Biosynthesis and Chemopreventive Potential of Jute (*Corchorus capsularis and C. olitorius)* Flavonoids and Phylogeny of Flavonoid Biosynthesis Pathways

OPEN ACCESS

Authors

Pratik Satya[1], Debabrata Sarkar[1], Amitava Chatterjee[2], Srikumar Pal[3], Soham Ray[1], Laxmi Sharma[1], Suman Roy[1], Amit Bera[1], Srinjoy Ghosh[1], Jiban Mitra[1], Gouranga Kar[1], Nagendra Kumar Singh[4]

Affiliations

1 ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, Kolkata, India
2 Faculty Centre of Integrated Rural Development & Management, Ramakrishna Mission Vivekananda Educational and Research Institute, Narendrapur, Kolkata, India
3 Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India
4 ICAR-National Institute for Plant Biotechnology, Pusa Campus, New Delhi, India

Key words

Corchorus spp., flavonoid, Malvaceae, MMP-2, pathway, phylogeny

received 10.07.2021
revised 08.10.2021
accepted 18.11.2021

Bibliography

Planta Med Int Open 2022; 9: e23–e33
DOI 10.1055/a-1712-7978
ISSN 2509-9264

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Correspondence

Pratik Satya
ICAR-Central Research Institute for Jute and Allied Fibres,
Barrackpore, Kolkata,
India
pscrijaf@gmail.com

Supplementary material is available under https://doi.org/10.1055/a-1712-7978.

ABSTRACT

Flavonoids are valuable phytochemicals for human health and nutrition. Jute (*Corchorus capsularis* and *C. olitorius*), a vegetable rich in phenolics and flavonoids, is globally consumed for its health benefit, but the biosynthesis pathways and metabolic profiles of its flavonoids are poorly characterized. Elucidating the flavonoid biosynthesis pathways would augment the broader use of jute, including targeted synthesis of its specific flavonoids. We reconstructed the core flavonoid biosynthesis pathways in jute by integrating transcriptome mining, HPLC and flavonoid histochemistry. In *C. capsularis* (white jute), the flavonoid biosynthesis pathways' metabolic flux was driven toward the biosynthesis of proanthocyanidins that mediate the acquisition of abiotic stress tolerance. However, higher levels of flavonols in *C. olitorius* (tossa jute) render it more suitable for nutritional and medicinal use. Jute flavonoid extract exhibited *in vitro* inhibition of matrix metalloproteinase-2, suggesting its potential chemopreventive and immunity-boosting roles. Using the flavonoid biosynthesis pathways profiles of 93 plant species, we reconstructed the flavonoid biosynthesis pathways phylogeny based on distance-based clustering of reaction paths. This reaction-path flavonoid biosynthesis pathways phylogeny was quite distinct from that reconstructed using individual gene sequences. Our flavonoid biosynthesis pathways-based classification of flavonoid groups corroborates well with their chemical evolution, suggesting complex, adaptive evolution of flavonoid biosynthesis pathways, particularly in higher plants.

ABBREVIATIONS

CHS	chalcone synthase
CHI	chalcone isomerase
DPBA	diphenylboric acid 2-aminoethyl ester
EGCG	epigallocatechin gallate
FBP	flavonoid biosynthesis pathway
FLS	flavonol synthase
HCT	shikimate-o-hydroxycinnamoil transferase
KEGG	Kyoto encyclopedia of genes and genomes
ML	maximum-likelihood
MMP	matrix metalloproteinase

Introduction

Flavonoids represent a wide variety of secondary metabolites ubiquitously distributed in the plant kingdom, contributing to diverse physiological functions including pigmentation, plant defense, stress response, nodulation, UV-protection, auxin transport, root development, pollen fertility, and fruit development [1, 2]. They are valuable for human health and nutrition as antioxidants, cardio-vascular disease protectants, and anti-allergen, anti-cancer, and anti-inflammatory agents [3]. Since they constitute an essential component of prophylactic measures for SARS-CoV-2 (COVID-19) patients [4], identifying plant dietary sources containing high concentrations of flavonoids has recently drawn renewed interest. Jute (white jute, *Corchorus capsularis* L.; tossa jute, *C. olitorius* L.; Malvaceae s. l.), a bast fiber crop, is valued worldwide as a health-promoting leafy vegetable for its high micronutrient, vitamin, phenolics and flavonoid content and is used in many countries as salad, cooked vegetable, soup and beverage [5]. It exhibits the highest antioxidant activity among major vegetables [6]. With a very high oxygen radical absorbance capacity, it is a component of the famous "Okinawa diet" historically recommended for prolonged lifespan and better health [7]. In addition, ethnomedicinal uses of jute as anti-aging, anti-hemorrhagic, anti-cystitis, hypoglycemic, anti-obesity, and gastro-protective elixir are well established [5, 8].

Though chemical analyses identified many flavonoids in jute [6, 9, 10], there are contradictory reports that could provide only an abstruse idea of the flavonoids synthesized in it, perhaps due to the presence of mucilaginous compounds that interfere in sample extraction [9]. Moreover, other factors, such as the presence of diverse derivatives with similar structures, metabolite channeling, analytical system and metabolite cut-off size also interfere with chemical analyses of phenolics and flavonoids [11]. Metabolic pathway identification has recently progressed beyond chemical analyses to overcome such limitations. Genomic and transcriptomic databases are increasingly utilized to reconstruct metabolic pathways combined with chemical analyses [12, 13]. However, a comprehensive reconstruction of the FBPs in a flavonoid-rich vegetable has not been reported yet.

