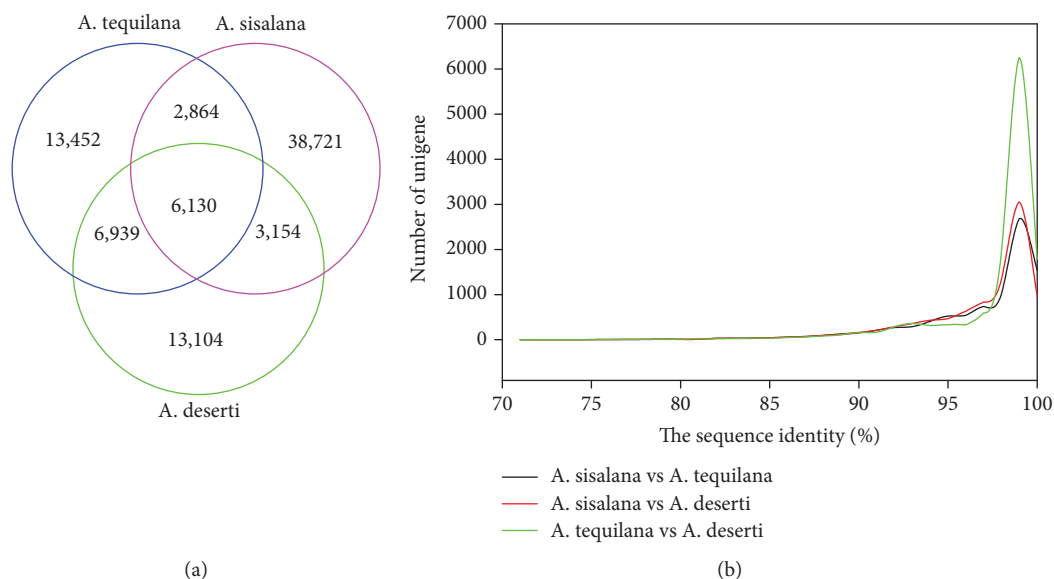


FIGURE 2: Sequence comparisons between domesticated *A. tequilana* and *A. sisalana* and wild *A. deserti* transcriptomes. (a) The sketch map showing 6130 unigene terms that were identified within the three transcriptomes. (b) The identity distribution of all orthologous unigene pairs.

downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/>) with the accession numbers GAHU01110124 and GAHT01022741, respectively. Partial chloroplast sequences for *A. sisalana* were obtained from a previous transcriptome database [24]. The whole chloroplast sequence of *A. americana* was obtained from NCBI under the accession number KX519714. The trnL+trnL-trnF sequences (about 900 bp) for the 4 and other 14 agave species were also downloaded according to the previous study [15].

2.2. Transcriptome Data and Gene Annotation. Three agave transcriptome databases were according to previous studies [23, 24]. Transcriptome assembly of the Illumina sequence was performed by Rnnotator for each species, respectively [26]. Only unigenes with corresponding predicted proteins in *A. deserti* transcriptome were used to search orthologous genes from *A. tequilana* and *A. sisalana* transcriptomes by the BBH method, respectively [27]. These orthologous unigenes in three transcriptomes were annotated in the public databases: NCBI nonredundant protein (Nr) and nonredundant nucleotide (Nt) databases (<http://www.ncbi.nlm.nih.gov/>), Swiss-Prot (<http://www.ebi.ac.uk/uniprot/>), and Gene Ontology (GO) (<http://www.geneontology.org/>), respectively.

2.3. Ka/Ks Analysis. It represents positive selection when the ratio of nonsynonymous (Ka) to synonymous nucleotide substitutions (Ks) is significantly higher than 1, whereas the ratios significantly less than or equal to 1 are subjected to purifying or neutral selection [28]. The Ka, Ks, and Ka/Ks values were estimated by the Codeml model of the program of phylogenetic analysis by maximum likelihood (PAML) between *A. deserti* unigenes with orthologous unigenes in *A. tequilana* or *A. sisalana*, respectively [29].

2.4. In Silico Gene Expression Analysis. The expression pattern of positive selected unigenes was subjected to in silico gene expression analysis in agave leaves. The reads per kilobase per million mapped read (RPKM) value of these unigenes in *A. deserti* and *A. tequilana* were calculated by RSEM software according to the previous study [23, 30]. The RPKM data was further normalized with two reference genes (tubulin and serine/threonine-protein phosphatase) in each agave species [31].

2.5. Selection Pressure Detection and Protein Structure Modeling. To detect the selection pressure on positive selection unigenes, Ka/Ks ratios were calculated in sliding window (30 bp under a step size of 6 bp) by using DnaSP 5.0 [32]. Translated protein sequences of positive selection unigenes were used for structure modeling by Swiss-Model (<https://swissmodel.expasy.org/>) [33].

3. Results

3.1. Phylogeny of Agave Species. Phylogenetic analysis was conducted to reveal the phylogenetic relation for agave species in this study. We obtained chloroplast sequences for *A. tequilana* (GAHU01110124), *A. sisalana* (CL7065.Contig2), and *A. deserti* (GAHT01022741), by using blast against three agave transcriptomes [23, 24]. The chloroplast sequence of *A. americana* was from GenBank under the accession number KX519714 (<http://www.ncbi.nlm.nih.gov/genbank/>). The trnL+trnL-trnF sequences (about 900 bp) from the 4 and other 13 agave species [15] were employed for phylogenetic analysis to reveal their evolutionary pattern. The results indicated that *A. tequilana*, *A. sisalana*, *A. americana*, and *A. deserti* (former, DQ500894+DQ500928) sequences were grouped together, while the *A. deserti* (GAHT01022741) sequence was in another group (Figure 1(a)). We further

