

Capturing Acyltransferase(s) Transforming Final Step in the Biosynthesis of a Major Iridoid Glycoside, (Picroside-II) in a Himalayan Medicinal Herb, Picrorhiza kurroa

ANJALI KHARB

Bennett University

Shilpa Sharma

Bennett University

Ashish Sharma

Bennett University

Neeti Nirwal

Bennett University

Roma Pandey

Bennett University

Dipto Bhattacharyya

Bennett University

Rajinder Singh Chauhan (✉ rajinder.chauhan@bennett.edu.in)

Bennett University <https://orcid.org/0000-0002-1997-6543>

Research Article

Keywords: Picroside-II, Acyltransferase, RNA-seq, qPCR, Protein structure modelling, Docking

Posted Date: November 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1009658/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Picrorhiza kurroa has been reported as an age-old ayurvedic hepatoprotection to treat hepatic disorders due to the presence of iridoids such as picroside-II (P-II), picroside-I, and kutkoside. The acylation of catalpol and vanilloyl coenzyme A by acyltransferases (ATs) is critical step in P-II biosynthesis. Since accumulation of P-II occurs only in roots, rhizomes and stolons, uprooting of this critically endangered herb has been the only source of this compound. Recently, we reported that P-II acylation likely happen in roots, while stolons serve as the vital P-II storage compartment. Therefore, developing an alternate engineered platform for P-II biosynthesis require identification of P-II specific AT/s.

Methods and results

In that direction, egg-NOG function annotated 815 ATs from *de novo* RNA sequencing of tissue culture based 'shoots-only' system and nursery grown shoots, roots, and stolons varying in P-II content, were cross-compared *in silico* to arrive at ATs sequences unique and/or common to stolons and roots. Verification for organ and accession-wise upregulation in gene expression of these ATs by qPCR has shortlisted six putative 'P-II-forming' ATs. Further, six-frame translation, *ab initio* protein structure modelling and protein-ligand molecular docking of these ATs signified one MBOAT domain containing AT with preferential binding to the vanillic acid CoA thiol ester as well as with P-II., implying that this could be potential AT decorating final structure of P-II.

Conclusion

Organ-wise comparative transcriptome mining coupled with reverse transcription real time qPCR and protein-ligand docking led to the identification of an acyltransferases, contributing to the final structure of P-II.

Introduction

Picrorhiza kurroa Royle ex Benth, Scrophulariaceae family has been reported as a major herbal therapeutic for liver diseases [1–3]. Annual trade of 1000-2000 metric tons of *P. kurroa* has been reported by the National Medicinal Plants Board (Ministry of AYUSH, Government of India <https://www.nmpb.nic.in/>), which include raw material for herbal drugs as well as for foreign export of dried *P. kurroa*. Unfortunately, traders whether legal or illegal, used to uproot the herb along with roots. This erratic procedure minimized chances of regeneration of new herbs to an extent of having officially been declared as an endangered or critically endangered species [4]. Some regions where *P. kurroa* naturally thrives are the Tungnath Mountains in the Garhwal region and in the Kumaon Mountains [5, 6]. *P. kurroa* is a perennial herb and is normally found growing inside rocky crevices and cliffy mountains in the Himalayan ranges [7]. Noticeable characteristics where *P. kurroa* grows include nearby infestation of bryophytic vegetation or matted shrub vegetation, mountains with comparatively high soil moisture and

high soil organic matter, but with lowest possible total soil nitrogen availability and spatial remoteness to higher land plants of vascular origin [6]. *P. kurroa* biosynthesize and harbour a rich repertoire of terpenoids having diverse decorations of the main backbone and one among the superior most terpenoid structure modifications is the acylation of bicyclic monoterpenes with phenylpropanoids or phenolics to form iridoid glycosides [1, 2]. The bryophytes too are recognized for their voluminous store of terpenoids [8]. It is not unusual to infer that these terpenoids act as allelopathic and/ or symbiotic ground microbial interactors and assist in the survival of the herb or the bryophytes [9]. More strikingly, inclusion of nitrogen in these terpenoid structures are less profound or missing in comparison to monoterpenoids indole alkaloid structures commonly found in medicinal plants like *Catharanthus roseus* that resides in warmer climates.

Picroside-II (P-II, amphicoside) is a major iridoid found in stolons and roots of *P. kurroa* [10]. P-II has a vanilloyl moiety acylated to C-6 of a catalpol (CAT) moiety and is also commercially referred to as 6-vanillyl catalpol (<https://www.sigmaaldrich.com/DE/en/search/amphicoside>). P-II serves as the dominant active principle in hepatoprotective herbal formulations of *P. kurroa*, which also includes picroside-I (P-I, 6'-cinnamoyl catalpol) and kutkoside as the other active chemical ingredients [1]. In addition P-II has been demonstrated to possess plethora of pharmacological activities (Table S1). The last committed step in P-II biosynthesis is the acylation of vanillic acid (VA) and CAT [11, 12]. Considering pharmacological viewpoint, this acylation between the monoterpene bicyclic iridoid CAT and the phenolic acid VA has conferred the bioactive properties in P-II. Therefore, identification of acyltransferase's (ATs) specific to P-II biosynthesis is important for engineering and developing an exclusive production platform for P-II.

Shoots grown in tissue culture at 15°C (PKS15) and 25°C (PKS25) and shoots from field-grown plants (PKSS) usually contain only P-I, while P-II is mainly detected in field-grown roots (PKSR) and stolons (PKSTS). *P. kurroa* stolon is the only organ that accumulates significant amounts of both P-I and P-II. *De novo* next generation RNA sequencing (RNA-seq) of PKS15, PKS25, PKSS, PKSR, and PKSTS samples has been recently reported [1]. In the present study, the annotated transcriptomes were thoroughly mined for transcripts annotated as acyltransferases (ATs). The criteria of *in silico* differential expression (transcripts per million values of RNA-seq reads, TPM) of ATs transcripts in transcriptome samples of shoot, root and stolons from *P. kurroa* coupled with their transcript abundance by reverse transcription real-time quantitative PCR (qPCR) have been recently utilized to identify gene paralogues of picrosides biosynthetic pathways [2]. Further, gene co-expression sub-networks generated from RNA-seq of organs varying in P-I and P-II accumulation in *P. kurroa* have revealed crucial components of picrosides biosynthetic machinery [1]. Using similar approach, bioinformatics sorting of ATs common or unique between any two transcriptomes varying in P-I and P-II contents have been carried out in the present investigation. ATs shortlisted for acylation of P-II were further analyzed for their real time organ-wise RNA expression profiles by qPCR. Abundance of gene expression of these transcripts was further verified in *P. kurroa* accessions collected from wild habitat of *P. kurroa* exhibiting 89.33 fold variations in P-II content (fold change determined from fresh weight %). Transcripts short-listed from above criterion were

translated *in silico* by six-frame translation and were modelled into three dimensional proteins by protein structure modelling to observe for protein to acyl donor, protein to acyl acceptor, and protein to P-I/ P-II interactions. This study is a step-wise approach of utilizing parameters for sorting ATs specific to P-II biosynthesis. Thus an initial comparison and TPM based bioinformatic sorting of ATs have been accompanied with additional reduction of possibilities by qPCR screening for higher transcript abundance and finally cross-verified by molecular 'protein to ligand' docking interactions. Both P-II and the acyl donor vanilloyl CoA exhibited preferential binding to protein structure modelled proteins with membrane-bound O-AT domain (membrane-bound O-acyltransferase family, MBOAT domain).

Materials And Methods

Mining of putative acyltransferase(s) from NGS transcriptome datasets

For extracting and identification of putative ATs that catalyse the final step of picosides biosynthesis in *P. kurroa*, five transcriptome datasets viz. PKS-25, PKS-15, PKSS, PKSTS, and PKSR varying for picosides content were computationally searched for sequences that have putative AT function.