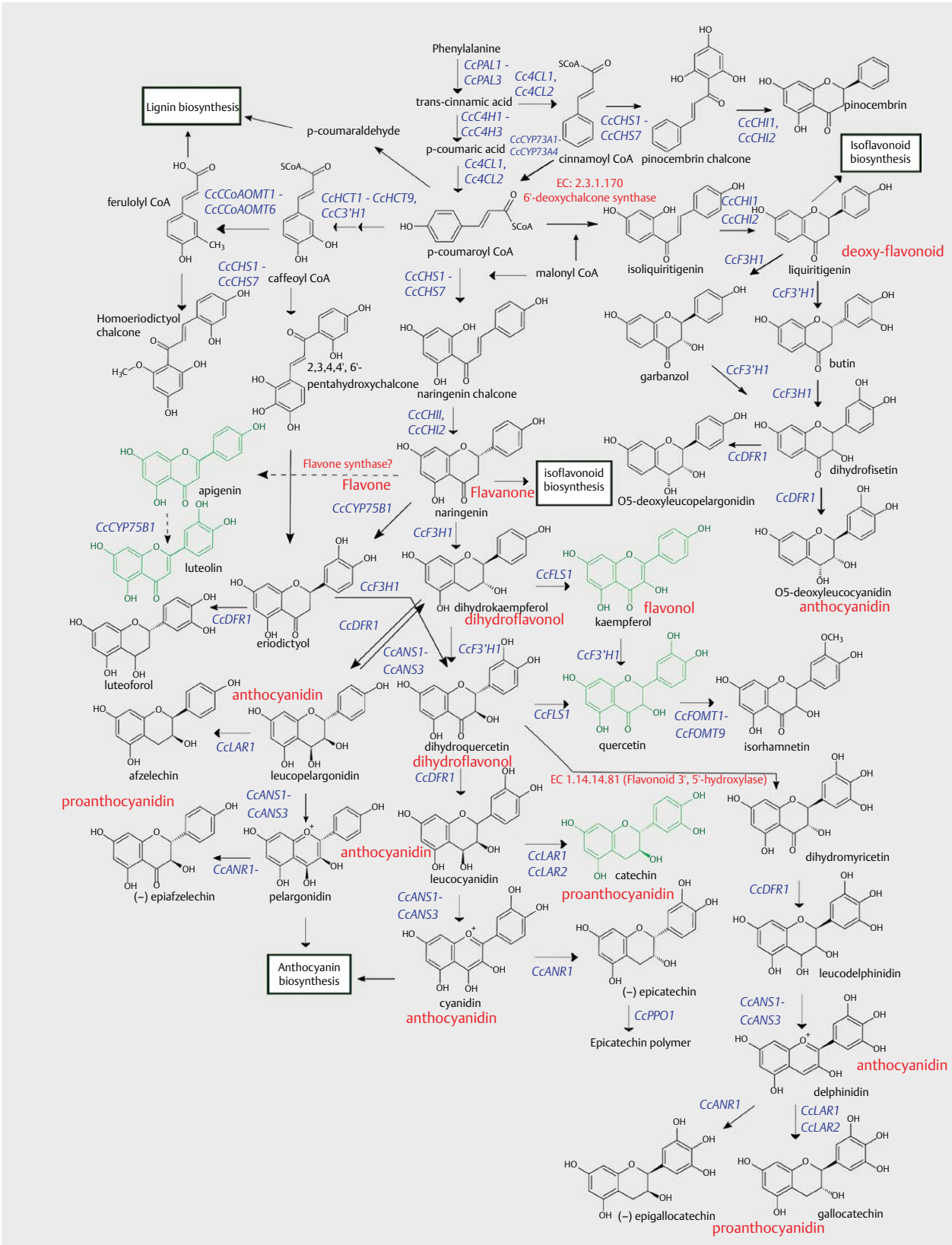
The jute plant extract was also very promising for growth inhibition and degradation of carcinoma cell lines [14, 15]. Since Taiwo et al. [15] suggested a role of jute phenolics in anti-tumorigenic activity, it will be interesting to investigate the potential role of its flavonoids. MMPs, a group of endopeptidases involved in tumor cell metastasis, are widely used as cancer biomarkers and therapeutic targets [16]. As flavonoid extracts, particularly green-tea catechin, have been shown to inhibit both pro- and active-MMP-2 proteins [17], the potential of jute flavonoids as an anti-MMP-2 agent would diversify their use in chemoprophylaxis and overall immunity development as well.

The core FBPs in higher plants generate 6 to 7 major groups of compounds, including flavone, flavonol, flavanone, proanthocyanidin, and anthocyanidin. In land plants, these groups have evolved over 500 million y to cope with stresses like UV radiation, drought, and novel herbivores [18]. However, the evolutionary paths of many of the FBP genes reveal significant discrepancies with the known plant phylogeny [19], showing different rates and patterns of evolution [20]. The presence of multiple copies of an FBP gene, with proven functional roles for only a few of them, further complicates the task of assigning their sequences to reaction paths and reconstructing a comprehensive gene sequence-based FBP phylogeny. Therefore, in recent years, species phylogeny based on metabolic diversity has proven to be an effective strategy for retracing the evolutionary courses of metabolic pathways [21]. Such a metabolic pathway-driven phylogeny captures the overall evolutionary signatures that are otherwise missed in single gene-based phylogeny [22]. Many strategies for converting the metabolic data into phylogenetic indices have been developed [21, 23, 24]. However, all these studies essentially dealt with the phylogenetic distribution of species, ignoring the very possibility of classifying the reaction paths to garner additional insights into the evolution of biosynthesis pathways. For a reaction path is controlled by an enzyme and hence its corresponding orthologous gene, such clustering can provide important clues to the evolution of flavonoid groups in land plants, for which we have little phylogenetic evidence.

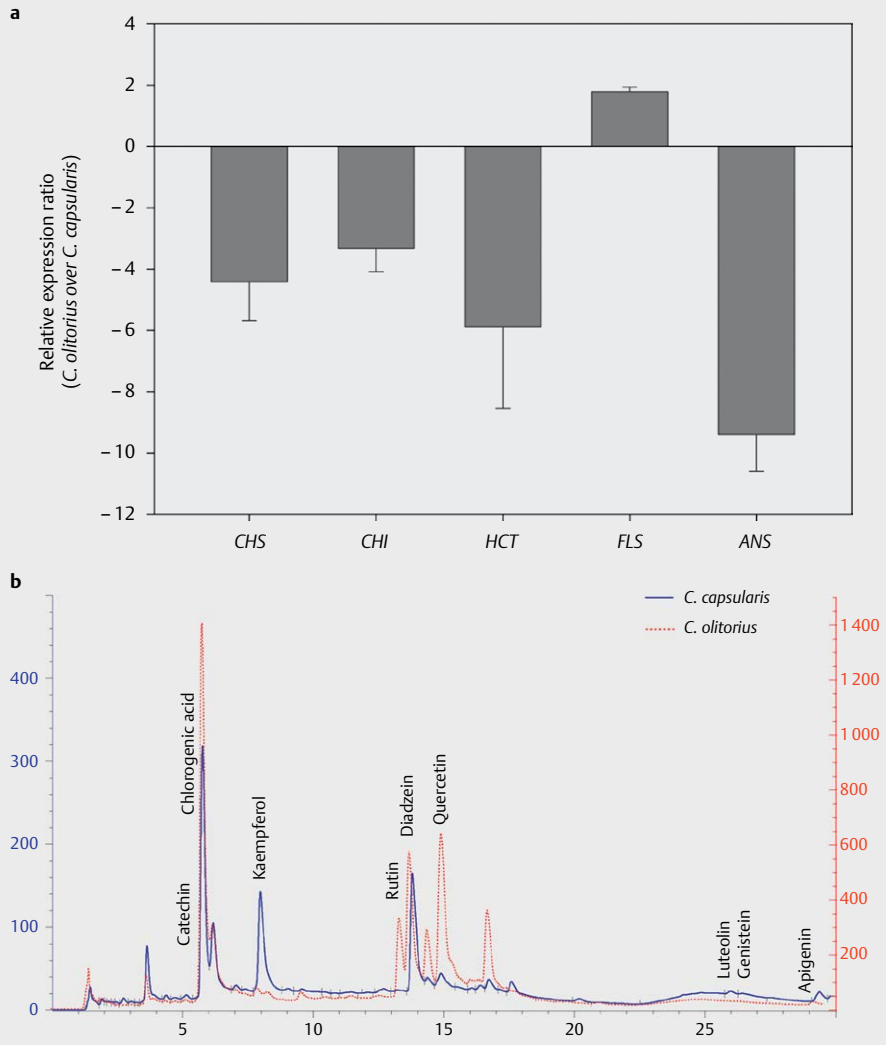
In this study, we reconstructed and validated the core FBPs in jute, with an objective to investigate the biosynthesis, distribution, and antitumor potential of its flavonoids. Further, we constructed a plant FBP-matrix using the reaction-paths information and analyzed it, in comparison with that of a sequence-based one, to understand the evolution of flavonoids in higher plants. Our results showed that distance-based clustering of pathway information could provide valuable clues to the evolution of metabolic pathways.

