



OPEN

## Transcriptional analysis reveals key insights into seasonal induced anthocyanin degradation and leaf color transition in purple tea (*Camellia sinensis* (L.) O. Kuntze)

Tony Kipkoech Maritim<sup>1,2,3</sup>, Mamta Masand<sup>1,2</sup>, Romit Seth<sup>1</sup> & Ram Kumar Sharma<sup>1,2</sup>✉

Purple-tea, an anthocyanin rich cultivar has recently gained popularity due to its health benefits and captivating leaf appearance. However, the sustainability of purple pigmentation and anthocyanin content during production period is hampered by seasonal variation. To understand seasonal dependent anthocyanin pigmentation in purple tea, global transcriptional and anthocyanin profiling was carried out in tea shoots with two leaves and a bud harvested during in early (reddish purple: S1\_RP), main (dark gray purple: S2\_GP) and backend flush (moderately olive green: S3\_G) seasons. Of the three seasons, maximum accumulation of total anthocyanin content was recorded in S2\_GP, while least amount was recorded during S3\_G. Reference based transcriptome assembly of 412 million quality reads resulted into 71,349 non-redundant transcripts with 6081 significant differentially expressed genes. Interestingly, key DEGs involved in anthocyanin biosynthesis [PAL, 4CL, F3H, DFR and UGT/UGT], vacuolar trafficking [ABC, MATE and GST] transcriptional regulation [MYB, NAC, bHLH, WRKY and HMG] and Abscisic acid signaling pathway [PYL and PP2C] were significantly upregulated in S2\_GP. Conversely, DEGs associated with anthocyanin degradation [Prx and lac], repressor TFs and key components of auxin and ethylene signaling pathways [ARF, AUX/IAA/SAUR, ETR, ERF, EBF1/2] exhibited significant upregulation in S3\_G, correlating positively with reduced anthocyanin content and purple coloration. The present study for the first-time elucidated genome-wide transcriptional insights and hypothesized the involvement of anthocyanin biosynthesis activators/repressor and anthocyanin degrading genes via peroxidases and laccases during seasonal induced leaf color transition in purple tea. Futuristically, key candidate gene(s) identified here can be used for genetic engineering and molecular breeding of seasonal independent anthocyanin-rich tea cultivars.

Tea, (*Camellia sinensis* (L.) O. Kuntze), is an economically important plantation crop rich in flavor enhancing and health beneficial metabolites that also confer resistance to biotic and abiotic stresses<sup>1</sup>. Most of the elite tea cultivars grown worldwide are suitable for processing of the three popular types of commercial teas, viz; non-fermented (green tea), semi-fermented (oolong tea), and completely fermented (black tea)<sup>2</sup>. Global production of black and green tea is consistently increasing and projected to reach 4.42 and 3.31 million tons, respectively by 2027<sup>3</sup>. Despite the significant increase in global production, tea prices have stagnated or declined causing reduced profit margins and returns to the farmers<sup>4</sup>. Therefore, diversification of tea products can be an alternative way-out to raise demand and shore up market prices. Nevertheless, diversification is greatly influenced by availability of suitable germplasm, wherein, anthocyanin-rich tea cultivars due to their pleasant taste and rich health beneficial compounds are getting global attention for production of “Specialty Tea” that can fetch premium prices to the farmers. Through breeding (natural or artificial hybridization) followed by clonal field selection, anthocyanin-rich tea cultivars (purple tea) have been developed for product diversification<sup>5,6</sup>. Anthocyanin, a water-soluble

<sup>1</sup>Department of Biotechnology, CSIR-Institute of Himalayan Bioresource Technology, P.O. Box No. 6, Palampur, HP 176061, India. <sup>2</sup>Academy of Scientific and Innovative Research (AcSIR), CSIR-HRDC Campus, Ghaziabad, Uttar Pradesh 201 002, India. <sup>3</sup>Tea Breeding and Genetic Improvement Division, KALRO-Tea Research Institute, P.O. Box 820-20200, Kericho, Kenya. ✉email: rksharma.ihbt@gmail.com

secondary metabolite synthesized in the cytosol and localized in vacuoles, is an important plant pigment responsible for the red, violet, purple and blue pigmentations in different plant tissues<sup>7</sup>. Their accumulations also play an important role in reducing damage caused by UV irradiation, photoinhibition and oxidative stress and more so enhance pollination<sup>8</sup>. Interestingly, various animal model studies have reported several health benefits including antioxidant, antiaging, anticancer and antidiabetic properties of anthocyanin-rich tea<sup>9,10</sup>. Therefore, anthocyanin-dependent purple coloration can be a potential morphological marker for breeding and selection of quality tea cultivars. Nevertheless, rapid degradation limits optimum utilization of anthocyanin-rich cultivars for production of specialty tea at commercial scale<sup>11,12</sup>. Therefore, understanding the molecular basis of anthocyanin degradation and leaf color transition is important for its genetic improvement. Previous metabolomic studies attributed the dynamic balance and ratio of different anthocyanin molecules, catechins and other plant pigments (chlorophyll and carotenoids) to variation in purple leaf coloration in tea<sup>13,14</sup>. Earlier molecular studies have also identified key genes/proteins and various transcription factors involved in anthocyanin biosynthesis in tea<sup>15,16</sup>. Likewise, few previous studies focused on developmental transition associated flavonoid/anthocyanin accumulation and leaf color change to attainment of maturity<sup>17–20</sup>. Additionally, recent studies in various plant species have revealed the involvement anthocyanin degradation enzymes [beta-glucosidases, polyphenol oxidase (PPO), class III peroxidases (POD) and laccases], and well-coordinated plant hormone signaling pathways in regulation of anthocyanin accumulation<sup>21,22</sup>. Interestingly, the role of PPO, POD and beta-glucosidases in oxidation of flavonol glycosides responsible for bitter-taste and astringency in tea was recently reported<sup>23</sup>. However, their role in anthocyanin accumulation in tea is not known. Moreover, the mechanism controlling seasonal induced anthocyanin degradation and purple leaf color transition remains elusive in tea. Therefore, transcriptome profiling of anthocyanin-rich tea during different seasons can provide deeper understanding of the molecular mechanism controlling seasonal induced leaf color transition in purple tea.

Annual tea production in India is broadly categorized into early flush (April to mid-June: S1), main flush (mid-June to mid-September: S2) and backend flush (mid-September to October: S3) seasons with availability of the best quality tea during S1<sup>24</sup>. Despite recording better yields, the main flush (S2) is compromised in terms of quality. Therefore, development and commercialization of “anthocyanin-rich tea” can be the potential alternative for increasing profitability of the tea sector. Interestingly, long-term field evaluation of anthocyanin-rich tea cultivars reveals purple leaf color transitions from reddish purple (RP) during S1 to dark gray purple (GP) during S2 and, subsequently turn moderately olive green (G) during S3.

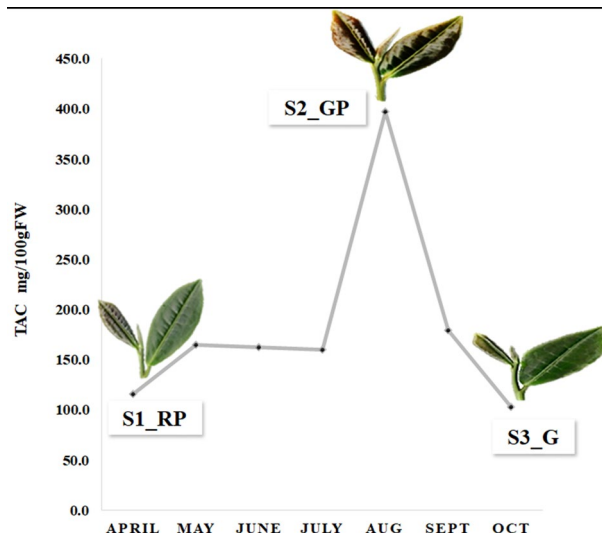
In the present study, seasonal dependent total anthocyanin accumulation (TAC) vis-à-vis global transcriptional analysis of anthocyanin-rich tea during early (S1\_RP), main (S2\_GP) and backend (S3\_G) flush seasons was performed to elucidate the dynamics of leaf color transitions in pigmented tea. Global gene expression analysis was performed for successful identification of key structural and regulatory genes involved in seasonal-induced biosynthesis/degradation of anthocyanin in purple tea. To the best of our knowledge, this is the first study to report on comprehensive elucidation of seasonal induced anthocyanin degradation, phytohormonal regulation and role of repressor TFs in anthocyanin accumulation in tea. This dataset forms a platform for advance understanding of molecular regulation of anthocyanin accumulation in different seasons and can be used futuristically for molecular breeding of seasonal independent anthocyanin-rich tea cultivars.