Alignment analysis of mined transcripts from 5 transcriptome datasets using BLASTn with criteria of >75% query coverage and >90% was performed to pull out sequences in different permutations and combinations of the datasets. Comparative analysis was performed using five transcriptome datasets to shortlist acyltransferase(s) corresponding to P-I and P-II biosynthesis. Comparison of stolon (P-I & P-II) and roots (P-II) have identified acyltransferase(s) that play possible role in P-II biosynthesis. Unique acyltransferase(s) were extracted and pooled together with ATs that were common in stolons (PKSTS) and roots (PKSR) with respect to P-II (Fig. 1).

Collection of plant material

Field-grown plants of *P. kurroa* were collected from experimental nursery established in experimental research station of Himalayan Forest Research Institute at Jagatsukh, H.P. India (2193 m, altitude; 32°12'0N and 77°12'0E). The plants were harvested, washed with tap water and segregated into shoots, stolons, and roots. All samples were stored at -80°C for use in future experiments (for example, in qPCR and high-performance liquid chromatography (HPLC)).

RNA extraction and cDNA synthesis

Total RNA was isolated from *P. kurroa* tissues by using Takara total RNA isolation kit (Nucleospin RNA Plant) by following the manufacturer's instructions. The isolated RNA was stored at -80°C and the quality of RNA was checked by 1% (w/v) ethidium bromide-stained agarose gel and visualized by observing UV fluorescence through the gel documentation system (Biorad). RNA was quantified in a Nano Drop Multiskan GO (Thermo Scientific) by measuring absorbance at 260 nm and 260/280 nm ratio. The synthesis of cDNA was carried out with 1µg RNA using a Takara cDNA synthesis kit according to the

manufacturer's protocol. The concentration of each cDNA sample was adjusted to equal volume and subjected to gene expression analysis.

Validation of gene expression using qPCR analysis

Gene expression profiles of ATs were checked by qPCR in PKSS, PKS15, PKS25, PKSTS, PKSR, PKR84 (*P. kurroa* accession having high P-II content in roots), and PKR31 (*P. kurroa* accession having low P-II content in roots). For qPCR, equal volumes of cDNA of each sample were subjected to qPCR on a CFX96 system (Bio-Rad Laboratories) using AT partial transcript-specific primers (Suppl. File 1A) and Hi-SYBr master mix (with Taq Polymerase) (Hi-media) in biological duplicates where each biological replicate consisted of three technical replicates. The qPCR reaction mix (10 μ L) for AT expression analysis consisted of 0.3 μ L forward primer (10 pmol), 0.3 μ L reverse primer (10 pmol), 5 μ L Hi-SYBr master mix (Hi-media), 1 μ L cDNA, and 3.4 μ L nuclease-free water. The qPCR cycles involved an initial denaturation of 3 min at 95°C, followed by 39 cycles of denaturation at 95°C for 10 seconds (s), annealing for 30 s at 50 – 65°C, and elongation for 20 s at 72°C. In the final step, melt curve analysis was done at 65-95°C with a 0.5°C increment at each 0.05 s to verify the amplification of single products. Expression levels of each sample were normalized relative to 26S rRNA as an internal control and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method [13].

Quantification of picrosides through HPLC analysis

P. kurroa accessions PKR84 and PKR31 were collected and stored at -80°C and were further evaluated for their organ-specific P-II content using reverse phase HPLC. Representative HPLC chromatograms showing P-II from root samples of PKR84 and PKR31 are given alongside a chromatogram of P-II standard obtained from Sigma (G0174). P-II retention times in the HPLC runs ranged at 10.5 \pm 0.5 min. Briefly, *P. kurroa* root samples were ground to fine powder using a mortar and pestle in liquid nitrogen. Powdered samples (100.0 mg) were extracted in 10.0 mL 80% HPLC grade aqueous methanol. The samples were then vortexed to mix properly and extracted overnight at room temperature (~25°C). Samples were centrifuged at 10,000 rpm for 15 min. and the supernatant was filtered through 0.22 μ m filter (Millipore) and used for HPLC analysis [14]. A reverse phase HPLC system consisting of Sunfire C18 μ m (Waters, 4.6 \times 250 mm, 5 μ m) and photodiode array detector (PDA; Waters 2998) was used for detection of P-II at 280 nm. The isocratic mobile phase consisted of HPLC grade water (Molecular grade water Aldo from Merck), methanol (Merck, USA), and acetonitrile (Merck, USA) in the ratio of 70:15:15, respectively and 0.03% trifluoroacetic acid. A flow rate of 1.0 mL min⁻¹ was maintained throughout the run. Sample volume of 20 μ L was injected and HPLC run was conducted for 30 min at 30°C. The extraction and quantification experiments were performed in triplicates.

Molecular Modelling and Docking

The nucleotide sequences (partial transcripts) were translated using Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/) and further used as input files for Pfam (<http://pfam.xfam.org/>) to identify domain and family of protein sequence. Three dimensional (3-

D) structures of these sequences/ domains were obtained by *ab initio* modelling using I-Tasser (<https://zhanglab.dcmf.med.umich.edu/I-TASSER/>). Models were chosen based on their Z-score, Qmean and C-value. After protein structure modeling, energy minimization was done using Swiss PDB Viewer and structure validation was performed on Ramachandran plot server (<https://zlab.umassmed.edu/bu/rama/>). The six protein structures and 8 ligands used in this study were prepared for docking analysis using AutoDock Tools (<http://autodock.scripps.edu/resources/adt>). Molecular Docking was performed in triplicates and average of corresponding values was considered as final score. All 8 ligands were screened simultaneously with six proteins separately. The grid was generated by looking into the active binding sites of template proteins as indicated by I-TASSER tool. Different conformations of Protein-ligand binding complexes were generated. Molecules with lowest binding energy (highest binding affinity) and comparatively predominant interactions were chosen for further analysis. Bond interactions within protein-ligand complexes were visualized using LigPlot.

Table 1 Total number of putative Acyltransferase transcripts retrieved from five transcriptome datasets from different tissues *Picrorhiza kurroa*

S. No.	Transcriptome Datasets	No. of Acyltransferase
1	PKS25	154
2	PKS15	161
3	PKSS	132
4	PKSTS	180
5	PKSR	188
	TOTAL	815

PKS15: *P. kurroa* tissue culture shoots grown at 15°C;

PKS25: *P. kurroa* tissue culture shoots grown at 25°C;

PKSR: *P. kurroa* roots from field-grown plants;

PKSS: *P. kurroa* shoots from field-grown plants;

PKSTS: *P. kurroa* stolons from field-grown plants

Results

In silico sorting of Picroside-II specific Acyltransferases in *P. kurroa*

The eggNOG function annotated RNA-seq files pertaining to five transcriptomes under study were thoroughly searched for the general term 'acyltransferase'. A sum-total of 815 AT FASTA (FAST All) sequences corresponding to unique identifiers were retrieved (Fig. 1, Fig. 2a, Table 1). BLAST alignment comparison taking pairs of individual transcriptome samples as query and subject, resulted in the identification of transcripts which were unique to stolons (Stolon uniseq) when compared with RNAseq samples corresponding to PKS15 (58 unique ATs), PKS25 (63 unique ATs), and PKSS (67 unique ATs) (Fig. 2b). Similarly BLAST alignment comparison taking pairs of individual transcriptome samples as query and subject, resulted in the identification of transcripts which were unique to PKSR (Root uniseq) when compared with RNAseq samples corresponding to PKS15 (69 unique ATs), PKS25 (70 unique ATs), and PKSS (79 unique ATs) (Fig. 2b). Additionally, 123 AT sequences were found to be common between root and stolon samples (Stolon & Root: comseq, Fig. 2a, b). An additional layer of *in silico* reductive screening was used by pair-wise cross-comparisons between unique and common AT sequences retrieved from root and stolon AT sequences (Fig. 2b). Thus, a total of 177 AT sequences specific to P-II were obtained from these comparative analyses. AT redundancy was removed by searching for duplicate sequence identifiers occurring more than one time. Manual short-selection of ATs were carried out to remove transcripts with insignificant TPM values (TPM values of transcripts > 0.5 were considered as significantly expressed and were proceeded for further qPCR analyses). A final list of 22 ATs with corresponding BLASTX identities is tabulated in Suppl. File 1B.