Results and Discussion

We provide a compendious description of the jute FBPs integrating transcriptomics and metabolite analyses, thereby generating the first FBP-profile of a flavonoid-rich plant species. The reconstructed jute FBPs (▶ **Fig. 1**) were characterized by 55 reaction paths involving 40 isoforms of 13 genes (size range: 277–2163 bp; mean coverage in KO0941: 95.8 %) that account for the biosynthesis of all the major flavonoid groups (Table 1S, text ST1, Supporting Information). The majority of these FBP genes had > 85 % sequence similarities with that identified in jute genomes (Table 2S, Supporting Information). We identified 2 downstream pathways responsible for anthocyanin biosynthesis in jute – one *via* cyanidin formation and the other *via* pelargonidin synthesis (▶ **Fig. 1**). Moreover, the proanthocyanidin biosynthesis pathways were also active in the young jute stem. While almost all the genes involved in major



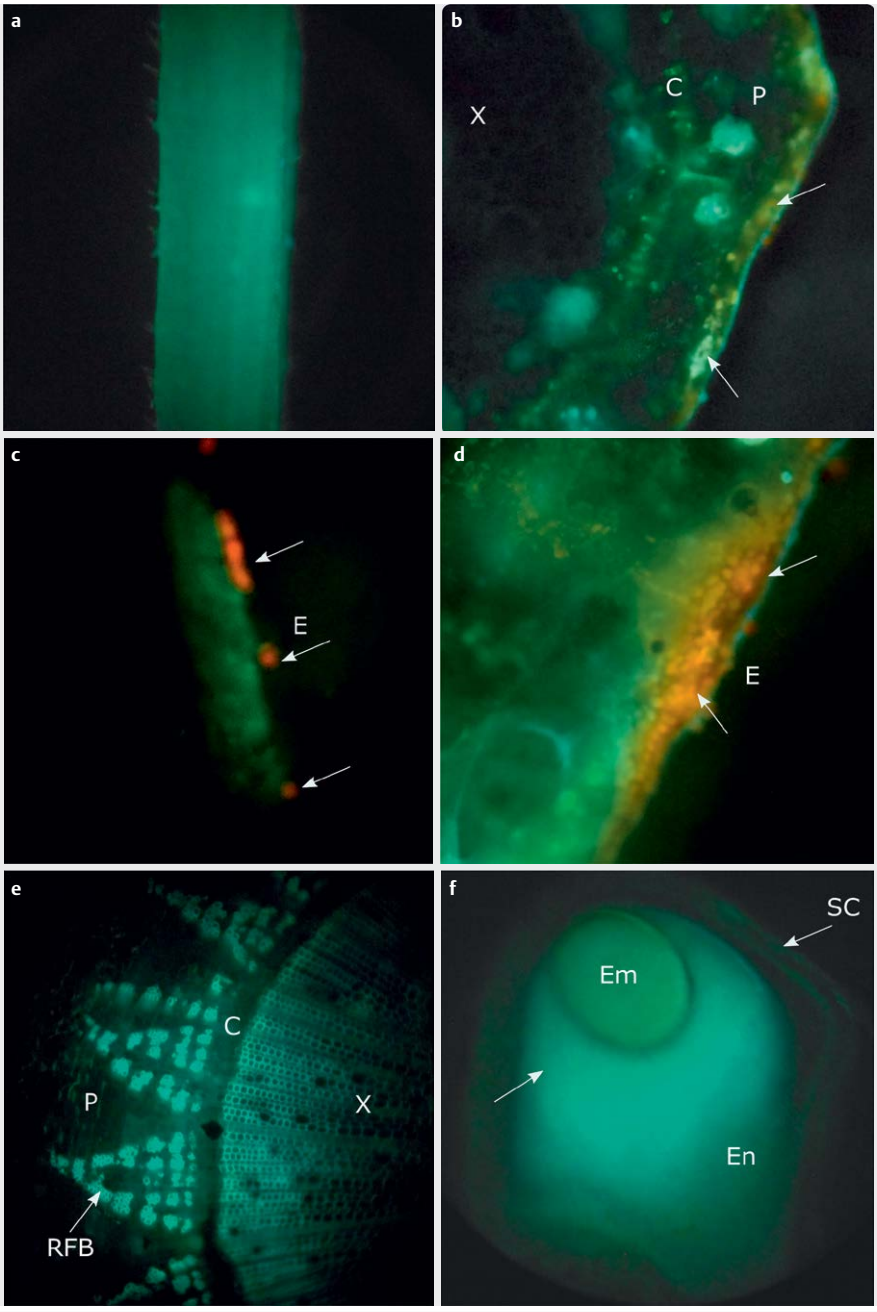
▶ **Fig. 1** Flavonoid biosynthesis pathways (FBPs) in jute reconstructed from transcriptome and metabolite characterization. Probable reaction paths are marked by dotted arrows. Green-colored metabolites were identified by HPLC in this study. *ANR*, anthocyanidin reductase; *ANS*, anthocyanidin synthase; C4H, cinnamate 4-hydroxylase; *CCoAOMT*, caffeoyl-CoA O-methyltransferase; *4CL*, 4-coumarate: CoA ligase; *CHS*, chalcone synthase; *CHI*, chalcone isomerase; *CYP73A*, trans-cinnamate 4-monooxygenase; *CYP75B*, flavonoid 3'-monooxygenase; *CYP98A3*, coumaroylquinate (coumaroylshikimate) 3'-monooxygenase; *DFR*, dihydroflavonol 4-reductase; *F3H1*, flavanone 3-dioxygenase; *FLS*, flavonol synthase; *FOMT*, flavonoid 3'-O-methyltransferase; *HCT*, shikimate O-hydroxycinnamoyltransferase; *LAR*, leucoanthocyanidin reductase; PAL, phenylalanine ammonia-lyase; *PPO*, polyphenol oxidase. For details, refer to Supporting Information (Table 1S).



▶ **Fig. 2** Gene expression of FBP enzymes and flavonoid profiles. **a**) Relative expression (log-fold change) of 5 FBP genes in *C. olitorius* over *C. capsularis*. **b**) Comparative HPLC chromatogram of flavonoids from methanolic extract of *C. capsularis* (solid line, blue) and *C. olitorius* (dotted line, red). The X-axis represents retention time; Y-axis represents peak length (left side scale: *C. capsularis*, right side scale: *C. olitorius*).

flavonoids biosyntheses were identified in jute (▶ **Fig. 1**), *flavone synthase II* (EC 1.14.19.76) and *flavonoid 3', 5'-hydroxylase* (EC 1.14.14.81) were surprisingly absent in bast as well as hypocotyl transcriptomes. We could, however, identify a candidate gene in the *C. olitorius* genome (LLWS01000724.1) that had 80.5 % sequence similarity with *Theobroma cacao*'s *flavonoid 3', 5'-hydroxylase 2*. However, it might not have expressed in jute bast or hypocotyl. Of the 5 key FBP genes (CHS, CHI, HCT, FLS, and ANS), expressions of all except *CcFLS1* were down-regulated in *C. olitorius* (▶ **Fig. 2a**), with *CcANS2* recording the highest level of down-regulation (9.4-fold). Since ANS catalyzes the biosyntheses of cyanidin and pelargonidin, we suggest that anthocyanin biosynthesis is more active in *C. capsularis* than in *C. olitorius*. Between the 2 species, *C. capsularis* exhibits wider variation in stem pigmentation and has a higher abiotic stress tolerance [25, 26], which may be due to its anthocyanin-specific metabolic drive. Down-regulations of *CcCHS2* (4.4-fold), *CcCHI2* (3.3-fold), and *CcHCT8* (5.9-fold) also indicated a possible surge of chalcone biosynthesis in *C. capsularis*. On the contrary, Cc*FLS1* was up-regulated in *C. olitorius* (1.8-fold), suggesting its FBP's metabolic flux was driven toward higher flavonol biosynthesis.