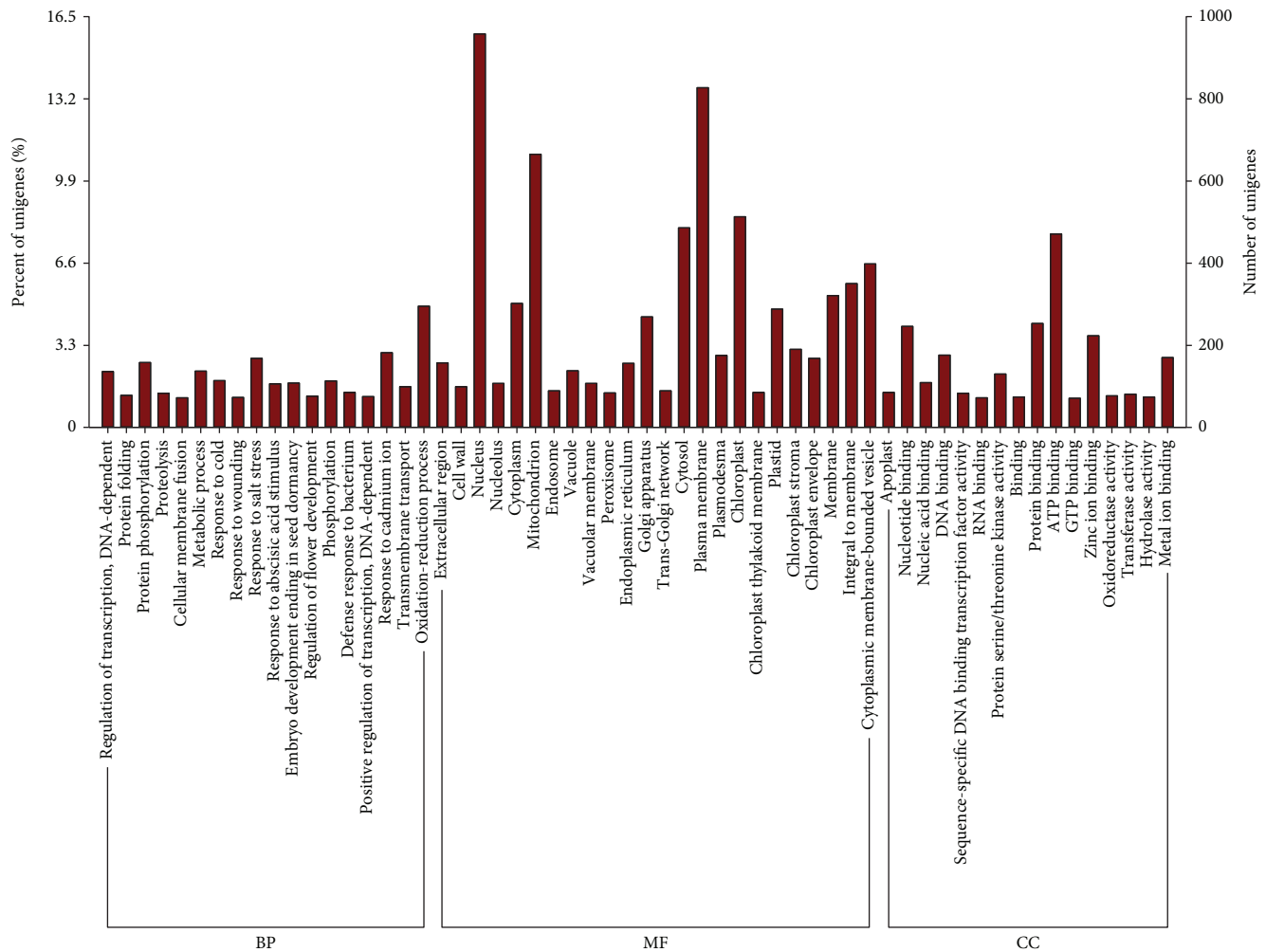


FIGURE 3: Gene ontology classifications of 6130 orthologous unigenes. The results are clustered in the three main categories as biological process (BP), cellular component (CC), and molecular function (MF).

selected the four chloroplast sequences (about 2490 bp) for phylogenetic analysis and found that *A. deserti* (GAHT01022741) was separated with the other 3 species (Figure 1(b)).

3.2. Sequence Comparison between Agave Transcriptomes. A summary for the three databases was described in Table 1. There were 29,367, 29,327, and 50,851 sequences employed for orthologous gene searching from *A. tequilana*, *A. deserti*, and *A. sisalana* transcriptomes, respectively. As a result, we identified 13,069 unigene pairs between *A. tequilana* and *A. deserti*, 8976 pairs between *A. sisalana* and *A. tequilana*, and 9284 pairs between *A. sisalana* and *A. deserti* (Figure 2(a)). Among these orthologous unigene pairs, more than 91% unigene pairs had an identity over 91% (Figure 2(b)). Furthermore, a total of 6130 unigene terms were obtained from the three agave transcriptomes. GO functional classification indicated that these genes were assigned to 30,405 functional terms. There were 14,915 terms in biological process (49.05%), 8546 in molecular function (28.11%), and 6944 in cellular component (22.84%) (Figure 3).

3.3. Identification of Genes Selected in the Domestication of Agaves. The K_a , K_s , and K_a/K_s values were calculated for 6130 orthologous unigene pairs in *A. tequilana* and *A. sisalana* separately with *A. deserti* (Figure 4(a)). The correlation between K_a and K_s values was also estimated in *A. tequilana* ($r = 0.515$, $P < 0.05$) and *A. sisalana* ($r = 0.206$, $P < 0.05$) unigene pairs, respectively. 393 unigenes (6.5%) in *A. tequilana* and 262 unigenes (4.5%) in *A. sisalana* showed a K_a/K_s ratio higher than 1, while the K_a/K_s ratio of more than 90% unigenes was lower than 1 (Figure 4(b)).

The significance of the K_a/K_s value for all 6130 unigene terms was analyzed, and the results indicated that 1117 unigenes were significantly selected at least in *A. tequilana* or *A. sisalana* (P value < 0.05) (Supplementary Table 1). These genes were further characterized into 15 classifications according to their annotations (Table 2). Among these, 54 unigenes were subjected to positive selection ($K_a/K_s > 1$), while the residues of 1064 unigenes were subjected to purifying selection ($K_a/K_s < 1$). Furthermore, 27 and 19 unigenes were positively selected in *A. tequilana* and *A. sisalana*, respectively. 8 unigenes were positively selected in both agave species (Supplementary Table 1).

