





Construction of a High-Density Genetic Map for Pitaya Using the Whole Genome Resequencing Approach

Zhijiang Wu^{1,†}, Haiyan Deng^{1,†}, Guidong Liang¹, Xiaoying Ye¹, Yonghua Qin^{2,3,*} and Lifang Huang^{1,*}

- ¹ Horticultural Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530007, China; zhijiangwu@gxaas.net (Z.W.); dhy212@gxaas.net (H.D.); liangguidong2005@126.com (G.L.); yexiaoying@gxaas.net (X.Y.)
- ² College of Horticulture, South China Agricultural University, Guangzhou 510642, China
- ³ State Key Laboratory for Conservation and Utilization of Subtropical Agrobioresources/Guangdong Provincial Key Laboratory of Postharvest Science of Fruits and Vegetables/Key Laboratory of South China Horticultural Crop Biology and Germplasm Enhancement, Ministry of Agriculture, Guangzhou 510642, China
- * Correspondence: qinyh@scau.edu.cn (Y.Q.); huanglifang@gxaas.net (L.H.)
- † Same contribution.

Abstract: Pitaya (*Hylocereus undatus*) is one of the most economic fleshy fruit tree crops. This study aimed at producing a high-density linkage genetic map of pitaya based on the whole genome resequencing (WGrS) approach. For this purpose, a bi-parental F1 population of 198 individuals was generated and genotyped by WGrS. High-quality polymorphic 6434 single polymorphism nucleotide (SNP) markers were extracted and used to construct a high-density linkage map. A total of 11 linkage groups were resolved as expected in accordance with the chromosome number. The map length was 14,128.7 cM with an average SNP interval of 2.2 cM. Homology with the sequenced reference genome was described, and the physical and genetic maps were compared with collinearity analysis. This linkage map in addition to the available genomic resources will help for quantitative trait mapping, evolutionary studies and marker-assisted selection in the important *Hylocereus* species.

Keywords: genetic map; pitaya; genome resources; crop improvement; marker assisted breeding

1. Introduction

The dragon fruit, commonly known as pitaya, belongs to the genus *Hylocereus* of Cactaceae family [1]. The *Hylocereus* species are diploid with 22 (2n = 2X) chromosomes. Due to its key characteristics, such as the attractive appearance and colors, great taste, high nutrient content, antioxidant capacity and antiproliferative activities [2–4], pitaya is emerging as an increasingly popular fleshy fruit worldwide. This is exemplified by the increasing planting area and production of pitaya (http://faostat3.fao.org/, accessed on 26 September 2021).

Fifteen species in this genus have been identified, five of which are cultivated for fruit production [5]. Among the cultivated pitayas, the red-skinned and red-fleshed pitaya (*Hylocereus monacanthus*) and red-skinned and white-fleshed pitaya (*Hylocereus undatus*) have been the focus of many studies. *Hylocereus undatus* is a perennial climbing plant originated from tropical areas of the northern, central, and southern Americas [1] and renowned for its attractive and nutritious fruits used as model for various studies on the *Hylocereus* species [1,6,7]. It has an excellent commercial value and is produced at a large scale as a new type of fruit crop [8,9]. The *H. monacanthus* has also been studied for its potential to resist against fungal diseases [10]. It has been used as a model plant to evaluate the effects of polyploidy in *Hylocereus* species. Recent, studies on pitaya have been reported for pigmentation [11], antioxidant determination [12], drought response [1], metabolic profiling [13] and post-harvest quality [14]. Genetic and genomic information-based breeding of this emerging fruit crop is greatly required to speed up the crop improvement.



Citation: Wu, Z.; Deng, H.; Liang, G.; Ye, X.; Qin, Y.; Huang, L. Construction of a High-Density Genetic Map for Pitaya Using the Whole Genome Resequencing Approach. *Horticulturae* **2021**, *7*, 534. https://doi.org/10.3390/ horticulturae7120534

Academic Editors: Renesh Bedre and Devi R. Kandel

Received: 10 October 2021 Accepted: 22 November 2021 Published: 2 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Hence, there is a dire need to increase the genetic and genomic resources of this crop. The genome of *Hylocereus undatus* was estimated on the basis of K-mer distribution assessment up to 1.58 GB, while only 1.41 GB of the assembled genome was reported [5]. Recently, a study revealed a genomic duplications and chromosomal co-localization of genes in dragon fruit by a chromosome-level genome study [15]. Hence, the whole genome sequence has yet to be completely sequenced and a complete genetic map is needed.

The development of linkage and/or association map is one of the major requirements for the identification of quantitative trait loci (QTL), genes associated to traits of interest and development of marker-assisted selection. The genetic mapping approaches and techniques are based upon arranged genetic and genomic information [16].

