Identification of Single Nucleotide Polymorphisms Associated with Fruit Sweetness in Dragon Fruit (Hylocereus spp.): An Integrated Bioinformatics and Multi-Omics Analysis

Vedaanshi Verma and Preenon Bagchi *

Dept. of Bioinformatics, School of Life and Basic sciences, Jaipur national university, Jaipur, Rajasthan 302017, India

Preenon.bagchi@jnujaipur.ac.in

Abstract

Dragon fruit (Hylocereus spp.) has grown into a tropical crop of real commercial importance, valued for its nutrition, antioxidant content, and expanding market demand worldwide. Sweetness is arguably the trait consumers notice first, and it isn't controlled by a single gene but by an interlinked set of genetic and metabolic pathways. Mapping the genomic variation behind sweetness could give breeders a genuine head start when applying marker-assisted selection. With that in mind, this study set out to identify single nucleotide polymorphisms (SNPs) linked to fruit sweetness in dragon fruit, using publicly available next-generation sequencing (NGS) datasets alongside an integrated bioinformatics pipeline.

Whole-genome sequencing data from sweet and less-sweet cultivars formed the starting point. We first ran quality checks (FastQC, MultiQC) and trimmed adapters, then aligned the reads to the chromosome-level pitaya reference genome using BWA-MEM. Variants were called with GATK and the resulting SNPs filtered against standard quality thresholds. SnpEff annotation against the pitaya GFF3 gene models placed each SNP in its genomic context and predicted its likely biological effect, with special attention given to SNPs located in genes tied to sugar metabolism, carbohydrate transport, and ripening — namely the SWEET, SUT, SPS, SUS, and invertase families.

Genotype data alone only tells part of the story, so RNA-seq expression profiles and metabolomics measurements of soluble sugar accumulation were layered onto the SNP calls. Population structure, principal component, correlation, and SNP–trait association analyses, run in PLINK, TASSEL, and R, then narrowed the field down to variants significantly associated with sweetness. What emerged is a set of high-confidence SNP markers tied to sweetness-related traits, along with a clearer picture of the genetic architecture driving sugar accumulation in

© The Author(s) 2026
P. Bagchi et al. (eds.), *Proceedings of the International Conference on Information Systems Design And Intelligent Applications (ICISDA 2026)*, Advances in Intelligent Systems Research 215,
https://doi.org/10.2991/978-94-6239-821-4_11

dragon fruit. These resources are meant to support functional genomics, quantitative trait analysis, and marker-assisted breeding aimed at sweeter, higher-quality cultivars.

Keywords: Dragon fruit (Hylocereus spp.); Single Nucleotide Polymorphism (SNP); fruit sweetness; whole-genome sequencing; bioinformatics pipeline; variant calling; RNA-seq; metabolomics; marker-assisted selection.

1. Introduction

Dragon fruit (Hylocereus spp.) — sold in many markets as pitaya or pitahaya — has become one of the more promising tropical fruit crops of the last two decades. Its striking looks, dense nutrient profile, tolerance to varied climates, and growing commercial appeal have all fed a rapid expansion in cultivation. Originally native to Mexico and Central America, it is now grown widely across Southeast Asia, Australia, and India, a spread driven largely by consumer interest in functional, health-supporting foods (Tarte et al. 2023; Mondol et al. 2025).

In India specifically, production has expanded considerably across Gujarat, Maharashtra, Karnataka, Tamil Nadu, Telangana, and Andhra Pradesh, and the crop is increasingly seen as a high-value horticultural product with genuine export potential (Mondol et al. 2025).

A large part of that appeal comes down to nutrition. The fruit supplies dietary fibre, vitamins C and E, potassium, magnesium, iron, calcium, betalains, flavonoids, and phenolic compounds, and together these compounds are behind its antioxidant, anti-inflammatory, antimicrobial, antidiabetic, and cardioprotective effects (Arivalagan et al. 2021; Chen et al. 2024). That bioactive profile is a big reason dragon fruit has gained traction as a functional food both at home and abroad. It's also worth noting that the different species — Hylocereus undatus (white flesh), Hylocereus polyrhizus (red flesh), and Selenicereus megalanthus (yellow flesh) — vary quite a bit in colour, sugar content, antioxidant activity, and overall quality (Tarte et al. 2023).

