RESEARCH

De novo transcriptome profiling unveils the regulation of phenylpropanoid biosynthesis in unripe *Piper nigrum* berries

Sweda Sreekumar[1,2,3], Kattupalli Divya[1,2], Nisha Joy[1,4] and E. V. Soniya[1*]

Abstract

Background: Black pepper (*Piper nigrum* L.) is rich in bioactive compounds that make it an imperative constituent in traditional medicines. Although the unripe fruits have long been used in different Ayurvedic formulations, the mechanism of gene regulation resulting in the production of the bioactive compounds in black pepper is not much investigated. Exploring the regulatory factors favouring the production of bioactive compounds ultimately help to accumulate the medicinally important content of black pepper. The factors that enhance the biosynthesis of these compounds could be potential candidates for metabolic engineering strategies to obtain a high level production of significant biomolecules.

Results: Being a non-model plant, de novo sequencing technology was used to unravel comprehensive information about the genes and transcription factors that are expressed in mature unripe green berries of *P. nigrum* from which commercially available black pepper is prepared. In this study, the key gene regulations involved in the synthesis of bioactive principles in black pepper was brought out with a focus on the highly expressed phenylpropanoid pathway genes. Quantitative real-time PCR analysis of critical genes and transcription factors in the different developmental stages from bud to the mature green berries provides important information useful for choosing the developmental stage that would be best for the production of a particular bioactive compound. Comparison with a previous study has also been included to understand the relative position of the results obtained from this study.

Conclusions: The current study uncovered significant information regarding the gene expression and regulation responsible for the bioactivity of black pepper. The key transcription factors and enzymes analyzed in this study are promising targets for achieving a high level production of significant biomolecules through metabolic engineering.

Keywords: Black pepper, Fruits, RNA-Seq, Developmental stages, Secondary metabolite

Background

Black pepper (*Piper nigrum* L.) is called 'Black Gold' or the 'King of Spices' and is a commonly used spice all over the world [1, 2]. The particular pungent taste and diverse medicinal properties of the fruits/berries of black pepper make it a treasured spice. Black pepper is prepared by first immersing the unripe green berries in hot water followed by sun-drying for many days [3]. This spice is effective in treating various diseases including cancer [4, 5].

Black pepper has an array of medicinally active compounds which is the reason for its high demand in Ayurveda. The commercially available black pepper is prepared from the mature green unripe berries and is largely used as a favourite spice as well as a health enhancer [6]. The phytochemicals in black pepper help to remove harmful free radicals that mutate DNA and

*Correspondence: evsoniya@rgcb.res.in
[1] Transdisciplinary Biology, Rajiv Gandhi Centre for Biotechnology (RGCB), Thiruvananthapuram, Kerala, India
Full list of author information is available at the end of the article

Sreekumar *et al. BMC Plant Biology*   (2022) 22:501

Page 2 of 14

hence help to protect from cancers and diseases [7]. It is also used as a preservative, perfumery and insecticide [2, 8]. Dietary intake of black pepper increases the secretion of bile, trypsin, chymotrypsin, pancreatic lipase and amylase indicating its beneficial effects on the gastrointestinal system [2]. The bioavailability of curcumin, the bioactive anti-cancer compound in turmeric, can be increased by 20-fold when administered along with piperine [9]. The generalised use of black pepper seems to be due to its ability to enhance the effect of various medicinal herbs.

Despite the immense medicinal benefits from the biochemicals in black pepper, the molecular network shaping the biosynthesis of these secondary metabolites is still largely unknown due to the lack of proper sequence information and fragmented data obtained [10]. In-depth analysis of the unknown mechanisms behind the biosynthesis of medicinally and biologically active secondary metabolites in black pepper has not been accomplished to date. Limited information on the regulation of biosynthesis of specialized metabolites is a hindrance in utilizing medicinally valuable herbs for commercial level production using synthetic biology [11, 12]. It is imperative to understand the genes responsible for the accumulation of biologically active secondary metabolites for the betterment of the quality parameters of the fruit [13]. Next Generation Sequencing (NGS) technology has revolutionized molecular biology by providing high-fidelity sequences and wholesome high-throughput information on transcriptomes [14]. High-throughput data helps to identify critical nodes including enzymes and transcription factors involved in the biosynthetic regulation of bioactive compounds.

Although minor experiments have been carried out based on the bioactive compounds present in pepper, high-throughput platforms have not been utilized yet to uncover the depth of the mechanisms and regulations that are involved in their biosynthesis. Only by dissecting these networks, a suitable metabolic engineering strategy can be adopted and applied in the near future to ensure high-level production of such therapeutically significant compounds. The major genes and regulatory agents should be tracked and a comprehensive understanding of the interplay between the key candidates that help in the biosynthesis of these principles should be portrayed to make the best use of this spice crop in the pharmaceutical industry. The present study has utilized high-throughput transcriptomics to obtain a better understanding of the under-studied aspects of the molecular network behind phenylpropanoid biosynthesis in *P. nigrum* fruits.

Methods

Selection of developmental stages of *P. nigrum*

Unripe berries of *P. nigrum* (variety—Panniyur 1) maintained in the green house of Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, were used for extracting RNA for transcriptome sequencing.The time of collection of fruits was recorded as days after flowering (DAF). Total RNA was extracted from five different developmental stages and chosen for the quantitative real-time PCR experiments (Fig. 1): Bud, Flower, Immature young fruit [stage 1(30 DAF)], Immature young fruit [stage 2 (60 DAF)], and Unripe berry (120 DAF). The transition from bud stage to fully flowering stage took approximately 12–15 days. The plant samples collected were flash-frozen using liquid nitrogen. They were stored at -80 °C until further use.

Isolation of RNA

High quality total RNA was extracted from 50-100 mg of tissues from the different developmental stages of *P. nigrum* as previously mentioned, using Spectrum™ Plant Total RNA Kit by standardized procedure
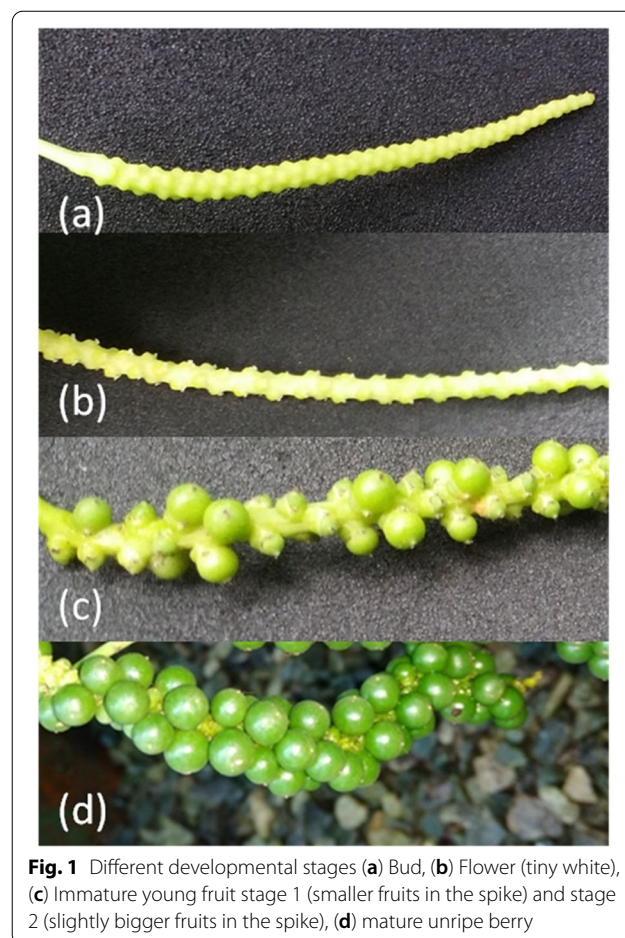

Fig. 1 Different developmental stages (**a**) Bud, (**b**) Flower (tiny white), (**c**) Immature young fruit stage 1 (smaller fruits in the spike) and stage 2 (slightly bigger fruits in the spike), (**d**) mature unripe berry

(Sigma-Aldrich). RNA isolation was carried out in an RNA-safe environment. Before adding the samples for grinding, pre-cooled the mortar by adding liquid nitrogen for maintaining the sample in a chilled condition. Ground the samples to fine white powder, intermittently adding liquid nitrogen to the mortar. Samples were homogenized with lysis buffer. The lysate was centrifuged to remove debris. Supernatant was mixed with binding buffer and loaded onto Spectrum column and further steps were followed as per manufacturer's guidelines. RNA was eluted in Nuclease free water (Ambion, USA). The quantity and quality of the RNA was assessed using microvolume spectrophotometer (Colibri, Germany).

High-throughput sequencing and analysis

De-novo transcriptome sequencing of *P. nigrum* unripe fruit was carried out. Workflow for the analysis is given in Additional file 1.The library preparation for RNA-Seq was performed at Genotypic Technology, Bangalore, India, according to Illumina-compatible NEBNext UltraTM Directional RNA Library Prep.

Illumina HiSeq sequencer was used for sequencing 150 bp length paired-end (PE) reads to produce a total of 2, 60, 90, 849 raw sequencing reads. The quality checking was carried out using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Processing of raw reads was done using Cutadapt tool [15] for adapters and low-quality bases trimming towards 3'-end, with average Phred quality score (Q) of ≥ 30 at each base position. A total of 2, 44, 45, 774 high-quality adapter-free data was used after pre-processing for assembly.