The development of application of genetic markers especially co-dominant markers such as simple sequence repeats (SSRs) and single nucleotide polymorphism (SNPs) are highly appreciated as they enable to estimate the (dominant and additive) allelic effects [17,18]. Researchers have developed and used the SSR [19], inter simple sequence repeats (ISSR) [20,21] and other types of molecular markers, such as random amplified polymorphic DNA (RAPD) [22,23] and chromosome scale sequencing-based markers [5], in *Hylocereus* species. Nonetheless, more markers and genetic maps are still needed for QTL mapping and their application in breeding programs.

The development and application of next-generation sequencing tools made SNPs as a widely available marker for high throughput genotyping. The detection, genotyping and selection of thousands of SNPs in various individuals through whole genome resequencing (WGrS) has become a common practice for higher crops [16]. The WGrS-based SNP markers allow for the detection and genotyping, resulting in an unparalleled and reduced cost per datapoint when mining for co-dominant polymorphisms in huge panels and for development of high-density genetic maps [16,24,25]. The aim of present study was to design a highly saturated linkage map to deal with the limitations in screening and identifying genomic regions linked to agronomical and commercially important traits of interest in *H. undatus* using the WGrS method.

2. Materials and Methods

2.1. Materials

The hybridization of white-fleshed *H. undatus* as the diploid male parent (4-09) and the red-fleshed *H. monacanthus* as the diploid female parent (13-7) was performed to develop the F_1 mapping population (Figure 1). The crossing was performed at the *Hylocereus* planting resources nursery in Nanning, Ministry of Agriculture, China. A population of a total of 198 F_1 individuals along with the parents was maintained in the nursery and used for subsequent analysis. The standard cultural practices such as watering, fertilizer applications and optimum temperature (20–25 °C) were kept constant [5].



Figure 1. Fruit appearance of the red-fleshed *H. monacanthus* as the diploid female parent (13-7) (**A**) and white-fleshed *H. undatus* as the diploid male parent (4-09) (**B**) were crossed to develop the F_1 mapping population.

2.2. Genomic DNA Extraction and Genotyping of Population by Whole Genome Resequencing

The tender stems and buds were selected from two-year-old Hylocereus parental plants and F_1 progeny. The samples were collected on 30 March 2021 and preserved in liquid nitrogen immediately after harvest. The samples were then stored in ultra-low temperature $(-80 \ ^{\circ}\text{C})$ refrigerator. The DNA secure Plant Kit (Tiangen Biotech, Beijing, China) was employed to extract the total genomic DNA from collected samples of pitaya. The quality of the isolated DNA was checked using agarose gel electrophoresis and used for processing the genome libraries. DNA samples qualified after inspection were broken into fragments of 350 bp length randomly by a Covaris breaker (Covaris Incorporation, 14 Gill Street, Unit H Woburn, MA, USA). The libraries were constructed by TruSeq Library Construction Kit (Illumina, San Diego, CA, USA). The use of reagents and consumables was performed as recommended in the instructions. The DNA fragments were prepared through end repair, addition of polyA tail, addition of sequencing adapters, purification and PCR amplification to complete the entire library. The constructed library was sequenced by Illumina HiSeqTM PE150 for each sample. The parental DNA samples were replicated to ensure the detection of SNPs and high-quality parental information to estimate the marker segregation types. The sequences were saved in fastq format [26]. The adapters were trimmed by Cutadapt v1.9 [27].

The original sequencing data generated contain connector information, low-quality bases, such as very low minor allele frequency, low read depth, undetected bases, etc. Strict filter threshold was applied to get clean reads from the raw sequences. The filter criteria include the following: (1) filter out reads pairs containing joint sequences; (2) if N-content in a single-ended sequence read exceeds 10% of the total length proportion of the read, the pair of paired reads was removed; and (3) if the base count of low-quality ($Q \le 5$ quality value) bases in a single-ended sequencing read were more than 50% of the length proportion of the read, this pair of paired reads was removed. After the above strict filtering of sequencing data, high-quality clean reads data were obtained. The sequencing data of all samples of parents and offspring were counted and analyzed for quality parameters including the number of sequencing reads, data yield, sequencing error rate, Q20, Q30, GC content, etc.

2.3. SNP Calling and Filtering

The SNPs were called by the process_reseq.1.0.py (python2) program and filtered by the site pre-filtering option in the VcfPreFilter.1.0.py (python2) program with default parameters. Both of these programs are the components of Vcf-Hunter package [28] (available at https://github.com/SouthGreenPlatform/VcfHunter/, accessed on 26 September 2021). A comparative analysis of the sequencing data against the reference genome [5] was performed, and coverage depth and coverage percentage were evaluated. The valid sequencing data were further compared to the available reference genome by using Burrows-Wheeler Aligner (BWA) software, and the comparison results were removed by Samtools. The Genome Analysis Tool Kit (GATK) program was used to detect the population SNP on the filtered bam file. The obtained SNPs were thus named in accordance to their respective position in the reference genome [5].