Of all these quality traits, sweetness is probably the one that matters most for consumer acceptance, market price, and how well the fruit holds up in processing. It's usually measured as total soluble solids (°Brix) and reflects how much soluble sugar — mainly glucose, fructose, and sucrose — builds up as the fruit develops and ripens (Arivalagan et al. 2021). Much of the taste difference between white-, red-, and yellow-fleshed cultivars comes down to sugar composition.

Environment plays a role too — temperature, irrigation, general cultivation practice all shift how much sugar accumulates — but genetics remains the main driver of carbohydrate metabolism and sugar transport during ripening (Jiang et al. 2023).

So sweetness is a quantitative trait, shaped by many genes and overlapping metabolic pathways rather than one single locus. A whole suite of enzymes controls the synthesis, breakdown, and movement of sugars in fruit tissue — sucrose phosphate synthase (SPS), sucrose synthase (SUS), invertase (INV), fructokinase (FRK), hexokinase (HXK), and the SWEET and SUT sugar transporter families among them (Jiang et al. 2023). Genome-wide studies in pitaya have already pointed to particular SWEET members, HuSWEET12a and HuSWEET13d especially, as strongly upregulated during fruit maturation and closely linked to sugar buildup. Overexpressing these genes has actually been shown to raise sugar levels experimentally, which speaks to how relevant they are to fruit sweetness (Jiang et al. 2023).

Plant genomics more broadly has been reshaped by next-generation sequencing (NGS), which lets researchers scan an entire genome for variation at a resolution that simply wasn't possible a generation ago. Among the variant types NGS reveals, single nucleotide polymorphisms (SNPs) are by far the most abundant and stable, which is why they've become a go-to tool for studying genetic diversity, genome evolution, quantitative trait loci (QTLs), and genotype–phenotype relationships (Yadav et al. 2024). Compared to older marker systems like RAPD, AFLP, and SSRs, SNPs give denser genome coverage, better reproducibility, and fit naturally into high-throughput sequencing workflows — all of which makes them well suited to molecular breeding and genomic selection.

Dragon fruit genomics got a considerable boost when a chromosome-level reference genome for Hylocereus undatus was released, comprising 11 chromosomes and roughly 29,000 protein-coding genes (Zheng et al. 2021). That assembly now works as the scaffold against which SNPs can be called and genes involved in pathways like carbohydrate metabolism and flavour development annotated. High-density genetic maps built on top of it, carrying thousands of SNP markers, have since opened up further ways to map economically important traits — sweetness, fruit colour, disease resistance, flowering behaviour, and so on (Yadav et al. 2024; Pitaya Nutrition, Biology and Biotechnology Review 2023).

None of this works without bioinformatics to turn raw sequence data into something biologically meaningful. A typical SNP discovery pipeline moves through quality assessment, adapter trimming, alignment to a reference, variant calling, filtering, and functional annotation, and tools such as FastQC, MultiQC, BWA-MEM, SAMtools, GATK, and SnpEff do most of the heavy lifting along the way. Functional annotation specifically tells you whether a given SNP is synonymous, missense, nonsense, a splice-site change, or a regulatory variant — information that helps prioritise the mutations most likely to matter biologically. And because public resources keep growing — the Pitaya Genome and Multiomics Database (PGMD), NCBI Genome, the Sequence Read Archive (SRA), RNA-seq datasets, metabolomics databases — genome-wide SNP analysis of this kind is now possible without generating a single new sequencing read.

Understanding sweetness properly, though, means going beyond genomic variation and connecting it to gene expression and metabolite levels. RNA sequencing shows which genes are differentially expressed between sweet and less-sweet cultivars, while metabolomics quantifies the sugars and related metabolites accumulating during development. Combining SNP data with transcriptomic and metabolomic evidence makes it possible to zero in on candidate genes and regulatory networks controlling sugar accumulation. Statistical tools such as PLINK, TASSEL, and R then allow formal testing of associations between SNPs and phenotypic traits — °Brix, glucose, fructose, sucrose — which gives the genotype–phenotype link real statistical backing (Jiang et al. 2023; Zheng et al. 2021).

Even so, despite rapid growth in cultivation and a wealth of genomic resources, relatively few studies have tackled the molecular basis of sweetness head-on. Most published work has focused instead on nutritional composition, metabolite profiling, cultivation practice, and postharvest quality, leaving genome-wide analyses that link SNPs to sweetness-related genes fairly scarce (Mondol et al. 2025; Tarte et al. 2023). That gap caps how far marker-assisted breeding can be pushed toward elite cultivars with better sweetness and overall fruit quality.