## Results

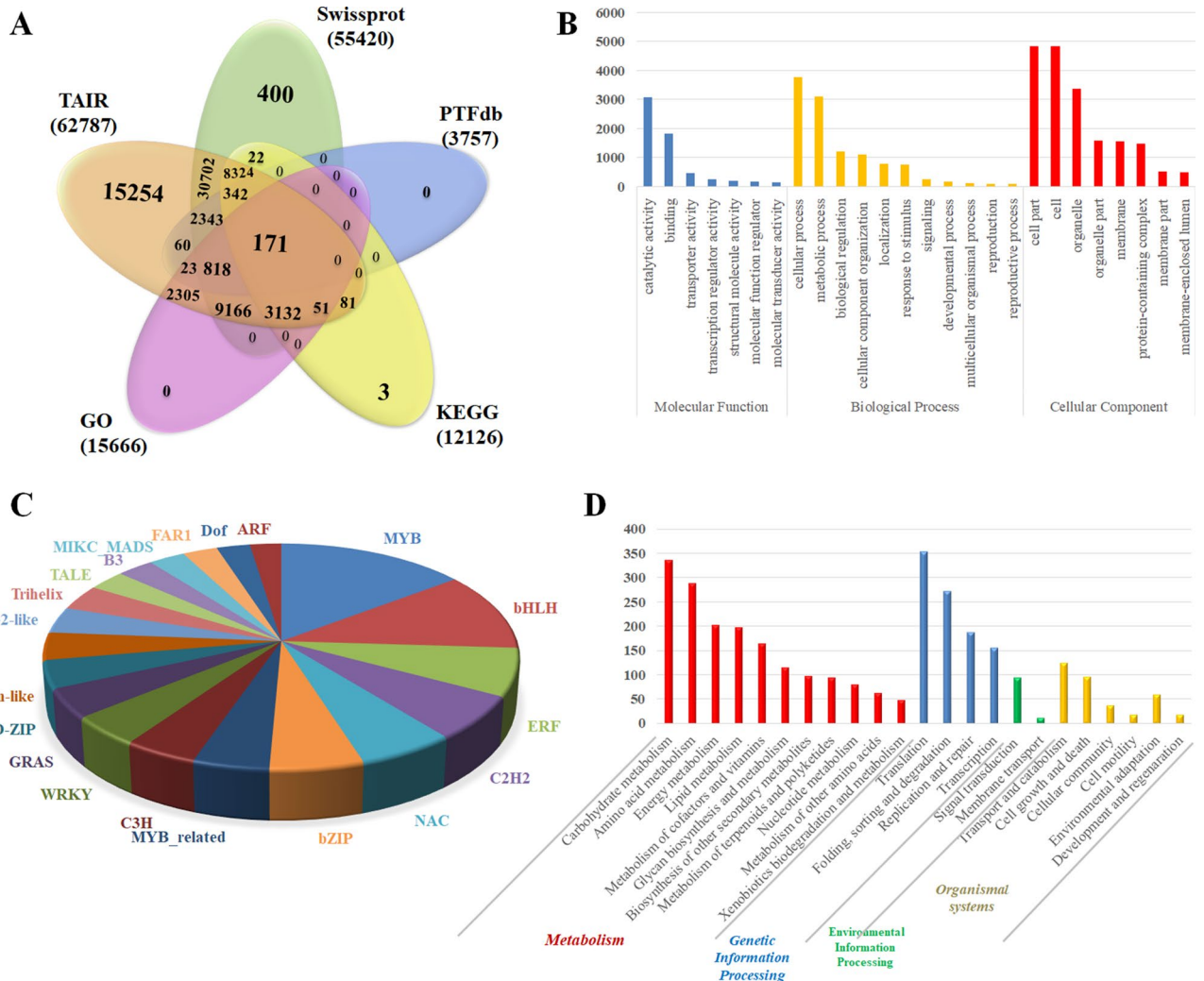
**Differential accumulation of total anthocyanin content (TAC).** To determine the variation in total anthocyanin dependent leaf color transitions in tea, spectrophotometric analysis of TAC was performed on samples collected monthly from April to October. Analysis of tea shoots with two leaves and a bud recorded clear difference in leaf color from reddish purple (RP) during April (S1\_RP) to dark-gray purple (GP) in August (S2\_GP) and moderately olive green (G) in October (S3\_G)]. Similarly, TAC recorded significant increase from 115.0 mg 100 g<sup>-1</sup> FW in April (S1\_RP) to 397.0 mg 100 g<sup>-1</sup> FW in August (S2\_GP) before declining to 103.0 mg 100 g<sup>-1</sup> FW in October (S3\_G) (Fig. 1).

**Transcriptome sequencing.** Transcriptome sequencing of three cDNA libraries representing reddish purple (S1\_RP), dark-gray purple (S2\_GP) and moderately olive green (S3\_G) leaf coloration yielded 412 million high quality reads with ~95% quality score. Reference based mapping of quality reads resulted into 88.4% (S1\_RP), 87.5% (S2\_GP) and 86.4% (S3\_G) of sample-wise reads that were successfully mapped to *Camellia sinensis* var. *sinensis* (CSS) genome<sup>25</sup>. Overall, 71,349 non-redundant transcripts with an average length of 1514.14 bp were identified in reference-based transcriptome assembly to the CSS<sup>25</sup>. Other assembly attributes such as N50 (1965 bp) and GC content (42.99%) provided high quality and quantity assurance of current RNA-seq data for downstream analysis (Supplementary Table S1).

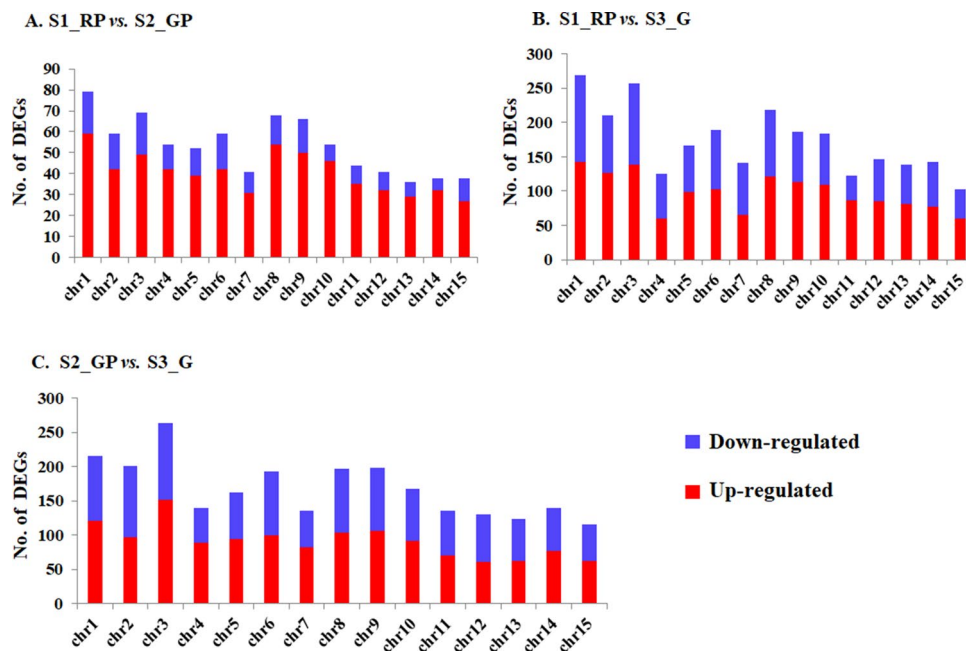
**Functional annotation.** To determine the putative functions of assembled transcripts, similarity search with multiple public databases using BLASTX algorithm (e-value cut-off ≤ 1e-5) annotated 62,788 (87%), 55,420 (77.7%), 12,126 (12%), transcripts in TAIR, SwissProt, and KEGG respectively (Fig. 2a). Gene ontology (GO) terms were assigned to 40,935 (57.4%) transcripts, classified into biological process (13,203; 32.2%), molecular function (13,019; 31.8%) and cellular component (14,713; 35.9%) (Fig. 2b). Furthermore, annotation with plant-TFdb identified 3757 transcripts factors representing 56 TF families with abundance of MYB (430), basic-helix-loop-helix class (330) and Ethylene response factors (ERF) (201) (Fig. 2c). NR transcripts successfully assigned to KEGG pathways were also classified into five functional groups with abundance of pathways involved in plant metabolism (Fig. 2d).



**Figure 1.** Leaf color transition and variation in total anthocyanin content.



**Figure 2.** (A) Venn diagram of annotated transcripts in multiple databases. (B) Gene ontology classification of transcripts. (C) Frequency distribution of transcription factors in transcriptome data. (D) Functional classification and pathway assignment of assembled transcripts by KEGG.



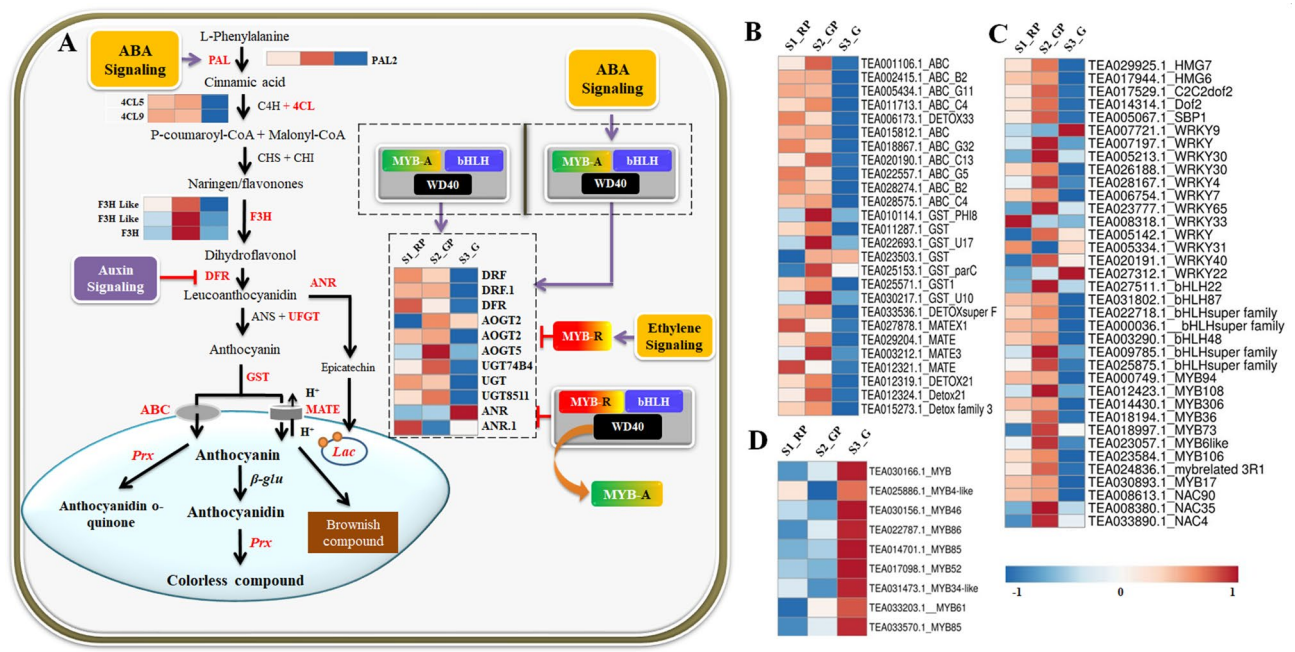
**Figure 3.** Differentially expressed genes identified by pairwise comparison and their chromosomal distribution in the CSS genome. Red and blue bars represent up-regulated and down-regulated DEGs, respectively.