Based on the genotype detection results of the *Hylocereus* parent genotypes, the development of polymorphic markers between the parents was carried out, and SNPs were filtered against the strict filters (including 10 < sequencing depth > 500, <20% missing data per site, <40% missing data per individual, >10% minor allele frequencies per site and >3 minimum read counts for heterozygous genotypes) to avoid false positives. The sites with missing data on parents and/or segregation patterns in offspring, which were not in accordance with the parental genotypes, were excluded as well. The filtering and then formatting of dataset were imported into the R environment by VcfR 1.5.0 package [29] and coded in R 3.4.4 [30].

2.4. Analysis of Linkage and Construction of Genetic Map

SNPs with Mendelian segregation ratio 1:1 (in case of segregating in one parent only) and 1:2:1 (in case of segregating in both parents) were used for linkage analysis. The significance of segregation deviation within families was tested by χ 2 test (p < 0.001), and the significantly deviated SNPs were eliminated. Finally, the SNP markers were mined to maintain a 100 bp marker to marker minimum spacing. The cross-pollinated model in JoinMap 4.1 software was used for linkage analysis and map constructions separately for each parent by population [31]. The LOD value was set in the range of 2~30, the severely unlinked markers were removed and the remaining markers were divided into linkage groups.

3. Results

3.1. Genotyping by Whole Genome Resequencing and Sequencing Quality Assessment

Overall, 200 genotypes including two parental genotypes, H. undatus as the male parent and *H. monacanthus* as the female parent, along with 198 progenies from the F_1 mapping population, were sequenced and genotyped. The minimum 6879.5 to maximum 11,338 million clean reads in progeny were obtained with an average 97% and 91% of the Q20 and Q30, respectively. For the female parent, the maximum was 29,982 million, while for male parent, 28,362.97 million clean reads were obtained with average of 96.45% and 90.8% Q20 and Q30, respectively (Table 1). The GC contents in the parental genome were 38.7%, while in the offspring population they were 38%. A recently reported genome sequence for *Hylocereus undatus* was used as reference genome [5]. The total length of the reference genome was 1,387.3 million bases with 36.9% GC contents and 0.5509% gap rate, while the length of N50 and N90 was 127.146 million and 108.398 million bases, respectively. An average 98% mapping rate on reference genome was observed with 82.47% 1X coverage rate and 5.1X average depth in the population. The average mapping depths 15.36X and 17.56X with 1X coverage of 89.45% and 90.26% were observed for male and female parents, respectively, while the average 4X coverage rate was 83.86% in both parents. The total mapping error rate was <0.03% (Table 1).

	Male Parent, H.	Female Parent, H.	F ₁ Progeny			
	undatus	monacanthus	Average	Maximum	Minimum	
Raw bases (bp)	28,362,977,400	29,982,174,300	7,853,867,311	11,374,635,600	6,902,555,400	
Clean bases (bp)	28,167,050,700	29,898,491,400	7,827,082,255	11,338,035,600	6,879,546,300	
Clean reads	187,780,338	199,323,276	199,323,276 52,180,548.36		45,863,642	
Mapped reads	185,486,093	194,135,772	51,026,958.15	73,621,450	44,944,080	
Effective rate (%)	99.31	99.72	100	99.86	99.04	
Error rate (%)	0.03	0.03	0	0.04	0.03	
Q20 (%)	96.02	96.89	97	97.84	95.37	
Q30 (%)	89.82	91.78	91	93.62	88.17	
GC content (%)	38.81	38.63	38	40.31	37.28	
Mapping rate (%)	98.78	97.40	98 0.9838		0.9652	
Average mapping depth	15.36	17.65	5.102 7.5		4.43	
Coverage 1X	89.45	90.26	82.468	86.32	77.87	
Coverage 4X	83.74	83.97	39.579	60.44	28.64	
Total SNPs	12,343,419	2,962,195	4,161,067.303	8,070,891	2,369,880	
Homozygous SNP	570,172	338,723	26,448.22222	85,498	15,662	
Heterozygous SNP	11,773,247	2,623,472	4,134,619.081	8,025,401	2,349,592	
Heterozygosity rate (%)	95.3808	88.5651	99.36101263	99.5461	98.3093	

Table 1. Genotyping summary for sequence filtering data set of Hylocereus species.