Shortlisting of P-II specific acyltransferases by qPCR

The partial AT transcripts were optimized for single band amplification with both genomic DNA and cDNA as templates, Hi-SYBR master mix (with Taq Polymerase) (Hi-media) was used in these PCR amplification reactions. Eight AT partial transcripts failed to amplify under all tested conditions, which is why qPCR studies were carried out with 14 partial AT transcripts. NODE_48592/618bp designated as AT_PII-10 exhibited a 6.22 ± 0.03 fold higher expression in PKSR compared to the PKSTS and PKSS samples (Fig. 3a). NODE_37534/628bp designated as AT_PII-1 exhibited 15.34 ± 0.05 fold higher expression in roots in comparison to stolon and leaf samples (Fig. 3a). NODE_59938/512bp named AT_PII-5 in Fig. 3a was found to have 3.23 ± 0.4 fold higher gene expression in root samples in comparison to stolon and leaf samples, respectively. NODE_34024/833bp designated as AT_PII-6 showed similar relative expression profiles in root and stolon samples (Fig. 3a). NODE_16506/1324bp designated as AT_PII-7 revealed 8.20 ± 0.32 fold higher expression in PKSR compared to PKSTS and PKSS (Fig. 3a). NODE_3996/2193bp or AT_PII-8 exhibited 6.03 ± 0.12 fold higher expression relative to stolon and leaf samples, respectively (Fig. 3a). NODE_11774/1549bp and NODE_107081/320bp designated as AT_PII-4 and AT_PII-14, respectively, exhibited insignificant expression in PKSR and PKSTS (Fig. 3a). Further, NODE_93830/351bp designated as AT_PII-13 displayed zero relative expression in PKSTS (Fig. 3a). Finally, six ATs were selected based on high expression in roots or stolons and BLASTN and BLASTX searches against viridiplantae these six ATs have been represented in Suppl. File 1. *P. kurroa* accessions PKR84 (2.68 % fresh weight P-II) and PKR31 (0.03 % fresh weight P-II) exhibited contrasting variation of around 89 fold in P-II content in roots (Fig. S1). qPCR analyses of the 6 selected transcripts from Fig. 3b with root cDNA of PKR84 and PKR31 exhibited significantly higher relative expression of five transcripts in PKR84 in comparison to PKR31

(Fig. 3b). AT_PII-1 exhibited 0.78 ± 0.18 fold higher expression in PKR84 in comparison to PKR31. AT_PII-5 exhibited 1.92 ± 0.30 fold higher expression in PKR84 in comparison to PKR31. AT_PII-6 exhibited 1.63 ± 0.08 fold higher expression in PKR84 in comparison to PKR31. AT_PII-7 exhibited 5.05 ± 0.02 fold higher expression in PKR84 in comparison to PKR31. AT_PII-8 exhibited 2.49 ± 0.09 fold higher expressions in PKR84 in comparison to PKR31. AT_PII-10 exhibited 2.21 ± 0.33 fold higher expression in PKR84 in comparison to PKR31 (Fig. 3b)

Protein structure modeling and molecular docking to verify specificity of acyltransferases to final structural decoration of Picroside-II

The shortest AT partial transcript spanned 512 bp (AT_PII-5) while the longest transcript was 2.193 kb in length (Suppl. File 1). Six translated coding frames were generated for each transcript using Transeq tool. Inclusion of conserved AT domain/s in the transcripts in the correct frame were confirmed by conserved domain database searches via Pfam (<http://pfam.xfam.org/>) (Suppl. File 2). *Ab initio* modeling of the translated amino acids was carried out using I-Tasser. It is imperative to mention here that since complete complementarity determining sequences (CDSs) was possibly not available in any of the partial AT transcripts, the downstream analyses after protein structure modeling will always have some degree of added inaccuracy, although since correct codon frames were carefully chosen before initiation of the *ab initio* modeling, the final 3-D protein structures will essentially retain the preferred binding pockets. Thus, the proteins with the selected partial AT transcript identifiers NODE_37534/628bp, NODE_59938/512bp, NODE_34024/833bp, NODE_16506/1324bp, NODE_3996/2193bp, and NODE_48592/618bp were individually docked with the supposed active acyl donors cinnamoyl-CoA, feruloyl CoA, and vanilloyl CoA one at a time to compare 3-D best-fit conformation as well as highest binding affinity (Suppl. File 2, Fig. 4, Fig. S2). Additionally, the probable acyl acceptor catalpol was docked to each protein separately (Suppl. File 2, Fig. S2). Further, the end products P-II, and feruloyl catalpol as well as the inactive acyl donor's cinnamic acid and VA have been docked separately (Suppl. File 2, Fig. S2). We argued that docking both the active acyl CoA thiol ester donor and acyl acceptor (CAT) onto the protein might increase inaccuracy levels of assumption; therefore, separate binding analyses might provide more adequate quantifications of pair-wise binding energies. Node_48592/618 bp (AT_PII-10, MBOAT) exhibited a highest comparative binding affinity for vanilloyl CoA, CAT, and P-II (Suppl. File 2, Fig. 4, Fig. S2). Evidently, binding affinity of vanillic acid was the lowest, suggesting absence of binding without acyl CoA activation (Suppl. File 2, Fig. 4, Fig. S2). Further, feruloyl catalpol exhibited similar binding energy values to P-II but the binding affinity of vanilloyl CoA was comparatively slightly higher than the end product iridoids, suggesting probable true binding; if we consider the fact that the activated acyl carrier vanilloyl CoA binds true to the protein with Node_48592/618 bp (AT_PII-10), while the end products docked to the protein pocket represent an instantaneous time when the bound products might be on the verge of escaping from the binding pockets after acylation (Suppl. File 2, Fig. 4, Fig. S2). It remains to be verified whether this protein is a vacuolar membrane bound protein as was predicted from root gene co-expression sub-network analysis [1], that acylation occurs in vacuoles inside the roots.

Discussion

The current study has utilized RNA-seq data from *P. kurroa* plants grown at an experimental farm located at around 1900 m altitude in Jagatsukh, Himachal Pradesh, India. A single AT could be sorted out after applying multi-step screening protocols which include *in silico* sorting of unique and common ATs from five transcriptomes exhibiting contrasting P-II contents; TPM-wise screening and removal of redundancy in transcripts, and screening by higher relative gene expression profiles of fourteen ATs in shoot, root and stolons of field-grown *P. kurroa*. ATs displaying significantly higher relative gene expression were manually picked up and transferred to prediction analysis by molecular protein-ligand docking after *ab initio* protein structure modelling of the respective proteins. Vanilloyl CoA exhibited highest binding preference to a partial AT transcript with the identifier NODE_48592/618bp. Strikingly, even P-II independently exhibited a close binding preference to NODE_48592/618bp similar to vanilloyl CoA, indicating high probability of this specific transcript to turn up as the P-II specific AT on the wet-lab bench.