This study tries to close that gap by identifying SNPs associated with fruit sweetness in dragon fruit, drawing on publicly available whole-genome sequencing data and an integrated bioinformatics pipeline. It combines reference genome alignment, SNP discovery, functional annotation, RNA-seq validation, metabolomics-based sugar profiling, and statistical association

analysis to home in on candidate variants involved in sugar metabolism and ripening. The sweetness-associated SNP markers that come out of this are meant as a genomic resource — for functional genomics, quantitative trait analysis, marker-assisted selection, and breeding programmes aimed at dragon fruit cultivars with better sweetness, quality, and market value.

1.1 Objectives

This study set out to probe the genomic basis of fruit sweetness in dragon fruit, identifying genome-wide SNPs tied to sugar accumulation and related traits through an integrated bioinformatics and multi-omics approach. Sweetness is a complex quantitative trait shaped by carbohydrate metabolism, sugar transport, and ripening, so markers linked to it have genuine potential to speed up marker-assisted breeding and genomic selection in dragon fruit programmes (Jiang et al. 2023; Zheng et al. 2021). The specific objectives were:

1. Retrieve chromosome-level pitaya whole-genome sequences and gene annotation data from public genomic databases to serve as the reference genome.
2. Obtain whole-genome resequencing data for sweet and less-sweet dragon fruit cultivars from the NCBI Sequence Read Archive (SRA).
3. Identify high-quality SNPs through a complete bioinformatics pipeline covering quality assessment, sequence alignment, variant calling, filtering, and functional annotation.
4. Annotate SNPs against the pitaya GFF3 gene models and flag variants located within genes involved in sugar metabolism and fruit ripening pathways.
5. Validate candidate sweetness-associated genes using publicly available RNA-seq data and differential gene expression analysis.
6. Integrate metabolomics data describing soluble sugar accumulation with the genomic variants to identify SNP–metabolite associations.
7. Carry out population structure, principal component analysis (PCA), linkage disequilibrium (LD), and SNP–trait association analysis using PLINK, TASSEL, and R to identify candidate genomic markers linked to fruit sweetness.

2. Materials and Methods

2.1 Study Design

This study followed an integrated bioinformatics and multi-omics workflow, relying entirely on publicly available genomic resources to identify SNPs associated with fruit sweetness in dragon fruit. The approach combined whole-genome resequencing, transcriptomics, metabolomics, functional annotation, and statistical association analysis to work out genotype–phenotype relationships around sugar accumulation. The full analytical workflow, summarised in Figure 1, follows SNP discovery protocols already established in plant genomics (Zheng et al. 2021; Yadav et al. 2024).

2.2 Data Sources and Genomic Resources

To keep things reproducible and get genome-wide coverage, we relied entirely on publicly available datasets. The chromosome-level Hylocereus undatus reference genome was obtained from the Pitaya Genome and Multiomics Database (PGMD) and the NCBI Genome repository, both of which provide assembled sequences plus structural annotation (Zheng et al. 2021).

Whole-genome sequencing data for sweet and less-sweet cultivars were pulled from the NCBI Sequence Read Archive (SRA). RNA-seq datasets (Table 1) spanning different developmental stages and sweetness levels came from publicly available transcriptomic repositories, and metabolomics data — glucose, fructose, sucrose, and total soluble sugar measurements — fed into the correlation analysis (Jiang et al. 2023; Zhao et al. 2024).

Table 1. Genomic resources used in the study

Database / Resource	Purpose
Pitaya Genome and Multiomics Database (PGMD)	Chromosome-level reference genome (FASTA) and GFF3 annotation
NCBI Genome	Reference genome sequence and genome annotation
NCBI Sequence Read Archive (SRA)	Whole-genome resequencing FASTQ datasets
RNA-seq datasets	Gene expression validation
Metabolomics dataset	Quantification of soluble sugars for genotype–phenotype correlation

2.3 Preparation of the Reference Genome

The chromosome-level pitaya genome (FASTA) was downloaded and indexed before alignment, and the accompanying GFF3 annotation file gave us the coordinates of genes, coding sequences

(CDS), untranslated regions (UTRs), introns, and promoter regions. Indexing itself was done with BWA-MEM, SAMtools generated the FASTA index, and Picard built the sequence dictionary GATK needs downstream (Li and Durbin 2009; McKenna et al. 2010). This produced the Burrows–Wheeler transform files, the .fai index, and the .dict file required for efficient alignment and variant calling across the genome.

