

# Molecular Characterization and Differential Expression of an Aromatic Heptaketide Producing Type III Plant Polyketide Synthase From Himalayan Rhubarb

Shahzad A. Pandith (✉ [drshahzad@uok.edu.in](mailto:drshahzad@uok.edu.in))

University of Kashmir <https://orcid.org/0000-0002-0849-9745>

Niha Dhar

National University of Singapore

Sumedha Bhosale

Savitribai Phule Pune University

Vitthal T. Barvkar

Savitribai Phule Pune University

Sumeer Razdan

Central University of Punjab

Manzoor A. Shah

University of Kashmir

Surrinder K. Lattoo

CSIR-Indian Institute of Integrative Medicine: Council of Scientific & Industrial Research Indian Institute of Integrative Medicine

---

## Research Article

**Keywords:** Aloesone synthase, heterologous expression, molecular docking, promoter analysis, qRT-PCR expression, Rheum australe, southern blotting

**Posted Date:** June 2nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-542003/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Plant Biotechnology Reports on January 15th, 2022. See the published version at <https://doi.org/10.1007/s11816-022-00741-5>.

# Abstract

*Rheum australe* (Himalayan Rhubarb, Polygonaceae) is an endangered medicinal and vegetable herb with known efficacy in different traditional medical systems. The age-old remedying properties of this perennial species are ascribed to the bio-active phytoconstituents viz anthraquinones, stilbenoids, chromones and dietary flavonoids. These metabolites share a major contribution of their synthesis from polyketide pathway which primarily involves the intricate Type III polyketide synthases (PKSs). In this context, present study was focussed on characterization of an important PKS member aloesone synthase (*RaALS*) known to catalyze six polyketide extensions of the starter acetyl CoA to generate aloesone. A full-length cDNA (1176 bp, ~ 42 kDa) was cloned and heterologously expressed in microbial hosts. Multiple sequence alignment and phylogenetic analyses revealed that *RaALS* shares high degree of similarity with orthologous PKSs. Model validation and donor-acceptor interactions were assessed using homology modeling and molecular docking studies. Further, southern blotting determined the existence of *RaALS* as single copy gene. qRT-PCR analyses revealed that higher expression of *RaALS* in leaves followed by stem and root corroborated well with the metabolite accumulation. Its expression was also found to be regulated by altitude. Additionally, an *in vitro* regeneration system was developed to evaluate the effect of various abiotic stressors viz methyl jasmonate, salicylic acid and UV light on the transcript abundance of *RaALS* vis-à-vis identified putative *cis*-regulatory promoter elements using genome walking approach. The study would act as a prelude to utilize this medical herb for prospective metabolic engineering attempts aimed at augmenting bio-active metabolite production for commercial purposes.

## Introduction

The polyketide natural products (PNPs) are a large family of chemical constituents which exhibit both structural and functional disparity. They show wide distribution across the tree of life and occur in microbes to higher kingdom levels (Gokhale et al. 2007). The significance of these metabolites is copiously demonstrated by varied compounds viz erythromycin, lovastatin, rapamycin, streptomycin, besides the low-molecular weight anthraquinones, stilbenoids, and flavonoids (Flores-Sanchez and Verpoorte 2009; Nair et al. 2012; Pandith et al. 2020). The PNPs are known for many biological as well as physiological functions which among others include the defense system in bacteria and plants (Dao et al. 2011; Zeng et al. 2012; Funabashi et al. 2008). Indeed, they are also associated with various pharmacological activities that have been extensively reviewed by our group (Pandith et al. 2020; Pandith et al. 2018). The PNPs are biosynthesized by successive condensation of simple malonyl-CoA-derived two-carbon acetate units with an acyl starter (CoA thioester) by multifunctional enzymes known as polyketide synthases (PKSs) which have three categories viz Type I, II and III based on their architectural configurations. The Type III PKSs are iteratively acting and condensing homodimeric (80-90 kDa in native homodimer state) enzymes with a single KS (ketosynthase) domain which performs all the functions as executed by the important domains of Type I and Type II PKSs (Hertweck 2009). These KS domains are members of the thiolase superfamily which involve Claisen condensation reaction to catalyze carbon-carbon bond formation. The thiolases are divided into four clades viz bacterial and eukaryotic thiolases,