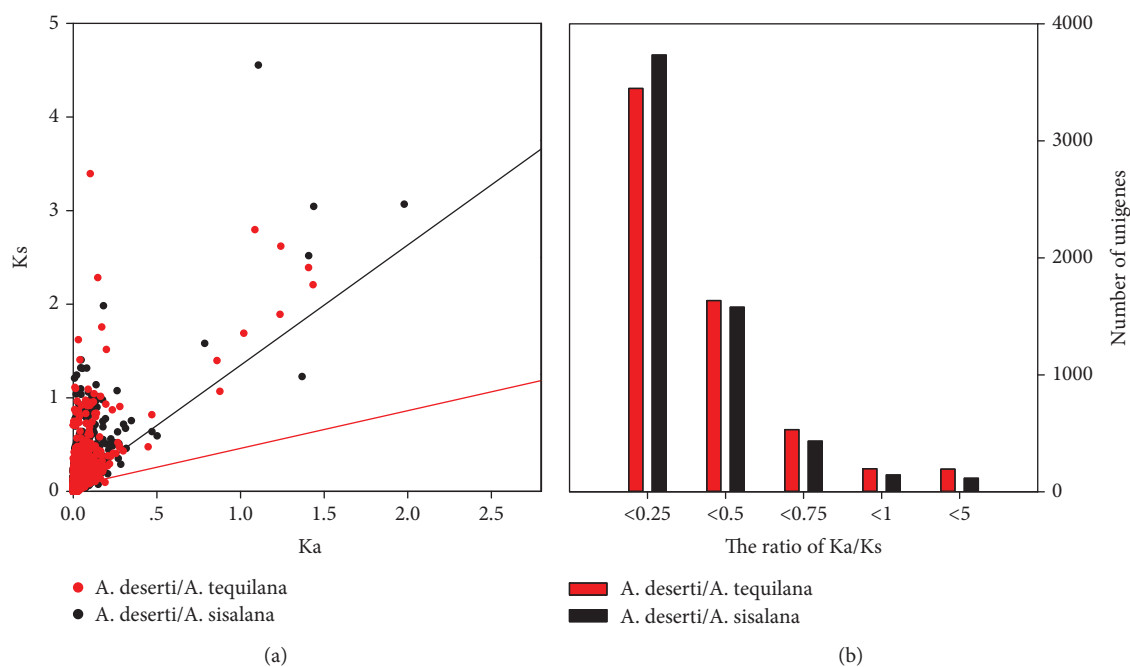


FIGURE 4: The nonsynonymous (Ka) and synonymous (Ks) nucleotide substitutions, as well as their Ka/Ks ratio. (a) The scatter diagram of Ka and Ks values of *A. tequilana* (red) and *A. sisalana* (black) compared with *A. deserti*, respectively. (b) The Ka/Ks distribution of *A. tequilana* (red) and *A. sisalana* (black) compared with *A. deserti*, respectively. The x-axis indicates the ratio of Ka/Ks.

TABLE 2: Characterization of significant selection unigenes.

Classification	Significant selection unigenes			
	Total	In <i>A. sisalana</i>	In <i>A. tequilana</i>	In both
Cell growth/division	78	30	37	11
Cell structure	37	12	22	3
Disease/defense	62	25	22	15
Energy	26	5	18	3
Intracellular traffic	6	2	3	1
Primary metabolism	94	27	55	12
Protein destination and storage	102	27	56	19
Protein synthesis	28	7	16	5
Secondary metabolism	56	16	33	7
Signal transduction	127	40	64	23
Transcription	82	26	40	16
Transcription factor	77	28	37	12
Transporter	56	17	31	8
Transposon	9	3	4	2
Unclassified	277	72	161	44
Total	1117	337	599	181

We further characterized the 54 unigenes subjected to positive selection and found that three unigenes were annotated as disease resistance protein (Table 3). In *A. tequilana*, four unigenes were annotated as 6G-fructosyltransferase, d-2-hydroxyglutarate dehydrogenase, gluconokinase, and 6-phosphofructokinase, respectively, and they

might be related to fructan production, while in *A. sisalana*, five unigenes were characterized as auxin-responsive protein IAA6, gibberellin receptor, PHD and RING finger protein, elongation of fatty acids protein, and zinc finger A20 and AN1 protein, respectively. These unigenes are probably related to the fiber development in *A. sisalana*. Besides, two unigenes, designated as RING-H2 finger protein and TOM1-like protein, were positively selected in both *A. tequilana* and *A. sisalana*.

3.4. In Silico Expression of Genes under Positive Selection in Agave Species. The expression patterns of genes under positive selection were analyzed according to the three transcriptome databases. In agave leaves, the 27 and 8 unigenes were grouped into two expression modes (Figure 5). However, the 19 unigenes in *A. sisalana* were not distinctly clustered into different expression modes. All these genes were differentially expressed in *A. tequilana* or *A. sisalana* when compared with *A. deserti*.

3.5. Selection Pressure and Structure Model of Putative Economic Trait-Related Genes. The sliding window analysis was used to examine the selection pressure of putative economic Trait-related genes. The results indicated the existence of different selection pressures in *A. tequilana* and *A. sisalana* genes (Figure 6). The three disease resistance genes all had a strong selection pressure in most sequence regions. Five unigenes showed a stronger selection pressure in *A. sisalana* (GAHT01109565, GAHT01002417, GAHT01054013, GAHT01038220, and GAHT01027892). Only GAHT01031288 showed a stronger selection pressure

TABLE 3: Unigenes subjected to positive selection in the domestication of Agave species.