3.2. SNP Markers Detection and Evaluation

Based on the genotype detection results of the dragon fruit parents, the development of polymorphic markers between the parents was carried out. We filtered out the sites with missing parental information, screened the sites with polymorphisms between the parents, and screened out sites that meet the mapping marker type of the population. Among them, the main marker types used for F1 population mapping were "hkxhk", "nnxnp" and "lmxll". Resultantly, a total of 4,601,639 polymorphic sites were obtained, of which the maximum available marker types for the F_1 population were "hkxhk", "nnxnp" and "lmxll", with 1,117,173, 482,398 and 3,001,321 total polymorphic markers, respectively (Table 2). The genome evaluation for the whole population and the SNP filtering based on genomic information quality (such as allele frequency and sequencing depth) revealed 4.16 million SNPs in F₁, while there were 2.96 and 12.34 million SNPs in male and female parents, respectively (Table 1). Among the parental genotypes, the female parent showed more heterozygosity (95.38%) than the male parent (88.57%). Thinning the SNPs at the rate of one SNP per 100 bp length reduced the SNP dataset. Thus, 6434 SNPs were used for linkage analysis in the population. The missing data threshold was kept constant for all progeny in the population.

Table 2. Allelic segregation and number of loci in each segregant type among F_1 progeny considered in this study and their genetic characteristics.

Segregant Type	Cross Type	Count	% †	Female Parent	Male Parent	Alleles Detectable	F1 Ratio Detectable
Type 1	lmxll	3,001,321	65.15	Heterozygous	Homozygous	2	1:1
Type 2	nnxnp	482,398	10.47	Homozygous	Heterozygous	2	1:1
Туре 3	hkxhk	1,117,173	24.25	Heterozygous	Heterozygous	2	1:2:1, 3:1, 1:3
Type 4a	abxcd	1	0.00002	Heterozygous	Heterozygous	4	1:1:1:1
Type 4b	efxeg	746	0.02	Heterozygou	Heterozygous	3	1:1:1:1
Type 5	aaxbb	5,146	0.11	Homozygous	Homozygous	2	
Туре 6	abxcc	186	0.0040	Heterozygous	Homozygous	3	2:1:1, 1:1
Type 7	ccxab	15	0.0003	Homozygous	Heterozygous	3	2:1:1, 1:1

+ Percentage of each segregant type with respect to total mapped loci.

3.3. Genetic Linkage Analysis and Construction of Maps

The linkage analysis was performed by a total of 6434 unique SNPs, covering a total distance of 14,128.7 cM. As the mapping population was developed from same parents, the SNPs were well distributed among each linkage group with a maximum of 1341 SNPs on the linkage group 4 with a total length of 1986.08 cM and a minimum of 159 SNPs on linkage group 8 with a total length of 255.81 cM (Table 3). Markers were mainly heterozygous in both parents (hk × hk), which represented a 99.36% heterozygosity rate in the population (Table 1). The genetic map with consensus of genotypic information from parental and offspring was built. Based on the linkage among markers, 11 linkage groups were clearly observable and were confidently defined (Supplementary Figures S1 and S2). This was the same as the base chromosome number (i.e., 2n = 2x = 22) [5]. The map of the population was dense as the maximum number of gaps (85.4%) were less than 5 cM, while only 0.12% of gaps were above 20 cM of distance. The average marker density ranged from a minimum of 0.65 cM per marker on linkage group 10 to a 3.39 cM length per marker on linkage group 6 (Table 3).

Linkage Group	Marker Number	Length Covered	Average Length	Maximum Gap	Gap < 5 cM	Gap 5 to 10 cM	Gap 10 to 20 cM	Gap > 20 cM	Ratio
LG01	708	2070.07	2.92	23.05	568	100	39	1	80.23
LG02	483	987.5	2.04	33.31	430	39	13	1	89.03
LG03	667	1828.85	2.74	20.05	540	94	32	1	80.96
LG04	1341	1986.08	1.48	12.84	1233	103	5	0	91.95
LG05	566	1459.75	2.58	14.12	448	100	18	0	79.15
LG06	255	864.52	3.39	15.44	184	59	12	0	72.16
LG07	836	1954.33	2.34	15.35	694	97	45	0	83.01
LG08	159	255.81	1.61	12.45	150	8	1	0	94.34
LG09	457	1049.96	2.3	26.89	382	58	15	2	83.59
LG10	390	255.1	0.65	7.46	382	8	0	0	97.95
LG11	572	1416.76	2.48	26.08	485	67	17	3	84.79
Total	6434	14128.7	2.2	33.31	5496	733	197	8	85.42

Table 3. Summary of genetic markers and mapping of 11 linkage groups of Hylocereus species.

3.4. Consensus Map and Collinearity Analysis of Genetic and Physical Map

In order to reveal the conserved gene order in genetic and physical maps, a collinearity analysis was performed by the position of the marker on the reference genome and the genetic map. Linkage groups were numbered as per the reference genome. A fairly accurate map integration was revealed by a good collinearity between the homolog linkage groups from the genetic and physical maps (Figure 2).