From the chromatograms of the methanolic leaf extracts (water: methanol, 1:10, v/v), we identified several key flavonoids, including kaempferol, quercetin, rutin, catechin, and luteolin in both jute species. In addition, several phenolics such as gallic acid, vanillic acid, and chlorogenic acid, as well as 2 isoflavonoids, viz., daidzein and genistein, were also present in the extract (▶ **Fig. 2b**). Previously, Ola et al. did not detect kaempferol, quercetin, or luteolin in *C. olitorius* extracts [9], while Arai et al. [27] detected kaempferol and quercetin but not luteolin. Instead, our results closely agree with the flavonoids detected in the ethanolic extract of *C. olitorius* leaf [10]. However, the methanolic solvent used in this study favored extraction of rutin, catechin, daidzein, and genistein, which was not detected in the ethanolic extract of jute leaf [10]. Although

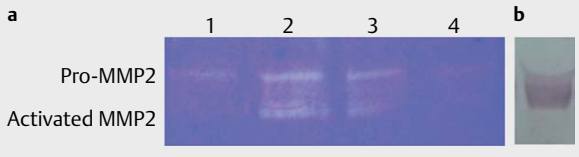


▶ **Fig. 3** Histochemical localization of flavonoid compounds in transverse sections of Corchorus plant tissues using diphenylboric acid 2-aminoethyl ester (DPBA). **a**) hypocotyl; **b**) young meristem; **c**) leaf; **d**) stem; **e**) root; **f**) pod. Arrows indicate deposition of flavonoid compounds. Flavonoids are present mostly under the epidermis in patchy areas. No flavonoid-specific staining could be observed in the root. Orange fluorescence indicates the presence of kaempferol and quercetin; bright yellow fluorescence indicates the presence of naringenin-chalcone. Blue-green fluorescence indicates the presence of other flavonoid and phenolic compounds. *X,* xylem; *P,* phloem; *C*, cambium; *E*, epidermis; *RFB*, root fiber bundle; *Em*, embryo; *En*, endosperm; *SC*, seed coat.

apigenin (3.8 µg/g DW) was detected only in *C. olitorius*, but detection of luteolin in *C. capsularis* may well suggest the biosynthesis of an apigenin precursor in it. A distinct sharp peak for kaempferol was identified from the chromatogram of *C. capsularis*. In contrast, in the case of *C. olitorius,* the peak for quercetin was more prominent (▶ **Fig. 2b**), indicating that kaempferol is converted readily to quercetin in *C. olitorius* but not in *C. capsularis*. The enzyme *FLS* drives biosynthesis of both kaempferol and quercetin from dihydrokaempferol and dihydroquercetin, respectively (▶ **Fig. 1**). Isoforms of *FLS* from *Camellia sinensis* show differential affinity to dihydrokaempferol and dihydroquercetin [28]; therefore, it may be possible that the *FLS* gene in *C. olitorius* has more affinity to dihydroquercetin. An upregulation of *CcFLS1* in *C. olitorius* suggests that the conversion rate of dihydroquercetin to quercetin might be higher in *C. olitorius*, resulting in more quercetin accumulation.

Since extraction and chromatographic identification of flavonoids in jute is challenging due to the presence of mucilaginous substances [9], we further complemented chemical analysis with flavonoid histochemistry identifying the major flavonoids and their distribution patterns in different organs. Both phenolics and flavonoids (sinapate esters) were localized in hypocotyl tissues as evident by evenly distributed green fluorescence (▶ **Fig. 3a**), with subsequent yellow-green and bright yellow spots under the epidermal layer in young stem indicating the accumulation of kaempferol and naringenin chalcone, respectively (▶ **Fig. 3b**). Their increasing intensity and change of color (to brilliant gold) were indicative of the accumulation of both quercetin and kaempferol in the leaf (▶ **Fig. 3c**) and bast tissues (▶ **Fig. 3d**). However, with just a pale-green autofluorescence of lignified fiber cells, the root tissues were characterized by the complete absence of flavonoids (▶ **Fig. 3e**). Though the embryo fluoresced bright green, indicating the presence of relatively high concentrations of phenolics, no such fluorescence was observed in the seed coat (▶ **Fig. 3f**). While both species exhibited similar phenolics and flavonoid accumulation patterns, total phenolics and flavonoids were higher in *C. olitorius* (48.5 ± 7.4 mg/g DW) than in *C. capsularis* (23.3 ± 4.1 mg/g DW), with the former producing more chlorogenic acid, quercetin, rutin, and catechin and the latter producing more kaempferol. The gene expression and chemical analyses posit a flavonol-oriented metabolic drive in *C. olitorius* with a higher quercetin reserve, making it a more valuable source of dietary and medicinally important flavonoids.