**Differential gene expression analysis.** To unravel global insight of differentially expressed genes during leaf color transitions, pairwise comparisons between S1\_RP, S2\_GP and S3\_G identified ~6081 DEGs, assigned to different chromosomes (*Chr*) of the CSS genome<sup>26</sup> (Fig. 3). 818 DEGs including 626 upregulated and 192 downregulated identified in S1\_RP vs. S2\_GP comparison were mostly mapped to *Chr1* (79 DEGs), *Chr3* (69 DEGs), *Chr8* (68 DEGs) and 38 DEGs each to *Chr14* and *Chr15*. Similarly, S1\_RP vs. S3\_G comparison revealed 2668 DEGs (1506 upregulated and 1162 downregulated) mainly assigned to *Chr1* (269 DEGs), *Chr3* (257 DEGs), and least in *Chr11* (122 DEGs) and *Chr15* (103 DEGs). Likewise, 2595 DEGs (1421 upregulated and 1174 down-regulated) identified in S2\_GP vs. S3\_G comparison recorded maximal mapping to *Chr3* (263 DEGs), and least number of DEGs (115) to *Chr15*.

**GO and KEGG enrichment analysis.** GO enrichment analysis using Tea Plant Information Archive (<http://tpia.teaplant.org/>) identified 136 significantly ( $FDR \leq 0.05$ ) enriched GO terms in S1\_RP vs. S2\_GP (50), S1\_RP vs. S3\_G (36) and S2\_GP vs. S3\_G (50) comparisons (Supplementary Table S2). In the biological process category, unique enrichment of 'phenylpropanoid metabolic process', 'secondary metabolite biosynthesis', 'reactive oxygen species metabolic process' and 'lignin biosynthetic process' in S1\_RP vs. S2\_GP comparison suggest their role in anthocyanin biosynthesis. While, significant enrichment of 'oxidation-reduction process', 'response to oxidative stress' and 'circadian rhythm' in S2\_GP vs. S3\_G suggest possible role in anthocyanin degradation. Likewise, in the molecular function category 'binding', 'peroxidase activity' and 'antioxidant activity' recorded optimum enrichment during reddish purple to dark grey purple leaf color transitions (S1\_RP vs. S2\_GP), whereas, enrichment of 'oxidoreductase activity', 'secondary active transmembrane transporter activity' and 'microtubule binding' in S2\_GP vs. S3\_G, suggesting their possible role in anthocyanin degradation. Similarly, cellular component associated with cell wall and membrane including 'Apoplast' were enriched significantly in S1\_RP vs. S2\_GP, while enrichment of 'chloroplast thylakoid membrane', 'photosynthetic membrane' and 'plastid thylakoid membrane' in S2\_GP vs. S3\_G suggests the critical role of cell wall and membrane organelles in regulating anthocyanin accumulation in tea. Additionally, 'glucan metabolic process', 'aminoglycan metabolic process' and 'cell wall macromolecule metabolic process' (biological processes), 'catalytic', 'hydrolase', 'transferase' and 'oxidoreductase' activities (molecular functions) and 'extracellular region', 'cell wall' and 'extracellular region' (cellular component) were significantly enriched irrespective of the pairwise comparisons.

Furthermore, KEGG enrichment analysis identified 55 metabolic pathways including 11 in S1\_RP vs. S2\_GP; 27 in S1\_RP vs. S3\_G and 17 in S2\_GP vs. S3\_G, (Supplementary Table S3). Metabolic pathways, such as 'cyanoamino acid metabolism', 'diterpenoid biosynthesis', 'N-glycan biosynthesis' and 'oxidative phosphorylation', 'photosynthesis' and 'pyruvate metabolism' were significantly enriched in S1\_RP vs. S3\_G, and S2\_GP vs. S3\_G suggesting potential role in anthocyanin degradation, while, 'isoflavonoid biosynthesis' uniquely enriched in S1\_RP vs. S2\_GP may be linked to anthocyanin biosynthesis. Interestingly, important metabolic pathways like 'plant hormone signal transduction', 'phenylpropanoid biosynthesis', 'starch and sucrose metabolism', 'flavonoid biosynthesis', 'MAPK signaling' and 'Sesquiterpenoid and triterpenoid biosynthesis' also recorded significant enrichment irrespective of the pair-wise comparisons. Among these, key pathways such as 'plant hormone





**Figure 4.** (A) Anthocyanin biosynthesis pathway. Highlighted genes (red) represent differentially expressed genes of the pathway. Purple pointed arrows represent activation, curved right arrow represent displacement while red bar-headed arrows represent inhibition. (B) Heatmap of differentially expressed transporters. (C) Transcription factors associated with anthocyanin biosynthesis in tea. (D) Putative repressor MYBs. Blue and red colors of heatmaps represent upregulated and downregulated expression levels, respectively. The color scale reflects  $\log_2$  transformed FPKM values.

signal transduction, 'phenylpropanoid biosynthesis' and 'flavonoid biosynthesis' were considered for downstream analysis to elucidate anthocyanin biosynthesis and degradation in tea.

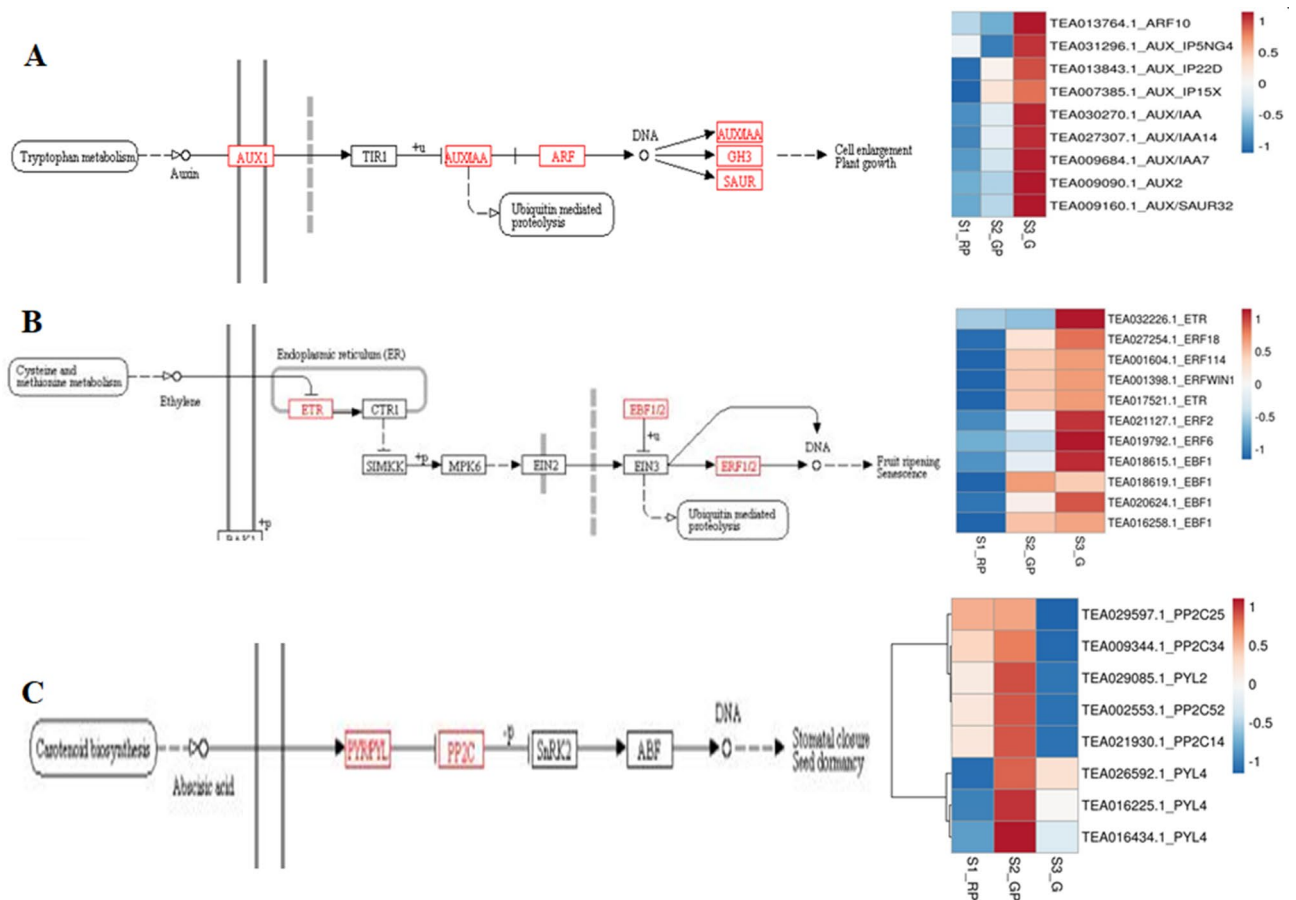
**Gene expression analysis of candidates related to anthocyanin biosynthesis.** Considering key role of phenylpropanoid/flavonoid pathway in anthocyanin biosynthesis, 20 transcripts encoding early biosynthetic genes (EBGs); phenylalanine ammonia-lyase (PAL), 4-coumarate-CoA ligase (4CL), flavanone 3-hydroxylase (F3H) and late-biosynthetic genes (LBGs); dihydroflavonol 4-reductase (DFR), anthocyanidin reductase (ANR) and anthocyanin 3'-O-beta-glucosyltransferase (AOGT), UDP-glycosyltransferase (UGT/UFGT) of the anthocyanin biosynthesis were differentially expressed (Supplementary Table S4; Fig. 4A). Their upregulation in S2\_GP and downregulation in S1\_RP and S3\_G, was strongly correlated to TAC (Supplementary Table S5). Interestingly, upregulated expression of anthocyanin reductase (ANR) reportedly involved in epicatechin biosynthesis<sup>27</sup> was negatively correlated with TAC during S3\_G indicating their important role in regulating metabolic flux towards anthocyanin biosynthesis in tea. Moreover, distribution of anthocyanin biosynthesis genes in 11 of 15 chromosomes (Fig. 5) including *Chr1*, *Chr2*, *Chr3* and *Chr10* successfully mapped with quality related QTLs, suggests close linkage between anthocyanin and other quality related parameters in tea<sup>26</sup>.