Construction of RNA sequencing libraries was done using Illumina-compatible NEBNext® Ultra™ Directional RNA Library Prep Kit (New England BioLabs, MA, USA).1 µg of total RNA was used for mRNA isolation. This was followed by fragmentation and priming. Further, first strand synthesis was done and then second strand synthesis was carried out. HighPrep magnetic beads (Magbio Genomics Inc, USA) were used for purifying double-stranded cDNA. The resulting cDNA was end-repaired, adenylated and ligated to Illumina multiplex barcode adapters according to the protocol of NEBNext® Ultra™ Directional RNA Library Prep Kit.

Illumina Universal Adapters used:

5'-AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCC

GATCT-3'

and Index Adapter:

5'-GATCGGAAGAGCACACGTCTGAACTCCAGT CAC[INDEX]ATCTCGTATGCCGT

CTTCTGCTTG-3'.

Here, [INDEX] is the unique sequence to identify sample-specific sequencing data.

After ligation with the adapter, cDNA was purified with HighPrep beads. Purified cDNA underwent 12 cycles of Indexing-PCR (37 ˚C for 15 min, then denaturation at 98 ˚C for 30 s, followed by cycle of 98 ˚C for 10 s and 65 ˚C for 75 s, then finally 65 ˚C for 5 min) for enriching the adapter-ligated fragments. The PCR product is the sequencing library, which was purified with HighPrep beads. Quality control checking was done. The quantification was done using Qubit fluorometer (ThermoFisher Scientific, MA, USA). The fragment size distribution was analyzed on Agilent 2200 Tapestation.

Assembly and annotation

Trinitysoftware package [16] was used to assemble the trimmed reads as contigs of the *P. nigrum* transcriptome. Long contig sequences were obtained by continuously joining overlapping reads of particular quality and length. The maximum and minimum length, N50 and average length were determined. CD-HIT open source program [17, 18] was used for clustering the transcripts based on similarity of sequences. The initial analysis included normalizing of read counts for the transcripts to get Reads Per Kilobase of transcript per Million mapped reads (RPKM).

Functional annotation was done using several databases including KAAS (http://www.genome.jp/tools/kaas/), MISA (https://webblast.ipk-gatersleben.de/misa/), GO (http://geneontology.org/) and Uniprot (https://www.uniprot.org/). NCBI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [19] was used for annotation based on sequence homology with protein sequences of 'Viridiplantae members' in Uniprot [20]. Cut-off value was set at more than 30% identity and E-value less than 1e-5. Pathway annotation was done by using KAAS server [21] against the reference dataset of Viridiplantae proteins. The reference organisms used were *Arabidopsis thaliana* (ath), *Theobroma cacao* (tcc), *Tarenaya hassleriana* (thj), *Eucalyptus grandis* (egr), *Glycine max* (gmx), *Fragaria vesca* (fve), *Vitis vinifera* (vvi), *Solanum lycopersicum* (sly), *Populus trichocarpa* (pop) and *Oryza sativa japonica* (osa). Simple sequence repeats were identified from assembled transcripts using MISA.

Expression analysis of selected genes in different developmental stages of fruits

Comparative quantitative real-time Polymerase Chain Reaction (qRT-PCR) experiment was done for selected genes in the different developing stages (bud, flower, young fruit stage 1 (YF1) and stage 2 (YF2), and unripe berries) (Fig. 1) to study the expression of the phenylpropanoid pathway-associated genes and transcription factors observed in the transcriptomic data of *P. nigrum*

berries. For this, the total RNA isolated from each developmental stage was used for cDNA synthesis. Forward and reverse primers were designed using IDT (www.idtdna.com) for the selected genes based on the contig sequences obtained from the transcriptomic data (Additional file 2). The self-complementarity of primers was checked using Oligocalc (http://biotools.nubic.northwestern.edu/).

The quantity and quality of the obtained RNA was checked with spectrophotometer (Colibri, Germany). cDNA was synthesized from 1 µg of RNA using a high capacity cDNA reverse transcription kit (Applied Biosystems, Life Technologies). Each reaction mixture containing 2.5 µl SYBR Green Master mix (Applied Biosystems), 1 µl cDNA template (50 ng), and 0.25 ul each of forward and reverse primer (10 ρM) was made up to a total volume of 5 µl with nuclease free water. Real-time experiments were performed in Biosystems 7900 HT Fast Real-Time system with Power SYBR Green qPCR Master Mix (ABI, Life Technologies) in384-well optical reaction plates (Applied Biosystems, USA). The PCR reactions were performed in triplicate under

following conditions: 10 min at 95 °C, and 40 cycles of 15 s at 95 °C and 45 s at 60 °C. The cDNA template was replaced with nuclease free water in the case of negative controls. 5.8S ribosomal RNA was used as the reference gene. The analysis was done using the formula 2^{-ddCT} and the standard deviation was represented as error bar [22]. The expression level in bud was arbitrarily set to 1 and the expression in different tissues was compared to that in the bud [23].

We compared the RPKM of selected genes from our data with that of [24] retrieved from NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/Traces/sra/; accession number SRS856941) in which they used a mixture of RNA from different stages of fruit (1–10 months post anthesis).

Results
NGS library preparation
Sequencing was performed in Illumina HiSeq sequencer with the total RNA extracted from unripe berry sample (Additional file 3).
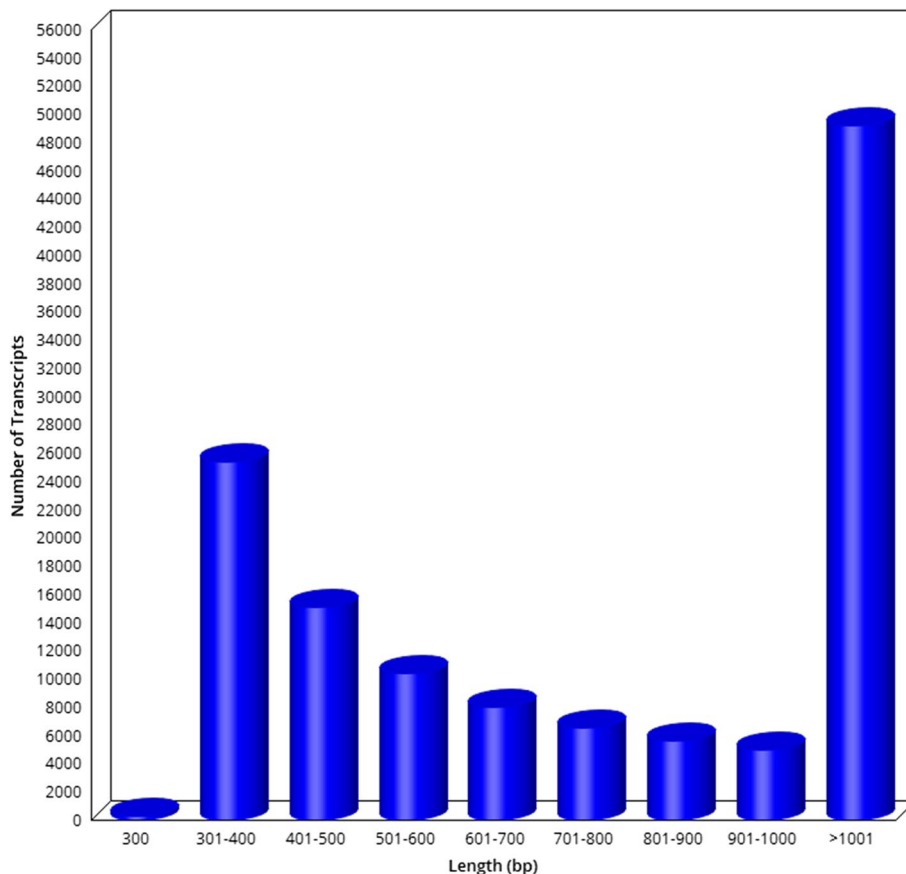

Fig. 2 Master unigene length distribution plot

Sreekumar *et al. BMC Plant Biology*   (2022) 22:501

Page 5 of 14

The sequencing resulted in a total of 2, 60, 90, 849 raw sequencing reads. Total of 24.45 million reads were used for the downstream analysis after pre-processing. On an average of 93.69% of high quality data was retained (Additional file 4).

Master transcripts were generated by clustering the transcripts of the sample with 95% similarity. The length distribution of master unigenes is given in Fig. 2.

Master transcripts were annotated by the homology search against "Viridiplantae" data from Uniprot containing 51,12,021 protein sequences downloaded on 08/01/2018, using BLAST. Thus 64.50% of the transcripts were functionally annotated. Top 10 abundant Gene Ontology (GO) terms were observed (Fig. 3).

Using Viridiplantae database, 47,085 transcripts with high similarity (E-value < 1e-60) were observed while using BLAST i.e., the possible hits of similar score for these transcripts that could be obtained by mere chance ranged from 1e-60 to 1e-5. This low E-value indicates high significance and were considered for further analysis. The E-value of the rest of the transcripts ranged from 1e-5 to 0 (Additional file 5). Since similarity of sequences is directly proportional to homology, the similarity distribution with reference was analysed. 15,709 of the sequences had a strong similarity of higher than 80% and 64,655 had below 80% similarity (Additional file 6).