Figure 2. Integration map collinearity analysis based on markers' positions on the physical map (bp) indicated in blue bars mapped to the positions on the genetic map (cM) indicated in red bars.

4. Discussion

The current study aimed at constructing a high-density genetic linkage map for *H. undatus* in order to facilitate further QTL/gene mapping and genome assembly-related

studies. Therefore, a mapping population of contrasting parents was first designed. Meanwhile, the polymorphic markers were identified, and a high-density linkage map was constructed with 6434 SNP markers spanning the 14,128.7cM genetic distance on 11 linkage groups.

The whole genome resequencing and genotyping are important steps to accelerate the genetic improvement of various traits in commercially and agronomically important species. Nevertheless, although *H. undatus* is an important economic fleshy fruited tree, few studies have been designed to investigate its genome [1,5,19]. In a recent study, the PacBio sequencing was performed and an assembled 1.41 Gb of genomic data were generated to design a reference genome for *H. undatus* [5]. The high ratio of repetitive/redundant sequences caused difficulties in the assembly procedure, and relatively short N50 contigs and scaffold sizes were achieved. Therefore, it may need deeper sequencing to get the maximum genome coverage. A high ratio of repetitive sequences with whole genome duplication and triplication has also been reported [5]. In this study, the comparative analysis of our genetic map with the recently reported reference genome showed 98% of sequence coverage with a very low gap rate, indicating the high quality of our sequenced data.

A maximum of 12,343,419 SNPs (Table 1) was revealed in the current study, accounting for only 0.875% of the total genome sequence as per a previously reported total genome size (1.41 Gb) [5]. Among the extracted SNPs, the discovered polymorphic markers were 6434 (Table 3). The ease to use transferability of SNP markers between different linkage maps and labs is their major advantage over amplified fragment length polymorphism (AFLP) and RAPD markers [32]. Each of the markers developed in the present study covered an average genetic distance of 2.2 cM. Therefore, it might be useful in the genome assembly and genomic evaluation studies.

The genomic studies always subjected to reduce the genotyping error rate and to enhance the read depth. The error rate could be ignored with the increasing read depth up to 12 [33]. For the generated raw data, the overall read depth of all revealed makers ranged from 4.43-fold to 17.65-fold, which was higher than the previously reported linkage map (16.3-fold) [5]. Therefore, to enhance the read depth of markers selected to construct the linkage map, data were filtered, and the makers with low read depth markers were removed prior to genotyping and mapping. Consequently, a high accuracy of marker-based genotyping was ensured by an optimum read depth.

There were 11 linkage groups in the linkage map constructed in this study, which was consistent with the karyotype of *H. undatus*. In various studies, the linkage maps constructed by SSR and/or AFLP markers revealed a different number of linkage groups than the actual chromosome numbers [34]. This might be due to the limited intermediate markers that linked the groups on same chromosome. In the present study, a sufficient number of (6,434) markers, which were randomly distributed across the genome, were mapped to 11 linkage groups and facilitated the complete map coverage. As a higher number of markers was used to genotype the 198 offspring, the clustering of several markers at one position was the result of limited meiotic events. The clustered markers may refer to as the "bin signatures". A bin signature comprises the consensus segregation pattern of marker loci that do not have any recombination and, thus, a marker interval of zero [35]. In the "bin signatures", the markers are normally known to cluster together; nonetheless, their orientation remains unknown, which may influence the scaffold orientation in genome assembly. The markers can be separated in "bin signatures" by genotyping the higher number of families and their progenies.

Using the polymorphic SNP markers from the two *Hylocereus* species only highlights the contrasting regions. This results in relatively uneven markers distribution along the linkage groups. Moreover, the screened SNPs covered only 0.875% of the total genome. The incomplete coverage or missing regions might be another limiting factor affecting the marker distribution. Even in this situation, more than 85% of marker interval spaces (gaps) were less than 5 cM. The larger gaps might be caused by the large number of repeated sequences in the *Hylocereus* genome. In the present study, the average distance between

adjacent markers was 2.2 cM in the linkage map, which is comparable to previous linkage maps [5,32,36–40]. Hence, this linkage map can be considered a high-density map for *Hylocereus*. The inter-marker distance compared to other previously reported linkage maps developed by using next-generation sequencing (NGS) technology for other species was also shorter [36–39].

The number of total SNPs in the male parent was much higher (12.343 million) than that of the female parent (2.962 million). As reported in previous linkage mapping studies, a limited number of common SNPs between the maps constructed from the male and female genome could be caused by the sex-linked genomic regions and have influenced the estimation accuracy of the male to female recombination ratios [16]. In the present study, a large number of common markers between the male and female genomes improved the estimated recombination frequency.

