agronomy

MDPI

Review

Understanding the Complex Functional Interplay between Glucosinolates and Cyanogenic Glycosides in *Carica papaya*

Insyirah-Hannah Ruhaizat-Ooi [1], Rabiatul-Adawiah Zainal-Abidin [2], Nur Syatila Ab Ghani [1], Nor Afiqah-Aleng [3], Hamidun Bunawan [1], Norfarhan Mohd-Assaad [1,4], Zeti-Azura Mohamed-Hussein [1,4] and Sarahani Harun [1,*]

1 Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
2 Biotechnology and Nanotechnology Research Centre, Malaysian Agricultural Research and Development Institute (MARDI), Serdang 43400, Malaysia
3 Institute of Marine Biotechnology, Universiti Malaysia Terengganu, Bandar Kuala Terengganu 21030, Malaysia
4 Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
* Correspondence: sarahani@ukm.edu.my

Abstract: Glucosinolates (GSLs) and cyanogenic glycosides (CGs) fulfil functions in plant defence and have been reported to be anticancer agents. Generally, GSL-containing plants do not produce CG, and vice versa, CG-containing plants do not synthesise GSLs. However, the production of both GSL and CG compounds was observed in *Carica papaya*. Additionally, several studies found both GSL glucotropaeolin and CG prunasin in papaya leaves. The advancement of genome technologies can be explored to elucidate the gene functions and other molecular discoveries in plants that might relate to GSLs and CGs. This review aims to discuss the complex interplay of the rare events whereby these two compounds (GSL and CG) co-occur in a bifurcation pathway in papaya. To our knowledge, this is the first review that highlights novel GSL and CG genes in papaya. Furthermore, species-specific pathways in papaya are also discussed and comprehensively described. The transcription factors involved in regulating GSL and CG biosynthesis pathways are also discussed, accompanied by relevant bioinformatic approaches that can help discover potential regulatory genes that control the production of prunasin and glucotropaeolin in papaya.

Keywords: glucosinolates; prunasin; glucotropaeolin; cyanogenic compounds; *Carica papaya*

Citation: Ruhaizat-Ooi, I.-H.; Zainal-Abidin, R.-A.; Ab Ghani, N.S.; Afiqah-Aleng, N.; Bunawan, H.; Mohd-Assaad, N.; Mohamed-Hussein, Z.-A.; Harun, S. Understanding the Complex Functional Interplay between Glucosinolates and Cyanogenic Glycosides in *Carica papaya*. *Agronomy* **2022**, *12*, 2508. https://doi.org10.3390/agronomy12102508

Academic Editors: Dezső Csupor and Javad Mottaghipisheh

Received: 26 July 2022
Accepted: 10 September 2022
Published: 14 October 2022

1. Introduction

Plants are sessile organisms that produce up to one million metabolites and compounds [1,2]. These metabolites can be used in various industries such as food, agriculture, medicine, and cosmetics [1–3]. Furthermore, the secondary (or specialised) metabolites contribute to an effective defence against biotic and abiotic stresses [4]. Terpenes, phenolics, and nitrogen-containing compounds (NCCs) are the three main chemical classes of plant secondary metabolites. Some of the compounds have bitter-tasting properties in nature [5]. Glucosinolates (GSLs) and cyanogenic glycosides (CGs) are examples of bitter-tasting NCCs that are known to participate in essential defence mechanisms in plants [6,7]. The co-occurrence of GSL and CG in a single plant species was initially thought to be due to the presence of contaminant extracts, as the secondary metabolites were perceived as mutually exclusive [8]. However, several articles reported the existence of both CG and GSL in *Carica papaya* [8–10].

GSLs have been found in 16 angiosperm plant species, especially in the *Brassicaceae* family, including *Arabidopsis thaliana*. GSLs are responsible for the bitter flavours of *Brassica* vegetables, including turnip (*Brassica rapa* ssp. *rapa*), broccoli (*Brassica oleracea* var. *italica*), and cauliflower (*Brassica oleracea* var. *botrytis*) [11]. Previous studies have also reported the GSL content in other *Cruciferae*, *Cleomaceae*, and *Caricaceae* species [12–16]. Currently, 130 GSL structures have been identified in GSL-containing plants [17]. GSLs are derived

from several amino acids, and depending on their amino acid precursor, GSLs are categorised into three types: aliphatic GSLs from methionine (Met), indolic GSLs from tryptophan (Trp), and benzyl GSLs from phenylalanine (Phe) and tyrosine (Tyr).

These sulfur- and nitrogen-containing compounds and their degradation products are known for their role in plant defence against fungi, bacteria, pests, and insects [12,18]. In the model plant *A. thaliana*, GSLs can be found in various organs, including seeds, roots, stems, and leaves [12,19]. Different GSL-containing plant organs have different GSL compositions, both quantitatively and qualitatively. Furthermore, the concentrations of GSL in roots are often higher than in shoots [20]. In these GSL-containing plants, a myrosinase enzyme known as β-thioglucosidase in different plant organelles will convert the inactive form of GSL into bioactive GSL hydrolytic products, such as isothiocyanates, nitriles, thiocyanates, and epithionitriles [21,22]. These bioactive compounds are produced upon pest infestation or physical processes, leading to tissue disruption [23].

Multiple studies have shown the therapeutic potential of GSL-derived compounds in various cancer treatments. For example, a bioactive compound, namely Indole-3-carbinol (I3C) produced from indolic GSL, is commercially used as a supplement for recuperating breast cancer patients [24]. Other studies have also provided supporting evidence on the capability of I3C to inhibit tumour invasion, metastasis, and cell-cycle progression in breast [25] and prostate [26] cancer cells. A couple of years later, I3C was shown to induce apoptosis in colorectal [27] and osteosarcoma [28] cell lines. In another example, sulforaphane, a bioactive compound derived from aliphatic GSL, has shown a potential to inhibit the carcinogenic cells of various malignant cancers, such as breast [29], prostate [30], liver, lung [31], and pancreas [32] cancers in rodents. Another type of bioactive GSL is benzyl isothiocyanate (BITC), which is produced from benzyl GSL or glucotropaeolin. This active compound is capable of suppressing cancer progression, including colon [33], breast [34–36], and pancreatic [37] tumours. Interestingly, BITC was suggested as a therapeutic agent to treat bacterial infection, potentially inhibiting the toxin production and growth of enterohemorrhagic *Escherichia coli* [38].

Cyanogenic glycosides (CGs) are another example of NCCs, known as glycosides of α-hydroxynitriles. CGs and GSLs have similar characteristics. Firstly, CGs are also produced from amino acids, including L-valine, L-isoleucine, L-leucine, L-phenylalanine, or L-tyrosine [39]. Secondly, CGs do not cause toxic effects in their intact form but will produce a bioactive compound, such as toxic cyanide (e.g., hydrogen cyanide) upon tissue damage due to the activity of a digestion enzyme [35], resulting in symptoms associated with cyanide poisoning when consumed in high amounts [40]. Ingestion of hydrogen cyanide can lead to intoxication and possible death if consumed at high concentrations. This mechanism protects plants from potential threats (e.g., insects and animals) and ensures plant survival [41]. However, while GSLs are found explicitly in *Brassica* and other related families, CGs are one of the most extensive classes of secondary plant metabolites found in more than 2600 plant species, with 112 CGs characterised [42]. The general chemical structures of GSL and CG are shown in Figure 1.

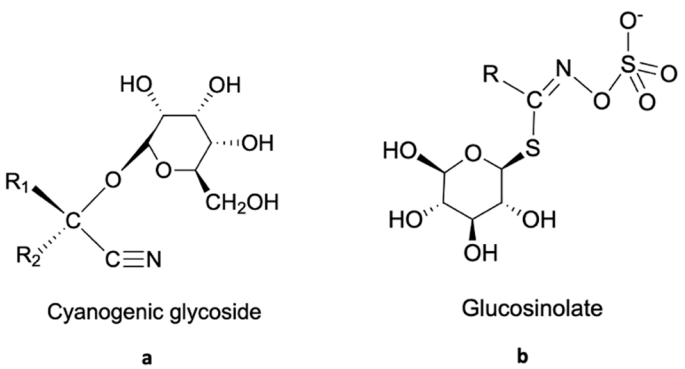


Figure 1. The general structure of (**a**) Cyanogenic glycoside and (**b**) Glucosinolate.

The occurrence of CGs is more widespread across the Plantae region than GSLs [43]. Several essential crops contain CGs, such as cassava (*Manihot esculenta*), sorghum (*Sorghum bicolor*), and barley (*Hordeum vulgare*) [41]. To date, dhurrin biosynthesis in sorghum is the most comprehensively studied tyrosine-derivative, also through the cloning and characterisation of the genes being involved [41]. Amygdalin and linamarin are other CGs primarily identified in sorghum and cassava plants [44,45]. Linamarin is a cyanogenic chemical present in the roots and leaves of cassava plants [46]. Thus, eating fresh cassava can produce cyanide poisoning in humans due to the degradation of linamarin by ß-glucosidase enzymes [47].

Amygdalin has been extensively studied for its potential application in human-targeted cancer treatment [48]. However, oral amygdalin treatment causes cyanide poisoning problems, such as decreased consciousness in pancreatic cancer patients [49]. While amygdalin can occasionally trigger cyanide poisoning, in vitro and in vivo evidence also suggests some therapeutic benefits in cancer treatment [50]. Furthermore, such cyanide toxicity can be utilised in innovative cancer-target therapy [51]. Targeted therapy objectively attempts to deliver drugs to particular tumour areas while limiting the side effects on healthy tissues [52].

Identifying genes involved in the GSL and CG biosynthesis pathways could provide further insights for the benefit of various applications, including genetic engineering, to manipulate compounds for relevant industry purposes, such as medical and agricultural uses [41]. For instance, CG compounds could be introduced as biopesticides to control diseases caused by the invasion of pests [8]. In addition, lowering the CG content in cassava would improve the food safety level for human consumption [53]. Similarly, molecular studies in GSL biosynthesis aimed to increase the synthesis of beneficial GSLs such as glucoraphanin [13,54] and reduce other GSLs that contribute to poor taste in crucifers [55]. Additionally, identifying genes would aid future research for cancer treatment [41]. A previous study found two essential enzymes, cytochrome P450 and UDP-glucosyl transferase, that have become vital factors to synthesise both compounds [41,56]. Additionally, some regulators and genes associated with the cytochrome P450 (CYP) enzymes have also been identified [56–58]. Furthermore, the CYP79 enzymes have been described in catalysing the rate-limiting steps in CG and GSL biosynthesis [58]. Hence, in this review, we highlight the co-occurrence of GSL and CG in papaya. We then construct the biosynthesis pathways of GSL and CG in papaya using bioinformatics to explore the upstream intermediates involved in their synthesis. Such information is valuable for basic plant science and genetic crop improvement and beneficial for food and agricultural industries and for medicine and cosmetics.