KAAS Server was used for pathway analysis. For identifying pathways, plant model organisms were considered as reference. Total of 14,063 transcripts were annotated against KAAS out of which "Ribosome" pathway was found to be most abundant followed by "Plant -Pathogen interaction", "Protein processing in endoplasmic reticulum" etc. (Fig. 4).

Total of 58% transcripts were annotated against KOG database. Top 25 KOG functions were plotted (Fig. 5). The homology search was carried out using NCBI BLAST 2.2.29. MISA perl script was used for the identification of Simple Sequence Repeats (SSRs) in the transcripts. It was found that 27.02% of the transcripts have SSRs. Total of 33,890 SSRs were predicted in 125,397 transcripts. However, 5732 transcripts had more than one SSR. Most of the SSRs were mono-nucleotide and tri-nucleotide repeats (Additional file 7). TA/AT motifs were observed to be the highest distributed SSR motif type reaching a total of 846, followed by CCG/GGC reaching a total of 608 (Fig. 6).

The gross RPKM for all the identified enzymes in each pathway was analyzed (Additional file 8). The most expressed biosynthetic pathway was thus found to be the phenylpropanoid pathway, followed by methylglyoxal degradation pathway. Terpenoid pathway was the third most expressed secondary metabolite pathway, followed by the flavonoid pathway.
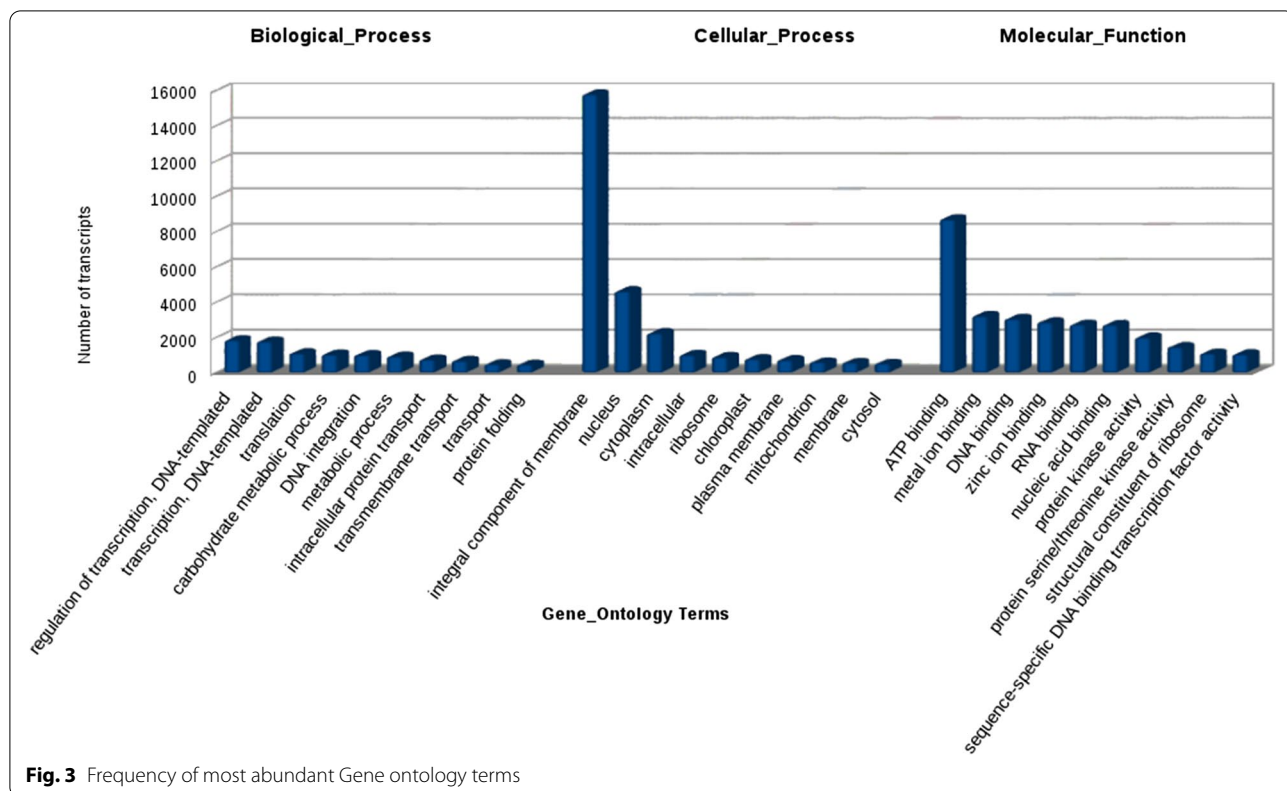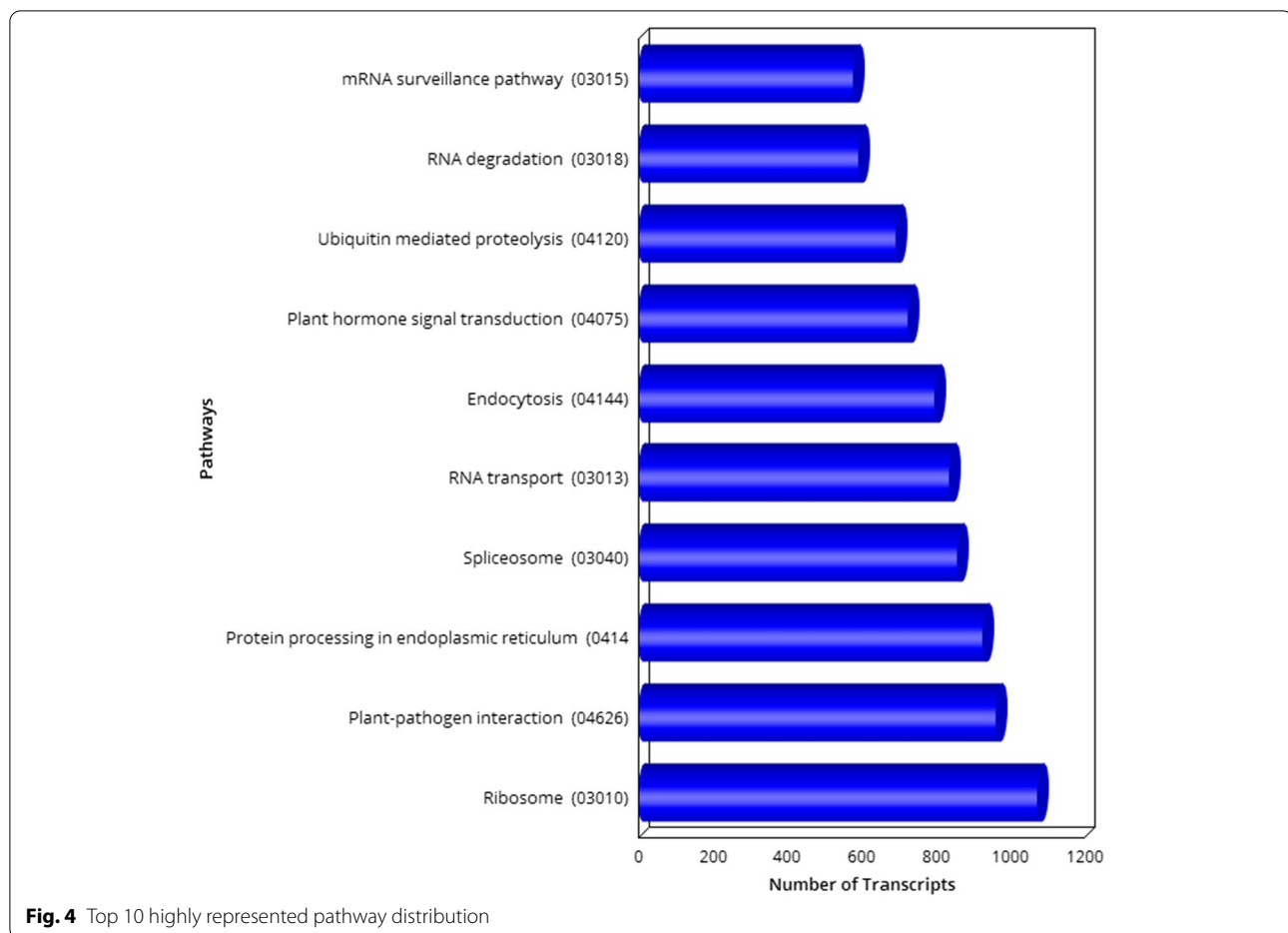