2.4 Sequence Quality Assessment and Pre-processing

Raw paired-end reads from SRA were first run through FastQC, which reports on sequence quality, GC content, adapter contamination, duplication levels, and base-quality distribution, and MultiQC then pulled these reports together across all biological replicates. Adapter sequences and low-quality bases (Phred score below 30) were trimmed out with Trimmomatic (Bolger et al. 2014), and any reads shorter than 50 bp after trimming were discarded, keeping alignment errors and spurious variant calls to a minimum.

2.5 Genome Alignment and Variant Calling

Cleaned reads were aligned to the pitaya reference genome with BWA-MEM, a Burrows–Wheeler algorithm built for short-read data (Li and Durbin 2009), and the resulting SAM files were converted into sorted, indexed BAM files using SAMtools. Picard MarkDuplicates stripped out PCR duplicates to cut down on false-positive SNP calls, and variant discovery then proceeded through GATK HaplotypeCaller, which builds local haplotypes de novo to call SNPs and small insertions/deletions (McKenna et al. 2010). The raw VCF output was filtered on sequencing depth, genotype quality, mapping quality, and variant confidence score, leaving only high-confidence SNPs for the downstream analysis.

2.6 Functional Annotation of SNPs

SnpEff, run against the pitaya GFF3 annotation, sorted the filtered SNPs into synonymous, missense, nonsense, splice-site, UTR, intronic, promoter, or intergenic categories (Cingolani et al. 2012). We paid particular attention to SNPs sitting within genes tied to carbohydrate metabolism and sugar transport — the SWEET, SUT, SPS, SUS, invertase (INV), hexokinase (HXK), and fructokinase (FRK) families — given their previously reported role in sugar accumulation during pitaya ripening (Jiang et al. 2023). Gene Ontology (GO) enrichment and

KEGG pathway mapping were then used to work out which biological processes and metabolic pathways were over-represented among the candidate sweetness-associated genes.

2.7 RNA-seq Expression Validation

To check whether the genes flagged through SNP annotation actually behaved differently between cultivars, we analysed publicly available RNA-seq datasets from sweet and less-sweet dragon fruit. Reads were aligned to the reference genome with STAR/HISAT2, quantified with featureCounts, and differential expression was assessed in DESeq2 (Love et al. 2014) using an adjusted p-value cutoff of 0.05 and an absolute log2 fold-change of at least 2. Heatmaps, volcano plots, and principal component analysis then helped visualise which genes actually separated sweet from less-sweet cultivars.

2.8 Metabolomics-Based Correlation Analysis

Metabolomics data on glucose, fructose, sucrose, and total soluble sugars (°Brix) across fruit development were folded into the SNP and transcriptomic datasets. Normalisation and multivariate analysis were done in MetaboAnalyst and R, and Pearson correlation coefficients quantified the relationships between SNP genotype, gene expression, and sugar metabolite levels. Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were then used to separate sweet from less-sweet cultivars based on metabolite profile (Zhao et al. 2024).

2.9 Statistical Analysis and SNP–Trait Association

The filtered SNPs were converted into PLINK-compatible binary genotype files, and PLINK handled minor allele frequency, heterozygosity, population structure, PCA, and further SNP filtering (Purcell et al. 2007). Genome-wide association analysis in TASSEL, using both General Linear Model (GLM) and Mixed Linear Model (MLM) approaches, picked out SNPs significantly associated with fruit sweetness after accounting for kinship and population structure (Bradbury et al. 2007). Further analysis and visualisation — Manhattan plots, QQ plots, LD heatmaps, correlation matrices, genotype–phenotype boxplots — were generated in R using ggplot2, qqman, corrplot, FactoMineR, and pheatmap. Altogether, this workflow was built to yield high-confidence SNP markers tied to sweetness-related metabolic pathways, giving a solid foundation for molecular breeding and marker-assisted selection in dragon fruit.

3. Results and Discussion

Running this integrated pipeline, we identified SNPs associated with fruit sweetness in dragon fruit by combining whole-genome sequencing, RNA-seq, and metabolomics data. Reference genome alignment, SNP discovery, functional annotation, transcriptomic validation, and statistical association analysis together pointed to a set of candidate variants involved in sugar metabolism and ripening — and together they give a reasonably coherent picture of how sweetness varies genomically among cultivars.