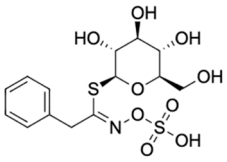
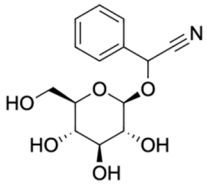
2. The Identification of Glucosinolate (GSL) and Cyanogenic Glycoside (CG) in Papaya

Carica papaya, generally known as papaya, is a significant tropical fruit consumed worldwide [59]. The essential nutrients produced by papaya facilitate the initial papaya development to secure several aspects, including the growth of various parts of the plant, including its foliage, trunks, and roots, leading to higher papaya productivity [60,61]. This tropical plant can be used as an alternative medicine since the leaves, fruits, stems, seeds, and roots could be used for an alternative medical treatment of various disorders, including cancers, ulcers, and gastritis [62]. Furthermore, secondary metabolites that provide essential nutrients for human health are abundant in papaya. Previous metabolomics studies have discovered carotenoids and tocopherols in the papaya seeds with antioxidant activity [63,64]. Papaya is a diploid plant, and the reported genome size is 372 Mb [65,66]. While GSL and CG biosynthesis pathways have been studied in cassava, sorghum, bitter almonds, and *Brassicaceae* vegetables, to the best of our knowledge, the biosynthesis pathways of both GSL and CG in papaya have not been described comprehensively in any accessible publication. However, Harun et al. [7] extensively reviewed a comprehensive inventory of GSL biosynthetic genes in *A. thaliana*, whereby the homologs in the *Brassicaceae* can be retrieved from SuCComBase [67]. Therefore, by using the reference genes from known GSL and CG

plants, the homologous genes could be identified in papaya using bioinformatic approaches. This information helps to construct the respective GSL and CG pathways in papaya.

Specifically, this crop has been shown to produce prunasin (CG) and benzyl GSL, also known as glucotropaeolin (GSL) [68]. Hence, it is suggested that the *Carica* species is unique, as these two compounds co-occur, synthesised in a bifurcation pathway [69]. Table 1 shows the chemical structure of glucotropaeolin and prunasin, which can also be found in *A. thaliana* and *Brassica oleracea*. However, the concentration level of this compound in different plant species can be variable. A study conducted in *Brassica oleracea* found a significant increment of glucotropaeolin concentrations in the organic plants compared to the conventional breeding approach [70]. For instance, the glucotropaeolin content in *A. thaliana* was only reported in transgenic *Arabidopsis* rosette leaves expressing *CYP79A2* under the control of the CaMV35S promoter. In contrast, the rosette leaves of wild-type plants did not contain detectable amounts of this type of GSL [71].

Table 1. Glucotropaeolin and prunasin in *Brassica* and papaya.

GSL Compound	Structural Formula	Plant Species
Glucotropaeolin		*Arabidopsis thaliana* [7] *Brassica oleracea* [70] *Carica papaya* [10]
Prunasin		

According to chromatograms, the initial report on the identification of CG and GSL was conducted on the dried papaya leaves, in which the CG was deficient [72]. A couple of years later, Bennet et al. [10] measured cyanide concentration as a proxy to estimate CG in the papaya plant. They reported variations in the distribution pattern of cyanide in papaya organs, in which the concentrations were highest in the young leaves and the tap root, with declining concentration as the leaves age. Similarly, the highest concentration of glucotropaeolin was recorded in the youngest papaya leaves. A similar compound was observed in other plant parts, such as the roots, leaf stalks, and stem internodes. [10]. In another study, a degradation product of glucotropaeolin, benzyl isothiocyanate (BITC), was identified in papaya. BITC was first found in the seeds of papaya [73] and later in its pulp [74,75]. BITC is a bioactive compound with medicinal and pharmacological properties [8]. Furthermore, several studies reported the capability of BITC to suppress mammalian carcinogenic cells [76].

Phenylalanine, a bitter amino acid [71], is the crucial precursor for papaya CG and benzyl GSL biosynthesis pathways. A recent study compared the correlation between phenylalanine, GSL, BITC, and CG and the bitter taste at different temperatures in papaya. The bitterness intensity is maximum in unripe fruit and gradually decreases as it matures. Furthermore, the bitterness intensity in cool-season fruits is more significant than in warm-season fruits. In their study, Jioe et al. [9] corroborated the previous findings where phenylalanine served as CG and GSL precursor. Based on their calculated correlation values, they also suggested that GSL was not the only component that generated a bitter taste in immature papaya fruits [9].

3. The Construction of Glucosinolate (GSL) Biosynthesis Pathway in Papaya

The GSL biosynthesis pathway has been elucidated by identifying various biological factors involved, such as regulators [77–82], enzymes [83–85], and protein transporters [86–89] that also seem to participate in the cross-talk with other essential metabolic processes such as phenylpropanoids, sulfur, and nitrogen in *A. thaliana* [90,91]. A comprehensive set of 113 known GSL genes were identified from the literature and using public pathway databases, encoding for transcriptional regulators, enzymes, and protein transporters [7]. Generally, GSL biosynthesis comprises several groups of genes initiated by the transcription factors that regulate the production of various secondary metabolites and GSL derivatives. For instance, six MYB genes originating from the R2R3-MYB transcription factor family have been described as regulating the production of GSL in *A. thaliana*. MYB34, MYB51, and MYB122 control the production of indolic GSLs, whereas MYB28, MYB29, and MYB76 control the production of aliphatic GSLs [92,93]. However, the transcription factor that controls benzyl GSL biosynthesis in *A. thaliana* is still unclear since the production of glucotropaeolin can only be detected in the engineered lines of *Arabidopsis* [71]. The ultimate step in GSL biosynthesis is the core structure synthesis, in which most of the biosynthetic genes involved in the indolic and benzyl GSLs are similar. In indolic GSL biosynthesis, CYP79B3 (tryptophan N-monooxygenase 2) catalyses the derivation of tryptophan, whereas CYP79A2 (phenylalanine N-monooxygenase) prepares the phenylalanine substrate in benzyl GSL (glucotropaeolin) biosynthesis [71].

Then, CYP83B1 (CYP83B1 monooxygenase) converts both tryptophan-derived and phenylalanine-derived acetaldoximes into aci-nitro compounds [94]. The remaining steps of GSL core structure formation involve several GSL biosynthetic enzymes that accommodate all GSL precursors regardless of their associated side chains. In this step, the S-alkylthiohydroximates are converted to thiohydroximic acids in a reaction catalysed by SUR1 (C–S lyase) [95]. In the glucosylation process, UGT74B1 was suggested to metabolise thiohydroximates based on the enzyme's in vitro and in vivo analysis [96]. The final step in the GSL core structure synthesis is the sulfation process of the desulfoglucosinolates to form intact glucosinolates involving the cytosolic sulfotransferase group (ST5a), such as SOT16, SOT17, and SOT18. The biochemical characterisation of sulfotransferases suggests the role of SOT16 to metabolise phenylalanine- and tryptophan-derived desulfoglucosinolates, such as glucobrassicin (3-indolylmethyl GSL) [97,98].

To identify GSL genes and reconstruct the biosynthesis pathway in papaya, we used the available information on the known gene-encoded enzymes in *A. thaliana* phenylalanine-derived GSL core structure biosynthesis, as it is an important precursor for pathways leading to both benzyl GSL and CG. The gene list was set as queries to identify the homologous genes in the papaya genome. Table 2 shows the identified homologous benzyl GSL genes using protein sequences searched for with the BLASTp program via the NCBI database (accessed on 16 February 2022).

Based on Table 2, the homologous GSL genes identified in papaya showed more than 40% sequence identity with an e-value ranging from 7.00×10^{-76} to 0.00 with the query sequence of GSL genes from *Arabidopsis*. The identified novel GSL biosynthetic genes were then used to construct the glucotropaeolin pathway generated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.kegg.jp/) (accessed on 18 February 2022). The primary biosynthetic genes encoding enzymes in GSL core structure synthesis are shown in Figure 2. However, CYP79A1 was identified in the BLASTp analysis instead of CYP79A2, which catalyses the phenylalanine substrate in the production of glucotropaeolin [71].

To elucidate the potential of regulatory mechanisms involved in GSL formation in papaya, bioinformatic analysis of *MYB* transcription factor genes was conducted with reference to the six known *MYB* genes from *A. thaliana*. In addition, the BLASTp (accessed on 22 February 2022) analysis of *Arabidopsis* against the genome sequences of papaya and its closely related species showed the existence of several *MYB* orthologous genes in papaya.

Table 2. Novel benzyl GSL biosynthetic genes identified in papaya using BLASTp.