Fig. 3 Frequency of most abundant Gene ontology terms

Sreekumar *et al. BMC Plant Biology*   (2022) 22:501

Page 6 of 14

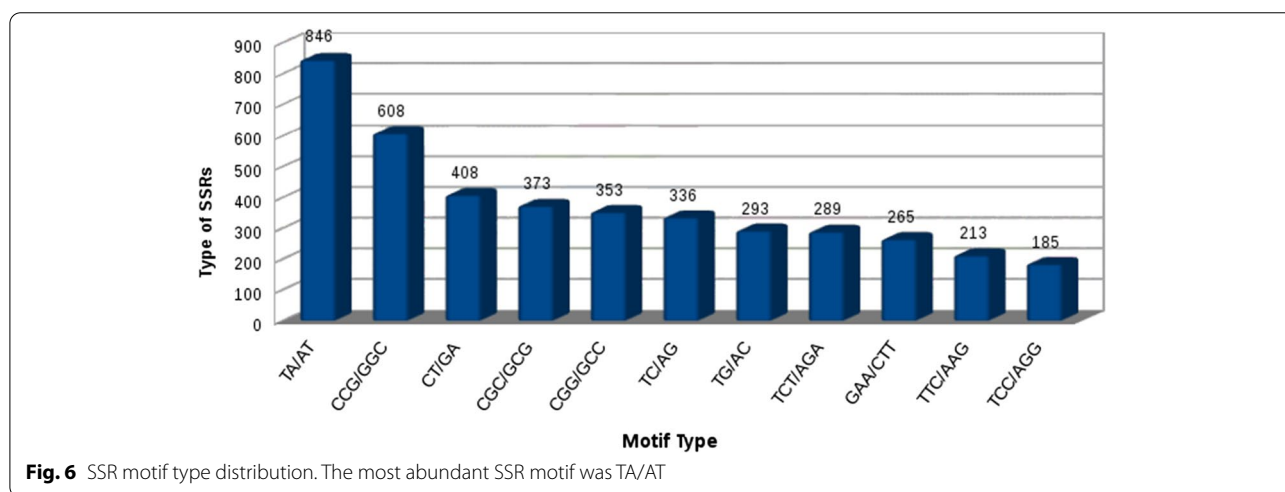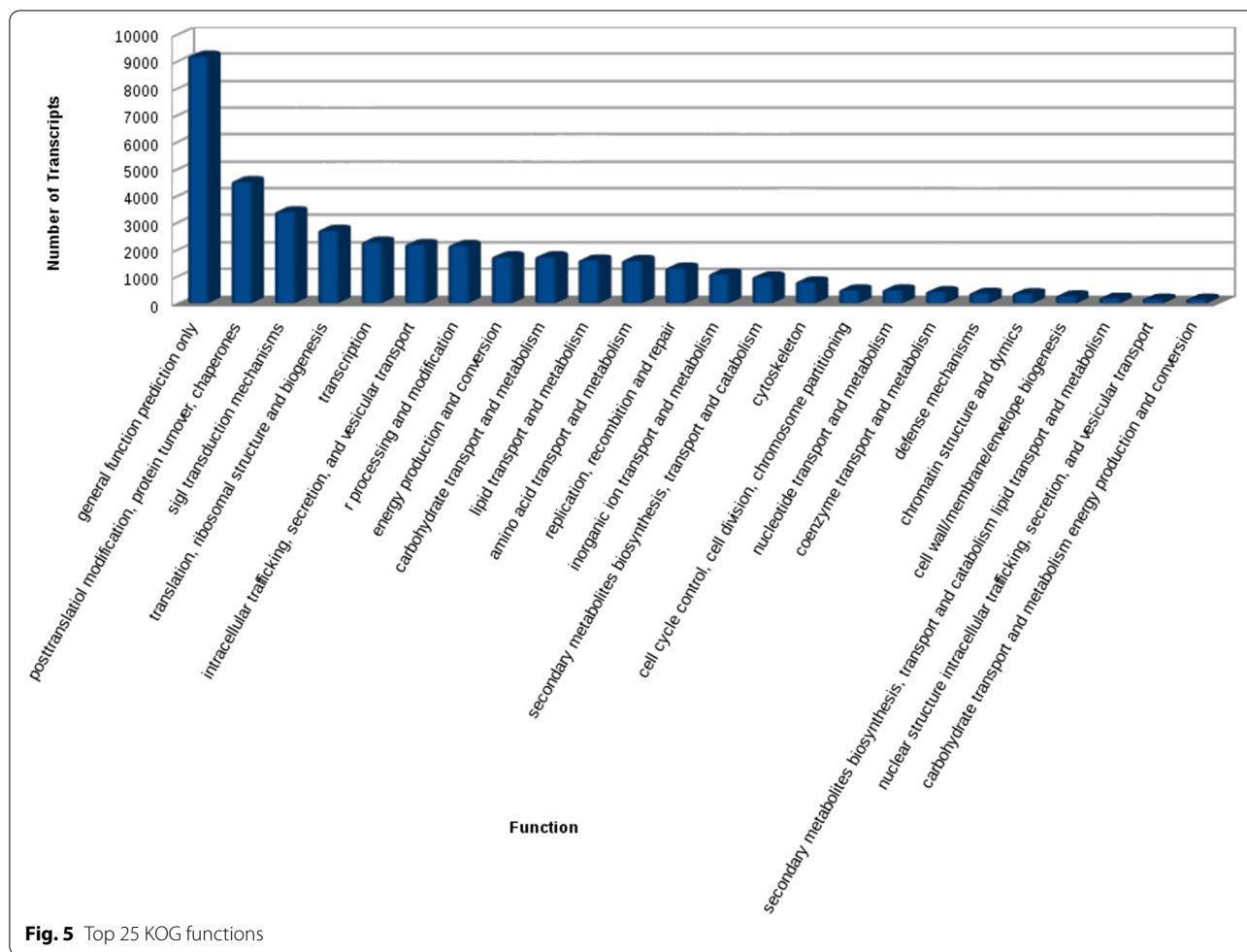**Fig. 4** Top 10 highly represented pathway distribution

Based on the highest RPKM observed for PAL and C4H, the corresponding reactions catalyzed by these enzymes are assumed to be the most prominent in the phenylpropanoid metabolic pathway in black pepper i.e., the trans-cinnamate biosynthesis and trans-4-coumarate biosynthesis. These two processes result in the biosynthesis of trans-cinnamate from L-phenylalanine and trans-4-coumarate from trans-cinnamate respectively [24, 25]. The most expressed pathway being the phenylpropanoid pathway, the enzyme participants of the pathway were identified from the transcriptome data, which included phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), 4-coumarate CoA Ligase (4CL), cinnamoylreductase (CCR), hydroxyl cinnamoyltransferase (HCT), cinnamyl alcohol dehydrogenase (CAD), Caffeic acid 3-O-methyltransferase (COMT) and p-coumaroyl ester 3'-hydroxylase (C3'H) [26]. Most of the enzymes identified belonged to the flavonone and monolignol biosynthetic branches in the phenylpropanoid pathway [27], and were selected for further study.
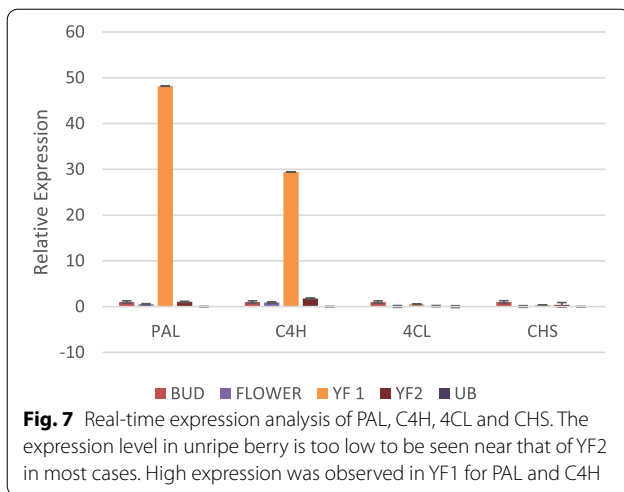
Real-time validation studies

The expression of the phenylpropanoid biosynthetic genes identified from the transcriptomic data were compared in the different developmental stages of *P. nigrum* berries using qRT-PCR. The details of the RNA samples used are given in Additional file 9.

The general phenylpropanoid pathway consist of the first four enzymes PAL, C4H, 4CL and CHS, and so it was significant to observe that the levels of these enzymes dropped in unripe berries (Fig. 7). The first enzyme of the phenylpropanoid pathway, PAL showed highest expression in the young fruits stage 1 as compared to bud, flower, young fruit stage 2 and unripe berry. PAL catalyses the deamination of phenylalanine to trans-cinnamate. Thus it connects the flux of carbon from shikimate pathway to the phenylpropanoid pathway [28]. C4H also showed similar expression level. However, 4CL and CHS showed abundance in bud stage. The succeeding genes of the pathway, such as CAD, CCR, HCT, p-coumarate 3-hydroxylase (C3H), Caffeoyl-CoA O-methyltransferase (CCOAOMT), Caffeic acid 3-O-methyltransferase