Integration of a high-density linkage genetic map with reference genome is beneficial for the improvement of genome assembly [5,16]. In the linkage map constructed in this study, the majority of markers were anchored to the physical map (Figure 2) and their relative distance between them was known. Collinearity analysis can reveal the conserved order of genes in genetic linkage and physical maps. The higher level of collinearity of anchored markers in this study will be helpful for future genome assembly and genomic studies, which may be further helpful for fine mapping of various parts of genome.

Above all the results related to genome assembly and the markers' integration, a preliminary arrangement of sequences along the physical length of chromosomes of *Hylocereus* was obtained. The level of integration also verified the accuracy of the linkage map constructed in this study. The markers that grouped together were also anchored to the scaffold with the same physical orientation, which partially validated the high accuracy of the linkage map. QTL mapping of traits of interest is one of the major applications of genetic maps. We anticipate that future studies will take advantage of our genetic map to uncover the genomic regions governing many growth and economically important traits in *Hylocereus*.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/ 10.3390/horticulturae7120534/s1, Figure S1: Consolidated map marker distribution map, X-axis: linkage group number; Y-axis: genetic distance (cM); blue lines indicate markers, Figure S2: Consolidated map marker distribution map, X-axis: chromosome number; Y-axis: genetic distance (cM); the left end of the horizontal line: marker genetic distance; the right end of the horizontal line: the name of the marker.

Author Contributions: Z.W., project design, conducted the experiments and article writing; H.D., conception, design of experiment and article writing; G.L., carried out experiments; X.Y., carried out nucleic acid extraction experiments; Y.Q., design, financial support, supervision and paper revision; L.H., design, supervision, revised draft and financial support. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by National Natural Science Foundation of China (31960578), Guangxi Science and Technology Project (GuiKe AD18281080), Science and Technology Development Foundation of Guangxi Academy of Agricultural Sciences (GuiNongKe 2021JM09, 31960578), Guangxi Innovation Team of National Modern Agricultural Industry Technology System (nycytxgxcxtd-17-06, nycytxgxcxtd-17-13), and Basic Scientific Research Business Special Project of Guangxi Academy of Agricultural Sciences (GuiNongKe 2021YT047).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and Supplementary Materials. The raw sequencing data have been deposited to NCBI SRA under the project number: PRJNA754205 (https://www.ncbi.nlm.nih.gov/bioproject/754205, accessed on 26 September 2021).

Conflicts of Interest: No conflict of interest.