archaeal thiolases, elongation enzymes and initiation enzymes wherein Type III PKSs belong to the later clade characterized by the presence of a Cys-His-Asn catalytic triad (Shimizu et al. 2017). The first experimental characterization of these initiation enzymes was reported by Reimold et al., (Reimold et al. 1983) in 1983 who pioneered in characterizing the symbolic Type III PKS (CHS, chalcone synthase) from cultured cells of *Petroselinum hortense* (Parsley). However, the actual existence vis-à-vis functional insight of CHS in plants was made by Ferrer et al. 1999 by publishing the first report describing its crystal structure from *Medicago sativa* (alfalfa) (Ferrer et al. 1999a). This discovery provided a platform to the researchers in the fields of natural products and biosynthetic engineering and led to the characterization of 58 orthologous groups (plants 28, fungi 5, amoeba and phaeophyceae 4, bacteria 21) of Type III PKSs from bacteria (Funa et al. 1999), amoebozoa (ME and Moore 2006), fungi (Hashimoto et al. 2014) and stramenopiles (Martinelli et al. 2020; Baharum et al. 2011) with about 1000 sequences recorded in the KEGG/GENES database (Kanehisa et al. 2016); the number goes on with recent advancements of genome sequencing technologies.

Studies have shown that the Type III PKSs exhibit their existence in nearly all the studied forms of life, barring animal (metazoan) and archaeal genomes, in a lineage-specific manner with at least a single representative of them that may possibly have paralogous forms as well. While encapsulating the existing functional and phylogenomic information about type III PKSs, and with an aim to cultivate improved computational tools for functional prediction of these intricate enzymes, Shimizu et al., recently proposed a new classification system wherein the plant Type III PKSs have been divided into three major groups (29 types) based on the nature of starter thioester. These three major categories include ring (aromatic) acyl-CoA starters, long-chain acyl-CoA starters, and short-chain acyl-CoA starters (Shimizu et al. 2017). Based on the mechanism of cyclization, later type of PKSs have been further categorized into four groups viz aldol cyclization-, Claisen cyclization-, lactonization- and multiple cyclization- type PKSs. Again, belonging to the later group, aloesone synthase (*RaALS*) is one of the important members of Type III PKS gene family which is known to catalyze six polyketide extensions of the starter acetyl CoA. The enzyme is known to share about 60% identity with other Type III PKSs while maintaining the conserved catalytic triad (Cys<sup>164</sup>, His<sup>303</sup> and Asn<sup>336</sup>) and Phe<sup>216</sup>—numbering in Alfalfa CHS. Nonetheless, unlike the well-characterized CHS and CHS-like enzymes, *RaALS* produces a heptaketide intermediate that undergoes decarboxylative C8-C13 aldol cyclization with an additional second ring formation to ultimately result in the formation of aloesone (2-acetyl-7-hydroxy-5-methylchromone), a biosynthetic precursor of the anti-inflammatory agent aloesin (aloesone 8-C-β-D-glucopyranoside) (Fig. 1). Pertinently, one of the distinctive features of *RaALS* which somehow separates it from the CHS family members is the respective replacement of three conserved residues (Thr<sup>197</sup>, Gly<sup>256</sup>, and Ser<sup>338</sup>; numbering in *Medicago sativa* CHS) with Ala, Leu, and Thr which affect the shape/volume of the initiation/elongation cavity of the active site. However, this ~43 kDa protein is generally known to maintain the overall architecture (fold) as that of the CHS superfamily enzymes (Abe et al. 2006; Lim et al. 2016).