Furthermore, differential expression of 26 candidate genes involved in anthocyanin transport (Supplementary Table S4) including 10 ATP-binding cassette transporters (ABC), 9 multidrug and toxic extrusion transporter (MATE) and 7 glutathione S-transferase (GST), mostly upregulated in S2\_GP, suggests the critical role of transporters in regulation of anthocyanin biosynthesis (Fig. 4B). Interestingly, strong positive correlation between upregulated expression of key transporters and TAC (Supplementary Table S5) suggest transporter mediated accumulation of anthocyanin in tea. Most of these transporters were assigned to *Chr7* and *Chr11* of CSS genome<sup>25</sup>.

**Dynamic expression of regulatory genes.** Anthocyanin biosynthesis is greatly influenced by dynamic expression of transcription factors (TFs) and phytohormones.

**Transcription factors (TFs) mediated regulation of anthocyanin.** Transcription factors (TFs) are master-proteins controlling key biological processes such as, metabolism, growth, biotic and abiotic stress responses<sup>28</sup>. In this study, 46 differentially expressed (DE) TFs including MYB and MYB-related (18), WRKY (12) and bHLH (7), NAC (3), C2H2 (2), HMG (2) and SBP (2) were identified during leaf color transitions of purple tea (S1\_RP, S2\_GP, S3\_G) (Supplementary Table S6; Fig. 4C). TFs including MYB (MYB306), bHLH (bHLH3, bHLH48, bHLH87 and bHLH94) exhibited strong positive correlation (>0.99) with key genes of anthocyanin biosynthesis (4CL5, ANR and AOGT) and TAC (Supplementary Table S5). Furthermore, strong positive correlation (>0.9) between HMG TFs (HMG-B7, B8) with EBGs (PAL2, 4CL5) and LBGs (UGT7AB4),





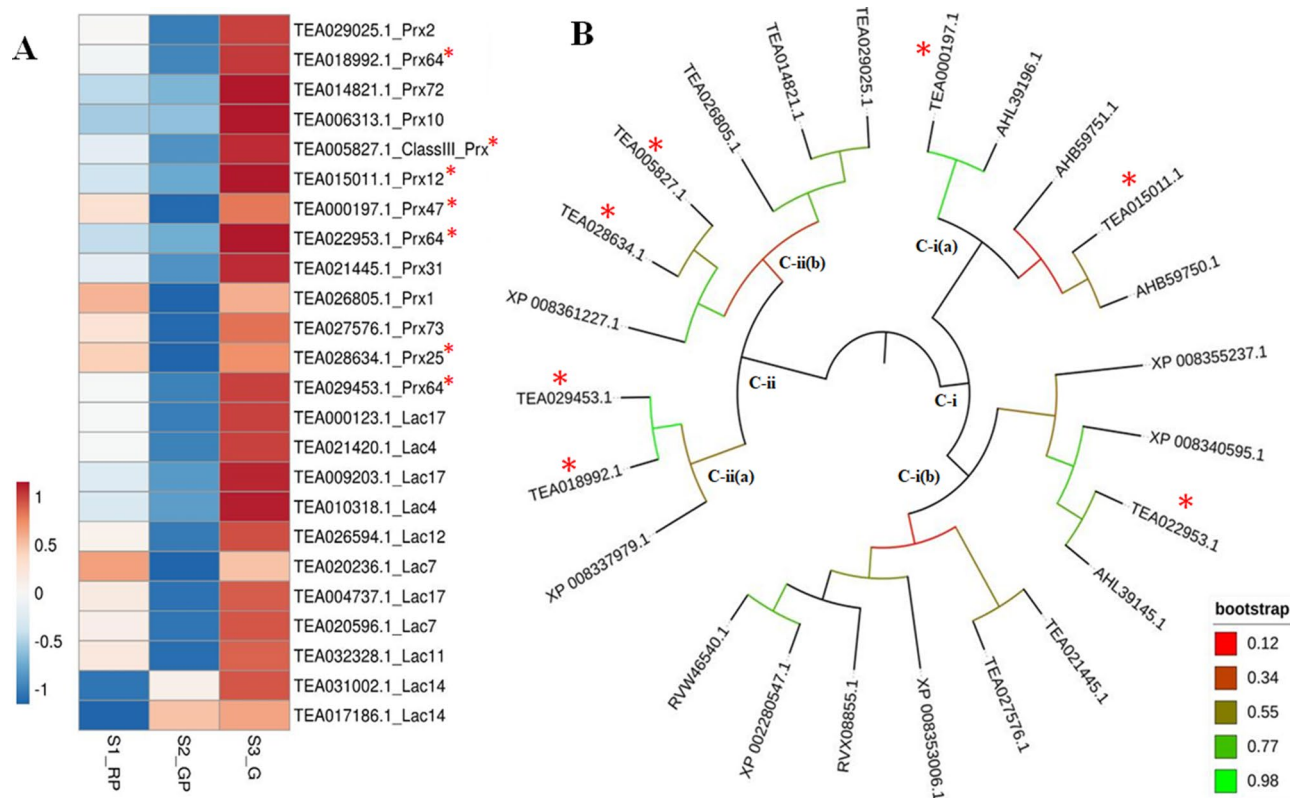
**Figure 6.** Differential expression of key genes involved in phytohormone signaling during leaf color transition. (A) Auxin acid signaling genes. (B) Ethylene signaling genes. (C) Abscisic acid signaling genes. Phytohormone signaling pathways were obtained from KEGG<sup>61</sup> (<https://www.kegg.jp/kegg/kegg1.htm>).

EBF and ERS regulators mapped to *Chr14*, *Chr15*, *Chr3*, *Chr2*, *Chr11*, *Chr8* and *Chr4* exhibited strong negative correlation ( $>0.999$ ) with TAC content (EBF1c and ERS1b), photoperiod (EBF1a), relative humidity (EBF1d) and precipitation (EBF1c and ERS1b) suggesting inhibitory role of ethylene and auxin in light-induced anthocyanin accumulation in tea (Supplementary Table S5). Contrarily, genes involved in ABA signaling [abscisic acid receptor (PYL4) and protein phosphatases (PPP2C\_34)] recorded upregulated and downregulated expressions during S2-GP and S3\_G, respectively (Fig. 6C). Strong positive correlation ( $>0.99$ ) with TAC, relative humidity (PYL4) and photoperiod (PP2C\_34) indicated their important role in ABA assisted anthocyanin accumulation during seasonal leaf color transitions in tea<sup>29</sup>.

**Expression analysis of candidate genes involved in anthocyanin degradation.** Intensity of purple leaf coloration is dependent on the net results of biosynthesis and degradation of anthocyanin. In this study, 24 DE key laccases (*Lac*; 11) and peroxidases (*Prx*; 13) involved in anthocyanin degradation were identified and mapped to *Chr1* and *Chr8* (Supplementary Table S4). The expression pattern of candidate anthocyanin degrading enzymes exhibited significant upregulation during S3\_G as compared to S2\_GP and S1\_RP (Fig. 7A). Nevertheless, their optimum expression in S3\_G along with negative correlation (0.769) with total anthocyanin content suggests their involvement in anthocyanin degradation and eventual loss of purple colour in tea cultivars (Supplementary Table S5). Moreover, strong negative correlation ( $>0.99$ ) identified between the expression of laccases (*lac12*, *lac17* and *lac4*) and peroxidases (*Prx31* and *Prx64*) with photoperiod also suggests light dependent anthocyanin degradation in tea. Furthermore, successful assignment of laccases and peroxidases preferably towards the ends of the chromosomes suggest their significant involvement in influencing genetic variation of desirable traits including leaf color transition in tea (Fig. 5).

**Phylogenetic analysis of differentially expressed peroxidases.** Phylogenetic analysis of 13 DE key putative peroxidases involved in anthocyanin degradation with corresponding functionally characterized peroxidases of *Populus*, *Apple* and *Brunfelsia* clustered into two main clades (I and II) each with sub-clades (a and b) (Fig. 7B). In sub-clade I(a), TEA000197.1 recorded significantly higher homology with known anthocyanin degrading class III peroxidase in *populus* [AHL39196.1], while TEA015011.1 was clustered with *Brunfelsia* peroxidase [AHB59750.1 BcPrx02]. Furthermore, significantly higher homology was revealed between





**Figure 7.** (A) The expression pattern of DE candidate peroxidases and laccases. (B) Phylogenetic tree of DE putative peroxidases and functionally characterized peroxidases. The \* indicates tea peroxidases exhibiting significant homology with known peroxidases.