Query Gene	Description	Gene	E-Value	Per. Identity	Accession
CYP79A2	tyrosine N-monooxygenase-like	*CYP79A1* *	0.00	55.78	XP_021889957.1
	tyrosine N-monooxygenase 2-like	*CYP79B3*	0.00	57.99	XP_021887085.1
	tryptophan N-monooxygenase 1-like	*CYP79B3*	0.00	54.13	XP_021887088.1
CYP833B1	cytochrome P450 83B1	*CYP83B1*	0.00	69.78	XP_021902795.1
	cytochrome P450 71B35-like	*CYP71B35*	4.00×10^{-135}	43.87	XP_021901245.1
	LOW-QUALITY PROTEIN: cytochrome P450 71B35-like	*CYP71B35*	1.00×10^{-133}	43.94	XP_021901202.1
SUR1	S-alkyl-thiohydroximate lyase SUR1	*SUR1*	0.00	64.09	XP_021900833.1
	tyrosine aminotransferase-like	*TAT*	0.00	62.79	XP_021900789.1
	probable aminotransferase TAT2	*At5g53970*	0.00	52.94	XP_021911536.1
UGT74B1	UDP-glycosyltransferase 74B1-like	*UGT74B1*	1.00×10^{-179}	56.89	XP_021908832.1
	UDP-glycosyltransferase 74B1-like	*UGT74B1*	7.00×10^{-162}	53.26	XP_021908582.1
	UDP-glycosyltransferase 74E2-like	*UGT74E2*	2.00×10^{-137}	48.28	XP_021894659.1
ST5a	LOW-QUALITY PROTEIN: cytosolic sulfotransferase 16-like	*SOT16*	1.00×10^{-148}	66.56	XP_021909632.1
	LOW-QUALITY PROTEIN: cytosolic sulfotransferase 15-like	*SOT15*	7.00×10^{-76}	41.99	XP_021912467.1

* *CYP79A2* should be the GSL biosynthetic gene encoding enzyme; however, the gene cannot be found in papaya.

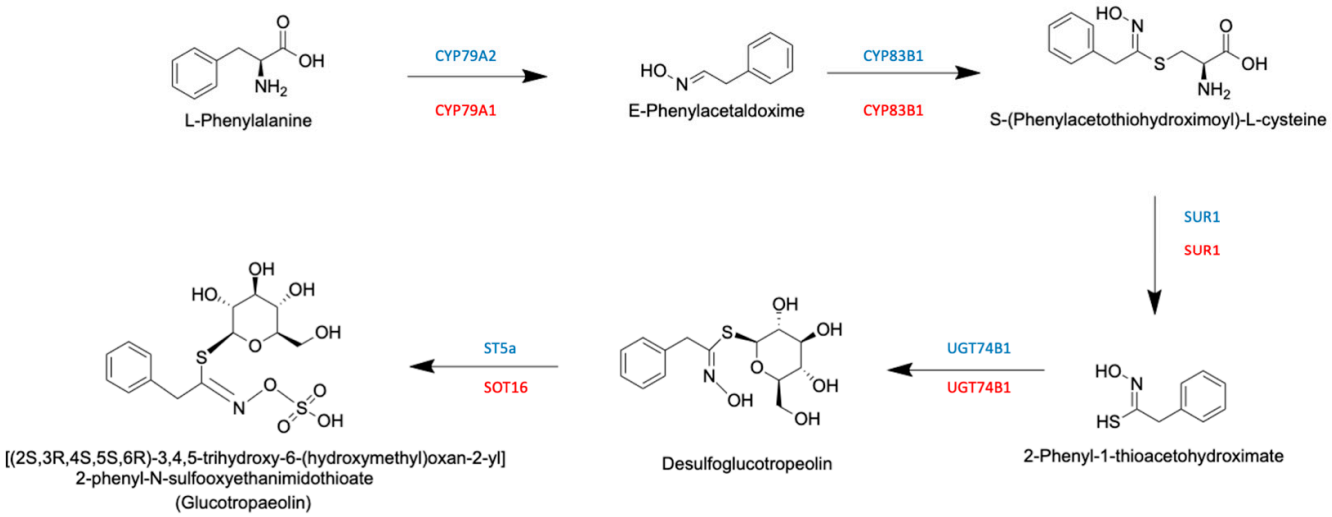


Figure 2. Proposed glucotropaeolin pathway in papaya. The fonts in blue represent known glucotropaeolin biosynthetic genes encoding enzymes in *A. thaliana* [7], and those in red represent the corresponding papaya enzymes identified using BLASTp.

To determine the evolutionary relationship among *MYB* genes in *A. thaliana*, *C. papaya*, and other closely related species, a phylogenetic tree was constructed for the 26 selected *MYB* genes in MEGA version 11.0 [99] using the Maximum-Likelihood method with 1000 bootstrap replicates (accessed on 22 February 2022). In addition, the conserved motifs of *MYB genes* from *A. thaliana*, *B. rapa*, *B. oleracea*, and *C. papaya* were also analysed using Multiple Expectation Maximisation for Motif Elicitation (MEME) version 5.4 (https://meme-suite.org/meme/tools/meme, accessed on 24 February 2022) [100] according to the following parameters: site distribution was set to zero or one occurrence (zoops), the maximum number of motifs for searching was set to 10, and the motif width was set between 6 and 50.

The phylogenetic analysis suggested that these *MYB* genes could be divided into three major clades: indolic GSL, aliphatic GSL, and MYB-like family (Figure 3). Motif identification revealed five conserved motifs (Motifs 1, 2, 3, 4, and 10) present in all *MYB* genes, highlighting the conservation of MYB motifs in these species (Figure 4). Motif 1 and Motif 2 were both 50 bp in length, containing the DNA-binding domains of Myb proteins

and the SANT domain family specifically involved in the transcriptional regulation [101]. The indolic GSL clade contained the evm.model.supercontig 3.239 gene sequence from *C. papaya* and several MYB orthologs from *Brassicaceae*, and the group can be represented by Motif 5, which belonged to the Myb-like DNA binding domain (Figure 4). Interestingly, the evm.model.supercontig 3.239, which has been annotated as Myb domain protein 122, also contained Motif 7 (Myb-like DNA-binding domain) specifically found in the aliphatic GSL clade. Based on this finding, papaya is speculated to possibly regulate the aliphatic GSL biosynthesis as well; further investigation is needed.

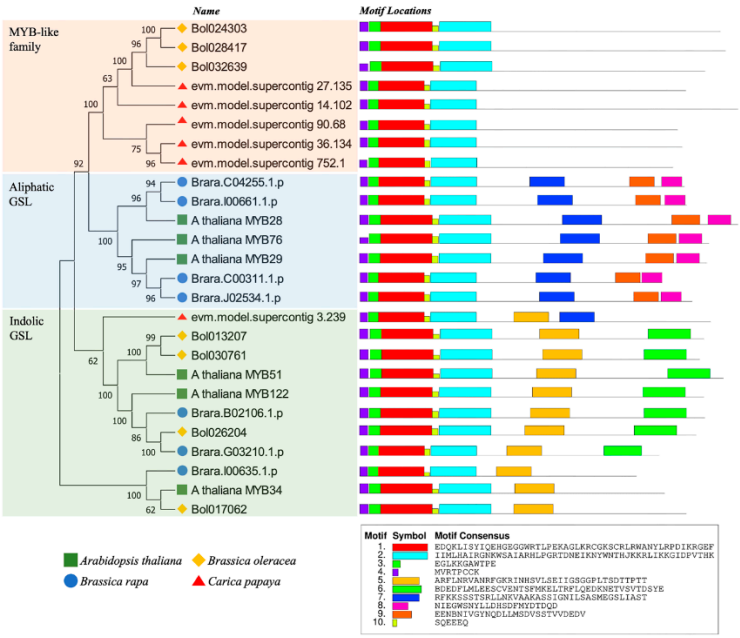


Figure 3. Phylogenetic relationship of MYB transcription factor genes and their corresponding motif structures in *Brassicaceae* and *C. papaya* using full-length protein sequences of 26 identified *MYB genes*.

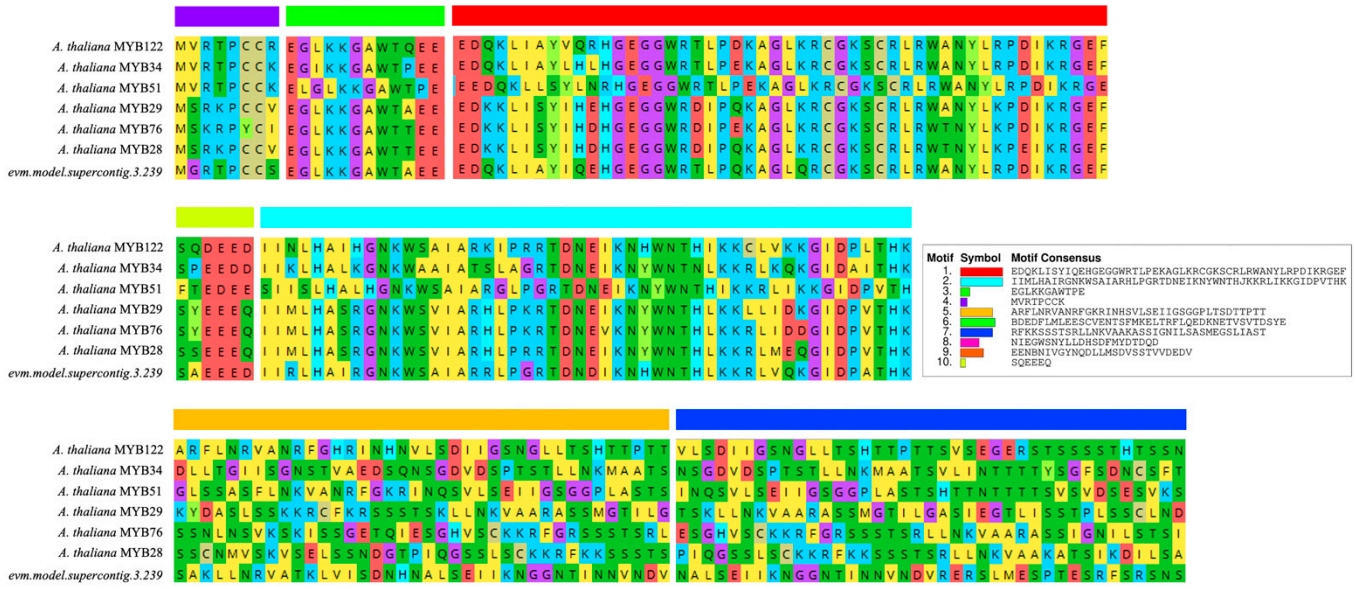


Figure 4. Motif structure of evm.model.supercontig 3.239 compared to *A. thaliana MYB* genes based on multiple sequence alignment, highlighting the presence of a unique MYB-like family motif (Motif 5) and unique aliphatic GSL motif (Motif 7) in the sequence. evm.model.supercontig 3.239 has been annotated as a probable *MYB* transcription factor-related gene using Phytozome 13 and SuCComBase databases.

