Analysis of Artemisia annua transcriptome for BAHD alcohol acyltransferase genes: Identification and diversity of expression in leaf, stem and root

ARTICLE *in* JOURNAL OF PLANT BIOCHEMISTRY AND BIOTECHNOLOGY · SEPTEMBER 2012

Impact Factor: 1.09 · DOI: 10.1007/s13562-012-0141-2

J. Plant Biochem. Biotechnol.
DOI 10.1007/s13562-012-0141-2

ORIGINAL ARTICLE

Analysis of *Artemisia annua* transcriptome for BAHD alcohol acyltransferase genes: Identification and diversity of expression in leaf, stem and root

Sudhakar Srivastava · Rajender Singh Sangwan

Received: 12 May 2012 / Accepted: 26 July 2012

Abstract Alcohol acyltransferases (AATs) of BAHD gene superfamily play diverse roles in plant secondary metabolism like synthesis of volatile esters, modified anthocyanins, terpene indole alkaloids, benzylisoquinoline alkaloids and capsacinoids. *Artemisia annua* is an important medicinal and aroma plant synthesizing monoterpene oil and several non-volatile sesquiterpenes including anti-malarial artemisinin. These AATs are characterized by two key structural motifs- HXXXD and DFGWG. ESTscan of *A. annua* trascriptome yielded 43 contigs homologous to BAHDs. Removal of redundant sequences led to identification of 16 discrete BAHD family AATs and assembly of 3 full length cDNA of BAHD acyltransferases. RT-PCR analysis confirmed the expression of these genes. The *in silico* assembled putative full length AATs were validated by PCR amplification, sequencing and matching the sequences with the assembled genes. Cross-species comparison of AAT amino acid sequences gave an updated account of conserved residues in BAHDs including the novel homologues identified in this study. Presence of several AATs in this plant underpins their co-necessity for acylation assisted extension of the diverse secondary metabolites. Functional characterization of the AATs are necessary to understand the subtle differences in motifs *vis- a-vis* catalytic specificities among these AATs.

Keywords *Artemisia annua* · Alcohol acyltransferase · BAHD gene family · Secondary metabolism · Glandular trichome genomics · Artemisinin

Abbreviations

BEAT	benzyl alcohol O-acetyltransferase
AHCT	anthocyanin O-hydroxycinnamoyltransferase
HCBT	N-hydroxycinnamoyl anthranilate benzoyl transferase
DAT	deacetylvindoline 4-O-acetyltransferase
AAT-RSs	Alcohol Acyl transferase RS1-2-3

Introduction

Acetyltransferases play diversely important roles in the life of an organism (D'Auria 2006; Sangwan et al. 2007; Sharma et al. 2005). Plant acyl-CoA dependent alcohol acyltransferases (AATs) constitute a superfamily called BAHD which derives its name from the first letter of the first four plant acyltransferases identified viz. BEAT (benzyl alcohol O-acetyltransferase from *Clarkia breweri*); AHCT (anthocyanin O-hydroxycinnamoyltransferase from *Gentiana triflora*); HCBT (N-hydroxycinnamoyl anthranilate benzoyl transferase from *Dianthus caryophyllus*); DAT (deacetylvindoline 4-O-acetyltransferase from *Catharanthus roseus*) (St-Pierre and De Luca 2000). They are characterized by the presence of a universally conserved amino acid motif HXXXD and a highly conserved amino acid sequence DFGWG (D'Auria 2006). The next generation sequencing (NGS) technology led to the addition of large amount of genomic and cDNA sequences to public domain databases that need to be mined, transformed into discrete genes, assigned to pertinent gene family and analyzed biochemically in terms of structural/sequence variation *vis-a vis* their catalytic deviation.

About 40 members of BAHD family acyltransferases characterized, as yet, pertain to mainly anthocyanin biosynthesis, alkaloid biosynthesis or volatile ester biosynthesis in fruits or flowers. To our knowledge, only two foliage expressed BAHDs are known that are also related to the wound inducible green leaf volatile ester biosynthesis in *Arabidopsis* (D'Auria et

Electronic supplementary material The online version of this article (doi:10.1007/s13562-012-0141-2) contains supplementary material, which is available to authorized users.

S. Srivastava · R. S. Sangwan (✉)
Central Institute of Medicinal and Aromatic Plants
(Council of Scientific and Industrial Research), P.O. CIMAP,
Kukrail Picnic Spot Road,
Lucknow 226015, India
e-mail: sangwan.lab@gmail.com

Published online: 11 September 2012

al. 2007a). *Arabidopsis thaliana* genome analysis alone has revealed the presence of >60 BAHD members but only a few of them have been characterized with respect to their expression or biochemical/catalytic function such as biosynthesis of anthocyanin (Luo et al. 2007), hydroxycinnamoyl spermidines (Grienenberger et al. 2009), cutin polyester (Panikashvili et al. 2009) etc. Growing number of BAHD members involved in synthesis of secondary or specialized metabolism (Brandt 2005; Dai et al. 2010; Dudareva et al. 1998; Grothe et al. 2001; Sharma et al. 2003; St-Pierre et al. 1998; Suzuki et al. 2001, 2003 and 2004; Walker and Croteau 2000a, b; Yang et al. 1997), are cue to their increasing recognition in metabolic diversity at different levels- inter-tssue, inter-species etc. Though, genomes and trascriptomes of several plants have been sequenced but dicrete gene members of BAHD are yet to be unambiguously -assigned to the family for their functional and diversity understanding. Furthermore, the transcriptome sequences available in databases as assembled reads (contigs) and/or more ambiguous shorter transcript sequences (ESTs) need to be analyzed to construct gene family and validated by a combination of both *in silico* and experimental approaches.

In this study, we have (i) identified non-redundant contigs of transcriptome of *A. annua* leaf that represented amino acid sequences homologous to those of BAHD members, (ii) successfully assembled some of the contigs to generate three putatively functional full length AATs (iii) validated the correctness of assembled AATs by direct full length cloning and sequencing, (iv) demonstrated the overlapping as well as differential expression of the AATs in a tissue-specific manner by semi-quantitative RT-PCR in leaf, stem and root and (v) carried out *in silico* analysis to identify the universally conserved characteristic motifs and residues in BAHDs and (vi) performed phylogentic analysis of assembled putative AATs of *Artemisia* .

Results

Artemisia annua produces a variety of secondary metabolites ranging from a complex mixture of monoterpene and other volatiles and their esters, non-volatile sesquiterpenes including anmti-malarial artemisinin and those pertaining to other chemical classes (Sangwan et al. 1998 and 2001; Sangwan et al. 1993 and 1999). Pyrosequencing of *A. annua* leaf transcriptome and subsequent clustering and assembly the transcript sequences into contigs has recently been reported (Wang et al. 2009); Graham et al. 2010).