<i>A. deserti</i> accession	<i>A. tequilana</i>			<i>A. sisalana</i>			Functional annotation		
	Accession	Ka	Ks	Ka/Ks	Accession	Ka		Ks	
GAHT01003963	GAHU01106881	0.0533	0.0493	1.0803	Unigene21947	0.0084	0.0104	0.8112	6G-Fructosyltransferase
GAHT01031288	GAHU01044798	0.0133	0.0121	1.1055	Unigene12882	0.0011	0.0269	0.0428	d-2-Hydroxyglutarate dehydrogenase
GAHT01109565	GAHU01096053	0.0165	0.0084	1.9594	CL119921.Contig1	0.0444	0.2201	0.2015	Gluconokinase
GAHT01002417	GAHU01001324	0.0214	0.0106	2.0164	Unigene18337	0.0639	0.1531	0.4171	6-Phosphofructokinase
GAHT01070676	GAHU01114859	0.0876	0.0797	1.0994	Unigene374	0.0929	0.2571	0.3613	NBS-type resistance protein RGC5
GAHT01099649	GAHU01173424	0.1314	0.1049	1.2523	CL112994.Contig3	0.1753	0.2373	0.7387	Disease resistance protein RGA1
GAHT01006596	GAHU01010948	0.0167	0.0132	1.2634	CL113186.Contig4	0.023	0.0139	1.6485	RING-H2 finger protein
GAHT01048216	GAHU01097154	0.1621	0.1363	1.1894	CL118667.Contig2	0.149	0.0737	2.0217	Disease resistance protein RPS2
GAHT01055399	GAHU01077165	0.0118	0.0043	2.7653	Unigene28650	0.0188	0.0065	2.9139	TOM1-like protein
GAHT01016426	GAHU01038186	0.0143	0.1124	0.1272	CL8882.Contig1	0.0289	0.0226	1.28	Auxin-responsive protein IAA6
GAHT01005221	GAHU01003524	0.0322	0.0583	0.5528	CL2916.Contig1	0.0301	0.0107	2.8275	Gibberellin receptor
GAHT01054013	GAHU01067136	0.009	0.0092	0.9716	Unigene233	0.0081	0.0059	1.383	PHD and RING finger protein
GAHT01038220	GAHU01063926	0.0116	0	—	Unigene4300	0.0156	0.0049	3.154	Elongation of fatty acids protein
GAHT01027892	GAHU01009883	0.0022	0.0281	0.0794	Unigene28305	0.0659	0.0551	1.1956	Zinc finger A20 and AN1 protein

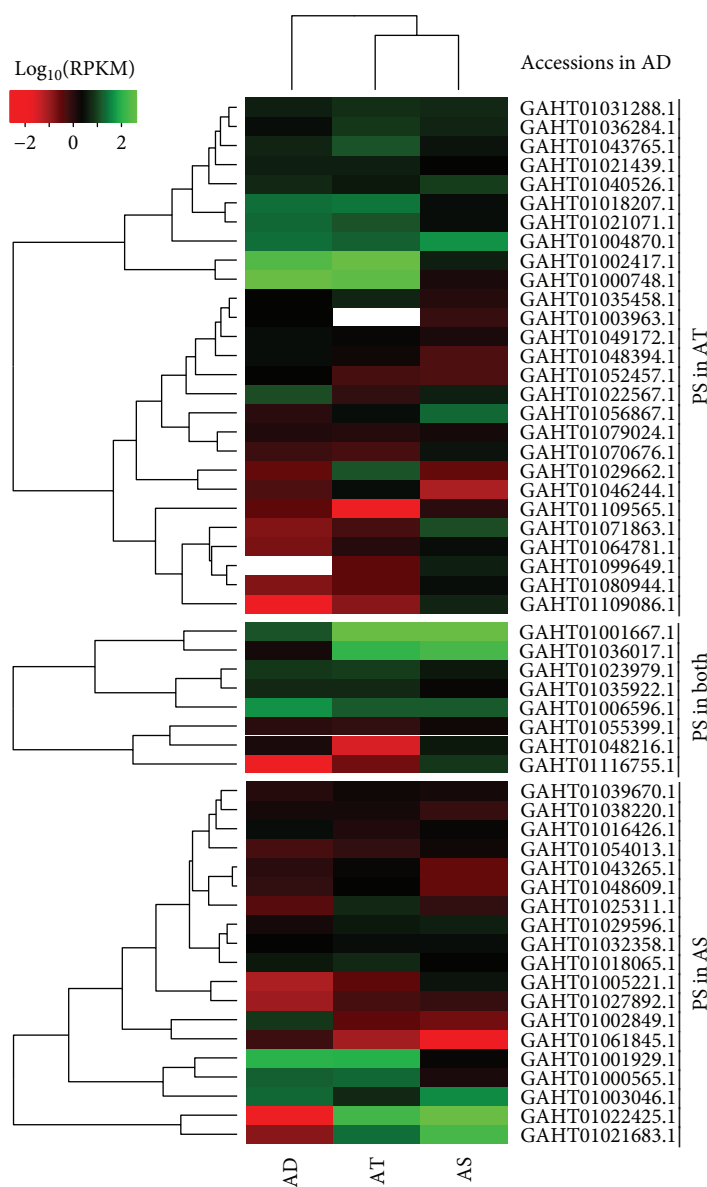


FIGURE 5: In silico expression pattern of positive selection (PS) unigenes in Agave leaves. Rows and columns in heat maps represent genes and Agave species. Species name of *A. deserti* (AD), *A. tequilana* (AT), and *A. sisalana* (AS) is shown under the heat maps. The colour bar at the top indicates the degree of expression: red—low expression; green—high expression.

in *A. tequilana*. The residue of five unigenes shared a strength-similar but region-different selection pressure.

The structure modeling for the 14 unigenes was further conducted to analyze the difference of agave proteins. According to the results, four unigenes designated as 6G-fructosyltransferase (Figure 7(a)), d-2-hydroxyglutarate dehydrogenase (Figure 7(b)), gluconokinase (Figure 7(c)), and 6-phosphofructokinase (Figure 7(d)), could match Swiss-Model sequences with identity > 30%, coverage > 75%, and appropriate description (Supplementary Table 2). Their structure models also showed significant differences in *A. tequilana* or *A. sisalana*. The unigene with a significantly distinct structure was differentially expressed when compared to their orthologous unigenes in other two species (Figure 5).

4. Discussion

A previous phylogenetic analysis suggested that *A. deserti* was closely related to *A. angustifolia* [15]. However, we compared the recently published chloroplast sequence (GAHT01022741) with the former one (DQ500894 + DQ500928) in *A. deserti* and found them having totally different sequences. It might be caused by sample collection because the two studies were separately conducted in Mexico and America. Wild agave species are usually identified by morphological traits, which is not as reliable as molecular identification. Our phylogenetic results also indicated a different evolution relation by both short and long chloroplast sequences (Figure 1). This indicated the closer evolution