Active fractions and pure compounds from *P. Kurroa* have been reported to exhibit anti-asthma, anti-inflammation, hepatoprotection, immunostimulation, and free radical scavenging properties (Table S1). These pharmacological properties are mostly attributed to the presence of a group of monoterpenoid bicyclic iridoid glycosides of catalposide class, namely, P-I, P-II, 6-feruloyl catalpol, and pikuroside [10, 15]. Among these compounds, P-II and pikuroside are found mainly in roots and rhizomes and structurally both of these compounds are constituted of two moieties, a catalpol and a vanillic acid moiety bonded in an ester linkage at C-6 of the backbone catalpol moiety. Pikuroside differs from P-II in having a 3, 10-acetal linkage in the catalpol backbone. Apart from *P. Kurroa*, similar iridoids with rigid three ring skeletons have been identified in *Catalpa bignonioides* and *Rehmannia glutinosa* [15]. After the formation of geraniol from geranyl pyrophosphate or farnesyl pyrophosphate, geraniol is cyclized to form epi-iridodial, the next step epi-iridotrial involves formation of an aldehyde group at C-4, the subsequent step is the glucosylation at C-1 and concurrent oxidation of the aldehyde to acid at C-4 of epi-iridotrial to form 8-epiloganic acid, and the next step involves the conversion to mussaenosidic acid by hydrolysis at C-8 [11, 12]. Mussaenosidic acid undergoes dehydration to yield deoxygeniposidic acid which is hydrolyzed at C-10 to form geniposidic acid [11, 12]. Geniposidic acid is decarboxylated at C-4 to form bartsioside, a penultimate hydroxylation at C-6 yields aucubin, and an epoxidation with the alcohol at C-10 leads to the biosynthesis of catalpol [11, 12].

Cinnamic acid with the formula $C_6H_5CH=CHCOOH$ or hydroxy cinnamic acids such as p-coumaric acid and caffeic acid, which are hydroxy derivatives of cinnamic acid, do not seem much suited to act as an acyl donor for P-II which is a vanilloyl catalpol (6-vanilloyl catalpol). Vanilloyl CoA (vanillic acid CoA thiol ester), obtained from the 3-methoxy-4-hydroxy derivative of ferulic acid (VA) supposedly, can act as activated acyl donor/s for acylation of CAT and VA to form P-II. Feruloyl CoA obtained from the 3-methoxy-4-hydroxy side chain unsaturated derivative of cinnamic acid (ferulic acid) will require an additional hydratase/ lyase/ dehydrogenase to form the VA moiety of P-II. According to an earlier report, feeding of radio-labelled ferulic acid and VA in *Vanilla planifolia* resulted in the bio-conversion of ferulic acid to vanillin by CoA-dependent β -oxidative cleavage-mediated pathway [16]. A modification of ferulic

acid to feruloyl-CoA, and further oxidation to vanilloyl-CoA followed by a reduction step yielded vanillin [16, 17]. Thus, vanilloyl CoA as a reported active acyl donor was utilized in our protein docking studies to investigate ATs capable of acylating CAT. Further, as an evidence for the existence of vanilloyl CoA, earlier reports have confirmed the presence of VA in roots and rhizomes of *P. kurroa* and a table (supplementary spreadsheet) representing compounds identified in *P. kurroa* roots, rhizomes and stems have been provided [1].

ATs aid natural product scaffold modification by acylating oxygen or nitrogen atoms to generate esters or amides. Among the two largely known AT families, the BAHD ATs accept CoA thioesters as a source of the acyl group, while the serine carboxypeptidase-like (SCPL) ATs frequently utilize the 1- O - β -ester of glucose (1- O - β -acyl acetals) as the acyl donor. The term BAHD is derived from the initials of the abbreviations of the first four biochemically characterized acyltransferases, Benzylalcohol O - acetyltransferase (BEAT), anthocyanin O -hydroxycinnamoyltransferase (AHCT), N - hydroxycinnamoyl/benzoyltransferase (HCBT), and deacetylvindoline 4- O -acetyltransferase (DAT). BEAT assists esterification to biosynthesize the floral volatile benzylacetate in *Clarkia breweri*, AHCT aids 5- O - acylation of anthocyanins in *Gentiana triflora*, HCBT catalyzes biosynthesis of anthramide phytoalexins in *Dianthus caryophyllus*, and DAT completes part of last step biosynthesis of vindoline in *C. roseus* [18]. Several BAHD ATs ranging from medium-chain-length aliphatic and benzyl-derived alcohols as substrates and acetyl CoA to other acyl CoAs as acyl donors have been biochemically and genetically characterized [19]. Two alcohol ATs VpAAT1 from *Vasconcellea pubescens* and CpAAT1 from *Carica papaya* that accept the substrates benzyl acetate and ethyl butanoate/ methyl butanoate, respectively, could be distinguished for their substrate size specificity based on the size and shape of the solvent channels in CpAAT1 and VpAAT1 proteins [20]. Around 110 diverse BAHD family enzymes have been reported of having been biochemically characterized and include acyl-CoA dependent acylations of substrates ranging from aliphatic/ aromatic aldehydes/ alcohols to terpenoids, alkaloids and phenolics [21]. Depending upon specific CoA thioester acyl donors or specific substrates, the BAHDs have been classified till date into eight distinct clades [22–25].

Antimicrobial triterpenoid saponin group of avenacins found in oats (*Avena* spp) have been reported to undergo O-Glc ester mediated acylation of either benzoate or N-methyl anthranilate moieties at C-21 of avenacins via SCPL ATs [26]. The authors also noted that the SCPL AT was found in a BAC contig spanning four genes within the avenacin gene cluster which included a glucosyl transferase and a N-methyltransferase responsible for the respective glucosylation and methylation of the avenacins [26]. Thus, gene machinery for biosynthesis and modification of cyclic compounds like iridoids might exist in adjacent gene clusters or in other words might possibly be co-expressed. Acylation can impact solubilization, stabilization, vacuolar uptake and color modulation of phenylpropanoids [27] as well as regulate the transport and storage of molecules [28–30] reported that ABC-type and proton gradient-driven transporters are most probably involved in vacuolar import of abscissic acid-glucose esters. Accordingly, a possibility crops up that activity tagging by glucose esterification of cyclic terpenoids like the iridoids might aid their transport into vacuoles where they might be acylated by SCPL or BAHD ATs for future storage and transport. Thus, these reports serve as vital clues to investigate gene co-expression

based networks in future for sorting out ATs required for P-II biosynthesis. Furthermore, engineering ATs in plants might lead to the biosynthesis of substrate-dependent acetates in plants inducing characteristic aromas such as in transgenic lisianthus plants, which are ornamental plants with beautiful but scentless flowers [31]. Moving even further, [32] had preconceived that it is highly likely that specific ATs might determine alkaloid patterns in particular plant species. Similar activities of alkaloid ATs were detected in *Lupinus* species and accessions largely varying in alkaloid contents and were interpreted accordingly that the allied gene expression might not affect concentrations of alkaloids in cells. Thus, specific ATs might determine presence of iridoids such as picosides in plants like *P. kurroa*, *C. bignonioides* and in *R. glutinosa*.

Normally, isotope studies and other chromatography-oriented procedures were used in the past decades to find out the nature of CoA thiol esters required for a particular scaffold generation in a secondary metabolite such as a monoterpene of interest. By far, p-coumaroyl CoA, cinnamoyl CoA, benzoyl CoA, feruloyl CoA, hydroxy cinnamoyl CoA, 3-hydroxybenzoyl-CoA, 3-hydroxy-3-phenylpropanoyl-CoA, 3-oxo-3-phenylpropanoyl-CoA, and anthraniloyl CoA are the known active acyl CoA donors in plants that have been reported to mediate acyl transfer reactions upon catalysis by specific ATs [33]. Kinetic study of a ion-exchange and gel-filtration chromatography purified vegetal acyclic monoterpene AT in *Palmarosa aroma* grass have been reported till date [34]. Recently, regulation of santalol sesquiterpene biosynthesis in sandalwood (*Santalum album*) were shown to be co-dependent on the elicitor methyl jasmonate and up-accumulation of two mevalonate pathway genes which include an acetyl-CoA C-acetyl transferase (AACT) and a 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) [35]. An alcohol AT (PpAAT1), exhibiting the biosynthesis of γ -decalactone was identified by a combined protein molecular docking and site-directed mutagenesis approach [36]. In this study, the authors docked acetyl-CoA and the four alcohols propanol, hexanol, benzyl alcohol, and decanol on to the active center of PpAAT1 to identify amino acids participating in enzyme-substrate recognition in the esterification reaction. During our docking studies to select an AT from the qPCR shortlisted ATs, we assumed that an AT committing to an esterification reaction between a phenylpropanoid or phenolic acid such as cinnamic acid or vanillic acid and the catalpol moiety, will exhibit a higher binding affinity to cinnamoyl CoA or vanilloyl CoA, and a slightly lower binding affinity to 6-vanilloyl catalpol (P-II) or feruloyl catalpol which will facilitate their release from the ATs after the completion of the esterification reaction. Another natural assumption to shortlist these ATs was that the binding preferences of either catalpol or cinnamic or vanillic acid as independent substrates must be substantially less.