Meanwhile, the MYB-like family contains MYB orthologs from both *Brassica* species and papaya but lacks representation of specific (aliphatic or indolic) GSL motifs. This could suggest that the eight *MYB* genes in the MYB-like family may undergo rapid functional divergence and do not carry any specific function in the GSL biosynthesis.

4. The Construction of Cyanogenic Glycoside (CG) Biosynthesis Pathway in Papaya

In general, the biosynthesis pathway of CG can be described in three processes. In the initial step, an enzyme from the cytochrome P450 family will convert the α-amino acids via N-hydroxylation to an N-hydroxylamino acid that will eventually be converted to aldoxime. Next is the conversion of the aldoxime to cyanohydrin as catalysed by another member of the cytochrome P450 family. Lastly, the cyanohydrin molecules are glycosylated by a soluble enzyme, UDP-glucosyltransferase a [39,41]. The detailed analysis of the biosynthesis pathway indicated that it has evolved from the pre-availability of CGs, compounds being widely distributed among plant kingdoms. The independent evolution of GSL from an ancient CG is suggested to occur through the evolution of ancestral CYP enzymes capable of metabolising reactive oximes [102].

Prunasin, one of the major CGs, is commonly found in stone fruits such as apricot (*Prunus armeniaca*) [103], peach (*Prunus persica*) [104], and bitter almond (*Prunus dulcis*) [105]. CGs such as amygdalin and prunasin contribute to the bitterness of almonds. The prunasin biosynthesis in *P. dulcis* involves three enzymatic reactions catalysed by three gene-encoded proteins: PdCYP79D16, Pd71AN24, and PdUGT94AF3 [57]. Here, we used the established CG genes from *P. dulcis* as queries to identify the homologous prunasin biosynthetic genes in papaya using BLASTp (accessed on 25 February 2022) (Table 3).

Table 3. Novel biosynthetic CG genes identified in papaya using BLASTP.

Query Gene	Description	Gene	E-Value	Per. Identity	Accession
CYP79D16	tryptophan N-monooxygenase 2-like	LOC110806501	0.00	57.7	XP_021887085.1
	tryptophan N-monooxygenase 1-like	LOC110806504	0.00	58.13	XP_021887088.1
	tyrosine N-monooxygenase-like	LOC110808696	0.00	56.82	XP_021889957.1
CYP71AN24	cytochrome P450 71A1-like	LOC110810705	5.00×10^{-162}	52.13	XP_021892654.1
	cytochrome P450 71B35-like	LOC110817138	5.00×10^{-136}	46.02	XP_021901245.1
	cytochrome P450 71D10-like	LOC110808646	7.00×10^{-132}	45.3	XP_021889891.1
UGT94AF3	beta-D-glucosyl crocetin beta-1,6-glucosyltransferase-like	LOC110819221	1.00×10^{-161}	55.11	XP_021904037.1
	putative UDP-rhamnose: rhamnosyltransferase 1	LOC110816080	6.00×10^{-76}	35.38	XP_021899800.1
	anthocyanidin 3-O-glucosyltransferase-like	LOC110821584	5.00×10^{-64}	32.02	XP_021907153.1

CYP79D16, *CYP71AN24* and *UGT94AF3* should be the CG biosynthetic gene encoding enzymes; however, the gene cannot be found in papaya.

Table 3 shows the identified homologous biosynthetic CG genes in papaya that are likely involved in the prunasin biosynthesis. The novel CG genes in papaya share more than 30% sequence identity and e-value ranging from 5.00×10^{-64} to 0.00 with the reference CG genes identified in *P. dulcis*. The three identified CG genes were mapped onto the prunasin biosynthesis pathway generated from the KEGG database (Figure 5) (accessed on 25 February 2022).

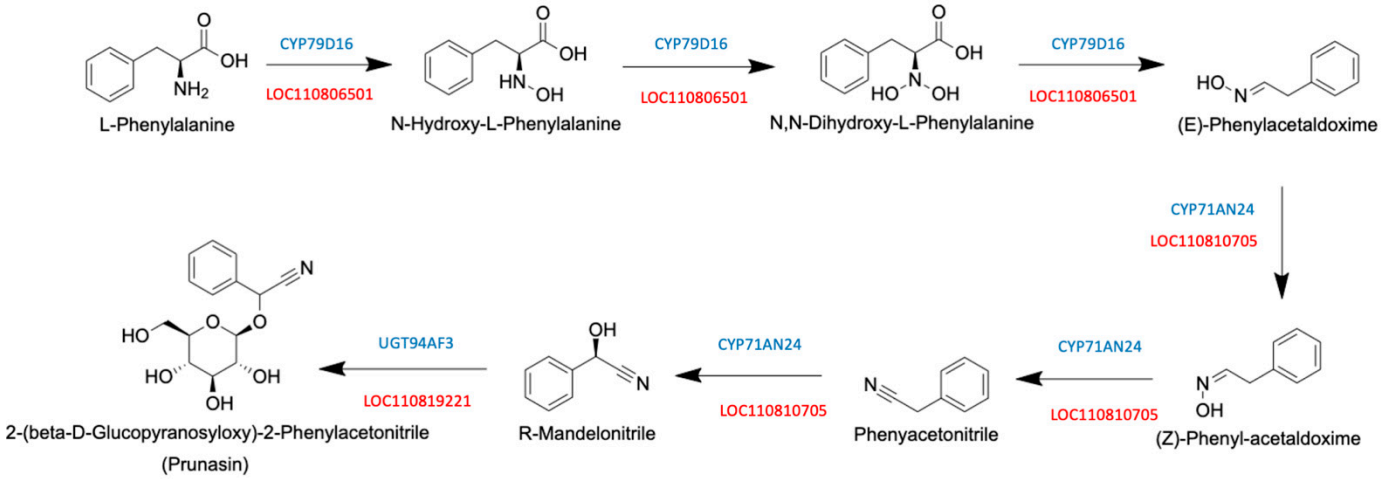


Figure 5. Proposed prunasin pathway in papaya. The fonts in blue represent known prunasin biosynthetic genes encoding enzymes in *P. dulcis* [105], and those in red represent the corresponding papaya enzymes identified using BLASTp.

The basic helix-loop-helix (bHLH) transcription factors are essential in regulating CG biosynthesis. The mutation of bHLH transcription factor gene clusters located at the Sweet kernel (Sk) locus of the *P. dulcis* disrupted the regulation of CG biosynthesis, thus leading to a reduced CG content in the plant and resulting in the sweet kernel trait.

To elucidate the potential involvement of prunasin biosynthesis in papaya and the possible evolutionary relationship of *bHLH* genes between Prunus species and papaya, the bioinformatics analysis of bHLH transcription factor genes was conducted in reference to five *bHLH* genes from *P. dulcis* [105]. The BLASTp analysis (accessed on 28 February 2022) of *P. dulcis bHLH* genes against the genome sequences of *P. persica* and papaya revealed two and seventeen orthologous *bHLH* genes in papaya and *P. persica*, respectively. The evm.model.supercontig_141.19 and evm.model.supercontig_1892.1 genes in papaya share 41–50% sequence identity and e-value, ranging from 1.8×10^{-128} to 3.6×10^{-80} with *bHLH1*, *bHLH2* and *bHLH4* from *P. dulcis*.

A phylogenetic tree was constructed for the five bHLH sequences from *P. dulcis* and ten non-redundant orthologous sequences from BLASTp hits in MEGA version 11.0 [99] using the Maximum-Likelihood method with 1000 bootstrap replicates (accessed on 28 February 2022). The protein sequences were previously aligned using ClustalW. Furthermore, the conserved motifs of *bHLH* genes from *P. dulcis*, *P. persica*, and papaya, were also analysed using Multiple Expectation Maximisation for Motif Elicitation (MEME) version 5.4 (accessed on 1 March 2022) according to the above-mentioned parameters (refer to Section 3). In addition, the conserved motifs for these genes were determined using the ScanProsite interface in the PROSITE web server [106].

The phylogenetic analysis suggested that these genes may be divided into three clades: *bHLH1-bHLH2* group, *bHLH4* group, and *MYC*-like family (Figure 6). The evm.model.supercontig_141.19 and evm.model.supercontig_1892.1 genes in papaya belong to the Myc-like family cluster with the absence of several motifs dominated by *bHLH* genes from *P. dulcis*. These *bHLHs* probably got lost during the process of evolution.