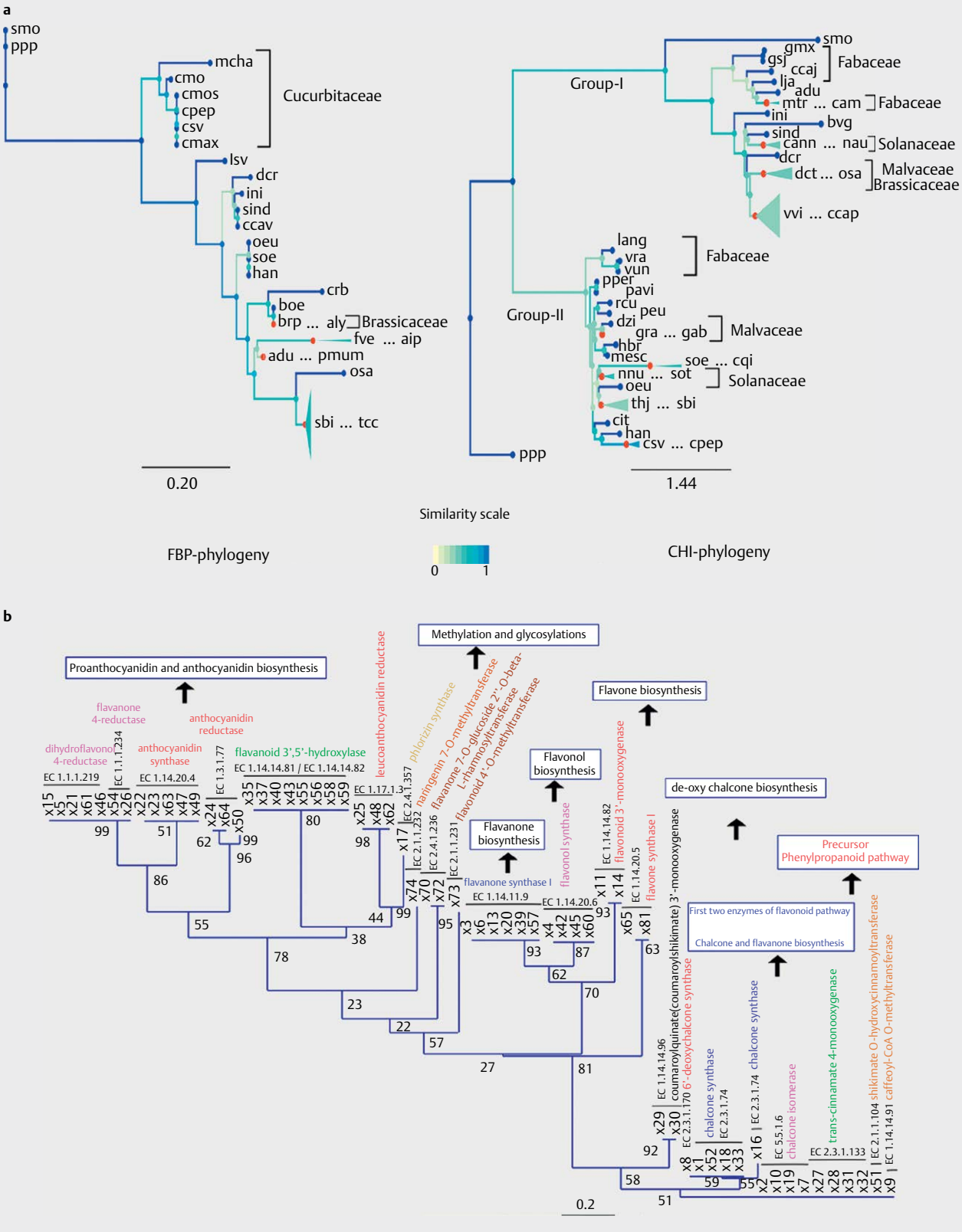
To further elucidate the medicinal importance of jute flavonoids, we examined the MMP-2 inhibitory potential of the flavonoid extracts (see Materials and Methods for details) of jute. As evident from the zymogram (▶ **Fig. 4**), we obtained higher inhibition of active MMP-2 (72 KD and 64 KD) using the *C. olitorius* flavonoid extract, and this inhibitory effect was comparable to that obtained using the tea EGCG (positive control). By comparison, the inhibitory effect of the *C. capsularis* flavonoid extract on active MMP-2 was less pronounced, suggesting partial inhibition of MMP-2 activity. We further confirmed the presence and composition of MMP-2 fractions by western blotting using human/mouse MMP-2-specific monoclonal antibodies. Flavonoids like catechin, apigenin, quercetin, and genistein block MMP activity [29]. Quercetin also downregulates MMP-2 activity in rats, thereby reducing hypertension [30]. Thus, stronger inhibition of MMP-2 by *C. olitorius* flavonoids might be attributed to higher quercetin content in *C. olitorius*. Previous studies have also reported the high anti-tumorigenic potential of jute leaf extract [14, 15]. Our results indicate that flavonoids may also play a vital role in the anti-tumorigenic potential of jute by inhibiting MMP-2 activity *in vivo*. Given that flavonoids like apigenin, luteolin, and quercetin are also potential COVID-19 inhibitors [4], the jute FBPs would further instigate experimentation on dietary intake of jute as a preventive/therapeutic agent against COVID-19. Altogether, the diversity of jute FBPs (▶ **Fig. 1**) suggests that jute's immune-boosting and chemopreventive potential, particularly *C. olitorius*, should be investigated in detail.



▶ **Fig. 4** Flavonoid-induced inhibition of MMP-2 activity. **a**) gelatin zymogram of human MMP-2 treated with different flavonoid inhibitors (1: MMP-2 + Tea epigallocatechin gallate, 2: MMP-2 + distilled water, 3: MMP-2 + *C. capsularis* flavonoids, 4: MMP-2 + *C. olitorius* flavonoids). **b**) Western blot of human MMP-2 with anti-human/mouse MMP-2 antibody.

Our gene expression and chemical analyses showed that the FBPs of the 2 jute species have different metabolic drives, suggesting species-level modifications during the evolution of the FBPs. In accordance with that, we noted that the reaction path-based FBP phylogeny did not agree with the Angiosperm Phylogenetic Group [31], although we could unequivocally distinguish the lower angiosperms from the higher ones (▶ **Fig. 5a**; **Fig. 1S**, Supporting Information). In our FBP-phylogeny, the Brassicaceous or the Cucurbitaceous species were clustered on specific nodes. Those belonging to Fabaceae, Malvaceae, Poaceae, or Solanaceae remained mostly ungrouped around multiple nodes (▶ **Fig. 5a**), suggesting adaptive evolution of FBPs in many higher plant families, in agreement with the chemical evidence-based evolution of the flavonoids [32]. In contrast, gene sequence-based CHI-phylogeny retrieved 2 distinct groups that diverged early from the bryophytes (▶ **Fig. 5a**; **Fig. 2S**, Supporting Information). Within each group, the Brassicaceous, Malvaceous, or the Solanaceous species formed separate clusters congruent with higher-plant phylogeny, but the Fabaceous species were distributed in multiple clusters. Despite high Robinson-Foulds distance ($> 10^4$) indicating significant differences in branching patterns between the reaction-path-based and CHI-based phylogenies, the evolutionary fates of the *CHI*-reconstructed Group-I species were more similar to that obtained by the FBP phylogeny (▶ **Fig. 5a**). Given that the individual gene-based evolutionary trees are difficult to integrate to obtain a holistic picture of FBP evolution [19, 20] and the trees themselves often suggest divergent evolutionary paths for different isoforms (as observed in the case of *CHI*), we showed that the reaction path-based phylogeny, being different from individual gene-based phylogeny can provide a comprehensive view of FBP evolution. This is not unexpected because phylogenetic trees based on a single gene follow a uniform rate of evolution and may disagree with metabolic pathway-based evolution [22].

However, the species phylogenies (▶ **Fig. 5a**) cannot help understand the evolution of specific flavonoid groups, as a single gene may be involved in the biosynthesis of multiple flavonoid groups (▶ **Fig. 1**). We could demonstrate a strong biological basis for reaction path-based NJ clustering of FBP enzymes because results mainly agreed with enzymatic classes that catalyze the biosyntheses of different groups of core flavonoids (▶ **Fig. 5b**; Table 4S, Supporting Information). As the clustering variables represented a continuous evolutionary gradient and the data points were generated based on the presence/absence of KEGG orthologs, we reasoned that the reaction path-based clusters could be linked with the evolution of the flavonoid groups. The clustering of the flavonoid groups agrees well with the chemical evidence of flavonoid groups



▶ **Fig. 5** Phylogenetic and cluster analysis of FBPs. **a**) Comparison of topological features of the FBP-phylogeny (computed using Jaccard's similarity coefficient and neighbor-joining) with the *CHI* phylogeny (computed using maximum likelihood) generated using phylo.io (https://phylo.io/). In the color-coded similarity scale (0 to 1), a score of 1 denotes identical subtree structures of corresponding nodes in the 2 trees. An expanded view of the 2 trees with bootstrap support and details of species notations used are provided in Supporting Information (Table 5S, Fig. 2S). **b**) NJ-clustering of reaction path-based FBP matrix using Pearson's correlation coefficient. Enzymes corresponding to the reaction paths (x1–x81, Table 4S) are indicated by names and Enzyme Commission (EC) numbers. Flavonoid groups biosynthesized from the reactions are marked within boxes. Details of the reaction path notations (x1 to x81) are provided in Table 4S, supporting information.

in plants [32] and orthogroup-based evolutionary pattern of anthocyanin biosynthesis pathway genes [33]. The grouping of the precursor phenylpropanoid pathway enzymes (CCoAOMT, CYP73A, and HCT) and FBP entry-point enzymes (CHI and CHS) at the root of over 90 species (Table 5S, Supporting Information)) including *P. patens* suggested their early evolution in lower plants. Interestingly, reaction paths involving flavonol (90 species) and flavanone biosynthesis (89 species) were found to be clustered close to these basal groups, thereby providing an evolutionary basis for the almost ubiquitous presence of these compounds in diverse plant lineage (▶ **Fig. 5b**). Since proanthocyanidins and anthocyanidins – characterized by 5 reaction-path classes representing 6 enzymes – were classified almost at a similar distance scale, these compounds might have evolved much later than the chalcones. Species belonging to Brassicaceae, Cleomaceae, Cucurbitaceae, Orchidaceae, and Pedaliaceae, including *P. patens,* were characterized by the complete absence of the LAR reaction paths (63 species). At the same time, the reaction path that converts kaempferol to quercetin was absent in 52 species, including the basal angiosperms (Amborellaceae, Funariaceae, and Selaginellaceae) and many eudicots (Asteraceae, Brassicaceae, and Cucurbitaceae). The ubiquitous presence of the genes encoding the upstream FBP enzymes (*CHS*, *CHI*, and *FS*) across taxa and family-specific loss of downstream FBP genes (*AS* and *LAR*) suggest that the upstream genes have a greater effect over FBP evolution than the downstream ones, which is a hallmark of adaptive evolution [34].