Recognition of multiple members of BAHD family in *A. annua* glandular trichomes

The ESTScan of transcriptome from *A. annua* leaf trichome, against the protein sequence of BAHDs of known function yielded a total of 43 out of 31,469 contigs of the transcriptome dataset showed *prima facie* resemblance to BAHDs (Table 1). Removal of redundant sequences led to scoring of 19 discrete contigs putatively related to BAHD family that shared significant homology (>50 %) with the known members of BAHD, affirming their assignment (Table 1). Sequences of these 19 contigs are given in Supplementary Fig. 1.

Individual analysis of the 19 contigs revealed that 11 of them (4127, 6873, 10186, 10891, 11590, 12508, 16809, 17424, 21793, 27794 and 28380) shared 50–84 % similarity with benzoyltransferases from different plants including benzoyl coenzyme A: benzyl alcohol benzoyltransferase (BEBT) of *Clarkia breweri* (Genebank: AAN09796). The BEBT of *Verbena* x *hybrida* (Genebank: BAE72881) was taken as a model against which contigs were aligned to locate the homologous regions (Supplemantal Fig. 2) because most of the contigs mentioned above shared in general maximum identity with BEBT of *Verbena* x *hybrida*. Among these 11 contigs, three (10186, 11590 and 28380) constituted parts of the same cDNA and rest of the eight cDNA were discrete and unlinked to each other (1+8=9 cDNAs). Five contigs (5201, 7738, 11768, 23763 and 27500) showed 50 % to 60 % protein sequence similarity with anthranilate-N-hydroxy-cinnamoyl/benzoyl transferases (HCBTs), another founder member of BAHD family. Besides, contig 5201 also shared 42 % similarity with anthocyanin malonyl transferase of *Salvia splendens* (Genebank: AAR26385). Among these 5 contigs, contigs 7738 and 27500 appeared to be part of the same cDNA as they had >98 % similarity in their 118 base overlapping region. Therefore, in all at least four different members of HCBT are present in *A. annua* (4 cDNAs). Among rest of the 3 contigs (2678, 9450 and 27700), 2678 and 9450 shared 45–60 % similarity with salutaridinol 7-O-acetyltransferase from *Papaver bracteatum* (Genebank: ACI45392) and *Papaver somniferum* (Genebank: ACI45392) whilst 27700 showed 51 % similarity with deacetylvindoline 4-O-acetyltransferase *Catharanthus roseus* (Genebank: AAC99311) indicating these three contigs could be part of three different cDNAs (3 cDNAs). The contig 4127 shared a minimum 200 bp overlapping region with 3 contigs- 10891, 16809, and 21793 but shared only up to 80 % overall identity at nucleotide level with each of the three contigs indicating 4127 is different from contigs 10891, 16809 and 21793. Contig 16809 shared approximately 400 bp region with 11590, which in turn formed a part of assembled full length cDNA-AAT-RS1. Contigs 12508 and 23763 were part of assembled complete cDNAs- AAT-RS2 and AAT-RS3, respectively, as described under ahead in this report. The uniqueness of non-overlapping contigs was further established by PCR using a 5′upstream (forward) primers of contigs annotated towards 5′regions and 3′ downstream (reverse) primers of contigs downstream to the upstream

Table 1 Contigs of Artemisia annua identified as member of BAHD superfamily. ESTScan was performed using protein sequences (query) of acyl transferases with known function against the transcriptome of *A. annua* glandular trichome (subject)

Query ID (Genebank)	Subject ID (Transcriptome)	Identity (%)	Alignment length	Query start	Query end	Subject start	Subject end	E-value	Score
AY563157.1	Contig12508	85.88	85	364	448	250	334	1.00E-12	73.8
AY563157.1	Contig11590	79.55	176	1138	1313	131	306	9.00E-10	63.9
AY563157.1	Contig10186	86.96	69	505	572	207	275	6.00E-08	58
AY563157.1	Contig28380	85.29	68	505	572	43	110	2.00E-07	56
AY563157.1	Contig27794	85	60	136	195	142	201	6.00E-05	48.1
AY563157.1	Contig16809	88.37	43	1055	1097	208	250	2.00E-04	46.1
AY563157.1	Contig4127	88.37	43	1055	1097	474	516	2.00E-04	46.1
AY563157.1	Contig10891	93.1	29	419	447	456	428	0.003	42.1
AF500202.1	Contig10186	85.39	89	319	407	52	140	8.00E-13	73.8
AF500202.1	Contig10186	82.83	99	472	569	207	305	7.00E-07	54
AF500202.1	Contig12508	82.35	85	331	415	250	334	1.00E-05	50.1
AF500202.1	Contig28380	87.23	47	472	518	43	89	2.00E-04	46.1
AF500202.1	Contig17424	86.27	51	1215	1265	62	112	2.00E-04	46.1
AY859053.1	Contig10186	84.11	107	460	566	34	140	6.00E-14	77.8
AY859053.1	Contig28380	87.23	47	628	674	40	86	2.00E-04	46.1
FJ548611.1	Contig28380	84.85	99	500	598	14	112	6.00E-14	77.8
FJ548611.1	Contig21793	92.31	39	879	917	9	47	9.00E-07	54
FJ548611.1	Contig10186	85.25	61	404	464	80	140	1.00E-05	50.1
FJ548611.1	Contig10186	88.89	45	529	573	207	251	1.00E-05	50.1
AY534530.1	Contig6873	84.48	58	54	111	296	239	7.00E-04	44.1
EU200366.1	Contig12408	89.74	39	229	267	13	51	2.00E-04	46.1
EU200366.1	Contig6873	89.74	39	229	267	131	93	2.00E-04	46.1
AY161302.1	Contig23763	89.47	38	295	332	46	9	8.00E-04	44.1
AY534531.1	Contig27700	96	25	1136	1160	119	143	0.003	42.1
AY534531.1	Contig11768	96	25	1136	1160	239	263	0.003	42.1
EU784138.1	Contig2025	100	21	1159	1179	96	76	0.003	42.1
NM_111219.2	Contig27794	88.89	45	117	161	105	149	1.00E-05	50.1
NM_111219.2	Contig6873	86.67	45	117	161	296	252	0.003	42.1
DQ886905.1	Contig13874	100	22	344	365	157	178	8.00E-04	44.1
DQ886905.1	Contig7064	100	22	344	365	95	74	8.00E-04	44.1
AAN09796.1	Contig10186	91.89	37	341	377	77	113	1.00E-05	50.1
AAN09796.1	Contig12508	81.48	81	352	432	274	354	0.003	42.1
AAF04787.1	Contig2692	96.15	26	680	705	37	62	7.00E-04	44.1
AF121854.1	Contig2692	100	23	679	701	37	59	1.00E-04	46.1
AF121853.1	Contig2692	100	26	680	705	37	62	3.00E-06	52
AF121851.1	Contig2692	100	23	680	702	37	59	2.00E-04	46.1
ACI45395.1	Contig5201	93.1	29	1268	1296	465	493	0.003	42.1
AAC99311.1	Contig2678	100	24	1141	1164	482	505	4.00E-05	48.1
AAC99311.1	Contig27500	96.3	27	1138	1164	69	43	2.00E-04	46.1
AAC99311.1	Contig9450	96.3	27	1138	1164	212	186	2.00E-04	46.1
AAC99311.1	Contig7738	96.3	27	1138	1164	200	174	2.00E-04	46.1
AAC99311.1	Contig27700	100	22	1136	1157	119	140	7.00E-04	44.1
AAC99311.1	Contig11768	100	22	1136	1157	239	260	7.00E-04	44.1

contigs (e.g. forward primer of 27794 and reverse primers of all the contigs in Supplemental Fig. 2, similarly all other possible combination were tested for rest of the contigs).