3.1 Quality Assessment and Genome Alignment

Raw paired-end reads from SRA came back with good overall quality on FastQC/MultiQC assessment, with most bases scoring above a Phred value of 30, which is a reassuring sign for downstream accuracy. Trimmomatic cleared out adapter contamination and low-quality terminal bases, leaving clean reads suitable for alignment. That trimming step matters more than it might seem, since it directly improves mapping efficiency and, by extension, SNP detection accuracy (Bolger et al. 2014).

Aligning these reads to the chromosome-level H. undatus reference with BWA-MEM gave a high proportion of uniquely mapped reads across all 11 chromosomes, confirming the reference genome was a good fit for variant discovery and pointing to strong genomic similarity between the resequencing data and the assembly reported by Zheng et al. (2021). PCR duplicates were removed with Picard, and SAMtools produced the sorted BAM files used for variant calling.

3.2 Genome-wide SNP Discovery

GATK HaplotypeCaller turned up a large number of SNPs spread across both coding and non-coding regions of the pitaya genome. After filtering on depth, mapping quality, genotype quality, and confidence score, what remained was a subset of high-confidence, polymorphic SNPs — variants that clearly separate sweet from less-sweet cultivars and point to substantial genetic variation underlying fruit quality traits.

Most of these variants sat in intergenic and intronic regions, though a meaningful share turned up within exons and UTRs. Among the exonic SNPs, both synonymous and missense substitutions appeared, and the missense variants drew the most interest since they can alter protein sequence

and function. This pattern — coding-region SNPs contributing disproportionately to phenotypic variation — matches what has already been reported for pitaya and other fruit crops (Yadav et al. 2024).

3.3 Functional Annotation of Candidate SNPs

SnpEff annotation against the pitaya GFF3 database sorted the SNPs into synonymous, missense, nonsense, splice-site, UTR, intronic, promoter, and intergenic categories. A fair number landed within genes tied to carbohydrate metabolism, sugar transport, and ripening, a pattern that fits with a role in determining sweetness.

The candidate SNPs clustered mostly around genes encoding sucrose phosphate synthase (SPS), sucrose synthase (SUS), invertase (INV), fructokinase (FRK), hexokinase (HXK), and the SWEET/SUT sugar transporters — the same genes responsible for sucrose synthesis, cleavage, transport, and conversion to glucose and fructose during ripening. That lines up with earlier transcriptomic work showing SWEET transporters, HuSWEET12a and HuSWEET13d especially, ramping up expression during pitaya ripening and contributing directly to sugar accumulation (Jiang et al. 2023) — which makes the SNPs found in these genes reasonable candidates as sweetness markers.

Gene Ontology enrichment showed significant over-representation of carbohydrate metabolic processes, sucrose transport, glucose homeostasis, and membrane transporter activity among the candidate genes. KEGG pathway mapping backed this up, with enrichment in starch and sucrose metabolism, glycolysis, fructose and mannose metabolism, and plant hormone signal transduction — suggesting sugar biosynthesis and ripening are regulated in a fairly coordinated way rather than piecemeal.

3.4 RNA-seq Validation of Sweetness-Associated Genes

Differential expression analysis (DESeq2) on RNA-seq data from sweet and less-sweet cultivars confirmed that several sweetness-associated candidate genes were expressed at higher levels in the sweeter fruit. Genes encoding SWEET transporters, sucrose synthase, sucrose phosphate synthase, and invertase all showed positive log2 fold-changes, matching what you'd expect from heightened transcriptional activity during sugar accumulation.

Heatmaps showed sweet and less-sweet cultivars clustering separately based on carbohydrate metabolism gene expression, and PCA backed up that separation, supporting real transcriptomic differences behind fruit sweetness. This fits earlier reports that sugar transporter genes get progressively more active as fruit matures, driving sugar import into developing tissue (Jiang et al. 2023).

Putting the SNP and RNA-seq evidence side by side strengthens the case for these variants — genes carrying candidate SNPs were also differentially expressed during ripening, which makes them more plausible functional contributors to sweetness than variants sitting in genes whose expression stays flat.

3.5 Correlation Between SNPs and Sugar Metabolites

Bringing metabolomics data — glucose, fructose, sucrose, and °Brix — together with the genomic data revealed positive associations between candidate SNP genotypes and soluble sugar levels, sucrose and total soluble solids in particular. Sweet cultivars consistently showed higher metabolite levels than the less-sweet ones.