*Rheum australe* (family Polygonaceae) is a high value and most sought-after medicinal herb of North West Himalayas which holds wide pharmacological significance as evidenced by various traditional

medical systems viz Chinese, Unani, Ayurvedic, etc. In fact, the swelling demand, nearly negligible cultivation and reckless harvesting from the wild has rendered the plant threatened (Pandith et al. 2014). The therapeutic properties of this broadly used high altitude perennial herb are ascribed to important class of secondary chemical constituents which mostly include anthraquinones and stilbenoids, besides the potent phytochemical flavonoids (Jiang et al. 2015; Pandith et al. 2020). As the major bio-active compounds are synthesized by Type III PKSs like CHS and CHS-like enzymes, the role of *RaALS* was also anticipated in influencing the biosynthesis of these metabolites. In this context, we characterized this heptaketide producing Type III PKS from *R. australe* for the first time as no significant reports of its effects on the overall growth and development of the plant are reported. It is rather essential to determine the role of *RaALS* in various biochemical and eco-physiological aspects of the plant. Indeed, we have previously reported the functional validation of two CHS paralogs exhibiting promiscuous behaviour with varied CoA thioester substrates to yield the final product naringenin-chalcone/naringenin (Pandith et al. 2016).

In present study, we describe the cloning and characterization of *RaALS*, a critical enzyme for the biosynthesis of chromones, from *R. australe*. The full-length gene was cloned and further expressed in the alternate/heterologous microbial (bacteria and yeast) hosts. The multiple sequence alignment and phylogenetic analyses were performed to determine the evolutionary relatedness of deduced *RaALS* sequence with the related Type III PKSs. Additionally, homology modeling and molecular docking studies were executed to confirm and validate the protein structure, and to gain insights into the donor and acceptor interactions at molecular level. The spatial (different tissues) and altitudinal (different locations along an altitudinal gradient) expression profiling of the transcripts of *RaALS* was also studied. Moreover, the exogenous elicitors viz methyl jasmonate (MeJA), salicylic acid (SA) and UV light, selected on the basis of isolated cis-acting promoter elements, were also evaluated in *in vitro* established cultures to study the expression profile of *RaALS* gene. Further, the gene copy number of the enzyme was determined using Southern-blotting technique. So far, this is the only report of *RaALS* characterization from the medicinal herb *R. australe* implicating its possible role in the regulation and biosynthesis of species-specific bio-active phytoconstituents. This work, therefore, emphasizes the importance of *RaALS* characterization to act as a potent prototype for future metabolic engineering attempts to enhance secondary metabolite production for commercial purposes.

## Methodology

### Plant Material

*In vitro* established cultures of *R. australe* were used as source material. The cultures were established from germinated saplings maintained under greenhouse conditions at Indian Institute of Integrative Medicine, CSIR, Jammu, India (32° 44' N longitude, 74° 55' E latitude; 305 m asl) as mentioned in our previous study (Pandith et al. 2016). In fact, a well-developed and reproducible *in vitro* regeneration system of *R. australe* with varying concentrations of several phytohormones was developed (data not

shown). These cultures were sustained in Murashige and Skoog medium and further used for elicitor (MeJA, SA) and UV-B treatments in addition to the molecular work (Fig. 2).

### RNA extraction and cDNA synthesis

Fresh or cryopreserved samples were used for total RNA isolation as described earlier (Pandith et al. 2016). The extracted RNA was incubated at 37 °C for 30 min with DNase I (Fermentas, Burlington, Canada) to remove the possible traces of genomic DNA. Quality of isolated RNA was assessed by electrophoresis on 1% formaldehyde agarose gel and by determining the absorbance ratio ( $A_{260/280}$ ) using spectrophotometer (AstraAuriga, Cambridge, UK).

For cDNA synthesis, 3-5 µg of DNase I treated total RNA was reverse transcribed using Revert-aid premium reverse transcription kit (Fermentas, Burlington, Canada) with a modified Adapter-oligo-(dT) primer according to the manufacturer's protocol. Briefly, 3-5 µg of purified RNA was mixed with 10 µM oligo(dT) primer and subjected to denaturation at 65 °C for 5 min followed by chilling on ice for 2 min. The reverse transcription reaction was carried out in a final volume of 20 µl reaction which contained 1× first strand buffer (250 mM Tris-HCl, pH 8.3; 250 mM KCl; 20 mM MgCl<sub>2</sub>; 50 mM DTT), 10 mM dNTPs and 1 µl of Moloney murine leukemia virus reverse transcriptase (200 units/µl). The reaction was incubated at 42 °C for 60 min followed by 70 °C for 5 min to inactivate the enzyme. The cDNA was stored at – 20 °C for further use.