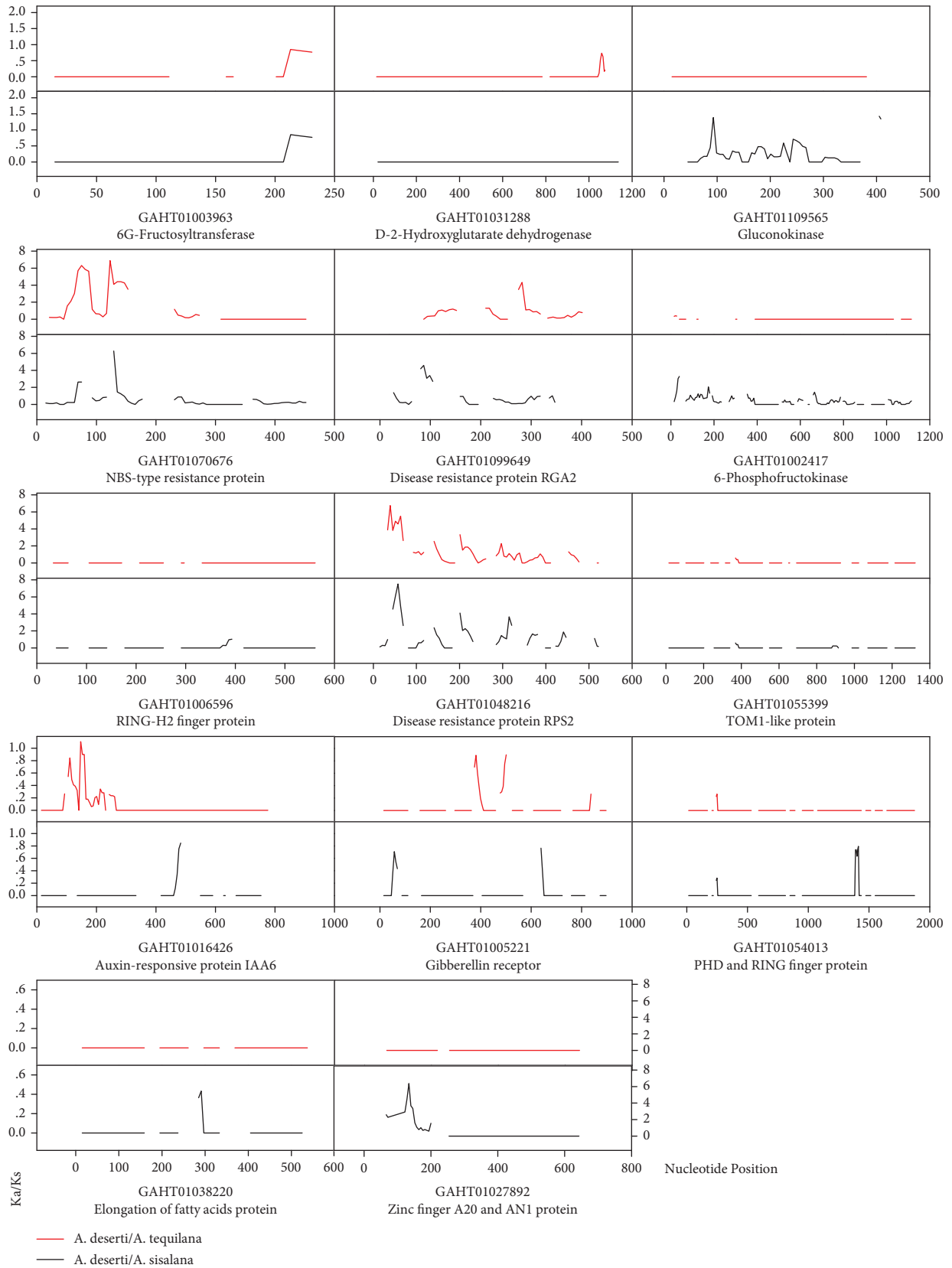


FIGURE 6: Sliding window analysis of positive selection unigenes in *A. tequilana* (red) and *A. sisalana* (black) compared with *A. deserti*, respectively. The window size is 30 bp, with a step size of 6 bp. The x-axis denotes the nucleotide position. The y-axis denotes the Ka/Ks ratio. The gaps represent Ka/Ks ratios that could not be computed.

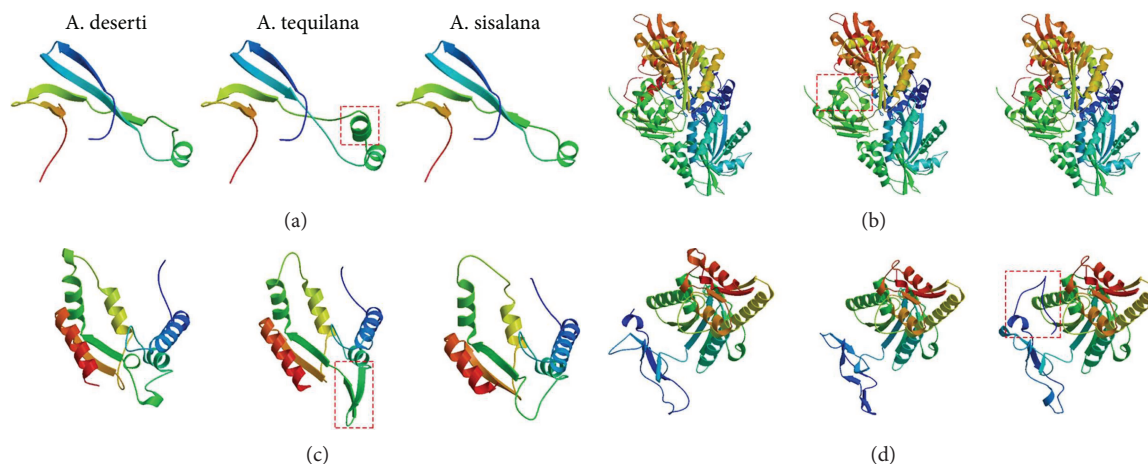


FIGURE 7: Protein structure model of 6G-fructosyltransferase (a), d-2-hydroxyglutarate dehydrogenase (b), gluconokinase (c), and 6-phosphofructokinase (d) in three *Agave* species by using Swiss-Model [33]. The significant structure difference within species was marked in red dotted line.

relationship between *A. tequilana* and *A. sisalana* than the evolution relationship between *A. tequilana* and *A. deserti* or between *A. sisalana* and *A. deserti*. It has been reported that *A. tequilana* and *A. sisalana* both had genetic relations with *A. angustifolia* [6, 7]. The phylogenetic result has also proved that, however, the two species actually possessed very different agronomic traits and applications. An important reason should be artificial selection even if the low-frequency serendipitous backyard hybridization would lead to distinct domestication of crops. This is a historical inheritance in Mexico and has significantly enriched the genetic diversity of crops [34].