Conclusion

Here, we demonstrated the potential of integrating genomic tools and metabolite screening in characterizing the FBPs. We showed that a complete core FBP is present in the 2 cultivated jute species. Modulation of gene expression created a flavonol-oriented metabolic drive in *C. olitorius,* thereby increasing its nutritional and medicinal value, particularly as an MMP-2 inhibitor. In contrast, *C. capsularis* FBP is oriented toward producing chalcones and anthocyanins, which provide higher abiotic stress tolerance and pigment variation. Phylogenetic investigations suggested an adaptive evolution of the FBP in higher plants that could account for the differences in FBP modulation in these two closely related species. Finally, we describe a novel reaction-path-based clustering approach and demonstrate its utility in resolving the evolution of the flavonoid groups.

Materials and Methods

Chemical and reagents

All the chemicals and reagents used in the experiments were of analytical purity. Compounds used as standards, DPBA, and EGCG were purchased from Sigma-Aldrich. FastStart Essential DNA Green Master was procured from Roche Life Science. Kits for RNA isolation and first-strand cDNA synthesis were obtained from ThermoFisher Scientific.

Plant material and growth conditions

We used the 2 cultivated jute species, *C. capsularis* cv. JRC-212 and *C. olitorius* cv. JRO-524, for gene expression studies and metabolite profiling. Their pure (breeder) seeds were obtained from Central Seed Research Station for Jute & Allied Fibres, Burdwan, India and selfed in isolation to develop single-plant progenies. Plants were grown in randomized blocks with 3 replications at the research farm of ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, India, and recommended crop management practices were followed to raise a healthy crop [13]. Leaves harvested from 30-day-old plants were dried in an incubator to constant weight at 50 °C for 24 h. Dried samples were ground to a fine mesh and processed for flavonoids extraction.

Flavonoids characterization

Flavonoids were extracted from leaf (50 mg dry weight) following Ola et al. [9] in 10 ml of water/methanol (1:10, v/v) as a solvent in a shaker-incubator (28 °C for 24 h, 50 rpm). The residue was re-extracted, and the extracted samples were pooled and stored at −20 °C until further analysis. Samples were filtered using a 0.2 µm nylon syringe filter (Phenomenex), and the filtrate was loaded into an HPLC (Agilent 1220) coupled with a C-18 RP column (125 × 4 mm) and photo-diode array detector (WL 280 nm). The separation was performed with a gradient of formic acid in water (0.1 %, v/v) (solvent A) and in acetonitrile (0.1 %, v/v) (solvent B), with a flow rate of 1 mL min^{-1} and column temperature of 35 °C. The gradient program was as follows: 15 % B A linear, 0–12 min; 50 % B A linear, 12–35 min; 85 % B A linear, 35–45 min; 15 % B A linear, 45–50 min, and a final plateau of 10 min. Chromatograms of 11 standard phenolic and flavonoid compounds (purity > 98 %) were developed for the identification of chromatographic peaks based on their comparative retention times and quantified using the calibration curve (standards) as suggested [29]. For localization of flavonoids in different tissues, free-hand cross-sections were treated with saturated (0.25 %) DPBA for 5–15 min in the dark [39], visualized under a fluorescent microscope (Olympus) fitted with an FITC filter (excitation 450–490 nm, suppression LP 515 nm) and photographed immediately with CCD camera (Teledyne Qimaging). Thirty representative samples were examined for each tissue, and the flavonoid classes were identified based on a color code [39].

Gene identification and pathway construction

Bast transcriptome of *C. capsularis* cv. JRC-212 (DDBJ/EMBL/GenBank accession: GBSD00000000.1), as characterized earlier [35], was used to identify the unigenes encoding the FBP enzymes followed by their validation using its corresponding hypocotyl transcriptome (DDBJ/EMBL/GenBank accession: GCNR00000000.1) [36]. Corresponding genomic regions were identified from *C. capsularis* cv. CVL-1 (accession: AWWV01000000.1) and *C. olitorius* cv. JRO-524 (accession: LLWS00000000.1) genome assemblies using NCBI BLAST (E-value < 10^{-5}). Conserved amino acid motifs, protein domains, active sites, and gene ontology functions of their annotated proteins were identified using InterPro (https://www.ebi.ac.uk/interpro). The selected unigenes were mapped to the KEGG database using BLASTx (E-value < 10^{-5}) [37]. Finally, the jute FBP was reconstructed using the KEGG pathway KO0941.

Gene expression analysis

Total RNA was extracted from 20-day-old JRC-212 and JRO-524 using TRIzol and PureLink RNA Mini Kit (Invitrogen), according to Chakraborty et al. [35]. After isolation, RNA was treated with RNase-Free DNase for 30 min at 37 °C, quantified using NanoDrop 8000 UV–Vis Spectrophotometer (Thermo Fisher Scientific) and tested for quality using Agilent's RNA 6000 Pico Kit. The RNA samples were reverse-transcribed using RevertAid HMinus First Strand cDNA Synthesis Kit, following the manufacturer's instructions (Thermo Fisher Scientific). Five randomly selected unigenes mapped in KEGG FBP, viz., *CcCHS2*, *CcCHI2*, *CcFLS1*, *CcHCT8*, and *CcANS2* were selected for differential gene expression studies. Their primers were designed (Table 3S, Supporting Information) using the default options of Primer3 [38], synthesized and tested for efficiency in a triplicate reaction set using 1.0, 2.0, and 5.0 ng primer and finally checked for primer-dimer formation by melting-curve analysis. Corresponding cDNAs were synthesized from DNase-treated RNA samples of both the species using a RevetAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher Scientific) and checked in agarose gel. The qRT-PCRs were performed in 20-µl reaction volumes (cDNA and primers: 1 µL each; FastStart Essential DNA Green Master (Roche Life Science): 10 µL), with 3 biological replicates for each treatment and 3 technical replicates per reaction on a LightCycler 480 (Roche Diagnostics Corporation) platform using 18 S rRNA (housekeeping) as an endogenous control. Relative gene expression in *C. olitorius* in comparison with *C. capsularis* was estimated according to Satya et al. [13].