### Cloning of *RaALS*

Degenerate primers (Table 1) were designed following the use of Blastn/Blastx (NCBI 2013) and ClustalOmega (Sievers and Higgins 2014) programmes. The primer sequences were based on highly conserved regions of nucleotide sequences of Type III plant polyketide synthases (PKSs) retrieved from the GenBank<sup>TM</sup> data base at NCBI (National Centre for Biotechnology Information). Using synthesized cDNA as template, RT-PCR (reverse transcriptase-polymerase chain reaction) for core amplification of *RaALS* was carried out with the following optimized cyclic conditions: 1 cycle of 94 °C for 3 min; followed by 35 cycles of 94 °C for 30 s, 60 °C for 45 s and 72 °C for 1:30 min; and a final extension of 72 °C for 10 min in a thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA). The screened amplicons of *RaALS* were cloned into pTZ57R/T vector (Fermentas, Burlington, Canada), transformed into an *E. coli* host strain (DH5<sup>TM</sup>; Invitrogen, Merelbeke, Belgium) and sequenced (ABI PRISM<sup>®</sup> 3130XL genetic analyzer; Applied Biosystems, Foster City, CA, USA). Sequences obtained were analyzed using the similarity search Blastn (NCBI 2013) programme to ensure homology and were further used to design gene specific primers (GSPs) for RACE PCR.

### 5' and 3' RACE

5' and 3' RACE was performed using Gene Racer cDNA amplification kit according to the instructions given in product manual (Invitrogen, USA). Respective cDNAs obtained were used to generate the flanking regions of the core amplicons from either side separately in two sets of PCRs. The primary reaction was

performed using 5'/3' RACE adapter primer (5'/3' RACE\_OUT) and 5'/3' ALS\_OUT as GSPs, whereas the secondary reaction was carried out using inner adapter primers (5'/3' RACE\_INN) and 5'/3' ALS\_INN as GSPs while using amplified products from first reaction set as template. All the PCR reactions were carried out in a 50 µl reaction volume containing 1 µl cDNA as template (except for secondary reaction which uses amplified products of initial reaction as template), 2.5 µl each of 10 µM adapter primers and GSPs for respective reactions and 44 µl of master mix (33.5 µl PCR grade water; 10 mM Tris HCl, pH 9.0; 50 mM KCl; 2.5 mM MgCl<sub>2</sub>; 0.2 µM dNTPs and 2.5 U of Taq DNA polymerase). The reaction thermo-profile for both initial and nested PCR amplifications was as follows: 3 min at 94 °C, 35 cycles (30 s at 94 °C, 30 s at 60-65 °C, 1:30 min at 72 °C) and 10 min at 72 °C followed by hold step at 4-15 °C. The obtained 5' and 3' nested amplicons were purified and sub-cloned into pTZ57R/T vector and further sequenced. The sequences acquired from core, 5' and 3' RACE fragments were refined (trimmed to eliminate bad quality bases with low confidence scores, QV<20), aligned and subsequently analyzed using Blastn/BlastX (NCBI 2013) tools to validate the prediction of targeted ALS.

### **Full-length cloning of *RaALS***

The sequences of core fragments and 5'/3' RACE products were compared and aligned to generate the full-length cDNA of *RaALS* which was then amplified using full length primers viz FulALS\_F and FulALS\_R (Table 1). A high fidelity proof-reading DNA polymerase (New England Biolabs, Herts, UK) was used for amplification of complete ORF of *RaALS* while using the respective primers. The reaction thermo profile was as follows: 1 cycle for 3 min at 94 °C; followed by 35 cycles of 94 °C for 30 s, 60 °C for 45 s, and 72 °C for 1:30 min; and a final extension at 72 °C for 10 min. Full length ORF was ligated into blunt-ended pJET vector and consequently subcloned into *E. coli* DH5α.