Sreekumar *et al. BMC Plant Biology*   (2022) 22:501

Page 7 of 14

Fig. 5 Top 25 KOG functions

Fig. 6 SSR motif type distribution. The most abundant SSR motif was TA/AT

(COMT) and Hydroxycinnamoyl-Coenzyme A shikimate/quinatehydroxycinnamoyltransferase (HST) that are involved in secondary metabolite production were not high in the unripe berries in comparison with other developmental stages of the fruit (Figs. 8 and 9). CAD was observed to be very high in the flowers. CCR, HCT, C3H
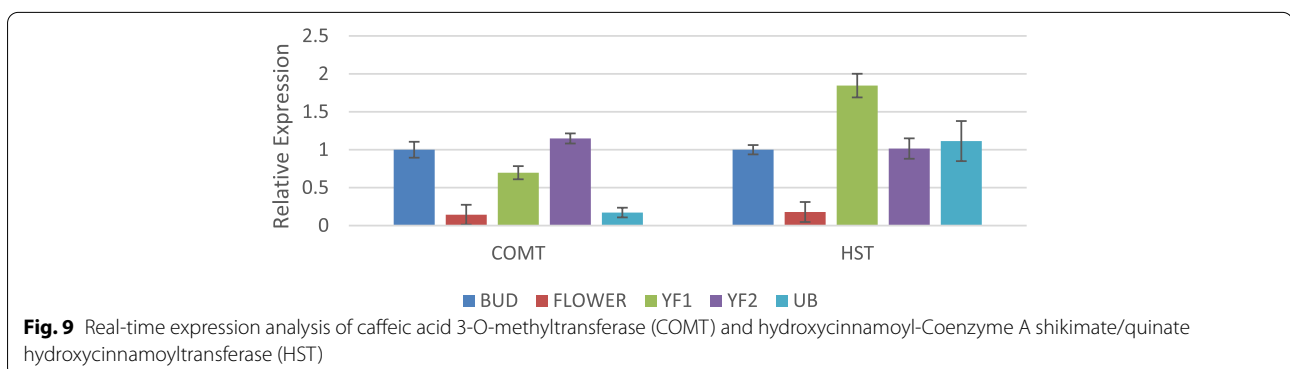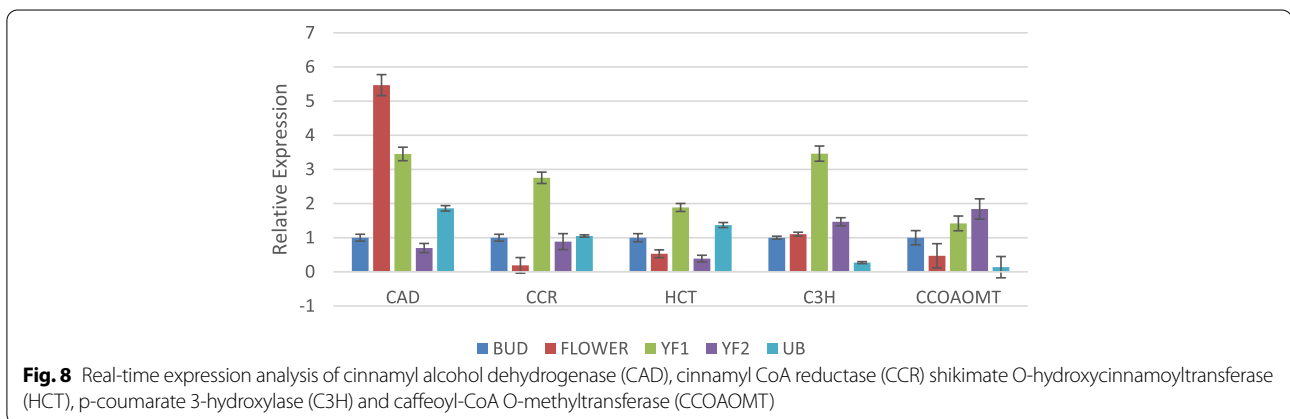
Fig. 7 Real-time expression analysis of PAL, C4H, 4CL and CHS. The expression level in unripe berry is too low to be seen near that of YF2 in most cases. High expression was observed in YF1 for PAL and C4H

and HST also showed abundance in young fruit stage 1. COMT and CCOAOMT were high in young fruit stage 2.

Expression analysis of genes encoding transcription factors

Based on the observation of gene expression, certain genes encoding transcription factors were also picked up from the transcriptome data that were associated with the phenylpropanoid pathway. The selected transcription factors included MYB1R1, R2R3MYB, WD40, WRKY2, WRKY33, and LIM. The repressing activity of MYB1R1 on PAL, C4H, HCT, C3H, COMT, CCOAOMT as well that of R2R3MYB on COMT and C4H was reported. WD40, WRKY2 and WRKY33 are known to be activators of CHS gene, promoter of C4H and defense-related genes respectively. LIM is a known transcriptional activator of PAL, 4CL and CAD [29–32]. The quantitative real-time expression analysis was performed for the genes encoding transcription factors. The expression level was analysed in comparison to bud, in all cases.

When we looked into the expression of transcription factors, we observed that R2R3MYB, LIM transcription factor and MYB1R1 were highly abundant in unripe berries. WD40 and WRKY33 showed high expression in bud, whereas WRKY2 in young fruit stage 1 (Figs. 10 and 11). The trend of expression of the phenylpropanoid pathway genes may be the reflection of the activity of the transcription factors. But in some cases, it was different from what was expected. For instance, LIM, which enhances the transcriptional activation of the genes PAL, 4CL and CAD, is high in UB compared to YF1. However, an increased expression of these genes was observed in YF1.
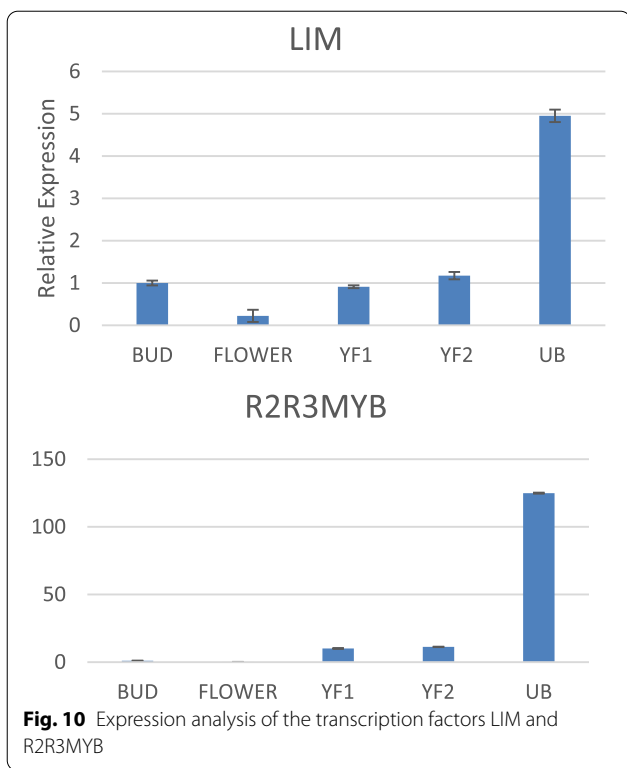

Fig. 8 Real-time expression analysis of cinnamyl alcohol dehydrogenase (CAD), cinnamyl CoA reductase (CCR) shikimate O-hydroxycinnamoyltransferase (HCT), p-coumarate 3-hydroxylase (C3H) and caffeoyl-CoA O-methyltransferase (CCOAOMT)

Fig. 9 Real-time expression analysis of caffeic acid 3-O-methyltransferase (COMT) and hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase (HST)

Fig. 10 Expression analysis of the transcription factors LIM and R2R3MYB

Discussion

The phenylpropanoid pathway is a complex network that has a set of important compounds undergoing modifications with the help of many enzymes. This results in a huge set of secondary metabolites spread over several branches [35, 36]. An enzyme can be responsible for synthesizing more than one metabolite that may be related to each other structurally [37].The complete information on the set of enzymes active at a particular stage can be uncovered using NGS technology. NGS technology helps in the identification of potential candidates that can be targeted to enhance the quality of fruits in *P. nigrum* [38].

When the fruits begin to develop, it is light green in colour and on maturation, colour darkens and the berry becomes harder [39]. The commercially available black pepper is prepared by a procedure involving hot water soaking and drying of the mature green unripe fruits. The high level of phenylpropanoids provides resistance to biotic and abiotic stress besides imparting aroma and spiciness [24]. Transcriptome data help to reveal the important enzymes and regulatory factors connected with the phenylpropanoid pathway in *P. nigrum* berries [40]. Several genes in a biosynthetic pathway can be regulated at the same time by over-expressing transcription factors through successful methods [41]. Transcription factors are reported to be promising candidates for crop improvement using transgenics [42].