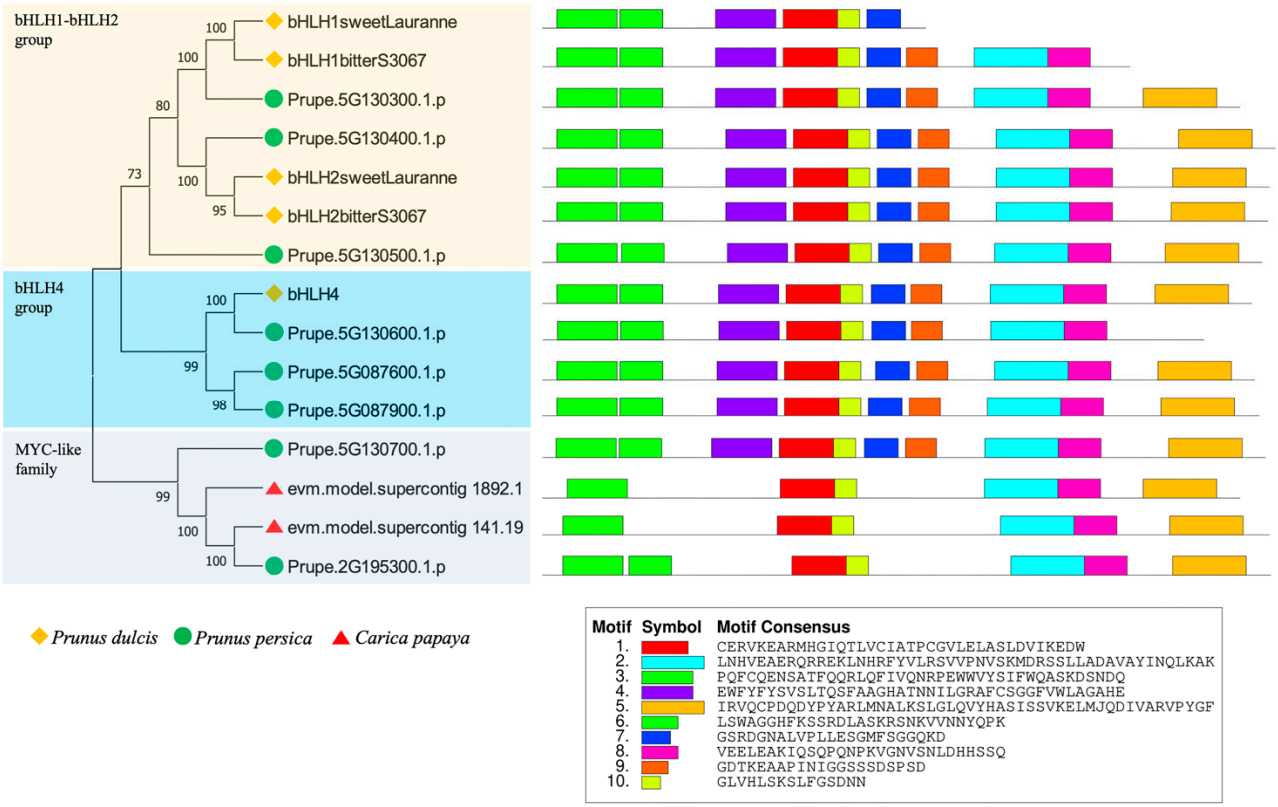


Figure 6. Phylogenetic relationship of *bHLH* genes encoding transcription factors and their corresponding motif structures in *P. dulcis*, *P. persica*, and papaya using full-length protein sequences of 15 identified *bHLH* genes.

Motif identification revealed the presence of Myc-type, basic helix-loop-helix (bHLH) domain profile (Prosite ID: PS50888) in all sequences except the bHLH1 from sweet almond variant (bHLH1sweetLauranne), which also reflected the presence of Motif 2 that contained 50 amino acids in MEME motif profiles (Figure 7, coloured in cyan). The absence of the bHLH domain in the bHLH1sweetLauranne gene sequence is in accordance with the point mutation occurring in the gene that prevents the transcription of P450 genes involved in the amygdalin biosynthesis pathway [105]. Thus, the presence of the bHLH domain in evm.model.supercontig_141.19 and evm.model.supercontig_1892.1 from papaya could potentially indicate the presence of a CG biosynthesis pathway. Interestingly, the evm.model.supercontig_141.19 also contained an additional domain, the ACT domain (Prosite ID: PS51671), which has not been detected in other sequences. The domain is predicted to be a regulatory domain for small ligand binding such as amino acids and is often involved in protein dimerisation [107], but the role of the ACT domain in papaya remains elusive.

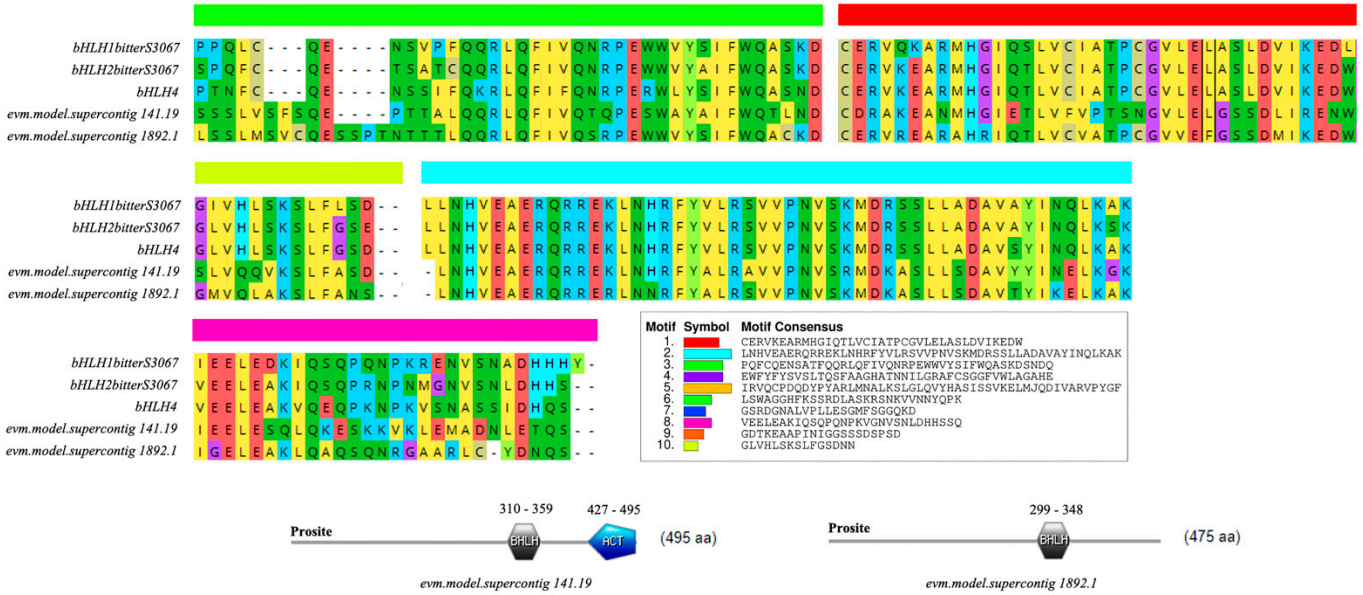


Figure 7. Motif structure of evm.model.supercontig_141.19 and evm.model.supercontig_1892.1 in papaya and *bHLH* genes in bitter almond (*P. dulcis*). The multiple sequence alignment of these gene sequences highlights the presence of the bHLH domain (Motif 2, coloured in cyan) and several other motifs in the sequences.

Additionally, the MEME motif profiles (Figures 6 and 7) also showed that these sequences exhibited three highly conserved motifs, Motif 1 (37 amino acids, red), Motif 3 (41 amino acids, green), and Motif 10 (15 amino acids, yellow); thus proposing the importance of these sequence fragments for the CG biosynthesis pathway, thus causing their preservation throughout the evolutionary process. Furthermore, further characterisation of Motif 3 in MEME predicted that the motif could represent a nitrilase/cyanide hydratase signature (ProSite ID: PS00921). The sequence signature is commonly found in enzymes associated with the degradation of nitriles and cyanides.

5. Future Perspectives and Concluding Remarks

Although GSL and CG have been extensively described in *Brassicaceae* and almonds, respectively, the molecular details of these compounds in papaya are still lacking. This review exemplifies the functional interplay of GSL and CG in papaya using the genomics data and bioinformatics approach. In our study, comparative genomics analysis and construction of the GSL and CG biosynthesis pathways have revealed several candidate genes and transcription factors (TFs) potentially involved in the GSL and CG biosynthesis pathways. Results from this study provide new insights into the biological process for the candidate genes and TFs that could be used to enhance the quality and quantity of papaya yields in agriculture and medication selection.

A continuous improvement in omics and computational technologies holds an essential key to unlocking valuable molecular information on GSL and CG in papaya, which could ultimately be used for treating crop diseases. This approach has successfully identified the MYB TFs responsible for GSL production in *B. oleracea* [108].

The next-generation sequencing (NGS) technology permits the researchers to design large-scale transcriptomics experiments to capture and enumerate the transcripts representing the GSL and CG. In addition, the gene co-expression networks would facilitate identifying potential key genes [109] that contribute to the molecular mechanism of GSL and CG biosynthesis in papaya. Previous efforts have generated the gene co-expression network of CG and performed the qRT-PCR analysis to investigate the regulatory mechanism of hydrogen cyanide (HCN) synthesis, which could provide a molecular basis for breeding new cultivars with low HCN content in common vetch [110].

The dynamic interplay of signalling and metabolic pathways governing GSL and CG in papaya tissues (i.e., peel, flesh, leaf) could also be unravelled via integrative analysis of transcriptomics, proteomics, and metabolomics datasets. For instance, CG compound (i.e., linamarin, lotaustralin) has been identified in flax seed by integrating genomics, transcriptomics, metabolomics and bioinformatics approaches [111]. In another study by Zhang et al. [112], metabolomics, qRT-PCR, and comparative genomics analysis were used to obtain insights into GSL profiles and accumulation patterns in a medicinal plant, *Isatis indigotica* Fort [112]. The findings laid a foundation to study further the accumulation and regulation of GSLs in medicinal plants. Integrating the multi-omics approach allows a detailed picture of interactions between two interconnected pathways and enables us to build predictive models on how different molecules interact to respond to various stresses, especially how defence mechanisms are triggered and activated.

Furthermore, a central knowledge-based repository is needed to ensure other researchers can use the molecular information on GSL and CG. Data curation of GSL and CG from multiple sources, including current literature, is essential to ensure comprehensive data is provided to serve the community scientifically. However, such a platform requires continuous effort to remain relevant to the community, particularly in terms of regular system maintenance and data updates.

We have discussed limitations and suggestions to conduct integrated research for studying GSL and CG in future work. Various studies have demonstrated using GSL and CG as plant food resources but lack their involvement as a bioresource in plant defence mechanisms. Hence, studying these two essential biosynthesis pathways will enhance GSL and CG's role, ultimately providing valuable biological resources for plant defence systems.

Author Contributions: Conceptualisation, R.-A.Z.-A. and S.H.; formal analysis, I.-H.R.-O., N.S.A.G. and S.H.; data curation, I.-H.R.-O., N.S.A.G. and S.H.; writing—original draft preparation, I.-H.R.-O. and S.H.; writing—review and editing, S.H., R.-A.Z.-A., N.S.A.G., N.A.-A., N.M.-A., H.B. and Z.-A.M.-H.; visualisation, I.-H.R.-O., N.S.A.G. and S.H.; supervision, S.H., R.-A.Z.-A. and Z.-A.M.-H.; funding acquisition, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universiti Kebangsaan Malaysia, Geran Galakan Penyelidik Muda (GGPM-2019-043) research grant awarded to Sarahani Harun. The APC was funded by Universiti Kebangsaan Malaysia (GP-2021-K021204 and GGPM-2019-042).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the Centre for Bioinformatics Research (CBR), Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, for the computational facilities.