Thus, assembly of contigs, their alignment and other bioinformatic analyses revealed the presence of 16 (9+4 +3) different cDNAs belonging to BAHD family in

A.annua glandular trichome. Expression profile of these individual ESTs (contigs) was investigated in other tissues such as leaf, stem and root, by using RTPCR.

Assembly of contig sequences into novel BAHD homologue from *A. annua* trichomes

The cDNA sequences of BAHD members from different plants available in National Center for Biotechnology Information (NCBI, USA) were downloaded. The sequences of the contigs of *A. annua* transcriptome scored above based on BLASTx based homology with BAHDs were individually used to carry out BLAST search of nucleotide databases using the nucleotide query (BLASTn) wherein they showed 90–100 % homology with *A. annua* contigs deposited as Transcriptome Shotgun Assembly (TSA) at NCBI by Graham and associates (2010). To extend the 5′ and 3′ region of these contigs by *in silico* approach, only those homologous sequences were considered that were identical or at least 98 % similar for at least 200 bases around overlapping region. The sequences were joined end to end by removing the overlapping region from either of the sequences. Subsequently, these assembled sequences were used to perform BLASTx to affirm that the complete cDNAs carrying open reading frame (ORF) and 5′ as well as 3′untranslated regions (UTRs) were in hand. Using this strategy, we were able to assemble three different complete cDNA sequences of AATs which were christened as AAT-RS1, AAT-RS2 and AAT-RS3. Sequence similarity among these assembly predicted AATs is given in Table 2. The deduced molecular weights of the three AATs (AAT-RS1, AAT-RS2 and AAT-RS3) were computed to be 50.8, 51.9 and 48.8 kDa, respectively, that correspond to the average molecular mass range (48–55 kDa) reported for the known BAHD family acyltransferases. These three new members of the family possessed all the signature motifs as well as structural and/or catalytic sequence motifs as discussed later in this article.

Isolation of complete cDNA sequences of assembled AATs from *Artemisia annua*

To validate the *in silico* assembled AATs using the contigs from *A. annua* transcriptome (shotgun) database, PCR was performed using gene specific primers and the tissue cDNA as template. The PCRs resulted into the amplicons of the expected sizes. Sequencing and subsequent analysis confirmed the identity and existence of the assembled AATs *in planta*.

Conserved residues analysis in putative AAT-RSs

Protein sequence of few members of BAHD superfamily members from the database and BAHDs of *A. annua* cloned in this study were taken to carry out a comprehensive and systematic *in silico* study. The alignment indicated that 16 residues appeared to be universally conserved and other 8–10 residues were conserved in around 90 % of the BAHDs (Fig. 1), which is further confirmed when a large number of BAHDs were taken to conduct similar analysis (Supplementary Fig. 2). Identity value of alignment of the three assembled BAHDs from *A. annua* is given in Table 2. Despite as low as 24 % sequence similarity at protein level among themselves, all the members belonging to all the four groups of BAHDs (BEAT, HCBT, ACHT and DAT) shared conservation of approximately 25 residues at similar positions, which appears to be a unique signature of this superfamily.

The signature motifs- HXXXD in the middle and DFGWG towards the C-terminus were also present in all the three AAT-RSs identified in this study. Besides, a highly conserved CGG motif upstream to HXXXD motif was also present in AAT-RSs. The conserved residues were sparsely distributed which are as follows with their approximate position (positions of residues referred with reference to benzyl alcohol benzoyl transferase, BEBT, from *Clarkia*; Genbank: AAN09796): Proline (P) that falls around 22 position; Y- 50; L-77; G-84, G-100 and G-310; Q-150; F-170, R-295; N-311. Further, there are few residues that are 80–90 % conserved in BAHDs: L-32; S-33; D-37; Y-75 and Y-76; A- 79; R-81; C-100; P-142; W-270, N-290. In addition, either leucine (L) or isoleucine (I) at around 60th position and an aromatic amino acid residue (mostly Y but also F) at around 75th position appears to be prevalent and conserved residue at similar positions in other BAHD members. These conserved residues are depicted in Fig. 1.

Table 2 Sequence identity among the three assembled complete cDNAs of *A. annua* predicted to encode BAHDs. Similarity for these putative AAT-RSs was also assessed against the BEBTs and HCBTs

Name	Length ORF (bp/aa)	Deduced molecular mass (Putative protein)	Amino acid identity (approx)				
			AAT-RS1	AAT-RS2	AAT-RS3	BEBT	HCBT
AAT-RS1	1386/463	50.8 Kda	–	70 %	26 %	70−83 %	25−30 %
AAT-RS2	1395/465	51.9 Kda	70 %	–	26 %	60−70 %	25−30 %
AAT-RS3	1323/441	48.8 Kda	26 %	26 %	–	25−30 %	70−80 %