Agave species originate from Central America with high tolerance to drought and temperature, which makes them a main and important kind of plants there [9]. Therefore, they would inevitably confront a series of abiotic and biotic stresses. Actually, we identified 62, 127, and 77 unigenes with classifications of disease/defense, signal transduction, and transcription factor, respectively. Among them, 13 significant selected unigenes were related to disease resistance and ten of them subjected to purifying selection (Supplementary Table 1). This might be accompanied with the process of agronomic Trait-derived domestication. Many disease resistance genes were also found to be lost during domestication in other crops [17, 22]. A previous study has already revealed a differentiated selection pressure on NBS-LRR genes in some agave species [35]. In this study, the sliding window analysis also showed a strong selection pressure with the three disease resistance unigenes (Figure 6). Furthermore, two of the three unigenes (GAHT01070676 and GAHT01099649) were subjected to purifying selection in *A. sisalana* (Table 3), which might be harmful for growth and development. This might also be responsible for the susceptibility to zebra disease caused by *Phytophthora nicotianae* in *A. sisalana* [36].

It has been reported that several transcription factor families play an important role in abiotic stress regulation, such as bHLH, zinc finger, MYB, AP2, NAC, WRKY, and bZIP families [37–42]. We also found 47 TFs subjected to purifying selection either in *A. tequilana* or in *A. sisalana*, from

the bHLH (8), zinc finger (23), MYB (6), AP2 (5), NAC (3), WRKY (1), and bZIP (1) families (Supplementary Table 1). For agave species, drought and high temperature are the main abiotic stresses. We speculated that different habitats should be an important natural selection pressure that affected the shape and size of the three agave species [43]. The purifying selection of the 47 stress-related candidate TFs might weaken the drought tolerance of *A. tequilana* and *A. sisalana* but enhance their biomass accumulations. The complex regulation and interaction of these TFs might be the key to reveal the mechanism of the remarkable drought tolerance in *A. deserti*. Much more molecular characterizations are still needed in future studies.

It has been reported that fruit/seed-related traits were subjected to high artificial selection pressure for their economic value [18, 19]. In *A. tequilana*, the most important economic trait focuses on fructan, and several studies have conducted functional characterization for fructosyltransferase genes [10, 11, 44]. In this study, we identified a fructan-related unigene and it was subjected to positive selection in *A. tequilana* (Table 3). Besides, three carbohydrate-related unigenes were also subjected to positive selection. Their positive selection might be responsible for the improvement of fructan yield in *A. tequilana*. In *A. sisalana*, fiber is the main economic purpose for its cultivation in tropical areas. In the present study, we found 5 unigenes subjected to positive selection only in *A. sisalana* (Table 3). A previous publication has reviewed the hormonal regulation of secondary cell wall formation [45]. The zinc finger family TFs are also proved to regulate cell wall development and cellulose biosynthesis [46, 47]. Besides, the elongation of fatty acids protein plays an important role in cell elongation [48]. Therefore, we speculated that the fiber-related traits in agaves are more likely controlled by hormonal and transcriptional regulation. And there are significant differences when compared with fructan-related traits, which might be mainly controlled by metabolic regulation in *A. tequilana*. Natural fiber is commonly generated as the result of secondary cell wall thickening in the main fiber crops such as cotton, ramie, flax, and

hemp [49–52]. As a constitutive structure of plant cells, there are many housekeeping and regulating genes during cell wall development, especially the secondary cell wall development [45, 53–55]. In contrast, fructans are mainly responsible for carbohydrate storage to vegetative tissues in many plant species [56]. They have also been increasingly considered protective agents against abiotic stresses [57]. However, the capacity to produce and store fructans in *A. tequilana* is much stronger than that in *A. deserti*, even if there is a much more severe environment for *A. deserti* [23, 44]. It is probably because fructan-related traits are regulated at the metabolic level. A more recent study has combined transcript, protein, and metabolite methods to reveal the molecular basis of the CAM process in agave [58]. The rapid development of high-throughput molecular methods has brought a great opportunity for the further understanding of the differences and evolution patterns among agave species.

5. Conclusion

This study represents the first transcriptome comparison within domesticated and wild agave species. The results revealed the importance of abiotic/biotic natural selection in agave evolution. Four unigenes related to fructan in *A. tequilana* and five unigenes related to fiber in *A. sisalana* were positively selected. These genes revealed the difference between *A. tequilana* and *A. sisalana* evolution, which would serve as a guidance for further studies on agave evolution, environmental adaptation, and improvement of economically important traits.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

XH and KY conceived and designed the experiments. XH analyzed the data and drafted the manuscript. BW carried out the Ka/Ks analysis. JX, JG, HC, SZ, JZ, and KY contributed to the transcriptome data. YZ revised the manuscript. CH and WW helped in the expression analysis. YL helped in the selection pressure detection. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Dr. Xiaohan Yang from the Oak Ridge National Laboratory (Oak Ridge, TN 37831-6407, USA), Dr. Thomas Nothnagel from Julius Kühn-Institut (Quedlinburg 06484, Germany), and Dr. Adel M. R. A. Abdelaziz from the Central Laboratory of Organic Agriculture, Agricultural Research Center (Giza 12619, Egypt), for the manuscript revision. This study was supported by the National Key R&D Program of China (2018YFD0201100),

the earmarked fund from the China Agriculture Research System (CARS-16-E16), the National Natural Science Foundation of China (31371679, 31771849), and the Central Public-Interest Scientific Institution Basal Research Fund (2017hzs1J014, 1630042018014).

Supplementary Materials

Supplementary 1. Table S1: the results of classification, Ka/Ks analysis, and functional annotation for the 6130 unigenes.

Supplementary 2. Table S2: details for structure modeling of putative economic Trait-related genes.