### **Computational and Phylogenetic Analyses of *RaALS* Sequence**

The complete cds fragments obtained were translated using Translate tool (Gasteiger et al. 2003), secondary structures predicted by SOPMA (Geourjon and Deleage 1995) and the properties of deduced amino acid sequences were estimated using ProtParam and Compute pI/Mw (Gasteiger et al. 2005) programmes. SignalP 4.1 (Petersen et al. 2011) was used to determine the presence of signal peptides and transmembrane helices predicted by TMHMM v. 2.0 (Möller et al. 2001). The amino acid sequences were aligned using Clustal Omega Multiple Sequence Alignment tool (Sievers and Higgins 2014) and the identification of functionally and structurally important residues in protein sequences was done by using ConSeq and ConSurf programmes (Ashkenazy et al. 2010). The deduced amino acid sequence of *RaALS* was aligned with multiple CHS plant sequences using Clustal W embedded in MEGA X (MEGA 2018). Maximum Likelihood Method was used to construct the phylogenetic tree with MEGA X software with 500 bootstrap replicates to obtain confidence levels with the branches.

### **Homology Modeling and Model Validation**

The primary sequence of *RaALS* was retrieved from UniProt (Accession number A0A0A0P5P9). Crystal structure of CHS1 (PDB ID: 4YJY, Resolution 1.86 Å) was chosen as template with 64% sequence identity

and 99% query coverage. Homology model was built using the SWISS-MODEL online server (Waterhouse et al. 2018). Model was energy minimized using Chimera version 1.13 (Pettersen et al. 2004) and amber ff12SB force field was used for parameterization. Indeed, hundred runs of steepest descent followed by conjugate gradient cycles were performed for energy minimization and QMEAN score was computed to analyse the overall quality of models (Benkert et al. 2011). The generated model was further validated by different structure assessment tools. Ramachandran plot was generated using PROCHECK tool (Laskowski et al. 1993) while ProSA (Wiederstein and Sippl 2007) made plot of local model quality. Further, molecular graphics and RMSD analyses was performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

### **Docking Simulations of Substrate and Product Molecule to Aloesone Synthase**

Model structure was subjected to protein-ligand docking by AutoDock Tools version 4.2 (Morris et al. 2009). Starter molecule acetyl-CoA and product aloesone were docked to the protein model. Prior to molecular docking all hydrogen atoms were added to the protein structure followed by Gastegier charges. AutoGrid 4.2 module of AutoDock 4.2 was used to produce grid maps. Hundred independent runs with step size of 0.2Å for translations and 5° per step for torsions and initial population of random individuals with population sizes of 300 individuals were set as fixed parameters for all the docking analyses using genetic algorithm (GA). For GA, the maximum number of energy evaluation was set to 2500000, maximum number of generations was set to 27000, and maximum number of top individual that automatically survived was set to 1 with the mutation rate of 0.02 and crossover rate of 0.8. The Lamarckian Genetic Algorithm was chosen for generating the best conformer and other docking parameters were set as default. Docked complexes were visualized by PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.

### **Plasmid Construction and Heterologous Expression in *E. coli* and yeast**

Full length coding sequence of *RaALS* gene was modified by adding restriction sites. Both forward (FulALS\_F) and reverse (FulALS\_R) primers were engineered to introduce *Bam*HI and *Eco*RI restriction sites at the beginning and end of the coding sequence, respectively. The primers were tailored for directional cloning in pGEX4T-2 vector and heterologous expression in *E. coli*. ORFs containing modified restriction sites were cloned and excised from pJET vector (Fermentas) with *Bam*HI and *Eco*RI and reconfirmed by sequencing before sub-cloning into the respective restriction sites of pre-digested and purified bacterial expression vector pGEX-4T-2. The cloned *RaALS* was expressed as fusion protein with GST-tag at N-terminus of the hybrid expression vector. Further, the heterologous expression of the recombinant protein was carried out as described by us earlier (Pandith et al. 2016). Briefly, the pGEX-*RaALS* expression cassette was transformed into *E. coli* BL21 (DE3) cells (Invitrogen, Merelbeke, Belgium) and single colony from the recombinant culture was cultured in Luria-Bertani (LB) medium containing ampicillin (100 mg/ml) at 37 °C on a shaker at 200 rpm for overnight. 1% culture was inoculated in fresh LB medium with respective antibiotic and incubated at 37 °C until optical density ( $A_{600}$ ) reached 0.4-0.6.