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

References

1. Afendi, F.M.; Okada, T.; Yamazaki, M.; Hirai-Morita, A.; Nakamura, Y.; Nakamura, K.; Ikeda, S.; Takahashi, H.; Altaf-Ul-Amin, M.; Darusman, L.K.; et al. KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* **2012**, *53*, e1. [CrossRef] [PubMed]
2. Saito, K. Editorial: The origin of plant chemodiversity—Conceptual and empirical insights. *Plant Cell Physiol.* **2020**, *11*, 890. [CrossRef]
3. Rai, A.; Saito, K.; Yamazaki, M. Integrated omics analysis of specialised metabolism in medicinal plants. *Plant J.* **2017**, *90*, 764–787. [CrossRef] [PubMed]
4. Wang, S.; Alseekh, S.; Fernie, A.R.; Luo, J. The structure and function of major plant metabolite modifications. *Mol. Plant* **2019**, *12*, 899–919. [CrossRef] [PubMed]
5. Izawa, K.; Amino, Y.; Kohmura, M.; Ueda, Y.; Kuroda, M. Human-environment interactions—Taste. *Compr. Nat. Prod. II Chem. Biol.* **2010**, *4*, 631–671. [CrossRef]
6. Taiz, L.; Zeiger, E. *Plant Physiology*, 5th ed.; Sinauer Associates, Inc.: Sunderland, MA, USA, 2010.
7. Harun, S.; Abdullah-Zawawi, M.R.; Goh, H.H.; Mohamed-Hussein, Z.A. A comprehensive gene inventory for glucosinolate biosynthetic pathway in *Arabidopsis thaliana*. *J. Agric. Food Chem.* **2020**, *68*, 7281–7297. [CrossRef]

8. Williams, D.J.; Pun, S.; Chaliha, M.; Scheelings, P.; O'Hare, T. An unusual combination in papaya (*Carica papaya*): The good (glucosinolates) and the bad (cyanogenic glycosides). *J. Food Compos. Anal.* **2013**, *29*, 82–86. [CrossRef]
9. Jioe, I.P.J.; Lin, H.-L.; Shiesh, C.-C. The investigation of phenylalanine, glucosinolate, benzylisothiocyanate (BITC) and cyanogenic glucoside of papaya fruits (*Carica papaya* L. cv. 'Tainung No. 2') under different development stages between seasons and their correlation with bitter taste. *Horticulturae* **2022**, *8*, 198. [CrossRef]
10. Bennet, R.N.; Kiddle, G.; Wallsgrove, R.M. Biosynthesis of benzylglucosinolates, cyanogenic glucosides and phenylpropanoids in *Carica papaya*. *Phytochemistry* **1997**, *45*, 59–66. [CrossRef]
11. Padilla, G.; Cartea, M.E.; Velasco, P.; de Haro, A.; Ordas, A. Variation of glucosinolates in vegetable crops of *Brassica rapa*. *Phytochemistry* **2007**, *68*, 536–545. [CrossRef] [PubMed]
12. Liu, Z.; Wang, H.; Xie, J.; Lv, J.; Zhang, G.; Hu, L.; Luo, S.; Li, L.; Yu, J. The roles of cruciferae glucosinolates in disease and pest resistance. *Plants* **2021**, *10*, 1097. [CrossRef] [PubMed]
13. Ishida, M.; Hara, M.; Fukino, N.; Kakizaki, T.; Morimitsu, Y. Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. *Breed. Sci.* **2014**, *64*, 48–59. [CrossRef] [PubMed]
14. Fahey, J.W.; Zalcmann, A.T.; Talalay, P. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **2001**, *56*, 5–51. [CrossRef]
15. Halkier, B.A.; Gershenzon, J. Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* **2006**, *57*, 303–333. [CrossRef] [PubMed]
16. Agerbirk, N.; Olsen, C.E. Glucosinolate structures in evolution. *Phytochemistry* **2012**, *77*, 16–45. [CrossRef]
17. Blažević, I.; Montaut, S.; Burčul, F.; Olsen, C.E.; Burow, M.; Rollin, P.; Agerbirk, N. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* **2020**, *169*, 112100. [CrossRef]
18. Clay, N.K.; Adio, A.M.; Denoux, C.; Jander, G.; Ausubel, F.M. Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science* **2009**, *323*, 95–101. [CrossRef] [PubMed]
19. Burow, M.; Halkier, B.A. How does a plant orchestrate defense in time and space? Using glucosinolates in Arabidopsis as case study. *Curr. Opin. Plant Biol.* **2017**, *38*, 142–147. [CrossRef] [PubMed]
20. van Dam, N.M.; Tytgat, T.O.G.; Kirkegaard, J.A. Root and shoot glucosinolates: A comparison of their diversity, function and interactions in natural and managed ecosystems. *Phytochem. Rev.* **2009**, *8*, 171–186. [CrossRef]
21. Koroleva, O.A.; Cramer, R. Single-cell proteomic analysis of glucosinolate-rich S-cells in *Arabidopsis thaliana*. *Methods* **2011**, *54*, 413–423. [CrossRef] [PubMed]
22. Sugiyama, R.; Hirai, M.Y. Atypical myrosinase as a mediator of glucosinolate functions in plants. *Front. Plant Sci.* **2019**, *10*, 1008. [CrossRef] [PubMed]
23. Herr, I.; Büchler, M.W. Dietary constituents of broccoli and other cruciferous vegetables: Implications for prevention and therapy of cancer. *Cancer Treat. Rev.* **2010**, *36*, 377–383. [CrossRef] [PubMed]
24. Bradlow, H.L. Indole-3-carbinol as a chemoprotective agent in breast and prostate cancer. *In Vivo* **2008**, *22*, 441–445. [PubMed]
25. Aronchik, I.; Bjeldanes, L.F.; Firestone, G.L. Direct inhibition of elastase activity by indole-3-carbinol triggers a CD40-TRAF regulatory cascade that disrupts NF-κB transcriptional activity in human breast cancer cells. *Cancer Res.* **2010**, *70*, 4961–4971. [CrossRef] [PubMed]
26. Jeong, Y.M.; Li, H.; Kim, S.Y.; Yun, H.Y.; Baek, K.J.; Kwon, N.S.; Myung, S.C.; Kim, D.S. Indole-3-carbinol inhibits prostate cancer cell migration via degradation of β-catenin. *Oncol. Res. Featur. Preclin. Clin. Cancer Ther.* **2011**, *19*, 237–243. [CrossRef]
27. Megna, B.W.; Carney, P.R.; Nukaya, M.; Geiger, P.; Kennedy, G.D. Indole-3-carbinol induces tumor cell death: Function follows form. *J. Surg. Res.* **2016**, *204*, 47–54. [CrossRef]
28. Lee, C.M.; Lee, J.; Nam, M.J.; Park, S.-H. Indole-3-carbinol induces apoptosis in human osteosarcoma MG-63 and U2OS Cells. *BioMed Res. Int.* **2018**, *2018*, 7970618. [CrossRef] [PubMed]
29. Zhang, Y.; Kensler, T.W.; Cho, C.-G.; Posner, G.H.; Talalay, P. Anticarcinogenic activities of sulforaphane and structurally related synthetic norbornyl isothiocyanates. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3147–3150. [CrossRef]
30. Singh, A.V.; Xiao, D.; Lew, K.L.; Dhir, R.; Singh, S.V. Sulforaphane induces caspase-mediated apoptosis in cultured PC-3 human prostate cancer cells and retards growth of PC-3 xenografts in vivo. *Carcinogenesis* **2004**, *25*, 83–90. [CrossRef]
31. Conaway, C.C.; Wang, C.-X.; Pittman, B.; Yang, Y.M.; Schwartz, J.E.; Tian, D.; McIntee, E.J.; Hecht, S.S.; Chung, F.-L. Phenethyl isothiocyanate and sulforaphane and their N-acetylcysteine conjugates inhibit malignant progression of lung adenomas induced by tobacco carcinogens in A/J mice. *Cancer Res.* **2005**, *65*, 8548–8557. [CrossRef]
32. Kallifatidis, G.; Rausch, V.; Baumann, B.; Apel, A.; Beckermann, B.M.; Groth, A.; Mattern, J.; Li, Z.; Kolb, A.; Moldenhauer, G.; et al. Sulforaphane targets pancreatic tumour-initiating cells by NF-κB-induced antiapoptotic signalling. *Gut* **2009**, *58*, 949–963. [CrossRef] [PubMed]
33. Lai, K.-C.; Huang, A.-C.; Hsu, S.-C.; Kuo, C.-L.; Yang, J.-S.; Wu, S.-H.; Chung, J.-G. Benzyl isothiocyanate (BITC) inhibits migration and invasion of human colon cancer HT29 cells by inhibiting matrix metalloproteinase-2/-9 and urokinase plasminogen (uPA) through PKC and MAPK signaling pathway. *J. Agric. Food Chem.* **2010**, *58*, 2935–2942. [CrossRef] [PubMed]
34. Warin, R.; Chambers, W.H.; Potter, D.M.; Singh, S.V. Prevention of mammary carcinogenesis in MMTV-neu mice by cruciferous vegetable constituent benzyl isothiocyanate. *Cancer Res.* **2009**, *69*, 9473–9480. [CrossRef] [PubMed]
35. Warin, R.; Xiao, D.; Arlotti, J.A.; Bommareddy, A.; Singh, S.V. Inhibition of human breast cancer xenograft growth by cruciferous vegetable constituent benzyl isothiocyanate. *Mol. Carcinog.* **2010**, *49*, 500–507. [CrossRef] [PubMed]