References

- Fan, Q.; Yan, F.; Qiao, G.; Zhang, B.; Wen, X. Identification of differentially-expressed genes potentially implicated in drought response in pitaya (*Hylocereus undatus*) by suppression subtractive hybridization and cDNA microarray analysis. *Gene* 2014, 533, 322–331. [CrossRef] [PubMed]
- Lee, E.J.; An, D.; Nguyen, C.T.; Patil, B.S.; Kim, J.; Yoo, K.S. Betalain and betaine composition of greenhouse-or field-produced beetroot (*Beta vulgaris* L.) and inhibition of HepG2 cell proliferation. J. Agric. Food Chem. 2014, 62, 1324–1331. [CrossRef] [PubMed]
- Suh, D.H.; Lee, S.; Heo, D.Y.; Kim, Y.-S.; Cho, S.K.; Lee, S.; Lee, C.H. Metabolite profiling of red and white pitayas (*Hylocereus polyrhizus* and *Hylocereus undatus*) for comparing betalain biosynthesis and antioxidant activity. J. Agric. Food Chem. 2014, 62, 8764–8771. [CrossRef]
- da Silva, D.V.T.; dos Santos Baião, D.; de Oliveira Silva, F.; Alves, G.; Perrone, D.; Del Aguila, E.M.; Paschoalin, V.M.F. Betanin, a natural food additive: Stability, bioavailability, antioxidant and preservative ability assessments. *Molecules* 2019, 24, 458. [CrossRef]
- Chen, J.; Xie, F.; Cui, Y.; Chen, C.; Lu, W.; Hu, X.; Hua, Q.; Zhao, J.; Wu, Z.; Gao, D.; et al. A chromosome-scale genome sequence of pitaya (*Hylocereus undatus*) provides novel insights into the genome evolution and regulation of betalain biosynthesis. *Hortic. Res.* 2021, *8*, 164. [CrossRef]
- Zhuang, Y.; Zhang, Y.; Sun, L. Characteristics of fibre-rich powder and antioxidant activity of pitaya (*Hylocereus undatus*) peels. *Int. J. Food Sci. Technol.* 2012, 47, 1279–1285. [CrossRef]
- 7. Cai, Y.; Xiang, Q.; Chen, J.; Peng, Y.; Wang, B. Analysis of nutritional components in pitaya fruit. *Nonwood For. Res.* 2008, 26, 253–256.
- 8. Hua, Q.; Chen, C.; Chen, Z.; Chen, P.; Ma, Y.; Wu, J.; Zheng, J.; Hu, G.; Zhao, J.; Qin, Y. Transcriptomic analysis reveals key genes related to betalain biosynthesis in pulp coloration of *Hylocereus polyrhizus*. *Front. Plant Sci.* **2016**, *6*, 1179.
- 9. Adnan, L.; Osman, A.; Abdul Hamid, A. Antioxidant activity of different extracts of red pitaya (*Hylocereus polyrhizus*) seed. *Int. J. Food Prop.* **2011**, *14*, 1171–1181. [CrossRef]
- 10. Taguiam, J.D.; Evallo, E.; Bengoa, J.; Maghirang, R.; Balendres, M.A. Pathogenicity of Epicoccum sorghinum towards dragon fruits (*Hylocereus* species) and in vitro evaluation of chemicals with antifungal activity. *J. Phytopathol.* **2020**, *168*, 303–310. [CrossRef]
- 11. Stintzing, F.C.; Schieber, A.; Carle, R. Identification of betalains from yellow beet (*Beta vulgaris* L.) and cactus pear [*Opuntia ficus-indica* (L.) Mill.] by high-performance liquid chromatography–electrospray ionization mass spectrometry. *J. Agric. Food Chem.* **2002**, *50*, 2302–2307. [CrossRef] [PubMed]
- 12. Pang, X.; Li, X.; Liu, X.; Cai, L.; Li, B.; Li, X. Transcriptomic analysis reveals Cu/Zn SODs acting as hub genes of SODs in *Hylocereus undatus* induced by trypsin during storage. *Antioxidants* **2020**, *9*, 162. [CrossRef] [PubMed]
- 13. Wu, Y.; Xu, J.; He, Y.; Shi, M.; Han, X.; Li, W.; Zhang, X.; Wen, X. Metabolic profiling of pitaya (*Hylocereus polyrhizus*) during fruit development and maturation. *Molecules* **2019**, *24*, 1114. [CrossRef] [PubMed]
- 14. Freitas, S.T.d.; Mitcham, E.J. Quality of pitaya fruit (*Hylocereus undatus*) as influenced by storage temperature and packaging. *Sci. Agric.* **2013**, *70*, 257–262. [CrossRef]
- 15. Zheng, J.; Meinhardt, L.W.; Goenaga, R.; Zhang, D.; Yin, Y. The chromosome-level genome of dragon fruit reveals whole-genome duplication and chromosomal co-localization of betacyanin biosynthetic genes. *Hortic. Res.* **2021**, *8*, 63. [CrossRef] [PubMed]
- 16. Cormier, F.; Lawac, F.; Maledon, E.; Gravillon, M.-C.; Nudol, E.; Mournet, P.; Vignes, H.; Arnau, G. A reference high-density genetic map of greater yam (Dioscorea alata L.). *Theor. Appl. Genet.* **2019**, *132*, 1733–1744. [CrossRef] [PubMed]
- 17. Wang, Y.; Rashid, M.A.R.; Li, X.; Yao, C.; Lu, L.; Bai, J.; Li, Y.; Xu, N.; Yang, Q.; Zhang, L. Collection and evaluation of genetic diversity and population structure of potato landraces and varieties in China. *Front. Plant Sci.* **2019**, *10*, 139. [CrossRef]
- 18. Rashid, M.A.R.; Zhao, Y.; Zhang, H.; Li, J.; Li, Z. Nucleotide diversity, natural variation, and evolution of Flexible culm-1 and Strong culm-2 lodging resistance genes in rice. *Genome* **2016**, *59*, 473–483. [CrossRef] [PubMed]
- Nashima, K.; Hosaka, F.; Shimajiri, Y.; Matsumura, M.; Tarora, K.; Urasaki, N.; Shoda, M.; Nishitani, C.; Sawamura, Y.; Yamamoto, T. SSR marker development and genetic identification of pitaya (*Hylocereus* spp.) collected in Okinawa Prefecture, Japan. *Hortic. J.* 2021, 90, 23–30. [CrossRef]
- Fan, Q.; Zheng, S.; Yan, F.; Zhang, B.; Qiao, G.; Wen, X. Efficient regeneration of dragon fruit (*Hylocereus undatus*) and an assessment of the genetic fidelity of in vitro-derived plants using ISSR markers. *J. Hortic. Sci. Biotechnol.* 2013, 88, 631–637. [CrossRef]
- Tao, J.; Qiao, G.; Wen, X.; Gao, G.; Liu, T.; Peng, Z.; Cai, Y.; Chen, N.; Yan, F.; Zhang, B. Characterization of genetic relationship of dragon fruit accessions (*Hylocereus* spp.) by morphological traits and ISSR markers. *Sci. Hortic.* 2014, 170, 82–88. [CrossRef]
- Rifat, T.; Khan, K.; Islam, M.S. Genetic diversity in dragon fruit (*Hylocereus* sp.) germplasms revealed by RAPD marker. J. Anim. Plant Sci. 2019, 29, 809–818.
- 23. Tel-Zur, N.; Abbo, S.; Bar-Zvi, D.; Mizrahi, Y. Clone identification and genetic relationship among vine cacti from the genera *Hylocereus* and *Selenicereus* based on RAPD analysis. *Sci. Hortic.* **2004**, *100*, 279–289. [CrossRef]
- 24. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **2012**, *7*, e32253.
- Ward, J.A.; Bhangoo, J.; Fernández-Fernández, F.; Moore, P.; Swanson, J.; Viola, R.; Velasco, R.; Bassil, N.; Weber, C.A.; Sargent, D.J. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genom.* 2013, 14, 2. [CrossRef] [PubMed]