References

- [1] A. M. Borland, H. Griffiths, J. Hartwell, and J. A. C. Smith, "Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands," *Journal of Experimental Botany*, vol. 60, no. 10, pp. 2879–2896, 2009.
- [2] P. S. Nobel, "Achievable productivities of certain CAM plants: basis for high values compared with C₃ and C₄ plants," *New Phytologist*, vol. 119, no. 2, pp. 183–205, 1991.
- [3] X. Yang, J. C. Cushman, A. M. Borland et al., "A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world," *New Phytologist*, vol. 207, no. 3, pp. 491–504, 2015.
- [4] M. Cedeño, "Tequila production," *Critical Reviews in Biotechnology*, vol. 15, no. 1, pp. 1–11, 1995.
- [5] K. R. Corbin, C. S. Byrt, S. Bauer et al., "Prospecting for energy-rich renewable raw materials: *Agave* leaf case study," *PLoS One*, vol. 10, no. 8, article e0135382, p. 10, 2015.
- [6] G. Lock, *Sisal*, Longmans, Green, London, UK, 1962.
- [7] S. H. Gentry, *Agaves of Continental North America*, University Of Arizona Press, Tucson, AZ, USA, 1982.
- [8] P. S. Nobel and T. L. Hartsock, "Temperature, water, and PAR influences on predicted and measured productivity of *Agave deserti* at various elevations," *Oecologia*, vol. 68, no. 2, pp. 181–185, 1986.
- [9] J. R. Stewart, "Agave as a model CAM crop system for a warming and drying world," *Frontiers in Plant Science*, vol. 6, p. 684, 2015.
- [10] C. Cortés-Romero, A. Martínez-Hernández, E. Mellado-Mojica, M. G. López, and J. Simpson, "Molecular and functional characterization of novel fructosyltransferases and invertases from *Agave tequilana*," *PLoS One*, vol. 7, no. 4, article e35878, p. 7, 2012.
- [11] E. M. Suarez-Gonzalez, M. G. Lopez, J. P. Délano-Frier, and J. F. Gómez-Leyva, "Expression of the 1-SST and 1-FFT genes and consequent fructan accumulation in *Agave tequilana* and *A. inaequidens* is differentially induced by diverse (a)biotic-stress related elicitors," *Journal of Plant Physiology*, vol. 171, no. 3–4, pp. 359–372, 2014.
- [12] P. S. Nobel, *Environmental Biology of Agaves and Cacti*, Cambridge University Press, Cambridge, 1988.
- [13] P. Y. Chen, C. H. Chen, C. C. Kuo, T. H. Lee, Y. H. Kuo, and C. K. Lee, "Cytotoxic steroidal saponins from *Agave sisalana*," *Planta Medica*, vol. 77, no. 09, pp. 929–933, 2011.
- [14] J. D. Santos, I. J. Vieira, R. Braz-Filho, and A. Branco, "Chemicals from *Agave sisalana* biomass: isolation and identification,"

- International Journal of Molecular Sciences*, vol. 16, no. 12, pp. 8761–8771, 2015.
- [15] S. V. Good-Avila, V. Souza, B. S. Gaut, and L. E. Eguiarte, “Timing and rate of speciation in Agave (Agavaceae),” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 24, pp. 9124–9129, 2006.
- [16] M. L. Robert, K. Y. Lim, L. Hanson et al., “Wild and agronomically important Agave species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy,” *Botanical Journal of the Linnean Society*, vol. 158, no. 2, pp. 215–222, 2008.
- [17] S. Guo, J. Zhang, H. Sun et al., “The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions,” *Nature Genetics*, vol. 45, no. 1, pp. 51–58, 2013.
- [18] L. Lu, D. Shao, X. Qiu et al., “Natural variation and artificial selection in four genes determine grain shape in rice,” *The New Phytologist*, vol. 200, no. 4, pp. 1269–1280, 2013.
- [19] J. Qi, X. Liu, D. Shen et al., “A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity,” *Nature Genetics*, vol. 45, no. 12, pp. 1510–1515, 2013.
- [20] J. Canales, R. Bautista, P. Label et al., “De novo assembly of maritime pine transcriptome: implications for forest breeding and biotechnology,” *Plant Biotechnology Journal*, vol. 12, no. 3, pp. 286–299, 2014.
- [21] X. Huang, J. Chen, Y. Bao et al., “Transcript profiling reveals auxin and cytokinin signaling pathways and transcription regulation during in vitro organogenesis of ramie (*Boehmeria nivea* L. gaud),” *PLoS One*, vol. 9, no. 11, article e113768, p. 9, 2014.
- [22] T. Liu, S. Tang, S. Zhu, Q. Tang, and X. Zheng, “Transcriptome comparison reveals the patterns of selection in domesticated and wild ramie (*Boehmeria nivea* L. gaud),” *Plant Molecular Biology*, vol. 86, no. 1-2, pp. 85–92, 2014.
- [23] S. M. Gross, J. A. Martin, J. Simpson, M. Abraham-Juarez, Z. Wang, and A. Visel, “De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*,” *BMC Genomics*, vol. 14, no. 1, p. 563, 2013.
- [24] P. Wang, J. Gao, F. Yang et al., “Transcriptome of sisal leaf pretreated with *Phytophthora nicotianae* Breda,” *Chinese J Tropical Crops*, vol. 35, pp. 576–582, 2014.
- [25] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, “MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods,” *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.
- [26] J. Martin, V. M. Bruno, Z. Fang et al., “Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-seq reads,” *BMC Genomics*, vol. 11, no. 1, p. 663, 2010.
- [27] M. Zhang and H. W. Leong, “Bidirectional best hit *r*-window gene clusters,” *BMC Bioinformatics*, vol. 11, Supplement 1, p. S63, 2010.
- [28] A. Doron-Faigenboim, A. Stern, I. Mayrose, E. Bacharach, and T. Pupko, “Selecton: a server for detecting evolutionary forces at a single amino-acid site,” *Bioinformatics*, vol. 21, no. 9, pp. 2101–2103, 2005.
- [29] Z. Yang, “PAML 4: phylogenetic analysis by maximum likelihood,” *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.
- [30] B. Li and C. N. Dewey, “RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome,” *BMC Bioinformatics*, vol. 12, no. 1, p. 323, 2011.
- [31] M. Hu, W. Hu, Z. Xia, X. Zhou, and W. Wang, “Validation of reference genes for relative quantitative gene expression studies in cassava (*Manihot esculenta* Crantz) by using quantitative real-time PCR,” *Frontiers in Plant Science*, vol. 7, p. 680, 2016.
- [32] P. Librado and J. Rozas, “DnaSP v5: a software for comprehensive analysis of DNA polymorphism data,” *Bioinformatics*, vol. 25, no. 11, pp. 1451–1452, 2009.
- [33] M. Biasini, S. Bienert, A. Waterhouse et al., “SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information,” *Nucleic Acids Research*, vol. 42, no. W1, pp. W252–W258, 2014.
- [34] C. E. Hughes, R. Govindarajulu, A. Robertson, D. L. Filer, S. A. Harris, and C. D. Bailey, “Serendipitous backyard hybridization and the origin of crops,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 36, pp. 14389–14394, 2007.
- [35] M. C. Tamayo-Ordóñez, L. C. Rodríguez-Zapata, J. A. Narvaez-Zapata et al., “Morphological features of different polyploids for adaptation and molecular characterization of CC-NBS-LRR and LEA gene families in *Agave* L.,” *Journal of Plant Physiology*, vol. 195, pp. 80–94, 2016.
- [36] J. Gao, Luoping, C. Guo et al., “AFLP analysis and zebra disease resistance identification of 40 sisal genotypes in China,” *Molecular Biology Reports*, vol. 39, no. 5, pp. 6379–6385, 2012.
- [37] G. Castilhos, F. Lazzarotto, L. Spagnolo-Fonini, M. H. Bodanese-Zanettini, and M. Margis-Pinheiro, “Possible roles of basic helix–loop–helix transcription factors in adaptation to drought,” *Plant Science*, vol. 223, pp. 1–7, 2014.
- [38] S. Ciftci-Yilmaz and R. Mittler, “The zinc finger network of plants,” *Cellular and Molecular Life Sciences*, vol. 65, no. 7-8, pp. 1150–1160, 2008.
- [39] K. J. Dietz, M. O. Vogel, and A. Viehhauser, “AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling,” *Protoplasma*, vol. 245, no. 1-4, pp. 3–14, 2010.
- [40] C. Dubos, R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, and L. Lepiniec, “MYB transcription factors in *Arabidopsis*,” *Trends in Plant Science*, vol. 15, no. 10, pp. 573–581, 2010.
- [41] S. Puranik, P. P. Sahu, P. S. Srivastava, and M. Prasad, “NAC proteins: regulation and role in stress tolerance,” *Trends in Plant Science*, vol. 17, no. 6, pp. 369–381, 2012.
- [42] K. Singh, R. C. Foley, and L. Onate-Sanchez, “Transcription factors in plant defense and stress responses,” *Current Opinion in Plant Biology*, vol. 5, no. 5, pp. 430–436, 2002.
- [43] N. J. Kooyers, “The evolution of drought escape and avoidance in natural herbaceous populations,” *Plant Science*, vol. 234, pp. 155–162, 2015.
- [44] E. A. D. Dios, A. D. G. Vargas, M. L. D. Santos, and J. Simpson, “New insights into plant glycoside hydrolase family 32 in Agave species,” *Frontiers in Plant Science*, vol. 6, p. 594, 2015.
- [45] V. Didi, P. Jackson, and J. Hejatko, “Hormonal regulation of secondary cell wall formation,” *Journal of Experimental Botany*, vol. 66, no. 16, pp. 5015–5027, 2015.
- [46] W. C. Kim, J. Y. Kim, J. H. Ko, H. Kang, J. Kim, and K. H. Han, “AtC3H14, a plant-specific tandem CCCH zinc-finger protein, binds to its target mRNAs in a sequence-specific manner and