36. Rao, C.V. Benzyl isothiocyanate: Double trouble for breast cancer cells. *Cancer Prev. Res.* **2013**, *6*, 760–763. [CrossRef]
37. Boreddy, S.R.; Pramanik, K.C.; Srivastava, S.K. Pancreatic tumor suppression by benzyl isothiocyanate is associated with inhibition of PI3K/AKT/FOXO Pathway. *Clin. Cancer Res.* **2011**, *17*, 1784–1795. [CrossRef]
38. Nowicki, D.; Rodzik, O.; Herman-Antosiewicz, A.; Szalewska-Pałasz, A. Isothiocyanates as effective agents against enterohemorrhagic *Escherichia coli*: Insight to the mode of action. *Sci. Rep.* **2016**, *6*, 22263. [CrossRef]
39. Vetter, J. Plant Cyanogenic Glycosides. In *Plant Toxins*; Carlini, C.R., Ligabue-Braun, R., Gopalakrishnakone, P., Eds.; Springer: Dordrecht, The Netherlands, 2017; pp. 287–317. ISBN 978-94-007-6464-4.
40. Cressey, P.; Saunders, D.; Goodman, J. Cyanogenic glycosides in plant-based foods available in New Zealand. *Food Addit. Contam. Part A* **2013**, *30*, 1946–1953. [CrossRef]
41. Ganjewala, D.; Kumar, S.; Devi, S.A.; Ambika, K. Advances in cyanogenic glycosides biosynthesis and analyses in plants: A review. *Acta Biol. Szeged.* **2010**, *54*, 1–14.
42. Yulvianti, M.; Zidorn, C. Chemical diversity of plant cyanogenic glycosides: An overview of reported natural products. *Molecules* **2021**, *26*, 719. [CrossRef]
43. Malka, S.K.; Cheng, Y. Possible interactions between the biosynthetic pathways of indole glucosinolate and auxin. *Frontiers in Plant Sci.* **2017**, *8*, 2131. [CrossRef] [PubMed]
44. White, W.L.B.; Arias-Garzon, D.I.; McMahon, J.M.; Sayre, R.T. Cyanogenesis in cassava: The role of hydroxynitrile lyase in root cyanide production. *Plant Physiol.* **1998**, *116*, 1219–1225. [CrossRef] [PubMed]
45. Kato, Y.; Terada, H. Determination method of linamarin in cassava products and beans by ultra high performance liquid chromatography with tandem mass spectrometry. *J. Food Hyg. Soc. Jpn.* **2014**, *55*, 162–166. [CrossRef] [PubMed]
46. Kuete, V. Health effects of alkaloids from African medicinal plants. In *Toxicological Survey of African Medicinal Plants*, 1st ed.; Elsevier Inc.: Amsterdam, The Netherlands, 2014; ISBN 9780128004753.
47. Rivadeneyra-Domínguez, E.; Vázquez-Luna, A.; Rodríguez-Landa, J.F.; Díaz-Sobac, R. Neurotoxic effect of linamarin in rats associated with Cassava (*Manihot esculenta* Crantz) consumption. *Food Chem. Toxicol.* **2013**, *59*, 230–235. [CrossRef] [PubMed]
48. Mosayyebi, B.; Imani, M.; Mohammadi, L.; Akbarzadeh, A.; Zarghami, N.; Edalati, M.; Alizadeh, E.; Rahmati, M. An update on the toxicity of cyanogenic glycosides bioactive compounds: Possible clinical application in targeted cancer therapy. *Mater. Chem. Phys.* **2020**, *246*, 122841. [CrossRef]
49. Dang, T.; Nguyen, C.; Tran, P.N. Physician beware: Severe cyanide toxicity from amygdalin tablets ingestion. *Case Rep. Emerg. Med.* **2017**, *2017*, 4289527. [CrossRef] [PubMed]
50. Shi, J.; Chen, Q.; Xu, M.; Xia, Q.; Zheng, T.; Teng, J.; Li, M.; Fan, L. Recent updates and future perspectives about amygdalin as a potential anticancer agent: A review. *Cancer Med.* **2019**, *8*, 3004–3011. [CrossRef]
51. Liczbiński, P.; Bukowska, B. Molecular mechanism of amygdalin action *in vitro*: Review of the latest research. *Immunopharmacol. Immunotoxicol.* **2018**, *40*, 212–218. [CrossRef]
52. Padma, V.V. An overview of targeted cancer therapy. *BioMedicine* **2015**, *5*, 19. [CrossRef]
53. Bolarinwa, I.F.; Oke, M.O.; Olaniyan, S.A.; Ajala, A.S. A review of cyanogenic glycosides in edible plants. In *Toxicology—New Aspects to This Scientific Conundrum*; IntechOpen: London, UK, 2016. [CrossRef]
54. Bell, L.; Wagstaff, C. Enhancement of glucosinolate and isothiocyanate profiles in Brassicaceae crops: Addressing challenges in breeding for cultivation, storage, and consumer-related traits. *J. Agric. Food Chem.* **2017**, *65*, 9379–9403. [CrossRef]
55. van Doorn, J.E. Development of Vegetables with Improved Consumer Quality: A Case Study in Brussels Sprouts. Ph.D. Thesis, University of Wageningen, Wageningen, The Netherlands, 1999.
56. Sønderby, I.E.; Geu-Flores, F.; Halkier, B.A. Biosynthesis of glucosinolates—Gene discovery and beyond. *Trends Plant Sci.* **2010**, *15*, 283–290. [CrossRef] [PubMed]
57. Thodberg, S.; Del Cueto, J.; Mazzeo, R.; Pavan, S.; Lotti, C.; Dicenta, F.; Neilson, E.H.J.; Møller, B.L.; Sánchez-Pérez, R. Elucidation of the amygdalin pathway reveals the metabolic basis of bitter and sweet almonds (*Prunus dulcis*). *Plant Physiol.* **2018**, *178*, 1096–1111. [CrossRef] [PubMed]
58. Jørgensen, K.; Morant, A.V.; Morant, M.; Jensen, N.B.; Olsen, C.E.; Kannangara, R.; Motawia, M.S.; Møller, B.L.; Bak, S. Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: Isolation, biochemical characterisation, and expression pattern of CYP71E7, the oxime-metabolising cytochrome P450 enzyme. *Plant Physiol.* **2011**, *155*, 282–292. [CrossRef] [PubMed]
59. Sekeli, R.; Hamid, M.H.; Razak, R.A.; Wee, C.Y.; Ong-Abdullah, J. Malaysian *Carica papaya* L. var. eksotika: Current research strategies fronting challenges. *Front. Plant Sci.* **2018**, *9*, 1380. [CrossRef] [PubMed]
60. Cruz, A.F.; de Oliveira, B.F.; de Carvalho Pires, M. Optimum level of nitrogen and phosphorus to achieve better Papaya (*Carica papaya* var. Solo) seedlings growth and mycorrhizal colonisation. *Int. J. Fruit Sci.* **2017**, *17*, 259–268. [CrossRef]
61. Zainal-Abidin, R.A.; Ruhaizat-Ooi, I.H.; Harun, S. A review of omics technologies and bioinformatics to accelerate improvement of papaya traits. *Agronomy* **2021**, *11*, 1356. [CrossRef]
62. Ming, R.; Yu, Q.; Moore, P.H.; Paull, R.E.; Chen, N.J.; Wang, M.L.; Zhu, Y.J.; Schuler, M.A.; Jiang, J.; Paterson, A.H. Genome of papaya, a fast growing tropical fruit tree. *Tree Genet. Genomes* **2012**, *8*, 445–462. [CrossRef]
63. Santana, L.F.; Inada, A.C.; Espirito Santo, B.L.S.D.; Filiú, W.F.; Pott, A.; Alves, F.M.; Guimarães, R.D.C.A.; Freitas, K.D.C.; Hiane, P.A. Nutraceutical potential of *Carica papaya* in metabolic syndrome. *Nutrients* **2019**, *11*, 1608. [CrossRef]
64. Jing, G.; Li, T.; Qu, H.; Yun, Z.; Jia, Y.; Zheng, X.; Jiang, Y. Carotenoids and volatile profiles of yellow- and red-fleshed papaya fruit in relation to the expression of carotenoid cleavage dioxygenase genes. *Postharvest Biol. Technol.* **2015**, *109*, 114–119. [CrossRef]