- 26. Herten, K.; Hestand, M.S.; Vermeesch, J.R.; Van Houdt, J.K. GBSX: A toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinform.* **2015**, *16*, 73. [CrossRef]
- 27. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011, 17, 10–12. [CrossRef]
- Garsmeur, O.; Droc, G.; Antonise, R.; Grimwood, J.; Potier, B.; Aitken, K.; Jenkins, J.; Martin, G.; Charron, C.; Hervouet, C. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 2018, *9*, 2638. [CrossRef] [PubMed]
- 29. Knaus, B.J.; Grünwald, N.J. VCFR: A package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 2017, 17, 44–53. [CrossRef]
- 30. RC Team. R: A Language and Environment for Statistical Computing; RC Team: Diepoldsau, Switzerland, 2013.
- 31. van Ooijen, J.W. *JoinMap*®4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations; Scientific Reserch: Wagening, Switzerland, 2006; Volume 93, p. 7778Walling.
- 32. Yu, Y.; Zhang, X.; Yuan, J.; Li, F.; Chen, X.; Zhao, Y.; Huang, L.; Zheng, H.; Xiang, J. Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp *Litopenaeus vannamei*. *Sci. Rep.* **2015**, *5*, 15612. [CrossRef]
- 33. Sun, X.; Liu, D.; Zhang, X.; Li, W.; Liu, H.; Hong, W.; Jiang, C.; Guan, N.; Ma, C.; Zeng, H. SLAF-seq: An efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* **2013**, *8*, e58700.
- 34. Campos-Ramos, R. Chromosome studies on the marine shrimps *Penaeus Uannamei* and *P. Californiensis* (Decapoda). *J. Crustacean Biol.* **1997**, 17, 666–673. [CrossRef]
- 35. Ren, Y.; Zhao, H.; Kou, Q.; Jiang, J.; Guo, S.; Zhang, H.; Hou, W.; Zou, X.; Sun, H.; Gong, G. A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. *PLoS ONE* **2012**, *7*, e29453. [CrossRef]
- 36. Zhang, L.; Yang, C.; Zhang, Y.; Li, L.; Zhang, X.; Zhang, Q.; Xiang, J. A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): Sex-linked microsatellite markers and high recombination rates. *Genetica* **2007**, *131*, 37–49. [CrossRef]
- Du, Z.Q.; Ciobanu, D.; Onteru, S.; Gorbach, D.; Mileham, A.; Jaramillo, G.; Rothschild, M. A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei. Anim. Genet.* 2010, 41, 286–294. [CrossRef]
- 38. Pérez, F.; Erazo, C.; Zhinaula, M.; Volckaert, F.; Calderón, J. A sex-specific linkage map of the white shrimp *Penaeus* (*Litopenaeus*) *vannamei* based on AFLP markers. *Aquaculture* **2004**, 242, 105–118. [CrossRef]
- 39. Andriantahina, F.; Liu, X.; Huang, H. Correction: Genetic map construction and quantitative trait locus (QTL) detection of growth-related traits in *Litopenaeus vannamei* for selective breeding applications. *PLoS ONE* **2013**, *8*, e75206.
- Alcivar-Warren, A.; Meehan-Meola, D.; Park, S.W.; Xu, Z.; Delaney, M.; Zuniga, G. ShrimpMap: A low-density, microsatellitebased linkage map of the pacific whiteleg shrimp, *Litopenaeus vannamei*: Identification of sex-linked markers in linkage group 4. *J. Shellfish Res.* 2007, 26, 1259–1277. [CrossRef]