- affects cell elongation in *Arabidopsis thaliana*,” *The Plant Journal*, vol. 80, no. 5, pp. 772–784, 2014.
- [47] D. Wang, Y. Qin, J. Fang et al., “A missense mutation in the zinc finger domain of OsCESA7 deleteriously affects cellulose biosynthesis and plant growth in rice,” *PLoS One*, vol. 11, no. 4, article e0153993, p. 11, 2016.
- [48] Y. M. Qin, C. Y. Hu, Y. Pang, A. J. Kastaniotis, J. K. Hiltunen, and Y.-X. Zhu, “Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis,” *The Plant Cell*, vol. 19, pp. 3692–3704, 2007.
- [49] M. Behr, S. Legay, E. Žižková et al., “Studying secondary growth and bast fiber development: the hemp hypocotyl peeks behind the wall,” *Frontiers in Plant Science*, vol. 7, p. 1733, 2016.
- [50] M. Chantreau, B. Chabbert, S. Billiard, S. Hawkins, and G. Neutelings, “Functional analyses of cellulose synthase genes in flax (*Linum usitatissimum*) by virus-induced gene silencing,” *Plant Biotechnology Journal*, vol. 13, no. 9, pp. 1312–1324, 2015.
- [51] J. Chen, Z. Pei, L. Dai et al., “Transcriptome profiling using pyrosequencing shows genes associated with bast fiber development in ramie (*Boehmeria nivea* L.),” *BMC Genomics*, vol. 15, no. 1, p. 919, 2014.
- [52] A. M. Schubert, C. R. Benedict, J. D. Berlin, and R. J. Kohel, “Cotton fiber development-kinetics of cell elongation and secondary wall thickening,” *Crop Science*, vol. 13, no. 6, pp. 704–709, 1973.
- [53] D. J. Cosgrove, “Growth of the plant cell wall,” *Nature Reviews Molecular Cell Biology*, vol. 6, no. 11, pp. 850–861, 2005.
- [54] S. G. Hussey, E. Mizrachi, N. M. Creux, and A. A. Myburg, “Navigating the transcriptional roadmap regulating plant secondary cell wall deposition,” *Frontiers in Plant Science*, vol. 4, p. 325, 2013.
- [55] M. Schuetz, R. Smith, and B. Ellis, “Xylem tissue specification, patterning, and differentiation mechanisms,” *Journal of Experimental Botany*, vol. 64, no. 1, pp. 11–31, 2012.
- [56] C. J. Nelson and W. G. Spollen, “Fructan,” *Physiologia Plantarum*, vol. 71, no. 4, pp. 512–516, 1987.
- [57] R. Valluru and W. V. D. Ende, “Plant fructans in stress environments: emerging concepts and future prospects,” *Journal of Experimental Botany*, vol. 59, no. 11, pp. 2905–2916, 2008.
- [58] P. E. Abraham, H. Yin, A. M. Borland et al., “Transcript, protein and metabolite temporal dynamics in the CAM plant *Agave*,” *Nature Plants*, vol. 2, p. 16178, 2016.