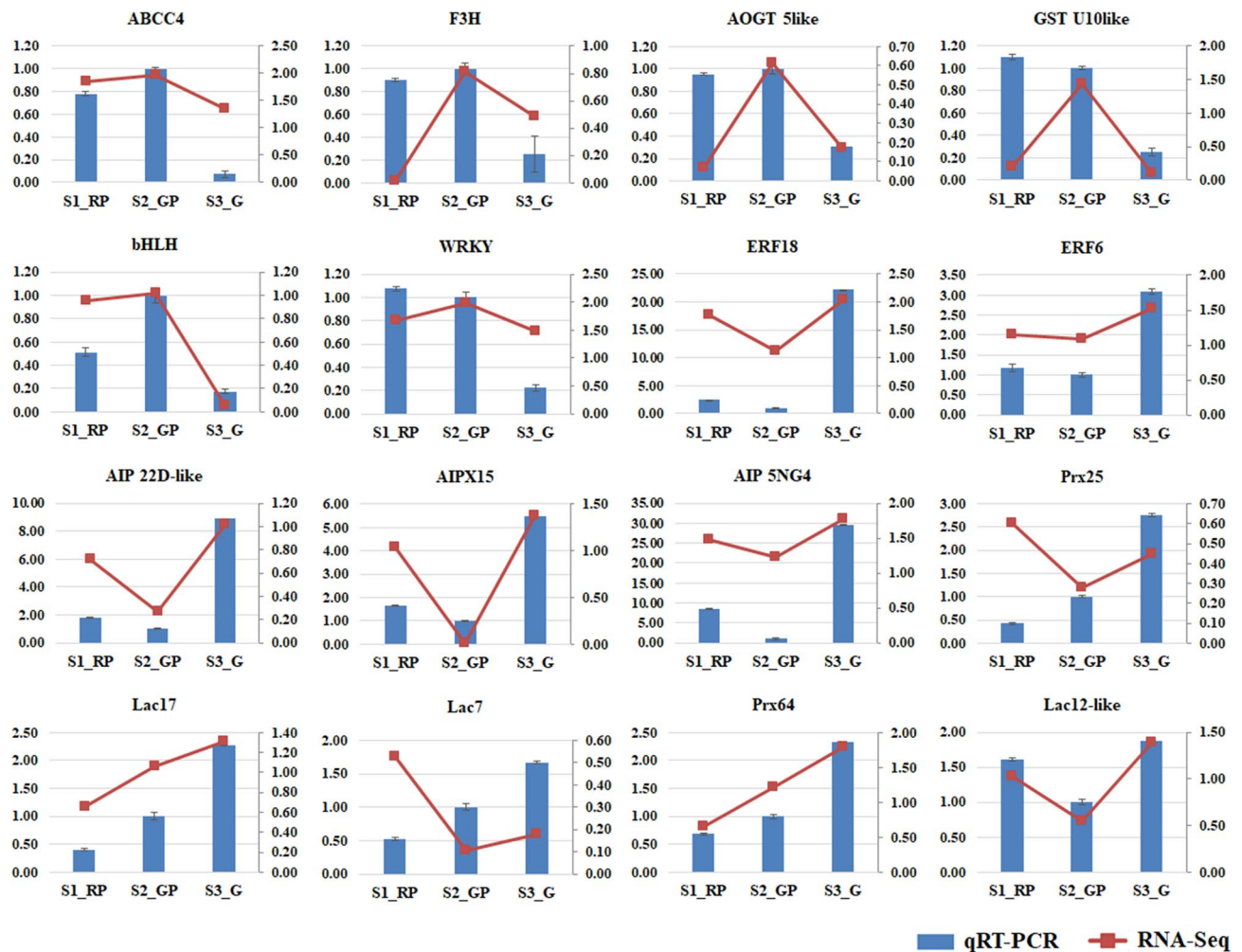
tea peroxidase [TEA022953.1] and class III peroxidase of populus [AHL39145.1] as grouped together in sub-clade I(b). Similarly, clade II revealed significant homology between tea peroxidases [TEA029453.1 and TEA018992.1] and apple (*Malus domestica*) peroxidase 64 [XP008337979.1] in sub-clade II (a), whereas, tea peroxidases [TEA028634.1 and TEA0005827.1] and apple (*Malus domestica*) peroxidase 25 [XP008361227.1] were grouped together in sub-clade II(b). Significant homology between candidate peroxidases with known anthocyanin degrading peroxidases (CIII Prx, BcPrx02, AHL39196.1, AHB59750.1) speculate key role of peroxidases in anthocyanin degradation in tea<sup>30</sup>.

**qRT-PCR validation of DEGs.** To confirm expression patterns of candidate genes identified in transcriptome dataset, sixteen key DEGs (Supplementary Table S7) involved in major pathways of anthocyanin accumulation were successfully validated in qRT-PCR analysis. The expression trends corresponded to RNA-seq expressions confirming the reliability of current RNA-seq data and computational analysis (Fig. 8). Interestingly, qRT expression patterns of validated genes were also consistent with TAC dynamics during leaf color transitions in tea (S1\_RP, S2\_GP, S3\_G).

## Discussion

The dynamics of biosynthesis and degradation of anthocyanin together may influence changes in composition and ratio of different anthocyanin molecules during leaf color transitions<sup>31</sup>. In this study, strong positive correlation between leaf color changes [reddish purple (S1\_RP), dark gray purple (S2\_GP) and moderate olive green (S3\_G)] and TAC, clearly distinguished dynamics of seasonal induced anthocyanin accumulation as recently reported in tea and, raspberry<sup>6,32</sup>. Optimum accumulation of anthocyanin during S2\_GP can be explored as “potential season” for production of “specialty anthocyanidin-rich green tea” that fetches premium price at commercial scale<sup>6</sup>. Moreover, key genes identified in this study can be validated and utilized for genetic improvement of TAC in tea plants during seasons of low synthesis. Therefore, genome-wide elucidation of molecular mechanism of seasonal induced leaf color transitions can be key to sustain anthocyanin accumulation and prevent degradation for seasonal independent commercial production of “specialty tea”. High throughput deep transcriptome sequencing generated ~ 412 million high quality/clean reads assembled into 71,349 transcripts with homology based functional annotation > 85% in multifarious public databases, indicating improved annotations and accuracy of the assembly during this study<sup>33</sup>. Furthermore, significant enrichment of biological functions (secondary metabolite biosynthesis, oxidoreductase activity) and pathways related to anthocyanin biosynthesis and degradation; ‘plant hormone signal transduction’ and ‘phenylpropanoid/flavonoid biosynthesis’ suggests their active contribution in anthocyanin accumulation and seasonal dependent leaf color transition in tea.





**Figure 8.** The expression profiles of 16 DEGs as determined by qRT-PCR and their correlation with RNA-seq expression data.

Additionally, significant positive correlation between qRT-PCR and RNA-seq expression data of 16 DE candidate genes confirm the reliability and utility of current RNA-seq data for genomic studies in tea.

Upregulated expression of PAL, 4CL, F3H, AOGT and UFGT during S2\_GP affirmed their critical role in anthocyanin biosynthesis<sup>34</sup>. Similarly, significant positive correlation between AOGT5, F3H\_Like, PAL2 genes and TAC suggest efficient performance of the pathway leading to enhanced anthocyanin accumulation during S2\_GP. Interestingly, current findings are in concordance with recent report on seasonal dependent responses of anthocyanin-rich tea cultivars in tea<sup>6</sup>. Additionally, dynamic expression of AOGT and UFGT, upregulation during S2\_GP, and downregulation during S3\_G, suggest their role as key regulators enhancing the hydrophilicity and stability of vacuolar accumulation of anthocyanin in tea<sup>35</sup>. Similarly, upregulated expression of ANRs, catalyzing the branching step leading to biosynthesis of epicatechin, during S3\_G suggest possible involvement in negative regulation of anthocyanin pathway due to limited/non-availability of substrate flux, (leucoanthocyanidin) for anthocyanin biosynthesis<sup>27</sup>. Moreover, upregulated expression of ANR during S3\_G may be associated with increasing substrate availability for laccase-mediated anthocyanin degradation in tea. Additionally, glutathione S-transferase (GST), ABC and MATE play key role in anthocyanin transportation from the cytosolic face of endoplasmic reticulum (biosynthesis site) through a glutathione S-transferase (GST) mediated transport mechanism in which anthocyanin is conjugated with glutathione by GST and moved to vacuolar surface, where ABC and MATE transporter proteins transfer them to the vacuole (storage site)<sup>36</sup>. ABC transporters are reportedly involved in ATP dependent transportation of glucosylated anthocyanidin (malvidin-3-O-glucoside), while MATE mediate transport of manonylated anthocyanidins in the presence of electrochemical gradients of H<sup>+</sup> or other ions created by V-ATPase or V-PPase<sup>37,38</sup>. Furthermore, ABC and MATE are considered primary and secondary active transport systems in plants, respectively<sup>39</sup>. In this study, upregulated expression of key transporters during purple leaf colour transitions of S1\_RP and S2\_GP and down regulated expression during S3\_G (moderately olive green) suggests well-coordinated seasonal dependent GST mediated anthocyanin accumulation in tea. Interestingly, functional validation GST in a mutant *Vitis vinifera* line (defective of GST) confirmed its key role in anthocyanin accumulation<sup>40</sup>. Additionally, significantly higher expression of ABC transporters during S2\_GP corroborates previous reports on their important role in anthocyanin accumulation in purple leaf tea cultivars<sup>41</sup>.

Furthermore, expression of important TFs including MYBs (18) and bHLH (7) may be associated with activation or repression of EBGs (MYBs) and LBGs (MYB-bHLH-WD40) of anthocyanin pathway<sup>42</sup>. Previous studies indicate that R2R3-MYB (MYB11, MYB12 and MYB111) are directly involved in the activation of EBGs (CHI, CHS, F3H and F3'H), while MBW (MYB-bHLH-WD40) complex activates LBGs such as DFR, ANS and UFGT<sup>43</sup>. Therefore, strong positive correlation between expression of key TFs (MYB and bHLH), anthocyanin biosynthetic genes and TAC during S2\_GP reported here, are in accordance to earlier studies in tea<sup>44</sup>.