In a published report concerning comparative acetylation of 19-hydroxytabersonine derivatives of monoterpene indole alkaloids like minovincinine, 19-hydroxytabersonine, and horhammericine or other structurally almost similar 16-hydroxytabersonine derivatives of monoterpene indole alkaloids like deacetylindoline from *C. roseus*, 19-O-acetyltransferase (TAT) was unable to acetylate the 16-hydroxytabersonine derivatives or any compounds outside the monoterpene indole alkaloid family, thus corroborating to the fact that all the three major participants in an acyl group transfer/ esterification reaction; the AT, the acyl donor and the acyl acceptor are highly structure-specific [21]. Thus, acylation predictions based on *in silico* docking studies require assessment of individual binding affinity scores for

the active acyl donor with the AT (a), the acyl acceptor with the AT (b), the acyl donor and acceptor with the AT (c) and the formed final product with the AT (d). We assumed that an equivalently solid approach would be to dock (a) and (d). In a conceived scenario of an instance where acylation has been achieved, the final product will have a close binding with the AT before being released from the binding protein pocket.

Activated acyl carrier formation in case of SCPL ATs involves the action of UDP-glucosyltransferases before SCPLs can function. We assumed that the active acyl donor vanilloyl CoA would be a better fit to mine for P-II-specific ATs. Our lab recently worked out 'no-direction' terpenoid-glycoside gene co-expression sub-networks in *P. kurroa* and hypothesized that the biosynthesis and accumulation of isoprene units stemmed from negative regulation of abscissic acid (ABA) or ABA degradation, with cautious inclusion of downstream terpenoid glycosides/ iridoids in the hypothesis, since their up-accumulation might not be concurrent to ABA degradation and is inevitably subject to further functional scrutiny [1]. Generation of P-II accumulation dependent organ-wise unique and common gene co-expression networks involving ATs in *P. kurroa* might bring in novel information supposedly missing from the current analysis. Our basic assumption in using this strategy would be that the required AT for final acylation of CAT and VA to form P-II might appear preferentially in either the terpenoid or the phenylpropanoid gene co-expression sub-networks. Moreover an AT gene co-expression network might provide an additional clue to P-II biosynthesis via gene-gene co-expressed interactions. The major drawbacks in this assumption are that firstly, P-II is an iridoid glycoside biosynthesized in *P. kurroa* to meet the requirements of *P. kurroa*, therefore there are ample possibilities that the candidate AT will not show up in any of the co-expressed gene sub-networks in as much as of the 'need to express' criteria and that it is a modification/ decoration of the secondary metabolite CAT which can exist in any forms such as P-I, P-II, or feruloyl catalpol, and additionally since VA as well as CAT can be stored or is accumulated in the roots and rhizomes of *P. kurroa* independently [1].

Conclusions

Bioactive fractions obtained from dried roots and rhizomes of *P. kurroa* are rich in the hepato-protectant iridoids P-II, P-I, and kutkoside. Acylation of CAT with VA forms P-II. In need to decipher the P-II specific AT, *de novo* RNA-seq data from five *P. kurroa* transcriptomes, significantly varying in P-II content have been sorted to pull out every transcript annotated as an AT. With regard to their possible functional aspect, these ATs have been subjected to pair-wise cross-comparisons with one transcriptome data being the main feed while the other data was used as a query. All possible cross-comparisons of unique and common ATs from root and stolon could be converted into a list of 177 P-II specific putative ATs. A second round of screening of these ATs based on redundancy removal and observation of significant TPM values resulted in further shortlisting of ATs to just 22. An efficient relative gene expression oriented sorting of these ATs further brought down the number of putative ATs to six. In the end, experimentation of protein structure modelling and protein-ligand docking of these putative 6 ATs favoured a single AT possessing MBOAT domain. RNAi and CRISPR-CAS9 mediated silencing of these putative ATs adjunct to biochemical experiments with the cloned and heterologously expressed putative AT protein will have the

final say on the observed *in silico* binding affinity and binding preferences of the putatively sorted and identified AT transcripts predicted to be specific to P-II biosynthesis.

Declarations

Data Availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files]

Acknowledgements

The authors thank the Department of Biotechnology, Ministry of Science and Technology, Government of India for providing financial support to RSC in the form of Programme Support on high altitude medicinal herbs. SS thank Bennett University for providing PhD research fellowship.

Funding

This work was financially supported by Department of Biotechnology, Ministry of Science and Technology, Government of India (Grant number: BT/COE/34/SP15268/2015) in the form of a Programme Support on high-value medicinal plants under Centres of Excellence.

Author information

Affiliations

**Department of Biotechnology, School of Engineering & Applied Sciences, Bennett University,
Greater Noida, Uttar Pradesh, 201310, India**

Anjali Kharb, Shilpa Sharma, Ashish Sharma, Neeti Nirwal, Roma Pandey, Dipto Bhattacharyya &

Rajinder Singh Chauhan

Contributions

RSC conceived and designed the central flow and objectives to be investigated in the current research. AK, SS, and AS performed more than 90% of the crucial experiments (*in silico* and wet lab) described in this study and compiled the results section. NN was involved during pursual of primary initiative work for this publication. RP provided the HPLC chromatograms. DB analysed the data and wrote the first complete draft of the manuscript. RSC analysed the data, meticulously edited and finalized the complete draft manuscript for communication.

Corresponding authors

Correspondence to Rajinder Singh Chauhan

Ethics declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval and consent of participate

Not applicable.

Consent for publication

Not applicable.

References

1. Sharma A, Bhattacharyya D, Sharma S, Chauhan RS (2021) Transcriptome profiling reveal key hub genes in co-expression networks involved in Iridoid glycosides biosynthetic machinery in *Picrorhiza kurroa*. *Genomics*. <https://doi.org/10.1016/J.YGENO.2021.07.024>
2. Kharb A, Chauhan RS (2021) Complexity of gene paralogues resolved in biosynthetic pathway of hepatoprotective iridoid glycosides in a medicinal herb, *Picrorhiza kurroa* through differential NGS transcriptomes. *Mol Genet Genomics*. <https://doi.org/10.1007/s00438-021-01787-w>
3. Mangal AK, Rath C, Tewari D, et al (2017) Microscopical and Preliminary Physicochemical Studies of Two Important Endangered Ayurvedic Medicinal Plants Kutki and Trayamana to establish their Identity. *J Drug Res Ayurvedic Sci*. <https://doi.org/10.5005/jp-journals-10059-0003>
4. Thani PR (2021) A comprehensive review on *Picrorhiza kurroa* Royle ex Benth. 10:307–313
5. Chandra H, Kumari P, Prasad R, et al (2021) Antioxidant and antimicrobial activity displayed by a fungal endophyte *Alternaria alternata* isolated from *Picrorhiza kurroa* from Garhwal Himalayas, India. *Biocatal Agric Biotechnol* 33:101955. <https://doi.org/10.1016/J.BCAB.2021.101955>
6. Bahuguna IM, Rathore BP, Brahmabhatt R, et al (2014) Are the Himalayan glaciers retreating? *Curr Sci*. <https://doi.org/10.18520/cs/v106/i7/1008-1013>
7. Debnath P, Rathore S, Walia S, et al (2020) *Picrorhiza kurroa*: a promising traditional therapeutic herb from higher altitude of western Himalayas. *J. Herb. Med.*
8. Chen F, Ludwiczuk A, Wei G, et al (2018) Terpenoid Secondary Metabolites in Bryophytes: Chemical Diversity, Biosynthesis and Biological Functions. *CRC Crit Rev Plant Sci*. <https://doi.org/10.1080/07352689.2018.1482397>