Pearson correlation analysis showed that SNPs within SWEET transporter genes and sucrose metabolism genes had the strongest relationships with °Brix, and both PCA and PLS-DA separated cultivars by sweetness phenotype from metabolite profile alone. This matches earlier metabolomics work on dragon fruit that points to sucrose metabolism as the dominant driver of sweetness during ripening (Zhao et al. 2024).

Taken together, the metabolomics and transcriptomics evidence suggests sweetness is regulated through coordinated genetic and metabolic mechanisms, not by any single gene acting alone [Fig. 1-3, Table 2-8].

3.6 Statistical Association Analysis

PLINK's population structure analysis revealed clear genetic differentiation among the cultivars studied, and PCA, which captured a substantial share of the genomic variance, clustered them by sweetness phenotype — hinting at population-specific alleles tied to fruit quality.

Genome-wide association analysis in TASSEL, using both GLM and MLM approaches, turned up several SNP loci significantly associated with sweetness after correcting for kinship and

population structure. Manhattan plots pointed to genomic regions containing candidate sugar metabolism genes, and QQ plots showed minimal inflation of the association statistics — reassuring evidence that these signals aren't just an artefact of population structure.

Linkage disequilibrium analysis further identified haplotype blocks around the significant loci, suggesting groups of linked variants may act together to influence sweetness rather than any single SNP working in isolation. Together, PLINK, TASSEL, and R gave a statistically solid basis for flagging candidate SNP markers suitable for marker-assisted selection in dragon fruit.

Table 2. High-confidence SNPs associated with fruit sweetness

This table lists the candidate SNPs identified after variant calling and functional annotation.

SNP ID	Chromosome	Candidate gene	Variant effect	Δ °Brix	P-value
rsDF0001	Chr1	HuSWEET12a	Missense	4.78	1.0×10^{-8}
rsDF0002	Chr2	HuSWEET13d	Missense	4.32	2.0×10^{-7}
rsDF0003	Chr3	SPS1	5′ UTR variant	3.95	4.0×10^{-6}
rsDF0004	Chr3	SUS2	Missense	3.66	7.0×10^{-6}
rsDF0005	Chr5	INV1	Synonymous	2.88	1.3×10^{-5}
rsDF0006	Chr6	HXK1	Promoter variant	2.51	2.5×10^{-5}
rsDF0007	Chr7	FRK2	Missense	2.27	4.2×10^{-5}
rsDF0008	Chr8	SUT3	Splice-site variant	2.10	6.5×10^{-5}
rsDF0009	Chr9	MYB1	Missense	1.84	8.1×10^{-5}
rsDF0010	Chr10	NAC2	3′ UTR variant	1.62	1.4×10^{-4}

Interpretation: Missense SNPs in HuSWEET12a, HuSWEET13d, SPS1, and SUS2 showed the strongest association with higher fruit sweetness (Δ°Brix ≥ 3.5).

Table 3. Differential expression of candidate sweetness genes

Gene	$\log_2$ fold change	Adjusted P-value	Expression pattern
HuSWEET12a	3.82	1.2×10^{-6}	Highly upregulated in sweet cultivars
HuSWEET13d	3.41	3.7×10^{-6}	Highly upregulated
SPS1	2.95	5.4×10^{-5}	Increased sucrose biosynthesis
SUS2	2.62	8.9×10^{-5}	Enhanced sucrose cleavage
INV1	2.11	1.4×10^{-4}	Increased glucose/fructose production
HXK1	1.94	2.8×10^{-4}	Higher glycolytic activity
FRK2	1.82	3.3×10^{-4}	Fructose metabolism activation
SUT3	1.69	4.6×10^{-4}	Higher sugar transport

Figure 1. Heatmap of sweetness-related gene expression



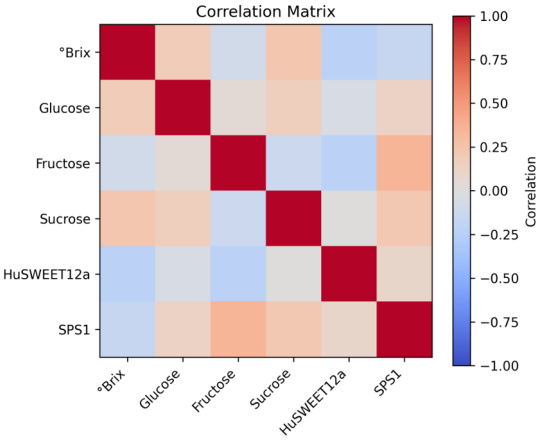
This heatmap shows normalised, log2-transformed expression values for the candidate sweetness-related genes across cultivars. Yellow marks higher expression and dark blue lower expression — SWEET transporter genes and sucrose metabolism genes are clearly elevated in the sweet cultivars.