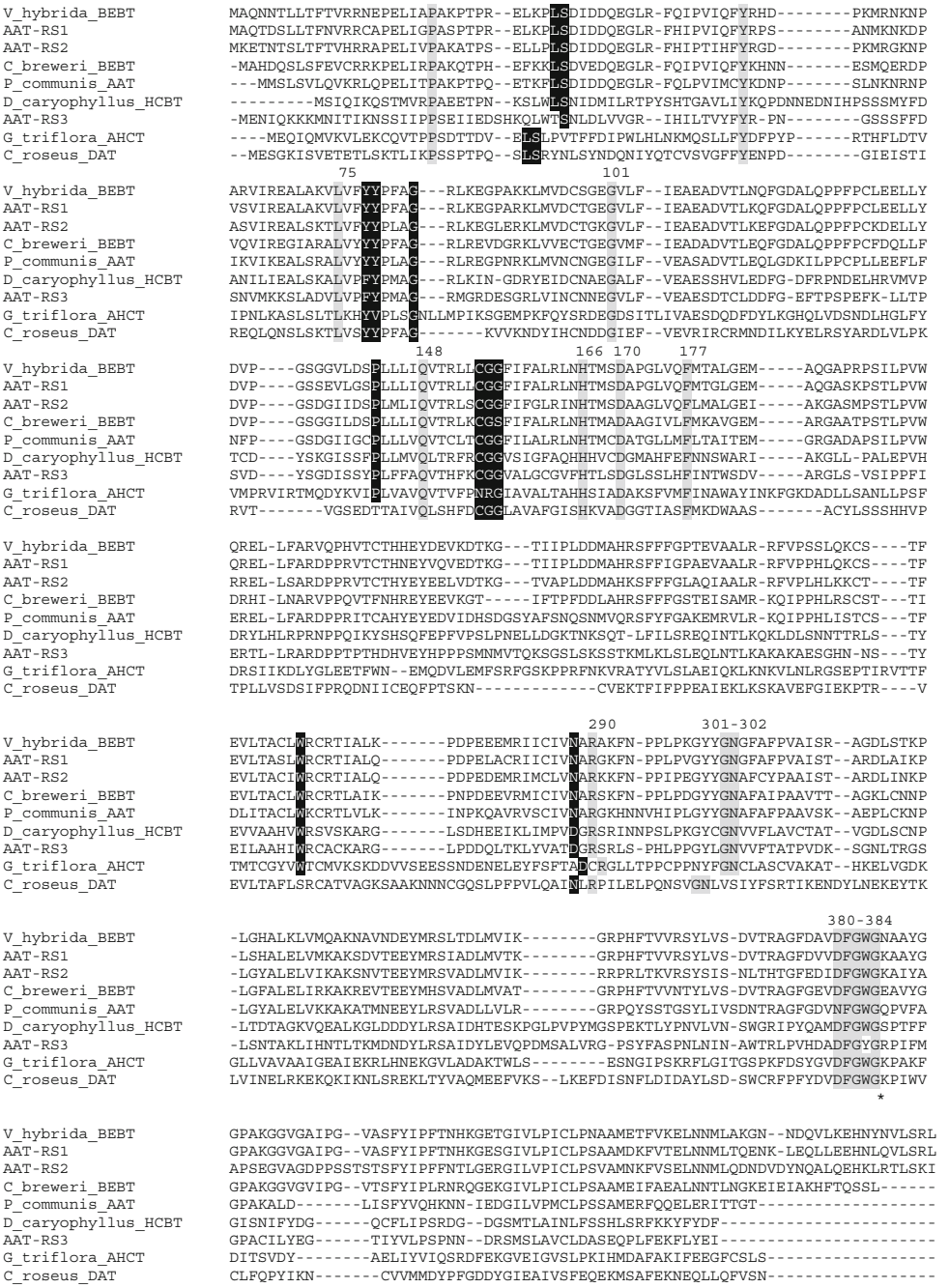

Fig. 1 Sequence alignment of selected BAHD family. The proteins sequence alignment of three putative BAHDs from *Artemisia annua* with some of the representative members of the BAHD family: BEBT from *Clarkia breweri* (GenBank: AAN09796); and *Verbena hybrida* (GenBank: AAN09796.1), AAT from *Pyrus communis* (GenBank: AAS48090.1), HCBT from *Dianthus caryophyllus* (GenBank: Z84383.1), AHCT from *Gentiana triflora* (GenBank: BAA74428.1) and DAT from *Catharanthus roseus* (GenBank: AAC99311.1). Universally conserved amino acid residues within the members of BAHD family are highlighted in *grey*. Residues highlighted in *black* are conserved in approximately 90 % of the members of BAHD. *Asterisks* indicate the positions of the natural variant/mutant residue in DFGWG motif in AAT-RS3

Expression analysis of AATs

Expression patterns could provide insight into the function of AAT genes. Using the contig-specific primers for each of the 19 AAT contigs expression profile was investigated semi-quantitatively by RT-PCR in the leaf, stem and root of *A. annua*. The comparative expression profiles are presented in Fig. 2. As expected, all the AATs were found to be expressed in leaves. Interestingly, almost all the contigs were also expressed in stem of the plant but none of them was found to be expressed in root tissue. Expression profile of contigs 10186, 11590 and 28380 were identical, which was obvious as these are part of same cDNA (AAT-RS1). Contigs 10891 and 27700 were expressed only in leaf and no expression could be detected in stem or root. Besides, the transcript accumulation of contigs 7738, 12508 and 16809 is less in stem cells as compared to leaf cells.

Phylogenetic analysis of putative members of BAHDs from *A. annua*

To work out the relationship among the assembled three putative BAHDs from *A. annua* viz., AAT-RS1, AAT-RS2

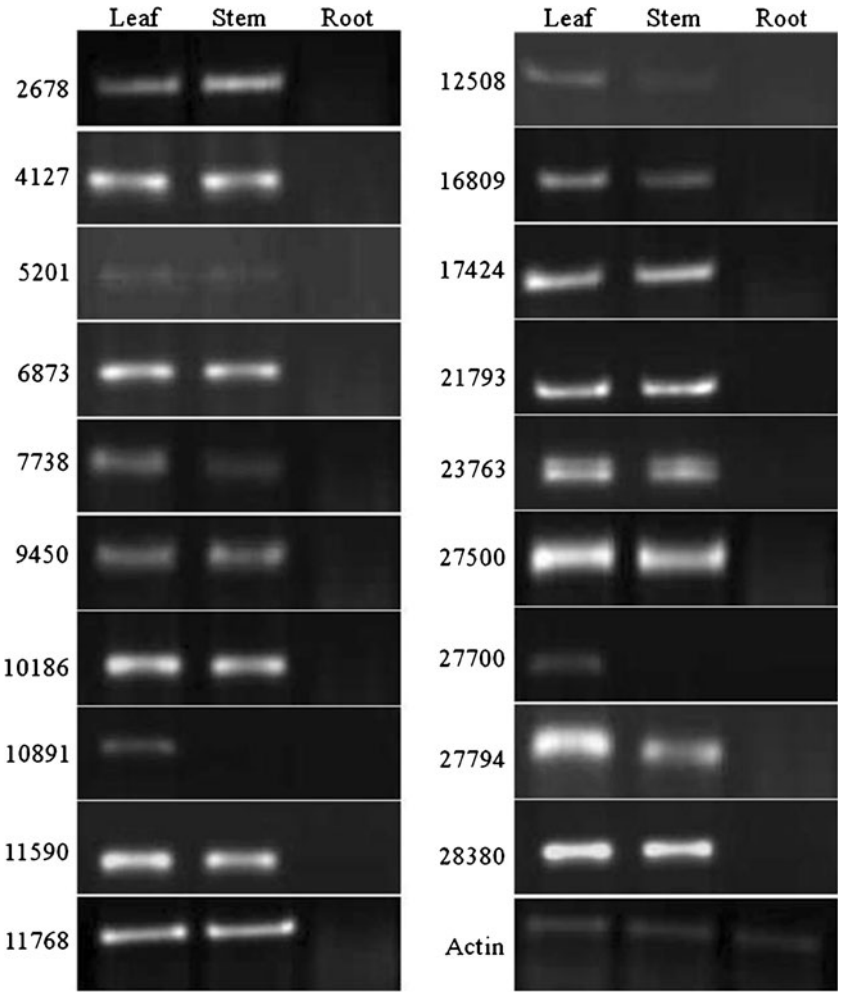

Fig. 2 Semi-quantitative RT-PCR analysis of *Artemisia annua* BAHD family AAT genes. Expression profile of contigs (identified by numbers) belonging to BAHD family in leaf, stem and root. Actin was used to equalize the quantity of cDNA