MMP-2 inhibition assay in vitro

The methanolic extract was dried in a vacuum evaporator and dissolved in an equal volume of sterile water. For MMP-2 inhibition assay *in vitro*, the efficacy of this aqueous extract was tested on MMP-2 previously isolated from saliva samples collected from breast cancer patients [40]. The MMP-2 activity was measured by gelatin zymography, according to Toth and Fridman [41] and Bhattacharyya et al. [40]. In brief, about 25 ng of human MMP-2 was mixed with 30 µg mL^{-1} of EGCG (positive control) (purity > 99 %) or jute leaf aqueous extract (treatment) or distilled water (negative control) and loaded into a gelatin (0.1 %)-impregnated PAGE (8 % polyacrylamide). The gel was run at 15 mA using Tris/Glycine/SDS buffer (pH 8.3), washed in 2.5 % Triton X-100 for 15 min, incubated overnight in a buffer solution (pH 7.4) containing 0.2 M NaCl, 4.5 mM $CaCl_2$ and 50 mM Tris at 37 °C followed by staining with Coomassie Brilliant Blue to develop the zymogram. The presence of MMP-2 was confirmed by immunoblotting using human/mouse MMP-2 (Santa Cruz) as primary and alkaline phosphatase-coupled Anti-MMP-2 (Santa Cruz) as secondary monoclonal antibodies [40, 42]. The color was developed using NBT/BCIP (Roche).

Construction of KEGG FBP-profile and phylogeny reconstruction using the FBP matrix

As distance-based phylogeny based on the presence/absence profiles of enzymes is particularly suitable for secondary metabolite pathways [21], we constructed a binary FBP matrix ***N*** (1 = presence of reaction enzyme and 0 = absence of reaction enzyme in the path) of the size $n_i \times n_j$, where n_i is the number of species present in the KEGG FBP repertoire (https://www.genome.jp/kegg-bin/show_pathway?map00941) as on 29.10.2020; $i = 1, ..., 93$ including the 2 jute species) and n_j is the number of reaction paths ($j = 1, ..., 81$), according to Heymans and Singh [24]. However, instead of their EC (enzyme commission) number-based matrix [24], we used the reaction path as a variable because different enzymes having the same EC might not be orthologous, while the enzymes catalyzing the same reaction path in the KEGG pathway are orthologous. The CHS-catalyzed reactions in each species were considered to be present (1) by default. We then built a species similarity matrix (***SS***) using Jaccard's similarity coefficient from ***N*** and reconstructed a phylogenetic tree (FBP-tree) using neighbor-joining (NJ), with 10,000 bootstraps using PAST v4.03 [43]. The species tree was rooted using *Physcomitrium patens* as an outgroup.

Gene sequence-based phylogeny reconstruction

To compare the pathway-based *versus* gene sequence-based evolution, we reconstructed a sequence-based phylogenetic tree of *CHI*, a key gene involved in the early FBP reaction paths. We selected 93 *CHI* orthologs as listed in KEGG (one from each species), aligned their amino acid sequences using MUSCLE manually removing the gaps, trimmed the alignment with BMGE and reconstructed an ML tree (CHI-tree), with *P. patens* as an outgroup, using discrete gamma model and 1,000 bootstraps in PhyML 3.0 as implemented in NGPhylogeny.fr [44]. The topological similarities between the FBP-tree and the CHI-tree were visually compared using tree comparison using the best corresponding nodes implemented in Phylo.io [45], a web-based tree comparison tool. Robinson-Foulds distance [46] was calculated to quantify the similarity of the 2 trees.

Clustering of flavonoid groups

To understand the evolution of the flavonoid groups, we clustered the reaction paths taking species as a variable. Since the number of reaction paths was different for each species, we created a normalized distance matrix (***N'***) of the reaction paths by dividing the species scores (n_{ij}) of the original distance matrix (***N)*** with the mean value of its path. As ***N'***was no longer binary, it was converted to a Pearson correlation coefficient-based distance matrix to compute a neighbor-joining (NJ) tree with 10,000 bootstraps using PAST v4.03 [43]. The tree was rooted from the entry-point reaction of the KEGG FBP (cinnamoyl-CoA to coumaroyl-CoA).

Statistical analysis

The gene identification and pathway characterization studies were performed using stringent e-values (expect value, $E < 10^{-5}$) following the BLAST program guideline. Gene expression (3 replicates) was compared using relative gene expression over the control using Pfaffl's "E-method" [36]. The results of HPLC analysis (4 replicates) are presented as the mean ± SD (SEM, 95 %). Statistical analyses for diversity and phylogeny were performed using likelihood test and bootstrap analysis, as described in the previous section.

Supporting Information

Text ST1: Outline of the jute FBP; Table 1S: KEGG mapped FBP genes identified from *C. capsularis* transcriptome; Table 2S: Genomic locations of KEGG mapped FBP genes; Table 3S. Primers used for ex-

pression analysis of selected FBP genes; Table 4S: FBP Reaction path notations; Table 5S: Linnaean classification of the species used for FBP matrix construction and phylogenetic analysis; Fig. 1S. Expanded view of the FBP-phylogeny; Fig. 2S: Expanded view of the *CHI* phylogeny.

Acknowledgments

This study was supported by the Indian Council of Agricultural Research under the project nos. ICAR-CRIJAF-JB 10.4 and ICAR-NPTC-3070.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] Falcone Ferreyra ML, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci 2012; 3: 222

[2] Kumar A, Kumar S, Bains S, Vaidya V, Singh B, Kaur R, Kaur J, Singh K. De novo transcriptome analysis revealed genes involved in flavonoid and vitamin c biosynthesis in *Phyllanthus emblica* (L.). Front Plant Sci 2016; 7: 1610

[3] Raffa D, Maggio B, Raimondi MV, Plescia F, Daidone G. Recent discoveries of anticancer flavonoids. Eur J Med Chem 2017; 142: 213–228

[4] Russo M, Moccia S, Spagnuolo C, Tedesco I, Russo GL. Roles of flavonoids against coronavirus infection. Chem-Biol Interac 2020; 328: 109211

[5] Nyadanu D, AduAmoah R, Kwarteng AO, Akromah R, Aboagye LM, Adu-Dapaah H, Dansi A, Lotsu F, Tsama A. Domestication of jute mallow (*Corchorus olitorius* L.): ethnobotany, production constraints and phenomics of local cultivars in Ghana. Genet Resour Crop Evol 2017; 64: 1313–1329

[6] Azuma K, Nakayama M, Koshioka M, Ippoushi K, Yamaguchi Y, Kohata K, Yamauchi Y, Ito H, Higashio H. Phenolic antioxidants from the leaves of *Corchorus olitorius* L. J Agric Food Chem 1999; 47: 3963–3966

[7] Maeda G, Takara K, Wada K, Oki T, Masuda M, Ichiba T, Chuda Y, Ono H, Suda I. Evaluation of antioxidant activity of vegetables from Okinawa prefecture and determination of some antioxidative compounds. Food Sci Technol Res 2006; 12: 8–14

[8] Park HY, Oh MJ, Kim Y, Choi I. Immunomodulatory activities of *Corchorus olitorius* leaf extract: Beneficial effects in macrophage and NK cell activation immunosuppressed mice. J Funct Foods 2018; 46: 220–226

[9] Ola SS, Catia G, Marzia I, Francesco FV, Afolabi AA, Nadia M. HPLC/DAD/MS characterisation and analysis of flavonoids and cynnamoil derivatives in four Nigerian green-leafy vegetables. Food Chem 2009; 115: 1568–1574

[10] Yakoub ARB, Abdehedi O, Jridi M, Elfalleh W, Nasri M, Ferchichi A. Flavonoids, phenols, antioxidant, and antimicrobial activities in various extracts from Tossa jute leave (*Corchorus olitorius* L.). Ind Crops Prod 2018; 118: 206–213

[11] Robinson AR, Gheneim R, Kozak RA, Ellis DD, Mansfield SD. The potential of metabolite profiling as a selection tool for genotype discrimination in *Populus*. J Exp Bot 2005; 56: 2807–2819