Considering key role of MBW complex and WRKY in vacuolar regulation of pH through controlled expression of proton pumps, upregulated expression of MYBs, bHLH and WRKY TFs may be associated with increased acidity in the vacuole enhancing purple colour pigmentation during S2\_GP<sup>7</sup>. Nevertheless, their downregulation possibly reduced acidity of the vacuole resulting in reduced intensity of anthocyanin purple coloration during S3\_G. Moreover, upregulation of novel NAC95, NAC4, NAC35, HMG6 & 7, DOF2 and SBP1 TFs during S2\_GP may be associated with transcriptional activation of gene cascades involved in anthocyanin biosynthesis in tea<sup>16</sup>.

MYB TFs are also involved in transcriptional repression in the biosynthesis of phenylpropanoid-derived compounds with ~ 37 repressor MYBs including those involved in regulation of anthocyanin biosynthesis already reported in plants<sup>43</sup>. Strong negative correlation between upregulated expression of candidate repressor MYBs (MYB4-like, MYB34, MYB46, MYB52, MYB61, MYB85 and MYB86) and TAC during S3\_G suggests their key role in seasonal induced regulation of anthocyanin accumulation. Moreover, MYB4-like TFs repress anthocyanin biosynthesis either through direct binding of promoter regions of ANS, DFR and UFGT, or displacing the activator MYB in the MBW complex hence reducing the complex to anthocyanin repressor complex<sup>43,45</sup>. Nevertheless, upregulated expression of MYB85 during S3\_G may be associated with either activation of repressor MYB4-like TF, or diversion of anthocyanin pathway to lignin biosynthesis at *P*-coumaroyl-CoA branching point, and hence regulating metabolic flux towards anthocyanin biosynthesis<sup>46</sup>. Additionally, novel putative repressors identified here can be further characterized at functional level and explored for genetic improvement of anthocyanin accumulation in tea.

Phytohormones are also important internal factors influencing anthocyanin biosynthesis in plants<sup>47</sup>. ABA regulates anthocyanin biosynthesis via serial activation of PYR/PYL leading to inhibition of PP2C, and triggering of SnRK2 which regulate ABF mediated activation of MBW complex<sup>47</sup>. Interestingly, strong positive correlation between the expression of PYL, PP2C and TFs (MYBs, bHLH), and TAC suggests ABA mediated transcriptional regulation of anthocyanin biosynthesis in tea (Fig. 5A) as earlier reported in blueberry and wolfberry<sup>48,49</sup>. Similarly, significant positive correlation between ABA signaling genes and phenylalanine ammonia lyase also suggest ABA-mediated direct regulation of anthocyanin biosynthesis pathway genes as reported in strawberry<sup>29</sup>. Contrarily, negative correlation between expression of auxin-signaling genes (AUX1, AUX/IAA and SAUR) and activator TFs (MYB, bHLH) during S3\_G, affirmed their inhibitory role in anthocyanin biosynthesis<sup>50</sup>. Furthermore, negative correlation recorded between genes involved in ethylene signaling (ETR, EBF1/2, ERF), TAC and day-length, highlighted photoperiod-dependent regulation of anthocyanin accumulation in tea<sup>29</sup>. Our findings corroborate recent study on light-mediated anthocyanin accumulation via induced expression of key regulatory and structural genes involved in anthocyanin biosynthesis<sup>51</sup>.

Moreover, concomitant upregulation of ethylene signaling genes along with repressor transcription factors suggest key role in negative regulation of anthocyanin accumulation in tea<sup>52</sup>. Likewise, negatively correlated expression dynamics of several putative peroxidases (13) and laccases (11) and TAC also suggests their crucial role in seasonal induced anthocyanin degradation in tea<sup>30,53,54</sup>. Interestingly, peroxidase mediated anthocyanin degradation via direct oxidation of anthocyanin molecules, or first deglycosylation of sugar moieties by  $\beta$ -glucosidases followed by peroxidase oxidation (Fig. 5a) has been reported in ornamental kales<sup>21</sup>. Similarly, epicatechin-mediated laccase degradation of anthocyanin via oxidation of epicatechin molecules to quinones followed by polymerization with anthocyanin leading its deglycosylation (Fig. 5a) is known in many fruits<sup>53</sup>. Therefore, positive correlation between upregulated expression of laccases and ANR (responsible for epicatechin biosynthesis) suggest active involvement in anthocyanin degradation during S3\_G. Similarly, their upregulation may also be linked to diversion of the substrate (*P*-coumaroyl CoA) to lignin biosynthesis, thus reducing metabolic flux towards anthocyanin biosynthesis<sup>55</sup>. Interestingly, significant homology of six candidate peroxidases with functionally validated peroxidases involved in anthocyanin degradation in brunsfellia, poplar and apple, possibly confirm their significant role in seasonal induced anthocyanin degradation in tea.

## Conclusion

In this study, enhanced purple leaf coloration during S2\_GP correlated with higher total anthocyanin content highlighting season's potential for production of "anthocyanin-rich specialty green tea". Identification of key DEGs associated with anthocyanin biosynthesis [PAL, 4CL, F3H, DFR and UGT/UFGT], vacuolar trafficking [ABC, MATE and GST] transcriptional regulation [MYB, NAC, bHLH, WRKY and HMG] including novel mechanism of phytohormone signaling genes [PYL and PP2C] and anthocyanin degrading genes [Prx and lac] suggest well-coordinated seasonal controlled metabolism and catabolism in tea. Current findings speculate molecular insights of seasonal induced anthocyanin biosynthesis and degradation in tea (Fig. 4). Wherein, downregulation of key activator TFs, transporters and ABA signaling genes possibly limits the expression of genes involved in anthocyanin biosynthesis reducing total anthocyanin accumulation causing loss of purple leaf coloration in tea. Nevertheless, upregulation of key repressor TFs (MYB4-like, and MYB85), genes involved in hormonal signaling (auxin and ethylene) may inhibit anthocyanin biosynthesis by binding promoter regions of structural genes repressing their activity. Moreover, upregulation of vacuolar peroxidases and laccases possibly degrade anthocyanin by removing glycosyl molecule converting into colorless anthocyanidin causing loss of purple leaf coloration in tea.

Genetic manipulation of key structural and transcription factor genes identified in this study can enhance seasonal independent anthocyanin accumulation and maintain prolonged purple leaf coloration. Furthermore, comprehensive genome-wide transcriptional data generated will provide a valuable genomic resource for development of functional molecular markers for expediting breeding and selection of elite anthocyanin-rich tea cultivars.

## Materials and methods

**Plant materials and sampling.** In this study, a mature purple tea cultivar (> 15 years old) ‘CSIR-IHBT-PT-17’ grown and maintained under full-sun field conditions at the CSIR-Institute of Himalayan Bioresource technology, Palampur, India (32° 6' 52" N, 76.533° 24' E) was selected. CSIR-IHBT-PT-17 is a purple-leaf tea cultivar, selected from a natural population of field grown pigmented tea cultivars. Major characteristics of the cultivar include: semi-arboreal plant type, purple-green young shoots with two leaves and a bud, which turns completely green as season changes with the exception of purple petiole. The leaves of ‘CSIR-IHBT-PT-17’ are also moderately elliptic with medium serration. Samples (bud with two leaves) were harvested monthly from April to October at 1100HRS-IST (Indian Standard Time), immediately frozen in liquid nitrogen and stored at – 80 °C until analysis. Subsequently meteorological data viz; precipitation, relative humidity, day length and air temperature were recorded during sampling (Supplementary Table S7; Supplementary information).

**Extraction and determination of total anthocyanin content.** Total anthocyanin content was determined using a pH differential method as described by Yuan et al.<sup>56</sup> with slight modifications. In brief, frozen shoots (100 mg) were crushed into fine powder in liquid nitrogen, extracted separately in 2 ml each of buffer A (25 mM KCl pH 1.0) and buffer B (400 mM sodium acetate pH 4.5), then centrifuged at 12,000g for 15 min at 4 °C. The supernatant was collected and diluted for direct measurement of absorbance at 510 nm and 700 nm. Total anthocyanin content was calculated using the following equation;

$$\text{Total anthocyanin (mg g}^{-1}\text{ FW)} = (A \times MW/\epsilon \times \text{dilution factor})$$

where A is the absorbance of the diluted sample calculated as follows;

$$A = (A_{510} - A_{700})_{\text{pH}1.0} - (A_{510} - A_{700})_{\text{pH}4.5}$$

MW is the molecular mass (484.8) of predominant anthocyanin (cyaniding-3-glucoside chloride), while  $\epsilon$  is the molar absorptivity (24,825) at 510 nm. Each sample was analyzed in triplicate and the results were expressed as the average of the three measurements.

**RNA extraction, library preparation and sequencing.** Based on leaf color transitions and significant variation in TAC, leaf samples harvested in April; S1\_RP (Reddish-Purple), August; S2\_GP (Dark Gray Purple) and October; S3\_G (olive-Green) were used for transcriptome analysis. Total RNA was extracted from each sample using IRIS method<sup>57</sup>. RNA quality and quantity was determined on 1% formaldehyde agarose gel and NanoDrop 2000 spectrophotometer, respectively. Further, RNA integrity (RIN) was determined using an Agilent Bioanalyzer 2100 (Agilent), and RNA samples with A260/A280 ratio > 1.8, A260/A230 ratio > 2.0 and RIN > 7 were considered for library preparation. Equimolar concentration of RNA from three replicates of each sample were pooled for cDNA library preparation using ILLUMINA RNA Sample preparation kit (Illumina Inc., CA, USA) as described by the manufacturer. Libraries were quantified using Qubit 2.0 fluorometer (Invitrogen, USA), while quality assessment was performed using AGILENT 2100 Bioanalyzer (Agilent Technologies, USA). The cDNA libraries were sequenced on ILLUMINA NOVASEQ 6000 (Illumina Inc., CA, USA). Paired-end sequencing format of 100 bp was adopted in this study.