and AAT-RS3 with BAHDs from other sources, phylogenetic analysis was performed. The sequences of putative members of BAHDs from *A. annua* were compared with 20 biochemically characterized members of BAHD superfamily from other plants to construct molecular phylogenetic tree using maximum parsimony, distance and maximum likelihood. All trees obtained by different methods showed more or less similar topology. Grouping of AAT-RS' with other BAHDs in Neighbour Joining (NJ) tree were consistent with similarity in sequences obtained by BLAST analysis. AAT-RS1 and AAT-RS2, which are 70 % similar with each other, are grouped with BEBTs from other plant species, while AAT-RS3, which shared 70–80 % similarity with HCBTs and approximately around 30 % similarity with BEBTs and other two AAT-RS', was grouped with HCBTs (Fig. 3). The analysis indicated that AAT-RS1 and AAT-RS2 were likely to be the enzymes playing a role in benzoyl transferase and AAT-RS3 might be involved in HCBTs related function. However, the speculations need to be verified by characterizing these proteins biochemically. Sizes of rest of the *A. annua* contigs, identified as members of BAHD family, were too small to assign specific function putatively specially owing to the sparse distribution of conserved residues or their stretches in BAHDs. Although grouping of all the 16 contigs with other BAHDs in phylogenetic tree presented here (Supplementary Fig. 4) and the similarity of these contigs with acyltransferases of known function in data base is largely consistent, there are some discrepancies. Like, based on BLAST analysis contigs 6873, 17424 and 27794 were more similar to BEBTs whereas these contig were grouped with HCBTs in phylogenetic tree.

Discussion

The acytransferases belongs to a superfamily BAHD. The molecular weight of these proteins ranges from 45 to 55 kDa (D'Auria 2006). From our EST sequencing data, we obtained 19 non redundant cDNAs belonging t BAHDs among which 3 cDNAs could be assembled to full length cDNA. The deduced molecular weights of these three full length AATs (AAT-RS1, AAT-RS2 and AAT-RS3) were computed to be 50.8, 51.9 and

48.8 kDa, respectively. Like many acyltransferases from other species, AAT-RSs also exhibit the signature motifs, HXXXD and DFGWG, and few other conserved amino acid residues. Ma et al. (2005) have proposed the conservation of the residues among BAHD members, however, identification and characterization of few more BAHDs subsequent to the study indicated some inconsistencies in assignment of the number of conserved residues. We took large number of BAHD superfamily sequences from the database plus the herein identified BAHDs of *A. annua* and carried out a comprehensive and systematic *in silico* study. The alignment indicated that 16 residues appeared to be universally conserved and other 8–10 residues were conserved in around 90 % of the BAHDs (Fig. 1). Identity value of alignment of the three assembled BAHDs from *A. annua* is given in Table 2. Despite as low as 24 % sequence similarity at protein level among themselves, all the members belonging to all the four groups of BAHDs (BEAT, HCBT, ACHT and DAT) shared conservation of approximately 25 residues at similar positions, which appears to be a unique signature of this superfamily. To update the knowledge on conserved residues and in motifs in AATs, we took large number of sample sequences for alignment (Supplementary Fig. 3) and found that 16 residues to be universally conserved in BAHDs. Among the motifs considered to be universally conserved in AATs, two pertain to be signature motifs- HXXXD in the middle and DFGWG towards the C-terminus apart from a highly conserved CGG motif upstream to HXXXD motif. Earlier studies have demonstrated that mutation in His or Asp in HXXXD motif led to loss of activity. Histidine involved in catalysis (acid–base) whilst Asp is considered to impart a momentous role for conformational suitability (St-Pierre and De Luca 2000;, Suzuki et al. 2003;, Bayer et al. 2004; Buglino et al. 2004; Gibbs et al. 1990). Presence of DFGWG also appears to be another characteristic feature of BAHDs and different roles have been attributed to this motif such as participation in