9. Cao W, Xiong Y, Zhao D, et al (2020) Bryophytes and the symbiotic microorganisms, the pioneers of vegetation restoration in karst rocky desertification areas in southwestern China. *Appl Microbiol Biotechnol*. <https://doi.org/10.1007/s00253-019-10235-0>
10. Kumar V, Bansal A, Chauhan RS (2017) Modular design of picroside-II biosynthesis deciphered through NGS transcriptomes and metabolic intermediates analysis in naturally variant chemotypes of a medicinal herb, *Picrorhiza kurroa*. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2017.00564>
11. Damtoft S (1994) Biosynthesis of catalpol. *Phytochemistry*. [https://doi.org/10.1016/S0031-9422\(00\)94819-2](https://doi.org/10.1016/S0031-9422(00)94819-2)
12. Shitiz K, Sharma N, Pal T, et al (2015) NGS Transcriptomes and Enzyme Inhibitors Unravel Complexity of Picrosides Biosynthesis in *Picrorhiza kurroa* Royle ex. Benth. *PLoS One*. <https://doi.org/10.1371/journal.pone.0144546>
13. Schmittgen TD, Zakrajsek BA, Mills AG, et al (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: Comparison of endpoint and real-time methods. *Anal Biochem*. <https://doi.org/10.1006/abio.2000.4753>
14. Pandit S, Shitiz K, Sood H, Chauhan RS (2013) Differential biosynthesis and accumulation of picrosides in an endangered medicinal herb *Picrorhiza kurroa*. *J Plant Biochem Biotechnol*. <https://doi.org/10.1007/s13562-012-0136-z>
15. Jia Q, Hong MF, Minter D (1999) Pikuroside: A novel iridoid from *Picrorhiza kurroa*. *J Nat Prod*. <https://doi.org/10.1021/np980493+>
16. Zenk MH, Schmitt JH. (1965) Purification and properties of acetyl-CoA: D-amino acid- α -N-acetyltransferase from yeast. *Biochem Z*. PMID: 5847960.
17. Kundu A (2017) Vanillin biosynthetic pathways in plants. *Planta*
18. Modolo L V., Reichert AI, Dixon RA (2009) Introduction to the different classes of biosynthetic enzymes. In: *Plant-derived Natural Products: Synthesis, Function, and Application*
19. D'Auria JC, Pichersky E, Schaub A, et al (2007) Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in *Arabidopsis thaliana*. *Plant J*. <https://doi.org/10.1111/j.1365-313X.2006.02946.x>
20. Morales-Quintana L, Moya-León MA, Herrera R (2012) Molecular docking simulation analysis of alcohol acyltransferases from two related fruit species explains their different substrate selectivities. *Mol Simul*. <https://doi.org/10.1080/08927022.2012.672738>
21. Carqueijeiro I, Dugé de Bernonville T, Lanoue A, et al (2018) A BAHD acyltransferase catalyzing 19-O-acetylation of tabersonine derivatives in roots of *Catharanthus roseus* enables combinatorial

synthesis of monoterpene indole alkaloids. *Plant J*. <https://doi.org/10.1111/tpj.13868>

22. D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.*
23. Luo J, Nishiyama Y, Fuell C, et al (2007) Convergent evolution in the BAHD family of acyl transferases: Identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *Plant J* 50:678–695. <https://doi.org/10.1111/j.1365-313X.2007.03079.x>
24. Yu XH, Gou JY, Liu CJ (2009) BAHD superfamily of acyl-CoA dependent acyltransferases in *Populus* and *Arabidopsis*: Bioinformatics and gene expression. *Plant Mol Biol*. <https://doi.org/10.1007/s11103-009-9482-1>
25. Tuominen LK, Johnson VE, Tsai CJ (2011) Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. *BMC Genomics*. <https://doi.org/10.1186/1471-2164-12-236>
26. Mugford ST, Louveau T, Melton R, et al (2013) Modularity of plant metabolic gene clusters: A trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell*. <https://doi.org/10.1105/tpc.113.110551>
27. Nakayama T, Suzuki H, Nishino T (2003) Anthocyanin acyltransferases: Specificities, mechanism, phylogenetics, and applications. *J. Mol. Catal. B Enzym*.
28. Gomez C, Terrier N, Torregrosa L, et al (2009) Grapevine MATE-type proteins act as vacuolar h⁺-dependent acylated anthocyanin transporters1[W][OA]. *Plant Physiol*. <https://doi.org/10.1104/pp.109.135624>
29. Zhao J, Dixon RA (2009) MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. *Plant Cell*. <https://doi.org/10.1105/tpc.109.067819>
30. Burla B, Pfrunder S, Nagy R, et al (2013) Vacuolar transport of abscisic acid glucosyl ester is mediated by ATP-binding cassette and proton-antiport mechanisms in *Arabidopsis*. *Plant Physiol*. <https://doi.org/10.1104/pp.113.222547>
31. Aranovich D, Lewinsohn E, Zaccai M (2007) Post-harvest enhancement of aroma in transgenic *Lisianthus* (*Eustoma grandiflorum*) using the *Clarkia breweri* benzyl alcohol acetyltransferase (BEAT) gene. *Postharvest Biol Technol*. <https://doi.org/10.1016/j.postharvbio.2006.09.001>
32. Suzuki H, Murakoshi I, Saito K (1994) A novel O-tigloyltransferase for alkaloid biosynthesis in plants. Purification, characterization, and distribution of *Lupinus* plants. *J Biol Chem*. [https://doi.org/10.1016/s0021-9258\(17\)40759-9](https://doi.org/10.1016/s0021-9258(17)40759-9)

33. Qualley A V., Widhalm JR, Adebessin F, et al (2012) Completion of the core β -oxidative pathway of benzoic acid biosynthesis in plants. *Proc Natl Acad Sci U S A*. <https://doi.org/10.1073/pnas.1211001109>
34. Sharma PK, Sangwan NS, Bose SK, Sangwan RS (2013) Biochemical characteristics of a novel vegetative tissue geraniol acetyltransferase from a monoterpene oil grass (Palmarosa, *Cymbopogon martinii* var. Motia) leaf. *Plant Sci*. <https://doi.org/10.1016/j.plantsci.2012.12.013>
35. Niu M, Yan H, Xiong Y, et al (2021) Cloning, characterization, and functional analysis of acetyl-CoA C-acetyltransferase and 3-hydroxy-3-methylglutaryl-CoA synthase genes in *Santalum album*. *Sci Rep*. <https://doi.org/10.1038/s41598-020-80268-3>
36. Song ZZ, Peng B, Gu ZX, et al (2021) Site-directed mutagenesis identified the key active site residues of alcohol acyltransferase PpAAT1 responsible for aroma biosynthesis in peach fruits. *Hortic Res*. <https://doi.org/10.1038/s41438-021-00461-x>