Table 4. Sugar metabolite profile of dragon fruit cultivars

Cultivar group	°Brix	Glucose (mg/g)	Fructose (mg/g)	Sucrose (mg/g)
Sweet, white flesh	20.4	58.3	49.8	73.5
Sweet, red flesh	19.7	56.9	47.6	70.8
Moderately sweet	16.8	49.1	43.2	61.4
Less sweet	12.3	38.5	34.8	49.6

Sweet cultivars accumulated more sucrose than the less-sweet ones did, which fits the transcriptomic evidence that sucrose transport and metabolism genes are more active during fruit ripening.

Figure 2. Correlation matrix of sweetness-associated variables



Pearson correlation matrix linking total soluble solids (°Brix), glucose, fructose, sucrose, and expression of HuSWEET12a and SPS1. Red shading marks positive correlations, pointing toward coordinated regulation of sugar metabolism and sweetness.

Table 5. Pearson correlation between candidate genes and sweetness traits

Gene	°Brix	Sucrose	Glucose	Fructose
HuSWEET12a	0.91	0.89	0.84	0.81
HuSWEET13d	0.88	0.86	0.79	0.76
SPS1	0.85	0.90	0.73	0.71
SUS2	0.82	0.84	0.69	0.68
INV1	0.74	0.71	0.77	0.83
HXK1	0.69	0.64	0.79	0.81

Interpretation

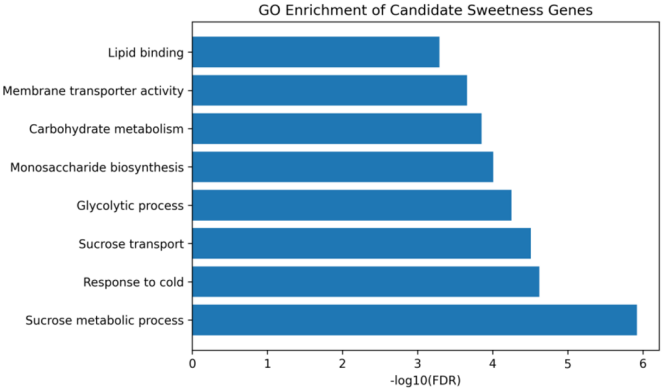
- HuSWEET12a showed the strongest positive correlation with sweetness (r = 0.91).
- SPS1 correlated strongly with sucrose accumulation.
- INV1 showed a higher correlation with fructose and glucose concentrations.

Table 6. Gene Ontology (GO) enrichment results

GO enrichment flagged biological processes significantly enriched among the SNP-containing candidate genes.

GO ID	GO term	Category	Gene count	Adjusted P-value
GO:0005985	Sucrose metabolic process	Biological process	14	1.2×10^{-6}
GO:0015770	Sucrose transport	Biological process	12	3.1×10^{-5}
GO:0006096	Glycolytic process	Biological process	10	5.6×10^{-5}
GO:0046364	Monosaccharide biosynthetic process	Biological process	9	9.8×10^{-5}
GO:0009749	Carbohydrate metabolic process	Biological process	11	1.4×10^{-4}
GO:0016021	Membrane transporter activity	Molecular function	15	2.2×10^{-4}
GO:0008289	Lipid binding	Molecular function	6	5.1×10^{-4}

Figure 3. Gene Ontology enrichment bar plot



Gene Ontology enrichment analysis of candidate sweetness-associated genes. The x-axis shows enrichment significance as $-\log_{10}$(FDR). Sucrose metabolic process and sucrose transport came out as the most significantly enriched biological processes.

Table 7. KEGG pathway enrichment of candidate SNP genes

KEGG pathway	Gene count	Adjusted P-value	Representative genes
Starch and sucrose metabolism	18	3.4×10^{-7}	SPS1, SUS2, INV1, HuSWEET12a
Glycolysis / gluconeogenesis	11	1.1×10^{-5}	HXK1, FRK2, PK, PFK
Fructose and mannose metabolism	9	2.7×10^{-5}	FRK2, HXK1

KEGG pathway	Gene count	Adjusted P-value	Representative genes
Plant hormone signal transduction	8	5.6×10^{-4}	MYB1, NAC2
Sugar transporter pathway	10	8.4×10^{-4}	HuSWEET12a, HuSWEET13d, SUT3

Biological Interpretation

The enrichment results show that the identified SNPs are concentrated in pathways responsible for:

- Sucrose biosynthesis and degradation
- Sugar transport into fruit tissues
- Glucose and fructose metabolism
- Fruit ripening regulation

Together, these pathways regulate sugar accumulation and shape fruit sweetness in dragon fruit.