65. Wei, F.; Wing, R.A. A fruitful outcome to the papaya genome project. *Genome Biol.* **2008**, *9*, 227. [CrossRef] [PubMed]
66. Arumuganathan, K.; Earle, E.D. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Report.* **1991**, *9*, 208–218. [CrossRef]
67. Harun, S.; Abdullah-Zawawi, M.-R.; A-Rahman, M.R.A.; Muhammad, N.A.N.; Mohamed-Hussein, Z.-A. SuCComBase: A manually curated repository of plant sulfur-containing compounds. *Database* **2019**, *2019*, baz021. [CrossRef]
68. Olafsdottir, E.S.; Bolt Jorgensen, L.; Jaroszewski, J.W. Cyanogenesis in glucosinolate-producing plants: *Carica papaya* and *Carica quercifolia*. *Phytochemistry* **2002**, *60*, 269–273. [CrossRef]
69. Clausen, M.; Kannangara, R.M.; Olsen, C.E.; Blomstedt, C.K.; Gleadow, R.M.; Jørgensen, K.; Bak, S.; Motawie, M.S.; Møller, B.L. The bifurcation of the cyanogenic glucoside and glucosinolate biosynthetic pathways. *Plant J.* **2015**, *84*, 558–573. [CrossRef] [PubMed]
70. Miranda Rossetto, M.R.; Shiga, T.M.; Vianello, F.; Pereira Lima, G.P. Analysis of total glucosinolates and chromatographically purified benzylglucosinolate in organic and conventional vegetables. *LWT Food Sci. Technol.* **2013**, *50*, 247–252. [CrossRef]
71. Wittstock, U.; Halkier, B.A. Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyses the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. *J. Biol. Chem.* **2000**, *275*, 14659–14666. [CrossRef] [PubMed]
72. Spencer, C.K.; Seigler, D.S. Cyanogenic glycosides of *Carica papaya* and its phylogenetic position with respect to the violates and capparales. *Am. J. Bot.* **1984**, *71*, 1444–1447. [CrossRef]
73. Ettlinger, M.G.; Hodgkins, J.E. The mustard oil of papaya seed. *J. Org. Chem.* **1956**, *21*, 204. [CrossRef]
74. Tang, C.S. Benzyl isothiocyanate of papaya fruit. *Phytochemistry* **1971**, *10*, 117–121. [CrossRef]
75. MacLeod, A.J.; Pieris, M. Volatile components of papaya (*Carica papaya* L.) with particular reference of glucosinolate products. *J. Agric. Food Chem.* **1983**, *31*, 1005–1008. [CrossRef]
76. Choi, H.S.; Cho, M.C.; Lee, H.G.; Yoon, D.Y. Indole-3-carbinol induces apoptosis through p53 and activation of caspase-8 pathway in lung cancer A549 Cells. *Food Chem. Toxicol.* **2010**, *48*, 883–890. [CrossRef] [PubMed]
77. Hirai, M.Y.; Sugiyama, K.; Sawada, Y.; Tohge, T.; Obayashi, T.; Suzuki, A.; Araki, R.; Sakurai, N.; Suzuki, H.; Aoki, K.; et al. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6478–6483. [CrossRef] [PubMed]
78. Frerigmann, H.; Berger, B.; Gigolashvili, T. bHLH05 is an interaction partner of MYB51 and a novel regulator of glucosinolate biosynthesis in Arabidopsis. *Plant Physiol.* **2014**, *166*, 349–369. [CrossRef] [PubMed]
79. Gigolashvili, T.; Berger, B.; Mock, H.; Mu, C. The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J.* **2007**, *50*, 886–901. [CrossRef] [PubMed]
80. Gigolashvili, T.; Yatusevich, R.; Berger, B.; Müller, C.; Flügge, U.-I. The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J.* **2007**, *51*, 247–261. [CrossRef] [PubMed]
81. Harun, S.; Rohani, E.R.; Ohme-Takagi, M.; Goh, H.H.; Mohamed-Hussein, Z.A. ADAP is a possible negative regulator of glucosinolate biosynthesis in *Arabidopsis thaliana* based on clustering and gene expression analyses. *J. Plant Res.* **2021**, *134*, 327–339. [CrossRef] [PubMed]
82. Ashari, K.S.; Abdullah-Zawawi, M.R.; Harun, S.; Mohamed-Hussein, Z.A. Reconstruction of the transcriptional regulatory network in *Arabidopsis thaliana* aliphatic glucosinolate biosynthetic pathway. *Sains Malays.* **2018**, *47*, 2993–3002. [CrossRef]
83. Knill, T.; Schuster, J.; Reichelt, M.; Gershenzon, J.; Binder, S.; Germany, T.K.; Institut, M.P. Arabidopsis branched-chain amino-transferase 3 functions in both amino acid and glucosinolate biosynthesis. *Plant Physiol.* **2008**, *146*, 1028–1039. [CrossRef] [PubMed]
84. Sawada, Y.; Kuwahara, A.; Nagano, M.; Narisawa, T.; Sakata, A.; Saito, K.; Yokota Hirai, M. Omics-based approaches to methionine side chain elongation in Arabidopsis: Characterisation of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. *Plant Cell Physiol.* **2009**, *50*, 1181–1190. [CrossRef] [PubMed]
85. Harun, S.; Afiqah-Aleng, N.; Karim, M.B.; Amin, M.A.U.; Kanaya, S.; Mohamed-Hussein, Z.-A. Potential *Arabidopsis thaliana* glucosinolate genes identified from the co-expression modules using graph clustering approach. *PeerJ.* **2021**, *9*, e11876. [CrossRef] [PubMed]
86. Sawada, Y.; Toyooka, K.; Kuwahara, A.; Sakata, A.; Nagano, M.; Saito, K.; Hirai, M.Y. Arabidopsis bile acid: Sodium symporter family protein 5 is involved in methionine-derived glucosinolate biosynthesis. *Plant Cell Physiol.* **2009**, *50*, 1579–1586. [CrossRef] [PubMed]
87. Gigolashvili, T.; Yatusevich, R.; Rollwitz, I.; Humphry, M.; Gershenzon, J.; Flu, U. The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*. *Plant Cell* **2009**, *21*, 1813–1829. [CrossRef] [PubMed]
88. Andersen, T.G.; Halkier, B.A.; Andersen, T.G.; Halkier, B.A. Upon bolting the GTR1 and GTR2 transporters mediate transport of glucosinolates to the inflorescence rather than roots. *Plant Signal. Behav.* **2014**, *9*, e27740. [CrossRef] [PubMed]
89. Harun, S.; Afiqah-Aleng, N.; Hadi, F.I.A.; Lam, S.D.; Mohamed-Hussein, Z.A. Identification of potential genes encoding protein transporters in *Arabidopsis thaliana* glucosinolate (GSL) metabolism. *Life* **2022**, *12*, 326. [CrossRef]
90. Jensen, L.M.; Jepsen, H.S.K.; Halkier, B.A.; Kliebenstein, D.J.; Burow, M. Natural variation in cross-talk between glucosinolates and onset of flowering in Arabidopsis. *Front. Plant Sci.* **2015**, *6*, 697. [CrossRef]
91. Kim, J.I.; Zhang, X.; Pascuzzi, P.E.; Liu, C.J.; Chapple, C. Glucosinolate and phenylpropanoid biosynthesis are linked by proteasome-dependent degradation of PAL. *New Phytol.* **2020**, *225*, 154–168. [CrossRef]

92. Frerigmann, H.; Gigolashvili, T. MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Molecular Plant* **2014**, *7*, 814–828. [CrossRef]
93. Gigolashvili, T.; Engqvist, M.; Yatusevich, R.; Müller, C.; Flügge, U.-I. HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytol.* **2008**, *177*, 627–642. [CrossRef]
94. Naur, P.; Petersen, B.L.; Mikkelsen, M.D.; Bak, S.; Rasmussen, H.; Olsen, C.E.; Halkier, B.A. CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolising oximes in the biosynthesis of glucosinolates in Arabidopsis. *Plant Physiol.* **2003**, *133*, 63–72. [CrossRef]
95. Mikkelsen, M.D.; Naur, P.; Halkier, B.A. Arabidopsis mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *Plant J.* **2004**, *37*, 770–777. [CrossRef]
96. Grubb, C.D.; Zipp, B.J.; Ludwig-mu, J.; Masuno, M.N.; Molinski, T.F.; Abel, S. Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J.* **2004**, *40*, 893–908. [CrossRef] [PubMed]
97. Piotrowski, M.; Schemenewitz, A.; Lopukhina, A.; Mu, A.; Janowitz, T.; Weiler, E.W.; Oecking, C. Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyse the final step in the biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* **2004**, *279*, 50717–50725. [CrossRef] [PubMed]
98. Klein, M.; Papenbrock, J. Kinetics and substrate specificities of desulfo-glucosinolate sulfotransferases in *Arabidopsis thaliana*. *Physiol. Plant.* **2009**, *135*, 140–149. [CrossRef] [PubMed]
99. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]
100. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [CrossRef]
101. Aasland, R.; Stewart, A.F.; Gibson, T. The SANT domain: A putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem. Sci.* **1996**, *21*, 87–88. [CrossRef]
102. Stauber, E.J.; Kuczka, P.; van Ohlen, M.; Vogt, B.; Janowitz, T.; Piotrowski, M.; Beuerle, T.; Wittstock, U. Turning the "mustard oil bomb" into a "cyanide bomb": Aromatic glucosinolate metabolism in a specialist insect herbivore. *PLoS ONE* **2012**, *7*, e35545. [CrossRef]
103. Deng, P.; Cui, B.; Zhu, H.; Phommakoun, B.; Zhang, D.; Li, Y.; Zhao, F.; Zhao, Z. Accumulation pattern of amygdalin and prunasin and its correlation with fruit and kernel agronomic characteristics during apricot (*Prunus armeniaca* L.) kernel development. *Foods* **2021**, *10*, 397. [CrossRef] [PubMed]
104. Fukuda, T.; Ito, H.; Mukainaka, T.; Tokuda, H.; Nishino, H.; Yoshida, T. Anti-tumor promoting effect of glycosides from *Prunus persica* seeds. *Biol. Pharm. Bull.* **2003**, *26*, 271–273. [CrossRef]
105. Sánchez-Pérez, R.; Pavan, S.; Mazzeo, R.; Moldovan, C.; Aiese Cigliano, R.; del Cueto, J.; Ricciardi, F.; Lotti, C.; Ricciardi, L.; Dicenta, F.; et al. Mutation of a bHLH transcription factor allowed almond domestication. *Science* **2019**, *364*, 1095–1098. [CrossRef]
106. Sigrist, C.J.A.; de Castro, E.; Cerutti, L.; Cuche, B.A.; Hulo, N.; Bridge, A.; Bougueleret, L.; Xenarios, I. New and continuing developments at PROSITE. *Nucleic Acids Res.* **2013**, *41*, D344–D347. [CrossRef] [PubMed]
107. Feller, A.; MacHemer, K.; Braun, E.L.; Grotewold, E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **2011**, *66*, 94–116. [CrossRef] [PubMed]
108. Araki, R.; Hasumi, A.; Nishizawa, O.I.; Sasaki, K.; Kuwahara, A.; Sawada, Y.; Totoki, Y.; Toyoda, A.; Sakaki, Y.; Li, Y.; et al. Novel bioresources for studies of *Brassica oleracea*: Identification of a kale MYB transcription factor responsible for glucosinolate production. *Plant Biotechnol. J.* **2013**, *11*, 1017–1027. [CrossRef] [PubMed]
109. Zainal-Abidin, R.-A.; Harun, S.; Vengatharajuloo, V.; Tamizi, A.-A.; Samsulrizal, N.H. Gene co-expression network tools and databases for crop improvement. *Plants* **2022**, *11*, 1625. [CrossRef]
110. Li, M.; Zhao, L.; Zhou, Q.; Fang, L.; Luo, D.; Liu, W.; Searle, I.R.; Liu, Z. Transcriptome and co-expression network analyses provide in-sights into the molecular mechanisms of hydrogen cyanide synthesis during seed development in common vetch (*Vicia sativa* L.). *Int. J. Mol. Sci.* **2022**, *23*, 2275. [CrossRef]
111. Dalisay, D.S.; Kim, K.W.; Lee, C.; Yang, H.; Rübel, O.; Bowen, B.P.; Davin, L.B.; Lewis, N.G. Dirigent protein-mediated lignan and cyanogenic glucoside formation in flax seed: Integrated omics and MALDI mass spectrometry imaging. *J. Nat. Prod.* **2015**, *78*, 1231–1242. [CrossRef]
112. Zhang, T.; Liu, R.; Zheng, J.; Wang, Z.; Gao, T.; Qin, M.; Hu, X.; Wang, Y.; Yang, S.; Li, T. Insights into glucosinolate accumulation and metabolic pathways in *Isatis indigotica* Fort. *BMC Plant Biol.* **2022**, *22*, 78. [CrossRef]