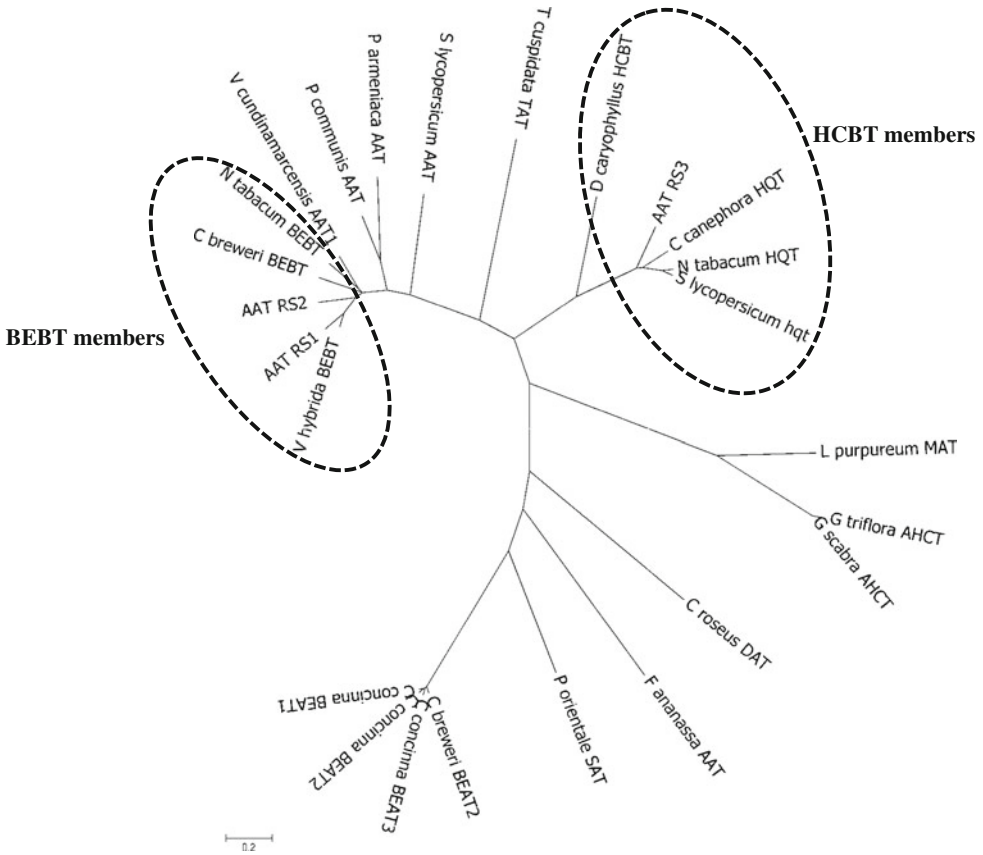

Fig. 3 Phylogram of functionally characterized members of BAHD family including the putative members viz., AAT-RS1, AAT-RS2 and AAT-RS3 from *A. annua*. Acyl Transefearses from *V. hybrida* BEBT (BAE72881.1), *N. tabacum* BEBT (AAN09798.1), *C. breweri* BEBT (AAN09796.1), *V. cundinamarcensis* AAT1 (ACT82248.1), *P. communis* AAT (AAS48090.1), *S. lycopersicum* AAT (AAS48091.1), *P. armeniaca* AAT (ACF07921.1), *F. ananassa* AAT (AAG13130.1), *C. breweri* BEAT-2 (AAF04787.1), *C. concinna* BEAT-1 (AAF04782.1), *C. concinna* BEAT-2 (AAF04783.1), *C. concinna* BEAT-3 (AAF04784.1), (AAF04785.1), *P. orientale* SAT (ACI45395.1), *C. roseus* DAT (AAC99311.1), *T. cuspidate* TAT (AAF34254.1), *N. tabacum* HQT (CAE46932.1), *S. lycopersicum* HQT (CAE46933.1), *C. canephora* HQT (ABO77957.1), *D. caryophyllus* HCBT (CAB06427.1), *G. triflora* AHCT (Q9ZWR8.1), *G. scabra* AHCT (BAD44688.1), *L. purpureum* MAT (AAS77404.1) were taken to construct phylogenetic tree. AAT-RS1 and AAT-RS2 were grouped with BEBTs and AAT-RS3 with HCBTs/HQTs, which are shown in *dotted circle*

catalysis and binding acyl CoA to catering of protein conformational requirement (D'Auria 2006; Ma et al. 2005; Buglino et al. 2004; Gibbs et al. 1990). Although, it has been shown earlier that mutation in Asp of DFGWG motif led to either complete inactivity or considerable reduction in activity (65 %) of these enzymes, natural variants of this motif are also present in a few members like Asn (N) was found to be present in place of Asp (D) residue and Val (V) in place of Phe (F) in *P. communis* (AAS48090.1) and *P. armeniaca*, AAT (ACF07921.1) respectively. In AAT-RS3 identified in this study, Tyr (Y) is present in place of prevalent Trp (W). However, the difference in activity of these BAHDs containing variable residues need to be tested by recreating the prevalent DFGWG sequence in the natural variants. Nevertheless, with the x-ray crystal structure available on at least two AATs has clearly demonstrated that the locale of DFGWG on the molecule is far away from the substrates binding site as well as the catalytic site (solvent channel) of the protein molecules (Ma et al. 2005; Unno et al. 2007). The overall lesser identity (25–35 %) in the protein sequences among the different groups of BAHDs but absolute and high conservation of residues (16 amino acids) in this superfamily is a very unique feature of these catalytic proteins. As low as 25 % similarity among the members show the versatility of this family which reflects that these proteins are likely to be involved in the production of an array of secondary metabolites. Although several AATs have been biochemically characterized, function of only few residues were conclusively elucidated and exact biological significance of most of the conserved residues in AATs are unknown. Site-directed mutagenesis at these residues versus and catalytic and/or structural features of the resultant proteins could be helpful in dissecting their roles that may give an impetus to understand the function structure relation of conserved residues. The knowledge can be utilized to characterize number of newly and earlier identified genes belonging to this family. Further, our analysis on conserved residues would be very instrumental in assigning newly identified genes to be grouped in BAHD family.

Sequence similarity and phylogenetic analysis strongly indicated that AAT-RS1 and AAT-RS2 were likely to be the enzymes playing a role in benzoyl transferase and AAT-RS3 might be involved in HCBTs related function. However, the speculations need to be verified by characterizing these proteins biochemically. No specific function for rest of the *A. annua* contigs could be putatively assigned owing to their partial sequences. Based on BLAST analysis contigs 6873, 17424 and 27794 were more similar to BEBTs whereas these contig were grouped with HCBTs in phylogenetic tree. Therefore, partial sequence seems to be an obvious bottleneck to assign any putative function with some certainty especially within superfamily where residues are sparsely conserved at certain positions in all the members, playing different roles, across the genera.

Trichomes are the sites of secondary metabolite production and AATs play significant role in their formation. The expression profile for all the 19 contigs representing unique AATs were studied by semi-quantitative RT-PCR in the leaf, stem and root of *A. annua*. As expected, all the AAT contigs were found to be expressed in leaves as the whole tissue included the trichomes also. Interestingly, almost all the contigs were also expressed in stem of the plant but none of them was found to be expressed in root tissue. Based on the ESTs of BAHD members identified from *Artemisia* glandular trichome and their almost similar expression profile in leaf and stem, and no expression in roots (plant part known to lack these structures), it may be speculated that stem and leaf contain functionally almost identical glandular trichomes with respect to acyltransferases but roots completely lack this set of catalytic functions. It seems leaf and stem AATs share a significantly similar metabolic requirement of ester formations while roots differ. It may be ecologically relevant for plant defense from herbivory, infections, insects etc. to which roots are relatively less exposed and/or less facile. Furthermore, this noted considerable expression of AATs in stem and none in roots is slightly at variance with the reported results on enzymes of terpenoid biosynthesis in *A. annua* as amorpha, 4,11-diene synthase (ADS) catalyzing committed step in artemisinin biosynthesis is most extensively transcribed in juvenile leaves and negligibly in stem and root (Kim et al. 2008). Even post-transcriptionally, levels of proteins (as evaluated by polyclonal antisera-based ELISA) of the three enzymes of artemisinin biosynthesis viz. ADS, cytochrome P_{450} monooxygenase (CYP71AV1) and cytochrome P450 reductase (CPR) are known to be much higher in leaf and stem than in roots (Zeng et al. 2009). Commonality of AAT expression in leaf and stem apart, the noted tissue-specificity of a few of the AATs in the two tissues (Fig. 2) may manifest corresponding specificity of secondary metabolites ester biogenesis synthesis in the plant. Several secondary metabolite esters are known to be synthesized commonly in both the leaf and stem of *A. annua* like linyl acetate, (Z)-α-*trans* bergamotal acetate, benzyl isovalorate, (Z)-3-hexenyl isovalerate etc. Nevertheless, some of the esters in the plant are synthesized in tissue-specific manner like cedr-8(15)-en-9-α -ol acetate in stem and hexyl tiglate, ethyl 2-methyl butyrate in the oil from leaf (Goel et al. 2007).