Figures

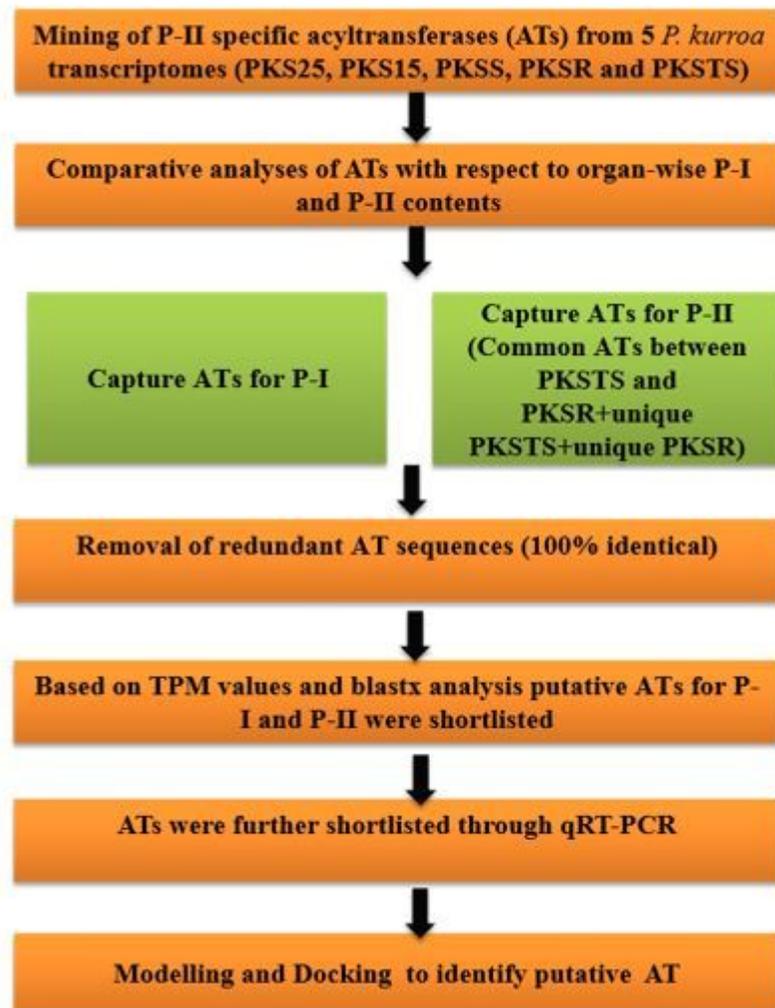
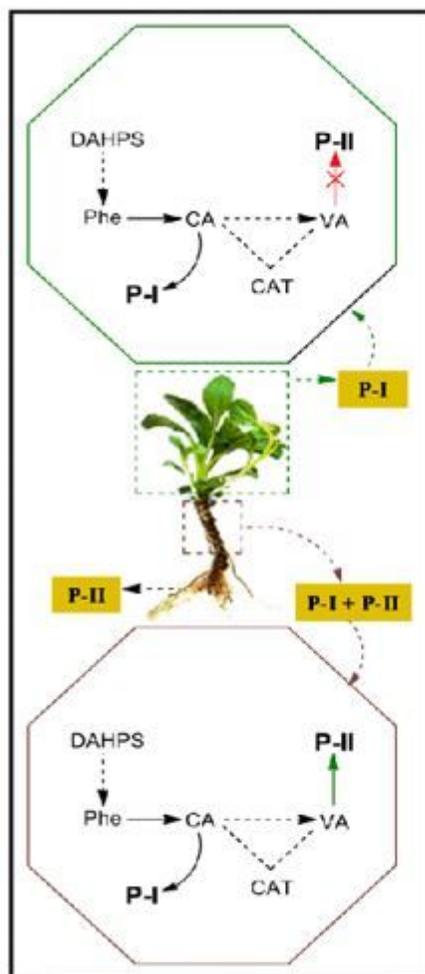


Figure 1

Diagram showing Picroside-I (P-I) is detected in shoots and stolons in *P. kurroa* while Picroside-II (P-II) is detected in stolons and roots (left panel). Right panel represents a schematic flow diagram for capturing Acyltransferase (ATs) specific for Picroside-II biosynthesis. (DAHPS, 2-dehydro-3-deoxyphosphoheptonatealdolase 3-deoxy-d-arabino heptulosonate7- phosphate synthetase; CAT, Catalpol; VA, Vanillic acid; CA, Cinnamic acid)

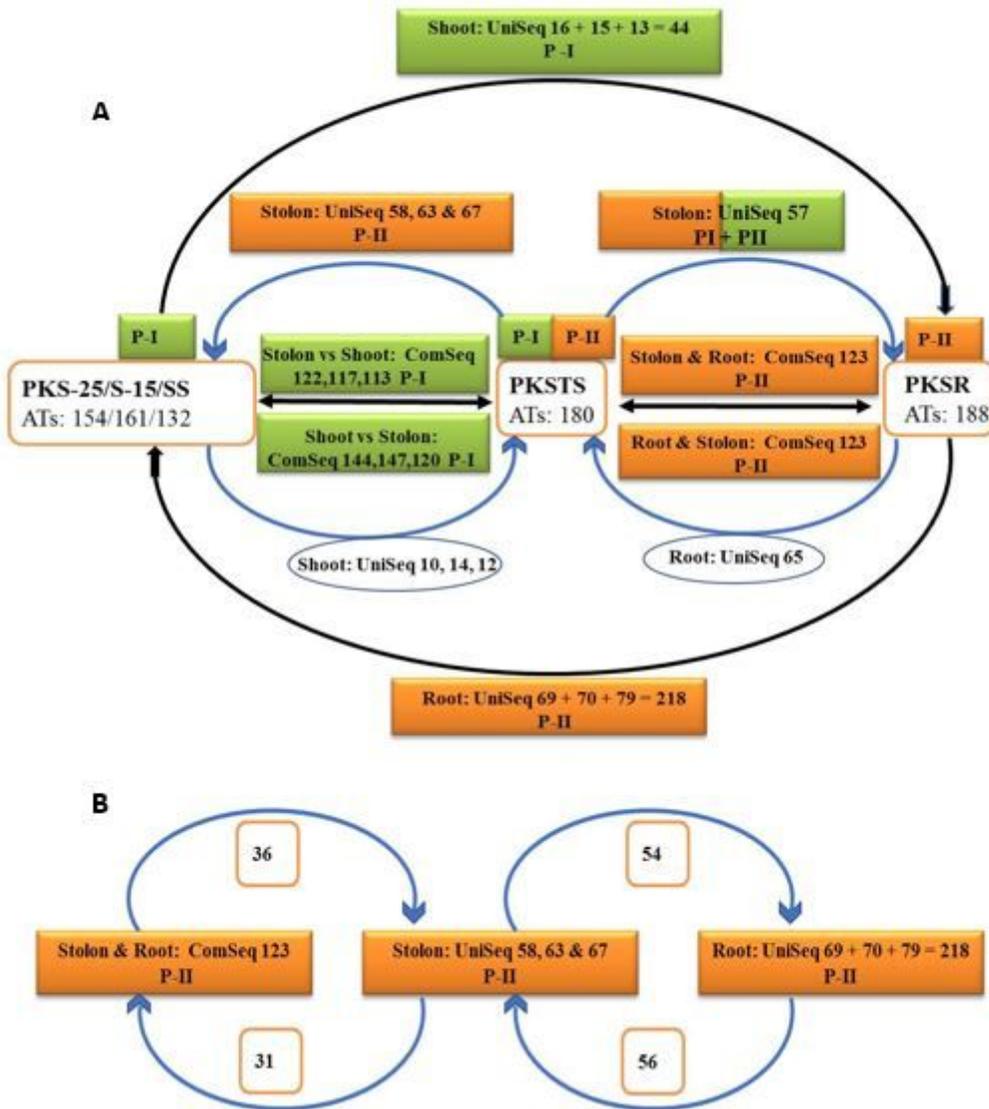


Figure 2

Comparative transcriptome analysis to identify putative Acyltransferases for the biosynthesis of Picrosides in *Picrorhiza kurroa*, total of 815 Acyltransferases have been identified from 5 different transcriptome datasets (PKS25, PKS15, PKSS, PKSTS and PKSR) (A); Identification of putative Acyltransferases commonly present in roots and stolons that might function in Picroside-II biosynthesis.