A wide array of genes implicated in secondary metabolism in *P. nigrum* berries were identified from the gene ontology annotation. PAL is the first enzyme in phenylpropanoid pathway [43]. CCoAOMT is a participant of monolignol biosynthetic process [44]. COMT, HCT, CCR and CAD are the key enzymes associated with the biosynthesis of lignin [18, 44–46]. The precursors of lignin are produced by the involvement of the enzymes C4H, 4CL and CAD [47]. C4H is necessary for the conversion of trans-cinnamic acid into p-coumaric acid. C4H also
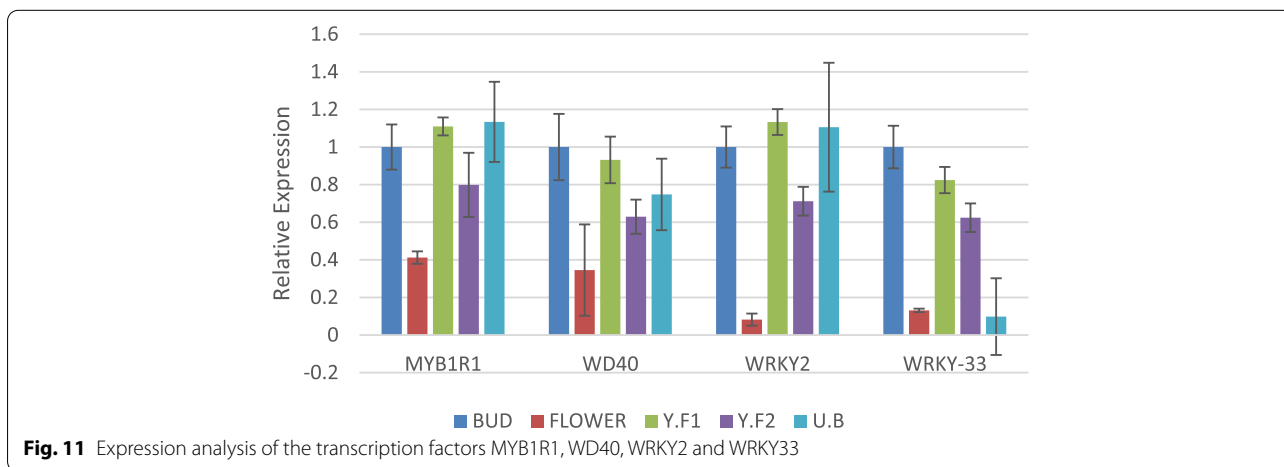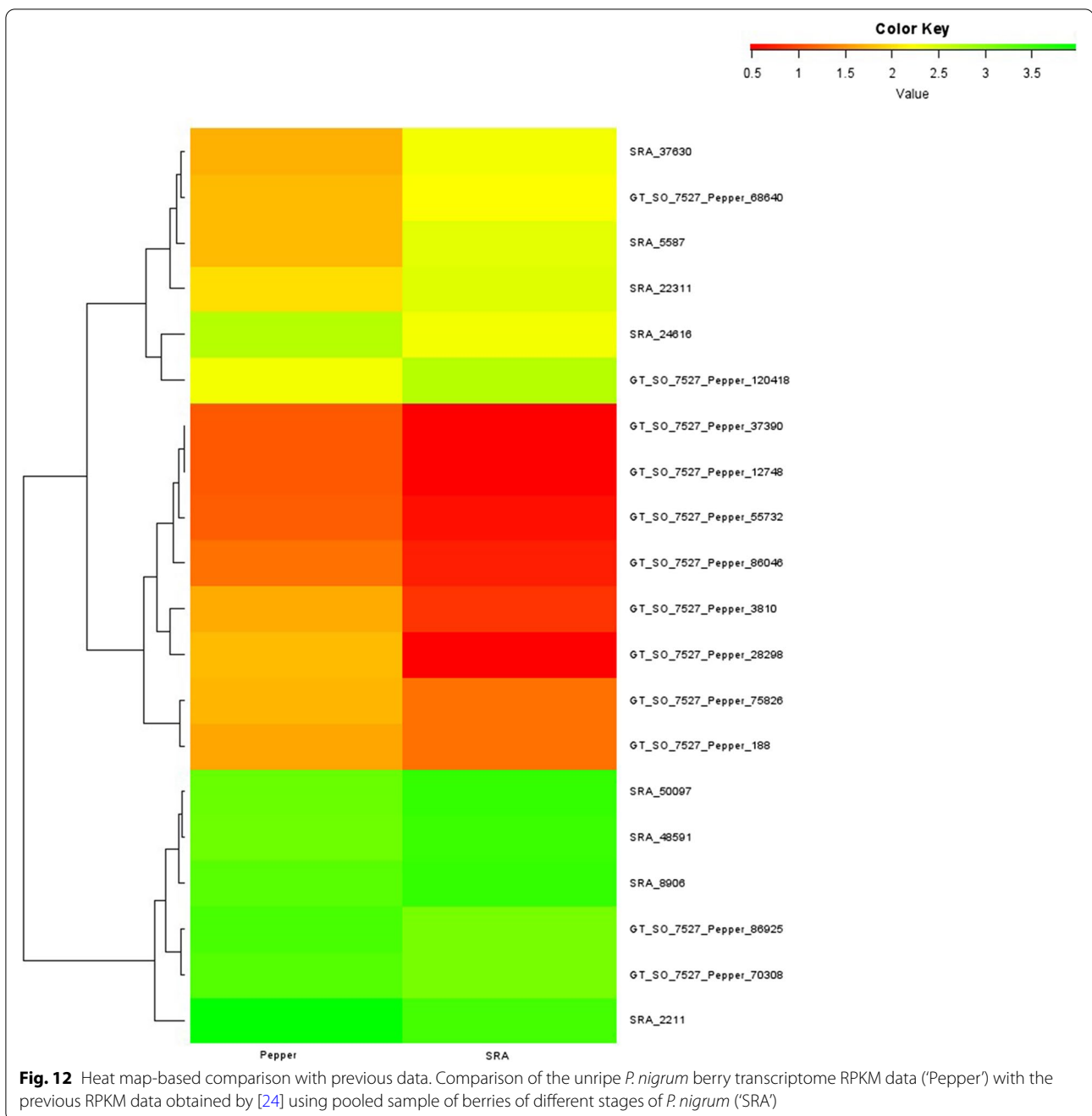
There are reports of previous transcriptome attempts in black pepper with variation in sample and other conditions [10, 23, 33, 34]. We compared the RPKM of selected genes from our data with that of [24] retrieved from NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/Traces/sra/; accession number SRS856941) in which they used a mixture of RNA from different stages of fruit (1–10 months post anthesis). None of the selected genes showed significant variation in RPKM (more than 2 fold) between the sample of the current study and that of the previous one (Fig. 12).

Fig. 11 Expression analysis of the transcription factors MYB1R1, WD40, WRKY2 and WRKY33

Fig. 12 Heat map-based comparison with previous data. Comparison of the unripe *P. nigrum* berry transcriptome RPKM data ('Pepper') with the previous RPKM data obtained by [24] using pooled sample of berries of different stages of *P. nigrum* ('SRA')

supervises the carbon flux for phytoalexins formed during stress conditions [48]. It is a key enzyme linked to the synthesis of many phytochemicals since p-coumarate serves as a common intermediate in several pathways [49]. 4CL, the last enzyme of phenylpropanoid metabolism, converts p-coumarate into its precursor ester to lead to the formation of lignins and flavonoids [35].

The monomers of lignin are called monolignols, a set of significant precursors formed through the phenylpropanoid pathway related to development and defense [35]. The branching point in the phenylpropanoid pathway into flavonoid pathway and monolignol biosynthetic pathway starts at the metabolite p-coumaroyl CoA. This compound can be a substrate for CHS resulting in the beginning of flavonoid pathway or for HCT, leading to monolignol synthesis [50]. HCT gene converts p-coumaroyl CoA into caffeoyl CoA and feruloyl CoA [51]. Negative regulation of one or a set of genes in the

Sreekumar *et al. BMC Plant Biology*   (2022) 22:501

Page 11 of 14

lignin biosynthetic branch can cause variation in the metabolic flux due to the lowered carbon flow towards the lignin pathway. Thus it depends on whether the key compound p-coumaroyl CoA serves as a substrate to HCT or CHS to determine the direction of the metabolic flux towards the monolignol or flavonoid branches of the phenylpropanoid pathway [50]. Any change in the regulation of a particular branch of the phenylpropanoid pathway is expected to immensely affect the other branches and their corresponding substrates [36]. Using metabolic engineering, it is possible to accumulate valuable metabolites by the extension of biosynthetic branches or addition of novel branches [52]. Better accumulation of desirable metabolites in *P. nigrum* fruit for industrial and pharmaceutical purposes can be obtained by hindering the competitive branches of biosynthesis [53].

Gene expression analysis helps to confirm the results obtained from RNA-Seq data [38]. The expression of selected genes was analysed in different developmental stages using qRT-PCR experiment. The expression pattern varied considerably in the different organs of *P. nigrum* i.e., bud, flower, young fruits stage 1 and 2 as well as unripe mature berries. In comparison with the bud stage, the expression of genes belonging to the phenylpropanoid pathway was lower in unripe berries in the majority of cases.

It is evident that the expression level of CHS is lower than HCT in the different developmental stages. This is in accordance with the trend in the transcriptomic data which showed the higher expression of monolignol biosynthetic genes in comparison with flavonoid pathway genes. It is to be noted that the most prominent pathway was identified as the phenylpropanoid pathway. Flavonoid pathway was only the 4th among the most active pathways in *P. nigrum*.

The qRT-PCR results showed that it was feasible to select a particular developmental stage for the metabolic engineering strategy to accumulate a specific bioactive compound. Thus the conventionally used developmental stages of fruits are not necessarily the only stages that have bioactivity. The other stages can also be considered for the isolation of specific bioactive compounds based on further confirmative studies in future.

Contrasting trends observed in gene expression may be a reflection of differential control of transcription [54]. To further look into this, we explored the putative transcription factors associated with the phenylpropanoid pathway in *P.nigrum*. Phenylpropanoid metabolism is reported to be regulated by highly conserved transcription factors like MYB and WD40 [55, 56]. In plants, MYB is the largest family of transcription factors with a MYB domain that is conserved in nature [57]. MYB regulates C4H and caffeic acid 3-O-methyltransferase (COMT)

genes [44]. MYB transcription factors show repressing activity to the expression of cinnamate 4-hydroxylase (C4H) gene by directly regulating corepressors that aid in the binding of MYB protein with the promoter of C4H gene [58]. If MYB transcription factor is repressed, it may direct the pathway towards lignin biosynthesis [59].