Methods

Plant material and chemicals

Plants of *Artemisia annua* L. (Asteraceae) cultivar Jeevan Raksha were raised in the glass-house of Central Institute of Medicinal and Aromatic Plants (Lucknow, India) following

standard practices. All oligo synthesis and DNA sequencing was resourced to Bioserve (Hyderabad, India).

ESTScan of *A. annua* transcriptome

The ESTScan programme, based on hidden Markov model (HMM) for prediction and reconstruction of coding regions, was used to identify the homologous sequences of BAHD family members from the transcriptome of *A. annua*.

Assembly of cDNAs sequences

All the contigs from *A.annua* identified as BAHDs were analyzed to get the 5′ and 3′ end extension using BLASTn tool. Only those sequences were considered for extension that shared at least 98 % similarity for at least 200 bases around overlapping region.

RNA isolation

Total RNA from *A. annua* leaf was isolated by Trizol method. The RNA was dissolved in 20 µl of nuclease free water. Quality and quantity of RNA was checked on agarose gel and by spectrophotometrically (Nanodrop).

Molecular phylogenetic analysis

Assembled three complete cDNAs from *A. annua* and other contigs identified as putative BAHDs, were compared with more than 20 biochemically characterized members of BAHD superfamily from other plants for phylogenetic analysis. The accession number of these sequences is given in legends of phylogram. Multiple sequences alignment and other processing of sequences were performed using Clustal W (Thompson et al. 1994) and other tools in MEGA5 software (Tamura et al. 2011). The phylogenetic analysis of sequences was performed using the maximum likelihood (ML), maximum parsimony (MP) and Neighbour-Joining (NJ) methods. For ML, Jones-Taylor-Thornton (JTT) model was used and Nearest Neighbour Interchange was opted for tree inference. Validation of contigs grouping in tree (where required), calculations of the distance matrix and construction of phylogenetic trees with bootstrap tests for 1000 replications were performed with MEGA 5.01 software.

cDNA preparation and semi-quantitative RT-PCR analysis

First strand cDNA synthesis was carried out using 5 µg of DNase-treated total RNA of *A. annua* leaf, shoot and root. The 15 cDNAs were made using cDNA synthesis kit (AMV Reverse Transcriptase, Fermentas) as per the manufacturer's protocol. Primers were synthesized based upon the sequence of contigs of each putative BAHD identified (Supplemental Table 1). Expression of BAHDs at transcript level was checked in leaf, shoot and root using semi-quantitative RT-PCR using gene specific primers. For the semi-qunatitative of PCR reactions, the constitutively expressed actin gene was used to normalize equal amounts of mRNA in the reaction for all the three tissues. The sizes of amplicons ranged from 197 to 360 bp (Supplemental Table 1). The PCR conditions were as follows: 94 °C for 2 min; 28 cycles of 94 °C for 15 s, 55 °C for 30 s and 72 °C for 30 s followed by a final extension at 72 °C for 5 min. The PCR products were ran in 1.2 % agarose gel in 1X TBE buffer. The amplification patterns were documented using a gel documentation system (Syngene, UK).

Table 3 Sequences of primers used to isolate the complete cDNAs of BAHDs from *A. annua*

cDNAs	Primers	Oligonucleotide sequence (5′–3′)	Primers synthesized from
AAT-RS1	AAT-RS1FLF1	CCACCCAAAGTCAACTACATCAAATCC	5′UTR
	AAT-RS1FLF2	CATATGGCACAAACTGACTCTCTCTTG	ORF including Initiation codon
	AAT-RS1FLR1	CACTTTAAAGCATGTTGCGTAGGCAC	3′ UTR
	AAT-RS1FLR2	GGATCCTTATAGTCTTGAAAGCACTTG	ORF including stop codon
AAT-RS2	AAT-RS2FLF1	TCGCGAGCACCATTTCACACAAAC	5′UTR
	AAT-RS2FLF2	CCATATGAAAGAAACCAACACTTCC	ORF including Initiation codon
	AAT-RS2FLR1	ACATGTTATGTAGTCGCCTAGTGAACG	3′ UTR
	AAT-RS2FLR2	GAATTCTCAAATTTTGGAAAGTGTGCG	ORF including stop codon
AAT-RS3	AAT-RS3FLF1	CTCGTATAATACACACATACCTCTGTG	5′UTR
	AAT-RS3FLF2	ATGGAAAATATCCAAAAAAAGAAAATG	ORF including Initiation codon
	AAT-RS3FLR1	CCTTTAGTTAAGACATCAGCAGCTTAC	3′ UTR
	AAT-RS3FLR2	TTAAATCTCGTACAAGAACTTCTC	ORF including stop codon

Isolation and cloning of full length AAT cDNAs (AAT-RSs) of *A. annua*

AAT-RSs have been isolated by PCR from a cDNA library of *A. annua* leaf. Details of the primers and their usage to isolate corresponding acyltransferases (AATs) have been given in Table 3. Primary PCR for all the three AATs viz., AAT-RS1, AAT-RS2 and AAT-RS3 was carried out with their respective gene specific FLF1 and FLRI primers and the PCR products were re-amplified with respective FLF2 and FLR2 primers. This reaction yielded amplicons of approximately 1.4 kb for all the three AAT-RS. The fragments were sliced out from the agarose gel, purified using Qiagen gel extraction columns and cloned in pTZ57R/T (Fermentas). These clones were confirmed by sequencing.

In silico analysis

The protein and nucleotide sequences of BAHDs were downloaded from National Center for Biotechnology Information (NCBI). The transcriptome of *Artemisia* glandular trichome was downloaded from the online (public domain) data files published by Wang and associates (2009). BLAST-x,-p,-n tools at NCBI were used to identify and assemble the sequences of contigs encoding BHADs. Alignments of the nucleotide and protein sequences were performed with ClustalW2 available online at www.ebi.ac.uk. Processing of sequences like cleaning, translation etc. was done using tools hosted on www.justbio.com.