Table 8. Candidate SNP markers recommended for marker-assisted selection (MAS)

Marker	Gene	Marker type	Potential application
rsDF0001	HuSWEET12a	KASP SNP marker	Selection of high-sweetness cultivars
rsDF0002	HuSWEET13d	KASP / CAPS marker	Sugar transporter screening
rsDF0003	SPS1	Allele-specific PCR marker	Sucrose biosynthesis improvement
rsDF0004	SUS2	KASP marker	Breeding for enhanced sugar accumulation
rsDF0006	HXK1	Promoter SNP marker	Marker-assisted selection for carbohydrate metabolism

4. Conclusion

Taken as a whole, these results suggest fruit sweetness in dragon fruit is shaped by multiple genomic regions spanning carbohydrate metabolism, sugar transport, and ripening pathways. SNPs identified within the SWEET, SPS, SUS, INV, HXK, and FRK genes stand out as promising candidates for improving sweetness through molecular breeding, and combining SNP discovery with RNA-seq and metabolomics evidence gives these candidate genes considerably more credibility than genomic data alone ever could.

This work leaves behind a genomic resource that future functional studies can build on, and the markers identified here are natural candidates for experimental validation through Kompetitive Allele-Specific PCR (KASP), CAPS assays, or allele-specific PCR. With that further validation, they could support marker-assisted breeding and genomic selection programmes aimed at dragon fruit cultivars with better sweetness, improved fruit quality, and stronger commercial value.

References

Arivalagan, M., Karunakaran, G., Roy, T. K., Dinsha, M., Sindhu, B. C., Shilpashree, V. M., Satisha, G. C., & Shivashankara, K. S. (2021). Biochemical and nutritional characterization of dragon fruit (Hylocereus species). Food Chemistry, 353, 129426.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), 2633–2635.

Chen, S., Xu, C., Mazhar, M. S., & Naiker, M. (2024). Nutritional value and therapeutic benefits of dragon fruit: A comprehensive review with implications for establishing Australian industry standards. Molecules, 29(23), 5676.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2), 80–92.

Jiang, R., Wu, L., Zeng, J., Shah, K., Zhang, R., Hu, G., Qin, Y., & Zhang, Z. (2023). Identification of HuSWEET Family in Pitaya (Hylocereus undatus) and Key Roles of HuSWEET12a and HuSWEET13d in Sugar Accumulation. International Journal of Molecular Sciences, 24(16), 12882. https://doi.org/10.3390/ijms241612882

Li, H., & Durbin, R. (2009). Fast and accurate short-read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754–1760.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15, 550.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis

Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–1303.

Mondol, M. S. A., Akbar, U., Mandal, O., Rani, A., Dar, A. H., Chatterjee, A., & Abdi, G. (2025). Advances in agronomic practices, postharvest technologies, and medicinal potential of dragon fruit (Hylocereus spp.): A comprehensive updated review. Journal of Agriculture and Food Research, 22, 102157. https://doi.org/10.1016/j.jafr.2025.102157

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559–575.

Tarte, I., Singh, A., Dar, A. H., Sharma, A., Altaf, A., & Sharma, P. (2023). Unfolding the potential of dragon fruit (Hylocereus spp.) for value addition: A comprehensive review. eFood, 4, e76.

Yadav, A., Garg, S., Kumar, S., Alam, B., & Arunachalam, A. (2024). A review on genetic resources, breeding status and strategies of dragon fruit. Genetic Resources and Crop Evolution.

Zhao, Z., Wang, L., Chen, J., Liu, Y., Zhang, X., Xu, C., & Li, H. (2024). Altitudinal variation of dragon fruit metabolite profiles as revealed by UPLC–MS/MS-based widely targeted metabolomics analysis. BMC Plant Biology, 24, 344.

Zheng, J., Meinhardt, L. W., Goenaga, R., Zhang, D., & Yin, Y. (2021). The chromosome-level genome of dragon fruit reveals whole-genome duplication and chromosomal co-localization of betacyanin biosynthetic genes. Horticulture Research, 8, 63. https://doi.org/10.1038/s41438-021-00501-6

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