As per previous reports, the downregulation of CCoAOMT in the MYB-over expressed condition results in low lignin production in tobacco [60]. VvMYB4a and VvMYB4b of *Vitis vinifera* may negatively regulate the synthesis of low molecular weight phenolic compounds by inhibiting certain genes involved in the phenylpropanoid biosynthesis [61]. MYB/bHLH/WD-repeat (MBW) family genes play significant role in the flavonoid pathway in Arabidopsis [44].

The regulation of phenylpropanoid pathway by R2R3-MYB transcription factors was evidently demonstrated [62]. Overexpression of R2R3-MYB transcription factor has adversely affected several genes in the phenylpropanoid pathway that reduced the accumulation of lignin [63]. These transcription factors repress flavonoid and phenylpropanoid pathway genes that help in the production of anthocyanins and hydroxycinnamic acid esters [64]. R2R3-MYB transcription factor functions in the regulation of flavonoid pathway and monolignol pathway [57, 65, 66], whereas it shows repressing activity on the phenylpropanoid enzymes COMT and C4H [30–32].

WD40 (also known as the WD or beta-transducin repeat) is implicated in the accumulation of phenylpropanoids in *Solanum tuberosum* [67]. WRKY transcription factors contain distinctive WRKY domains exclusive to plants and play indispensable role in defense [68]. The expression analysis of genes encoding WRKY transcription factors in response to abiotic stress was studied in different tissues of pearl millet, providing underpinning data for their application in improving crops [69, 70]. LIM helps in increasing lignin production by enhancing the activity of phenylalanine ammonia-lyase, 4-coumarate-CoA ligase and cinnamyl alcohol dehydrogenase genes of the phenylpropanoid pathway [71]. The comparatively lower expression of the genes observed in the real-time expression analysis may be a reflection of the activity of the transcription factors. However, in addition to transcription factors, post-transcriptional modifications also play an important role in deciding metabolic shifts [54]. Contradicting expression trends of some genes regulated by the transcription factors suggest the involvement of additional regulatory factors which needs further experimental evidence.

There are reports of previous transcriptome attempts in black pepper with variation in sample and other conditions [10, 23, 33, 34]. The comparison of the results obtained in this study needs to be confirmed using the

Sreekumar *et al. BMC Plant Biology*   (2022) 22:501

Page 12 of 14

recently published reference genome of black pepper [72]. However, we compared the expression of selected genes from our data with that of [24] with the help of a heat map (Fig. 12). They used pooled sample of different stages of fruit (1 to 10 months after flowering) from *P. nigrum* cv. Reyin 1 whereas we used the mature unripe berries alone from which black pepper is produced commercially. Though there was a general variation in the intensity, a similar trend of expression of genes and transcription factors was observed. Altered source–sink communications are assumed to have resulted in the observed trend of gene expression [54].

Conclusion

The *de novo* transcriptome profile of unripe *P. nigrum* berries brought out the prominent secondary metabolic pathways, of which the highest expressed genes belonged to the phenylpropanoid pathway. The differential expression of these genes and significant transcription factors regulating the pathway were analyzed in the various developmental stages from bud to the development of the green berries from which black pepper is prepared for commercial applications. The expression of the selected transcripts was compared with previously published transcriptome data of a pooled sample of *P. nigrum* fruits of different developmental stages. The significant information obtained from the study can be beneficial for the commercial production of natural pharmaceuticals as well as for metabolic engineering studies that aim for black pepper berries enriched in a particular secondary metabolite group.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-022-03878-1.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Additional file 8.

Additional file 9.

Authors' contributions
S.E.V and SS conceived and designed the experiments. SS, D.K. and NJ performed the experiments and analysis. SS and S.E.V wrote the article with contributions from all the authors. All authors have read and agreed to the final manuscript.

Availability of data and materials
"The datasets generated and/or analysed during the current study are available in the [NCBI SRANAME] repository, [PER PRJNA873239 and PRJNA-872302SISTENT WEB LINK OR ACCESSION NUMBER TO DATASETS]".

Declarations

Ethics approval and consent to participate
Experimental research on plants complies with relevant institutional, national, and international guidelines and legislation.

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Transdisciplinary Biology, Rajiv Gandhi Centre for Biotechnology (RGCB), Thiruvananthapuram, Kerala, India. [2]Research Centre, University of Kerala, Thiruvananthapuram, Kerala, India. [3]Biology Centre, Czech Academy of Sciences, Institute of Plant Molecular Biology, České Budějovice, Czech Republic. [4]Centre for Gene Regulation & Expression, School of Life Sciences, University of Dundee, Dundee, Scotland.

References

1. Salehi B, Zakaria ZA, Gyawali R, Ibrahim SA, Rajkovic J, Shinwari ZK, Khan T, Sharifi-Rad J, Ozleyen A, Turkdonmez E, Valussi M. Piper species: a comprehensive review on their phytochemistry, biological activities and applications. Molecules. 2019;24(7):1364.
2. Srinivasan K. Black pepper (*Piper nigrum*) and its bioactive compound, piperine. InMolecular targets and therapeutic uses of spices: Modern uses for ancient medicine. 2009. p. 25–64.
3. Thangaselvabal T, Gailce Leo Justin C, Leelamathi M. Black pepper (*Piper nigrum* L.)'the king of spices'–A review. Agric Rev. 2008;29(2):89–98.
4. Gorgani L, Mohammadi M, Najafpour GD, Nikzad M. Piperine—the bioactive compound of black pepper: from isolation to medicinal formulations. Compr Rev Food Sci Food Saf. 2017;16(1):124–40.
5. Al-Baghdadi OB, Prater NI, Van der Schyf CJ, Geldenhuys WJ. Inhibition of monoamine oxidase by derivatives of piperine, an alkaloid from the pepper plant *Piper nigrum*, for possible use in Parkinson's disease. Bioorg Med Chem Lett. 2012;22(23):7183–8.
6. Stojanović-Radić Z, Pejčić M, Dimitrijević M, Aleksić A, V Anil Kumar N, Salehi B, C Cho W, Sharifi-Rad J. Piperine-A Major Principle of Black Pepper: a review of its bioactivity and studies. Appl Sci. 2019;9(20):4270.
7. Zheng J, Zhou Y, Li Y, Xu DP, Li S, Li HB. Spices for prevention and treatment of cancers. Nutrients. 2016;8(8):495.
8. Damanhouri ZA, Ahmad A. A review on therapeutic potential of *Piper nigrum* L. Black Pepper): The King of Spices. Med Aromat Plants. 2014;3(3):161.
9. Kakarala M, Brenner DE, Korkaya H, Cheng C, Tazi K, Ginestier C, Liu S, Dontu G, Wicha MS. Targeting breast stem cells with the cancer preventive compounds curcumin and piperine. Breast Cancer Res Treat. 2010;122(3):777–85.
10. Joy N, Asha S, Mallika V, Soniya EV. *De novo* transcriptome sequencing reveals a considerable bias in the incidence of simple sequence repeats towards the downstream of 'pre-miRNAs' of black pepper. PLoS One. 2013;8(3):e56694.
11. Facchini PJ, Bohlmann J, Covello PS, De Luca V, Mahadevan R, Page JE, Ro DK, Sensen CW, Storms R, Martin VJ. Synthetic biosystems for the production of high-value plant metabolites. Trends Biotechnol. 2012;30(3):127–31.

Sreekumar *et al. BMC Plant Biology*   (2022) 22:501

Page 13 of 14

12. Gu L, Zhang ZY, Quan H, Li MJ, Zhao FY, Xu YJ, Liu J, Sai M, Zheng WL, Lan XZ. Integrated analysis of transcriptomic and metabolomic data reveals critical metabolic pathways involved in rotenoid biosynthesis in the medicinal plant *Mirabilis himalaica*. Mol Genet Genomics. 2018;293(3):635–47.

13. Savoi S, Wong DC, Arapitsas P, Miculan M, Bucchetti B, Peterlunger E, Fait A, Mattivi F, Castellarin SD. Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (*Vitis vinifera* L.). BMC Plant Biol. 2016;16(1):1–7.

14. Morozova O, Hirst M, Marra MA. Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet. 2009;22(10):135–51.

15. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):10–2.

16. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011;29(7):644.

17. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.

18. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2.

19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

20. The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480–9. https://doi.org/10.1093/nar/gkaa1100.

21. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35(suppl_2):W182–5.

22. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods. 2001;25(4):402–8.

23. Asha S, Sreekumar S, Soniya EV. Unravelling the complexity of microRNA-mediated gene regulation in black pepper (*Piper nigrum* L.) using high-throughput small RNA profiling. Plant Cell Rep. 2016;35(1):53–63.

24. Hu L, Hao C, Fan R, Wu B, Tan L, Wu H. De novo assembly and characterization of fruit transcriptome in black pepper (*Piper nigrum*). PLoS One. 2015;10(6):e0129822.