Acknowledgements One of us (RSS) is thankful to the Council of Scientific and Industrial Research (CSIR), New Delhi for the financial support in the form of a CSIR Inter-Laboratory Networked Project (NWP-09). Director, CIMAP is thanked for the research facilities and encouragements

References

Bayer A, Ma X, Stockigt J (2004) Acetyltransfer in natural product biosynthesis - functional cloning and molecular analysis of vinorine synthase. Bioorg Med Chem 12:2787–2795

Brandt SP (2005) Microgenomics: gene expression analysis at the tissue-specific and single-cell levels. J Exp Bot 56:495–505

Buglino J, Onwueme KC, Ferreras JA, Quadri LEN, Lima CD (2004) Crystal structure of PapA5, a phthiocerol dimycocerosyl transferase from mycobacterium tuberculosis. J Biol Chem 279:30634–30642

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol 9:331–340

D'Auria JC, Pichersky E, Schaub A, Hansel A, Gershenzon J (2007a) Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in Arabidopsis thaliana. Plant J 49:194–207

Dai X, Wang G, Yang DS, Tang Y, Broun P, Marks MD, Sumner LW, Dixon RA, Zhao PX (2010) TrichOME: a comparative omics database for plant trichomes. Plant Physiol 152:44–54

Dudareva N, D'Auria JC, Nam KH, Raguso RA, Pichersky E (1998) Acetyl-CoA:benzylalcohol acetyltransferase – an enzyme involved in floral scent production in Clarkia breweri. Plant J 14:297–304

Gibbs MR, Moody PCE, Leslie AGW (1990) Crystal structure of the aspartic-acid-199-asparagine mutant of chloramphenicol acetyltransferase to 2.35-Å Resolution- structural consequences of disruption of a buried salt bridge. Biochemistry 29:11261–11265

Goel DV, Singh M, Ali G, Mallavarupu KS (2007) Essential oils of petal, leaf and stem of the antimalarial plant *Artemisia annua*. J Nat Med 61:187–191

Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey D, Isaac PG, Khan AM, Larson TR, Li Y, Pawson T, Penfield T, Rae AM, Rathbone DA, Reid S, Ross J, Smallwood MF, Segura V, Townsend T, Vyas D, Winzer T, Bowles D (2010) The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. Science 327:328–331

Grienenberger E, Besseau S, Geoffroy P, Debayle D, Heintz D, Lapierre C, Pollet B, Heitz T, Legrand M (2009) A BAHD acyltransferase is expressed in the tapetum of Arabidopsis anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. Plant J 58(2):246–259

Grothe T, Lenz R, Kutchan TM (2001) Molecular characterization of the salutaridinol 7-O-acetyltransferase involved in morphine biosynthesis in opium poppy Papaver somniferum. J Biol Chem 276:30717–30723

Kim SH, Chang YJ, Kim SU (2008) Tissue specificity and developmental pattern of amorpha-4,11 diene synthase (ADS) proved by ADS promoter-driven GUS expression in the heterologous plant Arabidopsis thaliana. Plant Med 79:188–193

Luo J, Nishiyama Y, Fuell C, Taguchi G, Elliott K, Hill L, Tanaka Y, Kitayama M, Yamazaki M, Bailey P, Parr A, Michael AJ, Saito K, Martin C (2007) Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from Arabidopsis thaliana. Plant J 50(4):678–695

Ma X, Koepke J, Panjikar S, Fritzsch G, Stockigt J (2005) Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. J Biol Chem 280:13576–13583

Panikashvili D, Shi JX, Schreiber L, Aharoni A (2009) The Arabidopsis DCR encoding a soluble BAHD acyltransferase is required for cutin polyester formation and seed hydration properties. Plant Physiol 151:1773–1789

Sangwan RS, Agarwal K, Luthra R, Thakur RS, Sangwan NS (1993) Biotransformation of arteannuic acid into arteannuin-B and artemisinin in Artemisia annua L. Phytochemistry 34:1301–1302

Sangwan NS, Kumar S, Sangwan RS (1998) Isolation of genomic DNA from antimalarial plant Artemisia annua L. Plant Mol Biol Rep 16:365

Sangwan RS, Sangwan NS, Jain DC, Kumar S, Ranade S (1999) RAPD profile based analysis of genetic characterization of chemotypic variants of Artemisia annua L. Biochem Mol Biol Int 47:933–944

Sangwan NS, Farooqi AHA, Shabih F, Sangwan RS (2001) Regulation of essential oil production in plants. Plant Growth Regul 34:3–21

Sangwan NS, Sharma PK, Sangwan RS (2007) Geranyl acetate esterase is commonly present but linalyl acetate esterase occurrence is highly limited in plants. Flav Frag J 22:173–177

Sharma S, Sangwan NS, Sangwan RS (2003) Developmental process of essential oil glandular trichome collapsing in menthol mint. Curr Sci 84:544–550

Sharma PK, Sangwan NS, Sangwan RS (2005) TCA facilitated coenzyme A-SH stabilization based end point spectrophotometric DTNB-assay for plant terpene alcohol acetyl transferases (AATs) activity. Anal Biochem 346:176–178

St-Pierre B, De Luca V (2000) Evolution of metabolic pathways. Recent Adv Phytochemistry 34:285–315

St-Pierre B, Laflamme P, Alarco AM, De Luca V (1998) The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. Plant J 14:703–713

Suzuki H, Nakayama T, Yonekura-Sakakibara K, Fukui Y, Nakamura N, Nakao M, Tanaka Y, Yamaguchi M, Kusumi T, Nishino T (2001) Malonyl-CoA:Anthocyanin 5-O-Glucoside-6‴-O-Malonyltransferase from scarlet sage (Salvia splendens) flowers: Enzyme purification, gene cloning, expression, and characterization. J Biol Chem 276:49013–49019

Suzuki H, Nakayama T, Nishino T (2003) Proposed mechanism and functional amino acid residues of Malonyl-CoA: Anthocyanin 5-O-Glucoside-6‴-O-Malonyltransferase from flowers of Salvia splendens, a member of the versatile plant acyltransferase family. Biochemistry 42:1764–1771

Suzuki H, Sawada S, Watanabe K, Nagae S, Yamaguchi M, Nakayama T, Nishino T (2004) Identification and characterization of a novel anthocyanin malonyltransferase from scarlet sage (Salvia splendens) flowers: an enzyme that is phylogenetically separated from other anthocyanin acyltransferases. Plant J 38:994–1003

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution. http://www.megasoftware.net

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Unno H, Ichimaida F, Suzuki H, Takahashi S, Tanaka Y, Saito A, Nishino T, Kusunoki M, Nakayama T (2007) Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. J Biol Chem 282:15812–15822

Walker K, Croteau R (2000a) Molecular cloning of a 10-deacetylbaccatin III-10-O-acetyl transferase cDNA from Taxus and functional expression in Escherichia coli. Proc Natl Acad Sci USA 97:583–587

Walker K, Croteau R (2000b) Taxol biosynthesis: molecular cloning of a benzoyl- CoA:taxane 2α-O-benzoyltransferase cDNA from taxus and functional expression in Escherichia coli. Proc Natl Acad Sci USA 97:13591–13596

Wang W, Wang Y, Zhang Q, Qi Y, Guo D (2009) Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. BMC Genom 10:465

Yang Q, Reinhard K, Schiltz E, Matern U (1997) Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, Dianthus caryophyllus L. Plant Mol Biol 35:777–789

Zeng QP, Zeng XM, Yin LL, Yang RY, Feng LL, Yang XQ (2009) Quantification of three key enzymes involved in artemisinin biogenesis in *Artemisia annua* by polyclonal antisera-based ELISA. Plant Mol Biol Rep 27:50–57