**Transcript assembly and differential gene expression analysis.** Raw reads were trimmed to remove adapter sequences, ambiguous nucleotides (N) and low-quality reads (Phred quality score  $\leq$  Q30) using NGS QC Toolkit<sup>58</sup>. High quality clean reads were analyzed according to Tuxedo protocol<sup>59</sup>. In brief, high quality clean reads were assembled using the reference tea genome (*Camellia sinensis* var. *sinensis*<sup>25</sup>) using TOPHAT ver2.1.0. Functional annotation of assembled transcripts was performed by BLAST alignment of the sequences against public databases such as Swiss-Prot, KEGG and TAIR10. GO and KEGG annotations were performed using AgriGO toolkit and KEGG pathway database, respectively<sup>60,61</sup>. The Transcription factor (TF) families were identified using Plant transcription factor database (PTFDB)<sup>62</sup>.

**Identification of differentially expressed genes.** Transcripts abundance was estimated in FPKM (Fragments Per Kilobase of transcript per Million mapped reads) and differentially expressed genes with  $\log_2$  Fold Change (absolute value)  $\geq$  2 and  $p \leq 0.05$  were considered differentially expressed (DE) using cuffdiff tool in Cufflinks version 2.2.1 with default parameters. GO and KEGG enrichment of differentially expressed genes (DEGs) was performed using publicly available Tea Plant Information Archive (<http://tpia.teaplant.org/>) database<sup>63</sup>. The corresponding chromosomal locations of DEGs were also determined by mapping them to the chromosomes of *C. sinensis* var. *sinensis* based on their physical distances in the GFF genome files, downloaded from chromosome-scale genome of tea<sup>25</sup>. The physical chromosome maps were visualized using TBtools (a Toolkit for Biologist integrating various biological data handling tools) VS 1.051 (<https://www.tbtools.com>) with default parameters<sup>64</sup>.

**Phylogenetic analysis of DEGs involved in anthocyanin degradation.** To determine the phylogenetic relationship between DE putative anthocyanin degrading genes and functionally characterized genes from apple, populus and brunsfelia, phylogenetic analysis was performed using MEGA5.2<sup>65</sup>. Sequences were aligned using ClustalW<sup>66</sup> while phylogenetic tree was constructed using the Neighbor-joining method based on the Tamura-Nei model with bootstrap value kept at 1000. The resulting tree was visualized using iTOL v5 (<https://itol.embl.de/>)<sup>67</sup>.

**qRT-PCR validation of RNA-seq data.** In order to verify the reliability of RNA-seq data, 2 µg of total RNA used in transcriptome sequencing was reversed transcribed using RevertAid H minus First Strand cDNA Synthesis Kit (Thermo Scientific, USA). Gene specific primers were designed with BatchPrimer3 (<http://probe.spw.usda.gov/batchprimer3/>) and primer sequences are presented in Supplementary Table S8. The qRT-PCR was performed in a StepOnePlus Real-Time PCR system (Applied Bio-system, USA) using SYBR Green Master mix (Applied Biosystem). Considering its consistent expression pattern in a recent study<sup>68</sup>, GAPDH gene was used as internal control. All experiments were performed in three independent biological and technical replicates. Relative expression was calculated using  $2^{-\Delta\Delta CT}$  method<sup>69</sup>.

**Statistical analysis.** All statistical analysis was performed using GenStat V15.1 (VSN Intl.).

### Data availability

The nucleotide sequence(s) reported in this article have been submitted to NCBI SRA database under accession numbers; SRR12153423, SRR12153422 and SRR12153421.

Received: 12 October 2020; Accepted: 17 December 2020

Published online: 13 January 2021

### References

- Xia, E. H. *et al.* Tea plant genomics: Achievements, challenges and perspectives. *Hortic. Res.* <https://doi.org/10.1038/s41438-019-0225-4> (2020).
- Feng, Z. *et al.* Tea aroma formation from six model manufacturing processes. *Food Chem.* <https://doi.org/10.1016/j.foodchem.2019.01.174> (2019).
- Food and Agriculture Organization: The FAO Intergovernmental Group on Tea. Emerging trends in tea consumption: Informing a generic promotion process. In: *Twenty-third Session of the Intergovernmental Group on Tea. Hangzhou, China, 17–20 May 2018.* [www.fao.org/ccp/tea23/en/](http://www.fao.org/ccp/tea23/en/).
- Hajra, N. G. & Yang, C. W. M. Diversification of the tea products—global scenario. *J. Tea Sci. Res.* <https://doi.org/10.5376/jtsr.2015.05.0003> (2015).
- Muoki, C. R., Maritim, T. K., Oluoch, W. A., Kamunya, S. M. & Bore, J. K. Combating climate change in the Kenyan Tea Industry. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2020.00339> (2020).
- Hu, J. G. *et al.* Screening tea hybrid with abundant anthocyanins and investigating the effect of tea processing on foliar anthocyanins in tea. *Folia Hort.* <https://doi.org/10.2478/fhort-2020-0025> (2020).
- Passeri, V., Koes, R. & Quattrocchio, F. M. New challenges for the design of high value plant products: Stabilization of anthocyanins in plant vacuoles. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2016.00153> (2016).
- Zhu, H. *et al.* Anthocyanins function as a light attenuator to compensate for insufficient photoprotection mediated by nonphotochemical quenching in young leaves of *Acmena acuminatissima* in winter. *Photosynthetica* <https://doi.org/10.1007/s11099-017-0740-1> (2018).
- Kerio, L. C., Wachira, F. N., Wanyoko, J. K. & Rotich, M. K. Total polyphenols, catechin profiles and antioxidant activity of tea products from purple leaf coloured tea cultivars. *Food Chem.* <https://doi.org/10.1016/j.foodchem.2012.09.066> (2013).
- Mbuthia, K. S. *et al.* Tea (*Camellia sinensis*) infusions ameliorate cancer in 4TI metastatic breast cancer model. *BMC Complementary Altern. Med.* <https://doi.org/10.1186/s12906-017-1683-6> (2017).
- Oren-Shamir, M. Does anthocyanin degradation play a significant role in determining pigment concentration in plants?. *Plant Sci.* <https://doi.org/10.1016/j.plantsci.2009.06.015> (2009).
- Maritim, T. K. *et al.* Physiological and biochemical response of tea [*Camellia sinensis* (L.) O. Kuntze] to water-deficit stress. *J. Hort. Sci. Biotechnol.* <https://doi.org/10.1080/14620316.2015.11513200> (2015).
- Shen, J. *et al.* Metabolic analyses reveal different mechanisms of leaf color change in two purple-leaf tea plant (*Camellia sinensis* L.) cultivars. *Hortic. Res.* <https://doi.org/10.1038/s41438-017-0010-1> (2018).
- Lai, Y. S. *et al.* The dark-purple tea cultivar “Ziyan” accumulates a large amount of delphinidin-related anthocyanins. *J. Agric. Food Chem.* <https://doi.org/10.1021/acs.jafc.5b04036> (2016).
- Li, J. *et al.* Transcriptome analysis reveals the accumulation mechanism of anthocyanins in ‘Zijuan’ tea (*Camellia sinensis* var. assamica (Masters) kitamura) leaves. *Plant Growth Regul.* <https://doi.org/10.1007/s10725-016-0183-x> (2017).
- Wang, L. *et al.* Regulation of anthocyanin biosynthesis in purple leaves of zijuan tea (*Camellia sinensis* var. kitamura). *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18040833> (2017).
- Kumari, M. *et al.* Regulation of color transition in purple tea (*Camellia sinensis*). *Planta* <https://doi.org/10.1007/s00425-019-03328-7> (2020).
- Zhou, Q., Sun, W. & Lai, Z. Differential expression of genes in purple-shoot tea tender leaves and mature leaves during leaf growth. *J. Sci. Food Agric.* <https://doi.org/10.1002/jsfa.7308> (2015).
- Zhou, Q., Chen, Z., Lee, J., Li, X. & Sun, W. Proteomic analysis of tea plants (*Camellia sinensis*) with purple young shoots during leaf development. *PLoS One* <https://doi.org/10.1371/journal.pone.0177816> (2017).
- Wu, L. Y. *et al.* Transcriptomic and translomic analyses reveal insights into the developmental regulation of secondary metabolism in the young shoots of tea plants (*Camellia sinensis* L.). *J. Agric. Food Chem.* <https://doi.org/10.1021/acs.jafc.0c03341> (2020).
- Ren, J., Liu, Z., Chen, W., Xu, H. & Feng, H. Anthocyanin degrading and chlorophyll accumulation lead to the formation of bicolor leaf in ornamental kale. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms20030603> (2019).
- Kumar, D., Sarmah, B. K. & Das, P. K. Ethylene mediates repression of anthocyanin accumulation in black rice pericarps in the absence of light. *J. Plant Physiol.* <https://doi.org/10.1016/j.jplph.2019.02.011> (2019).
- Guo, X. Y. *et al.* Polyphenol oxidase dominates the conversions of flavonol glycosides in tea leaves. *Food Chem.* <https://doi.org/10.1016/j.foodchem.2020.128088> (2020).