25. Wang C, Xu N, Cui S. Comparative transcriptome analysis of roots, stems, and leaves of *Pueraria lobata* (Willd.) Ohwi: identification of genes involved in isoflavonoid biosynthesis. PeerJ. 2021;9:e10885.

26. Hübner S, Hehmann M, Schreiner S, Martens S, Lukačin R, Matern U. Functional expression of cinnamate 4-hydroxylase from Ammi majus L. Phytochemistry. 2003;64(2):445–52.

27. Pascual MB, El-Azaz J, de la Torre FN, Cañas RA, Avila C, Cánovas FM. Biosynthesis and metabolic fate of phenylalanine in conifers. Front Plant Sci. 2016;13(7):1030.

28. Gomez-Cano L, Gomez-Cano F, Dillon FM, Alers-Velazquez R, Doseff AI, Grotewold E, Gray J. Discovery of modules involved in the biosynthesis and regulation of maize phenolic compounds. Plant Sci. 2020;1(291):110364.

29. MacDonald MJ, D'Cunha GB. A modern view of phenylalanine ammonia lyase. Biochem Cell Biol. 2007;85(3):273–82.

30. Weisshaar B, Jenkins GI. Phenylpropanoid biosynthesis and its regulation. Curr Opin Plant Biol. 1998;1(3):251–7.

31. Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MS, Wang L. The phenylpropanoid pathway and plant defence—a genomics perspective. Mol Plant Pathol. 2002;3(5):371–90.

32. Guillaumie S, Mzid R, Méchin V, Léon C, Hichri I, Destrac-Irvine A, Trossat-Magnin C, Delrot S, Lauvergeat V. The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco. Plant Mol Biol. 2010;72(1):215–34.

33. Yang L, Zhao X, Ran L, Li C, Fan D, Luo K. PtoMYB156 is involved in negative regulation of phenylpropanoid metabolism and secondary cell wall biosynthesis during wood formation in poplar. Sci Rep. 2017;7(1):1–4.

34. Hao C, Xia Z, Fan R, Tan L, Hu L, Wu B, Wu H. De novo transcriptome sequencing of black pepper (*Piper nigrum* L.) and an analysis of genes involved in phenylpropanoid metabolism in response to *Phytophthora capsici*. BMC Genomics. 2016;17(1):1–4.

35. Vogt T. Phenylpropanoid biosynthesis. Mol Plant. 2010;3(1):2.

36. Qiu J, Gao F, Shen G, Li C, Han X, Zhao Q, Zhao D, Hua X, Pang Y. Metabolic engineering of the phenylpropanoid pathway enhances the antioxidant capacity of *Saussurea involucrata*. PLoS One. 2013;8(8):e70665.

37. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007;26(1):51–78.

38. Khew CY, Harikrishna JA, Wee WY, Lau ET, Hwang SS. Transcriptional sequencing and gene expression analysis of various genes in fruit development of three different black pepper (*Piper nigrum* L.) varieties. Int J Genomics. 2020;2020:1540915.

39. Joy N, Abraham Z, Soniya EV. A preliminary assessment of genetic relationships among agronomically important cultivars of black pepper. BMC Genet. 2007;8(1):1–7.

40. Bhattacharyya D, Sinha R, Hazra S, Datta R, Chattopadhyay S. De novo transcriptome analysis using 454 pyrosequencing of the Himalayan Mayapple, *Podophyllum hexandrum*. BMC Genomics. 2013;14(1):1–3.

41. Farré G, Blancquaert D, Capell T, Van Der Straeten D, Christou P, Zhu C. Engineering complex metabolic pathways in plants. Annu Rev Plant Biol. 2014;29(65):187–223.

42. Century K, Reuber TL, Ratcliffe OJ. Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. Plant Physiol. 2008;147(1):20–9.

43. Higuchi T. Biochemistry and molecular biology of wood. Berlin and Heidelberg: Springer-Verlag and GmbH & Co. KG; 2011.

44. Zhu H, Xia R, Zhao B, An YQ, Dardick CD, Callahan AM, Liu Z. Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. BMC Plant Biol. 2012;12(1):1–8.

45. Humphreys JM, Chapple C. Rewriting the lignin roadmap. Curr Opin Plant Biol. 2002;5(3):224–9.

46. Hamberger B, Ellis M, Friedmann M, de Azevedo Souza C, Barbazuk B, Douglas CJ. Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the Populus lignin toolbox and conservation and diversification of angiosperm gene families. Botany. 2007;85(12):1182–201.

47. Labeeuw L, Martone PT, Boucher Y, Case RJ. Ancient origin of the biosynthesis of lignin precursors. Biol Direct. 2015;10(1):1–21.

48. Xu H, Park NI, Li X, Kim YK, Lee SY, Park SU. Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in *Scutellaria baicalensis*. Biores Technol. 2010;101(24):9715–22.

49. Croteau R, Kutchan TM, Lewis NG. Natural products (secondary metabolites). Biochem Mol Biol Plants. 2000;24:1250–319.

50. Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu C, Pu Y, Rudis MR, Tang Y, Xiao X, Jackson L, Li G. A genomics approach to deciphering lignin biosynthesis in switchgrass. Plant Cell. 2013;25(11):4342–61.

51. Wu J, Wang D, Liu Y, Wang L, Qiao X, Zhang S. Identification of miRNAs involved in pear fruit development and quality. BMC Genomics. 2014;15(1):1–9.

52. Hidalgo D, Georgiev M, Marchev A, Bru-Martínez R, Cusido RM, Corchete P, Palazon J. Tailoring tobacco hairy root metabolism for the production of stilbenes. Sci Rep. 2017;7(1):1–1.

53. Tiwari M, Sharma D, Trivedi PK. Artificial microRNA mediated gene silencing in plants: progress and perspectives. Plant Mol Biol. 2014;86(1):1–8.

54. Ibánez AM, Martinelli F, Reagan RL, Uratsu SL, Vo A, Tinoco MA, Phu ML, Chen Y, Rocke DM, Dandekar AM. Transcriptome and metabolome analysis of citrus fruit to elucidate puffing disorder. Plant Sci. 2014;1(217):87–98.

55. Grotewold E. The genetics and biochemistry of floral pigments. Annu Rev Plant Biol. 2006;2(57):761–80.

56. Xie DY, Sharma SB, Wright E, Wang ZY, Dixon RA. Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. Plant J. 2006;45(6):895–907.

57. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15(10):573–81.

58. Zhou M, Zhang K, Sun Z, Yan M, Chen C, Zhang X, Tang Y, Wu Y. LNK1 and LNK2 corepressors interact with the MYB3 transcription factor in phenylpropanoid biosynthesis. Plant Physiol. 2017;174(3):1348–58.

59. Lin JS, Lin CC, Lin HH, Chen YC, Jeng ST. Micro R 828 regulates lignin and H 2 O 2 accumulation in sweet potato on wounding. New Phytol. 2012;196(2):427–40.

60. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O. Widespread translational inhibition by plant miRNAs and siRNAs. Science. 2008;320(5880):1185–90.

61. Cavallini E, Matus JT, Finezzo L, Zenoni S, Loyola R, Guzzo F, Schlechter R, Ageorges A, Arce-Johnson P, Tornielli GB. The phenylpropanoid pathway is controlled at different branches by a set of R2R3-MYB C2 repressors in grapevine. Plant Physiol. 2015;167(4):1448–70.

62. Flores FP, Singh RK, Kerr WL, Pegg RB, Kong F. Antioxidant and enzyme inhibitory activities of blueberry anthocyanins prepared using different solvents. J Agric Food Chem. 2013;61(18):4441–7.

63. Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DG, Wang H, Jackson L, Tang Y, Neal Stewart Jr C, Chen F. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. New Phytol. 2012;193(1):121–36.

64. Ma D, Reichelt M, Yoshida K, Gershenzon J, Constabel CP. Two R2R3-MYB proteins are broad repressors of flavonoid and phenylpropanoid metabolism in poplar. Plant J. 2018;96(5):949–65.

65. Jin H, Martin C. Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol. 1999;41(5):577–85.

66. Chen L, Xin X, Yuan Q, Su D, Liu W. Phytochemical properties and antioxidant capacities of various colored berries. J Sci Food Agric. 2014;94(2):180–8.

67. Payyavula RS, Singh RK, Navarre DA. Transcription factors, sucrose, and sucrose metabolic genes interact to regulate potato phenylpropanoid metabolism. J Exp Bot. 2013;64(16):5115–31.

68. Eulgem T, Somssich IE. Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol. 2007;10(4):366–71.

69. Tuteja N, Gill SS, Tiburcio AF, Tuteja R, editors. Improving crop resistance to abiotic stress. U.S.: Wiley; 2012.

70. Chanwala J, Satpati S, Dixit A, Parida A, Giri MK, Dey N. Genome-wide identification and expression analysis of WRKY transcription factors in pearl millet (*Pennisetum glaucum*) under dehydration and salinity stress. BMC Genomics. 2020;21(1):1–6.

71. Kawaoka A, Kaothien P, Yoshida K, Endo S, Yamada K, Ebinuma H. Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis. Plant J. 2000;22(4):289–301.

72. Hu L, Xu Z, Wang M, Fan R, Yuan D, Wu B, Wu H, Qin X, Yan L, Tan L, Sim S. The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. Nat Commun. 2019;10(1):1–1.

Publisher's Note