[12] Stushnoff C, Ducreux LJM, Hancock RD, Hedley PE, Holm DG, McDougall GJ, McNicol JW, Morris J, Morris WL, Sungurtas JA, Verrall SR, Zuber T, Taylor MA. Flavonoid profiling and transcriptome analysis reveals new gene-metabolite correlations in tubers of *Solanum tuberosum* L. J Exp Bot 2010; 61: 1225–1238

[13] Satya P, Sarkar D, Vijayan J, Ray S, Ray DP, Mandal NA, Roy S, Sharma S, Bera A, Kar CS, Mitra J, Singh NK. Pectin biosynthesis pathways are adapted to higher rhamnogalacturonan formation in lignocellulosic jute (*Corchorus spp*.). Plant Growth Regul 2020; 93: 131–147

[14] Furumoto T, Wang R, Okazaki K, Hasan AFMF, Ali MI, Kondo A, Fukui H. Antitumor promoters in leaves of jute (*Corchorus capsularis* and *Corchorus olitorius*). Food Sci Technol Res 2002; 8: 239–243

[15] Taiwo BJ, Taiwo GO, Olubiyi OO, Fatokun AA. Polyphenolic compounds with anti-tumour potential from *Corchorus olitorius* (L.) Tiliaceae, a Nigerian leaf vegetable. Bioorganic Med Chem Lett 2016; 26: 3404–3410

[16] Roy R, Yang J, Moses MA. Matrix metalloproteinases as novel biomarkers and potential therapeutic targets in human cancer. J Clin Oncol 2009; 27: 5287–5297

[17] Demeule M, Brossard M, Pagé M, Gingras D, Béliveau R. Matrix metalloproteinase inhibition by green tea catechins. Biochim Biophys Acta 2000; 1478: 51–60

[18] Markham KR. Distribution of flavonoids in the lower plants and its evolutionary significance. In: Harborne JB, ed. The Flavonoids. Advances in Research Since 1980. London: Chapman and Hall; 1988: 427–468

[19] Campanella JJ, Smalley JV, Dempsey ME. A phylogenetic examination of the primary anthocyanin production pathway of the Plantae. Bot Stud 2014; 55: 10

[20] Rausher MD, Miller RE, Tiffin P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol 1999; 16: 266–274

[21] Weißenborn S, Walther D. Metabolic pathway assignment of plant genes based on phylogenetic profiling – a feasibility study. Front Plant Sci 2017; 8: 1831

[22] Hong SH, Kim TY, Lee SY. Phylogenetic analysis based on genome-scale metabolic pathway reaction content. Appl Microbiol Biotechnol 2004; 65: 203–210

[23] Forst CV, Schulten K. Phylogenetic analysis of metabolic pathways. J Mol Evol 2001; 52: 471–489

[24] Heymans M, Singh AK. Deriving phylogenetic trees from the similarity analysis of metabolic pathways. Bioinformatics 2003; 19: i138–i146

[25] Basak SL. Quantitative genetics of fibre yield and its components. In: Denton IR, ed. Review on the Genetics and Breeding of Jute. Dhaka: International Jute Organization; 1993: 51–95

[26] Roy S, Lutfar LB. Bast fibres: jute. In: Kozlowski RM, ed. Handbook of Natural Fibres: Volume 1 Types, Properties and Factors Affecting Breeding and Cultivation. Cambridge: Woodhead Publishing Limited; 2012: 24–46

[27] Arai Y, Watanabe S, Kimira M, Shimoi K, Mochizuki R, Kinae N. Dietary intakes of flavonols, flavones and isoflavones by Japanese women and the inverse correlation between quercetin intake and plasma LDL cholesterol concentration. J Nutr 2000; 130: 2243–2250

[28] Jiang X, Shi Y, Fu Z, Li WW, Lai S, Wu Y, Wang Y, Liu Y, Gao L, Xia T. Functional characterization of three flavonol synthase genes from *Camellia sinensis*: roles in flavonol accumulation. Plant Sci 2020; 300: 11632

[29] Kong CS, Kim YA, Kim MM, Park JS, Kim JA, Kim SK, Lee BJ, Nam TJ, Seo Y. Flavonoid glycosides isolated from *Salicornia herbacea* inhibit matrix metalloproteinase in HT1080 cells. Toxicol in Vitro 2008; 22: 1742–1748

[30] Pereira SC, Parente JM, Belo VA, Mendes AS, Gonzaga NA, do Vale GT, Ceron CS, Tanus-Santos JE, Tirapelli CR, Castro MM. Quercetin decreases the activity of matrix metalloproteinase-2 and ameliorates vascular remodeling in renovascular hypertension. Atherosclerosis 2018; 270: 146e153

[31] APG III An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc 2009; 161: 105–121

[32] Rausher MD. The evolution of flavonoids and their genes. In: Grotewold E, ed. The Science of Flavonoids. Springer; 2006: 175–211

[33] Piatkowski BT, Imwattana K, Tripp EA, Weston DJ, Healey A, Schmutz J, Shaw AJ. Phylogenomics reveals convergent evolution of red-violet coloration in land plants and the origins of the anthocyanin biosynthetic pathway. Mol Phylogenet Evol 2020; 151: 106904

[34] Wright KM, Rausher MD. The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics 2010; 184: 483–502

[35] Chakraborty A, Sarkar D, Satya P, Karmakar PG, Singh NK. Pathways associated with lignin biosynthesis in lignomaniac jute fibres. Mol Genet Genomics 2015; 290: 1523–1542

[36] Satya P, Chakraborty A, Sarkar D, Karan M, Das D, Mandal NA, Saha D, Datta S, Ray S, Kar CS, Karmakar PG, Singh NK. Transcriptome profiling uncovers *β*-galactosidases of diverse domain classes influencing hypocotyl development in jute (*Corchorus capsularis* L.). Phytochemistry 2018; 156: 20–32

[37] Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 2007; 35: W182–W185

[38] Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3-new capabilities and interfaces. Nucleic Acids Res 2012; 40: e115

[39] Peer WA, Brown DE, Tague BW, Muday GK, Taiz L, Murphy AS. Flavonoid accumulation patterns of transparent testa mutants of *Arabidopsis*. Plant Physiol 2001; 126: 536–548

[40] Bhattacharyya N, Mondal S, Ali MN, Mukherjee R, Adhikari A, Chatterjee A. Activated salivary MMP-2 – a potential breast cancer marker. Open Confer Proc J 2017; 8: 22–32

[41] Toth M, Fridman R. Assessment of gelatinases (MMP-2 and MMP-9) by gelatin zymography. Methods Mol Med 2001; 57: 163–174

[42] Mondal S, Bardhan K, Dutta A, Chatterjee A. Identification of vertebrate MMP-2 and MMP-9 like molecules in the aqueous extract of nasturtium (*Tropaeolum majus*) flowers, *Bambusa balcooa* Leaves and nayantara (*Catharanthus roseus*) flowers. J Tumor 2018; 6: 540–544

[43] Hammer Ø, Harper DAT, Ryan PD. Past: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 2001; 4: 1–9

[44] Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 2018; 556: 452–456

[45] Robinson O, Dylus D, Dessimoz C. Phylo.io: Interactive viewing and comparison of large phylogenetic trees on the web. Mol Biol Evol 2016; 33: 2163–2166

[46] Robinson DF, Foulds LR. Comparison of phylogenetic trees. Math Biosci 1981; 53: 131–147