24. Sud, R. G. & Baru, A. Seasonal variations in theaflavins, thearubigins, total colour and brightness of Kangra orthodox tea (*Camellia sinensis* (L.) O. Kuntze) in Himachal Pradesh. *J. Sci. Food Agric.* [https://doi.org/10.1002/1097-0010\(200007\)80:9%3c1291::AID-JSFA633%3e3.0.CO;2-K](https://doi.org/10.1002/1097-0010(200007)80:9%3c1291::AID-JSFA633%3e3.0.CO;2-K) (2000).
25. Wei, C. *et al.* Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.1719622115> (2018).
26. Chen, J. D. *et al.* The chromosome-scale genome reveals the evolution and diversification after the recent tetraploidization event in tea plant. *Hortic. Res.* <https://doi.org/10.1038/s41438-020-0288-2> (2020).
27. Kim, Y. B. *et al.* Transcripts of anthocyanidin reductase and leucoanthocyanidin reductase and measurement of catechin and epicatechin in Tartary Buckwheat. *Sci. World J.* <https://doi.org/10.1155/2014/726567> (2014).
28. Czechowski, T., Bari, R. P., Stitt, M., Scheible, W. R. & Udvardi, M. K. Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: Unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J.* <https://doi.org/10.1111/j.1365-313X.2004.02051.x> (2004).
29. Jiang, Y. & Joyce, D. C. ABA effects on ethylene production, PAL activity, anthocyanin and phenolic contents of strawberry fruit. *Plant Growth Regul.* <https://doi.org/10.1023/A:1022539901044> (2003).
30. Zipor, G. *et al.* Planta anthocyanin degradation by a vacuolar class III peroxidase in *Brunfelsia calycina* flowers. *New Phytol.* <https://doi.org/10.1111/nph.13038> (2015).
31. Liu, T. *et al.* Transcriptome analyses reveal key genes involved in skin color changes of 'Xinlimei' radish taproot. *Plant Physiol. Biochem.* <https://doi.org/10.1016/j.plaphy.2019.04.006> (2019).
32. Kassim, A. *et al.* Environmental and seasonal influences on red raspberry anthocyanin antioxidant contents and identification of quantitative traits loci (QTL). *Mol. Nutr. Food Res.* <https://doi.org/10.1002/mnfr.200800174> (2009).
33. Wang, D. *et al.* Transcriptome analysis of *Sclerotinia ginseng* and comparative analysis with the genome of *Sclerotinia sclerotiorum*. *Physiol. Mol. Plant Pathol.* <https://doi.org/10.1016/j.pmpp.2018.11.003> (2019).
34. He, F. *et al.* Biosynthesis of anthocyanins and their regulation in colored grapes. *Molecules* <https://doi.org/10.3390/molecules15129057> (2010).
35. Boss, P. K., Davies, C. & Robinson, S. P. Expression of anthocyanin biosynthesis pathway genes in red and white grapes. *Plant Mol. Biol.* <https://doi.org/10.1007/BF00019111> (1996).
36. Gu, K. D., Wang, C. K., Hu, D. G. & Hao, Y. J. How do anthocyanins paint our horticultural products?. *Sci. Hortic.* <https://doi.org/10.1016/j.scienta.2019.01.034> (2019).
37. Francisco, R. M. *et al.* ABC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-glucosides. *Plant Cell* **25**(5), 1840–1854. <https://doi.org/10.1105/tpc.112.102152> (2013).
38. Zhao, J. *et al.* MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in *Medicago truncatula*. *Plant Cell* **23**(4), 1536–1555. <https://doi.org/10.1105/tpc.110.080804> (2011).
39. Shitan, N. & Yazaki, K. New insights into the transport mechanisms in plant vacuoles. *Int. Rev. Cell Mol. Bio.* **305**, 383–433. <https://doi.org/10.1016/B978-0-12-407695-2.00009-3> (2013).
40. Conn, S., Curtin, C., Bézier, A., Franco, C. & Zhang, W. Purification, molecular cloning, and characterization of glutathione S-transferases (GSTs) from pigmented *Vitis vinifera* L. cell suspension cultures as putative anthocyanin transport proteins. *J. Exp. Bot.* **59**(13), 3621–3634. <https://doi.org/10.1093/jxb/ern217> (2008).
41. Wang, L. *et al.* Regulation of anthocyanin biosynthesis in purple leaves of zijuan tea (*Camellia sinensis* var. *kitamura*). *Int. J. Mol. Sci.* **18**, 4. <https://doi.org/10.3390/ijms18040833> (2017).
42. Gonzalez, A., Zhao, M., Leavitt, J. M. & Lloyd, A. M. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J.* **53**(5), 814–827. <https://doi.org/10.1111/j.1365-313X.2007.03373.x> (2008).
43. Chen, C. *et al.* MYB transcription repressors regulate plant secondary metabolism. *Crit. Rev. Plant Sci.* **38**(3), 159–170. <https://doi.org/10.1080/07352689.2019.1632542> (2019).
44. Sun, B. *et al.* Purple foliage coloration in tea (*Camelliasinensis* L.) arises from activation of the R2R3-MYB transcription factor CsAN1. *Sci. Rep.* <https://doi.org/10.1038/srep32534> (2016).
45. Pérez-Díaz, J. R. *et al.* New member of the R2R3-MYB transcription factors family in grapevine suppresses the anthocyanin accumulation in the flowers of transgenic tobacco. *Plant Mol. Biol.* **90**(1–2), 63–76. <https://doi.org/10.1007/s11103-015-0394-y> (2016).
46. Geng, P. *et al.* MYB20, MYB42, MYB43, and MYB85 regulate phenylalanine and lignin biosynthesis during secondary cell wall formation. *Plant Physiol.* **182**(3), 1272–1283. <https://doi.org/10.1104/PP.19.01070> (2020).
47. Chung, S. W. *et al.* Transcriptional regulation of abscisic acid biosynthesis and signal transduction, and anthocyanin biosynthesis in 'Bluecrop' highbush blueberry fruit during ripening. *PLoS One* **14**, 7. <https://doi.org/10.1371/journal.pone.0220015> (2019).
48. Oh, H. D., Yu, D. J., Chung, S. W., Chea, S. & Lee, H. J. Abscisic acid stimulates anthocyanin accumulation in 'Jersey' highbush blueberry fruits during ripening. *Food Chem.* **244**, 403–407. <https://doi.org/10.1016/j.foodchem.2017.10.051> (2018).
49. Li, G. *et al.* ABA mediates development-dependent anthocyanin biosynthesis and fruit coloration in Lycium plants. *BMC Plant Biol.* **19**, 1. <https://doi.org/10.1186/s12870-019-1931-7> (2019).
50. Ji, X. H., Zhang, R., Wang, N., Yang, L. & Chen, X. S. Transcriptome profiling reveals auxin suppressed anthocyanin biosynthesis in red-fleshed apple callus (*Malus sieversii* f. *niedzwetzkyana*). *Plant Cell Tissue Organ Culture* **123**(2), 389–404. <https://doi.org/10.1007/s11240-015-0843-y> (2015).
51. Li, W. *et al.* The effects of ultraviolet a/b treatments on anthocyanin accumulation and gene expression in dark-purple tea cultivar 'Ziyan' (*Camellia sinensis*). *Molecules* **25**(2), 354. <https://doi.org/10.3390/molecules25020354> (2020).
52. Jeong, S. W. *et al.* Ethylene suppression of sugar-induced anthocyanin pigmentation in Arabidopsis. *Plant Physiol.* **154**(3), 1514–1531. <https://doi.org/10.1104/pp.110.161869> (2010).
53. Fang, F. *et al.* An intracellular Laccase is responsible for epicatechin-mediated Anthocyanin degradation in litchi fruit Pericarp. *Plant Physiol.* **169**(4), 2391–2408. <https://doi.org/10.1104/pp.115.00359> (2015).
54. Giardina, P. *et al.* Laccases: A never-ending story. *Cell. Mol. Life Sci.* **67**, 369–385. <https://doi.org/10.1007/s00018-009-0169-1> (2010).
55. Schuetz, M. *et al.* Laccases direct lignification in the discrete secondary cell wall domains of protoxylem. *Plant Physiol.* **166**, 798–807. <https://doi.org/10.1104/pp.114.245597> (2014).
56. Yuan, Y., Chiu, L. W. & Li, L. Transcriptional regulation of anthocyanin biosynthesis in red cabbage. *Planta* **230**(6), 1141–1153. <https://doi.org/10.1007/s00425-009-1013-4> (2009).
57. Ghawana, S. *et al.* An RNA isolation system for plant tissues rich in secondary metabolites. *BMC Res. Notes* <https://doi.org/10.1186/1756-0500-4-85> (2011).
58. Patel, R. K. & Jain, M. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. *PLoS One* **7**, 2. <https://doi.org/10.1371/journal.pone.0030619> (2012).
59. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**(3), 562–578. <https://doi.org/10.1038/nprot.2012.016> (2012).
60. Tian, T. *et al.* AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**(W1), W122–W129. <https://doi.org/10.1093/nar/gkx382> (2017).

61. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
62. Guo, A. Y. *et al.* PlantTFDB: A comprehensive plant transcription factor database. *Nucleic Acids Res.* **36**(1), D966–D969. <https://doi.org/10.1093/nar/gkm84> (2007).
63. Xia, E. H. *et al.* Tea plant information archive: A comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnol. J.* **17**(10), 1938–1953. <https://doi.org/10.1111/pbi.13111> (2019).
64. Chen, C. *et al.* TBtools-an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* <https://doi.org/10.1016/j.molp.2020.06.009> (2020).
65. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**(10), 2731–2739. <https://doi.org/10.1093/molbev/msr121> (2011).
66. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673> (1994).
67. Letunic, I. & Bork, P. Interactive TREE OF LIFE (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259. <https://doi.org/10.1093/nar/gkz239> (2019).
68. Seth, R. *et al.* Global transcriptional insights of pollen-pistil interactions commencing self-incompatibility and fertilization in tea [*Camellia sinensis* (L.) O. Kuntze]. *Int. J. Mol. Sci.* **20**(3), 539. <https://doi.org/10.3390/ijms20030539> (2019).
69. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>ΔΔCT method. *Methods* **25**(4), 402–408 (2001).

## Acknowledgements

This study was implemented through financial support from the Council of Scientific and Industrial Research (CSIR), India, under projects BSC-0301 and MLP-0146 to CSIR-IHBT. TKM is grateful to TWAS-CSIR for the award of Postgraduate Fellowship. This manuscript represents IHBT communication number 7378.

## Author contributions

R.K.S. and T.K.M.: conceived and designed the study. T.K.M.: performed the experiments. T.K.M., M.M. and R.S.: analyzed the data. T.K.M.: wrote the manuscript. R.K.S.: edited and approved the final version of the manuscript. All the authors reviewed and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80437-4>.

**Correspondence** and requests for materials should be addressed to R.K.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021