Total of 177 Acyltransferases have been identified that shows organ-wise Picroside-II content association (B) (UniSeq: Unique sequences, ComSeq: Common sequences)

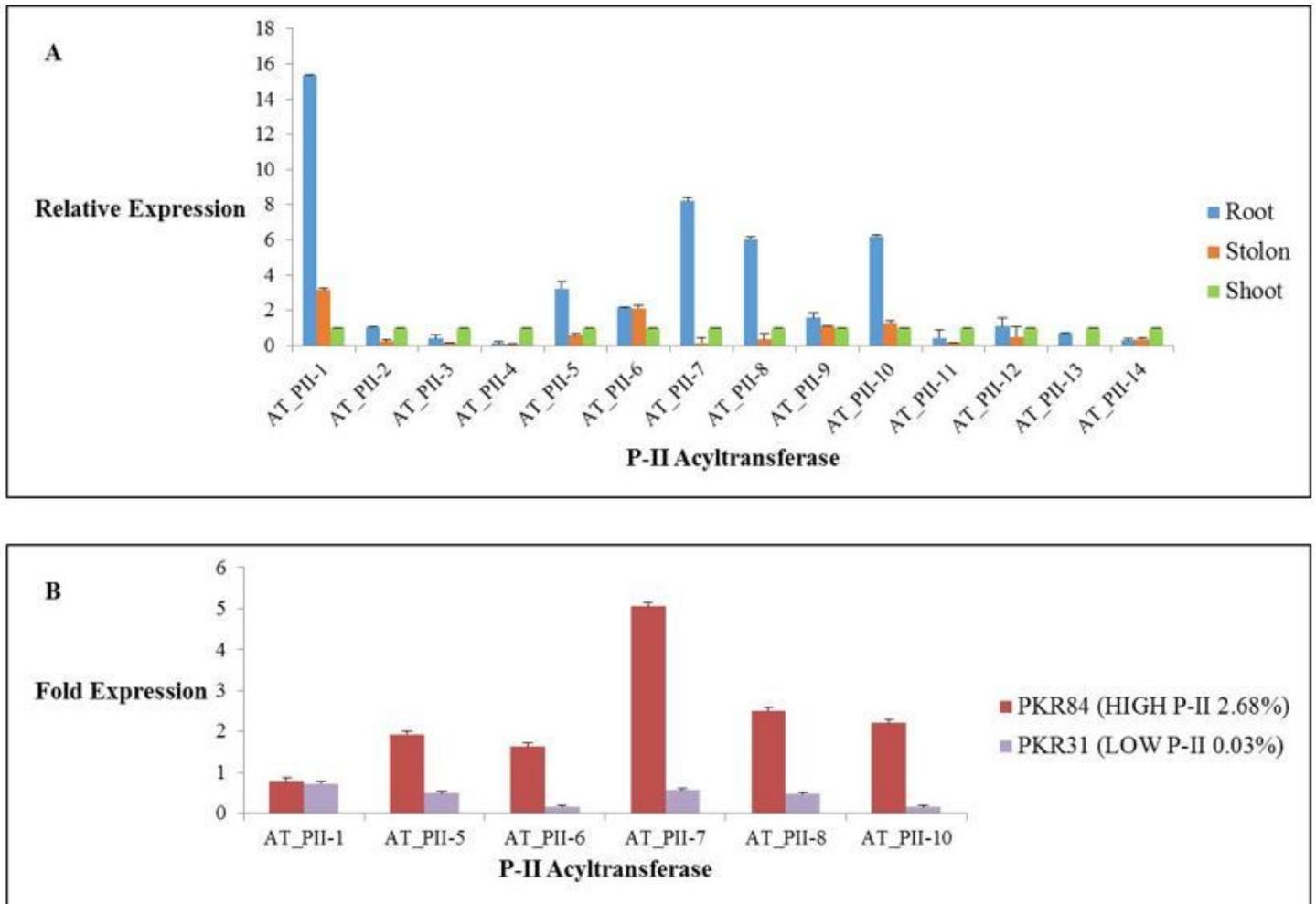


Figure 3

Relative expression status of Acyltransferases in shoots (PKSS), stolons (PKSTS), and roots (PKSR) of *Picrorhiza kurroa*, relative expression is calculated for picroside-II containing roots and stolons with respect to shoots (PKSS) as shoots do not produce picroside-II (A); Fold expression of Acyltransferases in *Picrorhiza kurroa* accessions PKR84 (High P-II containing roots) and PKR31 (Low P-II containing roots) (B).

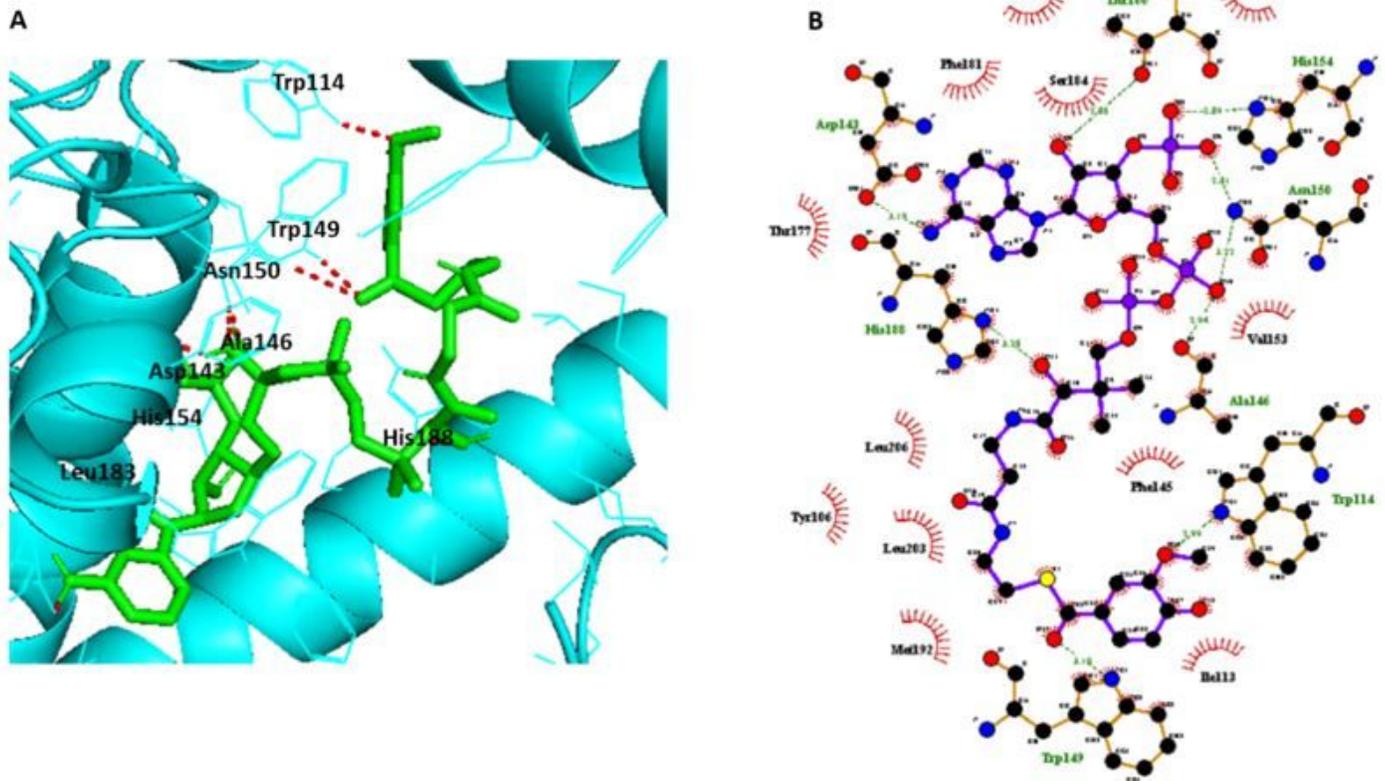


Figure 4

Vanilloyl Coenzyme A showing interaction with Acyltransferase MBOAT family (identifier Node_48592/618 bp (AT_PII-10)) (A); 2D representation of H-bonds, Hydrophobic and Salt Bridge interactions (B).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Suppl.Fig.docx](#)
- [Suppl.File1.xlsx](#)
- [Suppl.File2.xlsx](#)
- [SupplementaryTable.docx](#)