The protein expression was induced by adding varied concentrations (0.2 to 1 mM) of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG; Fermentas, Burlington, Canada) into the cultures. The cells growing at same condition without IPTG induction were used as control. The cultures were constantly incubated at 30 °C for 8-12 h and cells were harvested after every two hours by centrifugation. The cell pellets obtained were resuspended in 6X SDS loading buffer (0.375 M Tris, pH 6.8; 12% SDS; 60% glycerol; 0.6 M DTT; 0.06% bromophenol blue) and the expression was determined by running the samples on 10% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE).

Additionally, *RaALS* gene was also expressed in the fission yeast, *Schizosaccharomyces pombe* as described in one of the earlier investigations from our lab (Dhar et al. 2014). *Bgl*II restriction endonuclease (Table 1) was employed to design, clone (directional cloning), digest and further express the *RaALS* gene in *S. pombe* using the expression vector pDS472a under the control of *nmt1* promoter.

### Genomic Southern Blot Analysis

The genomic DNA was isolated from mature leaves of *R. australe* using the standard cetyl-trimethyl-ammonium bromide (CTAB) method with slight modifications to increase the yield of DNA extracted (CTAB DNA extraction buffer 2009). About 20  $\mu$ g of extracted, purified and quantified (at A260/280 nm) DNA was digested for 16-20 h at 37°C with restriction endonucleases *Bam*HI, *Eco*RI (non-cutters) and *Not*I (single cutter). The digested product was separated by electrophoresis on 0.7-0.8% (w/v) agarose gel in Tris-borate/EDTA buffer and then transferred onto a positively charged nylon membrane (Roche, Germany) and hybridized with the digoxigenin (DIG)-labeled DNA probe as described in our previous studies (Pandith et al. 2016; Rana et al. 2013; Dhar et al. 2014). The probe for *RaALS* was synthesised by PCR using full length primers (FulALS\_F and FulALS\_R; Table 1) that amplified the coding sequence of the gene. The steps of probe labelling, hybridization, blocking, washing, and signal detection were done as per the instructions given in the user manual of the DIG-DNA labelling and detection kit (Roche).

### Tissue-Specific and Site-Specific Transcript Study

The tissue- and site- specific expression profiling of *RaALS* was determined by qRT-PCR (quantitative real-time PCR) analysis. To study this relative expression, total RNA was isolated from different plant parts viz root, stem and leaf, and from the leaf tissues of plant samples collected from four different geographic locations (L<sub>1</sub>, L<sub>2</sub>, L<sub>3</sub> and L<sub>4</sub>) of North Western Himalayas (Table 2) as reported earlier (Pandith et al. 2016). For every sample, 3  $\mu$ g of DNase-treated RNA was reverse transcribed using the iScript cDNA synthesis kit (Bio-Rad) to synthesize the first strand cDNA as per the product manual. The qRT-PCR reactions were performed in triplicates following SYBR-based chemistry (Takara, Japan) while using the ABI StepOne real-time quantitative PCR system (Applied Biosystems, USA). Concisely, the standard reaction contained 0.5  $\mu$ l of cDNA as template, 200  $\mu$ M each primer (Table 1), 5  $\mu$ l of SYBRPremixExTaq and nuclease-free water to make up the final volume to 10  $\mu$ l. The manufacturers manual was followed for cycling conditions: holding stage of one cycle at 95 °C for 10 min, followed by the cycling stage (40 cycles) of 95 °C for 15 s and 60 °C for 1min, and finally melting curve stage of 95 °C for 15 s, 60 °C for 1

min, and 95 °C for 15 s. The real-time primers were designed by Primer Express version 3.0 (Applied Biosystems) and were further confirmed by a dissociation curve (observation of a single peak for each primer pair). A housekeeping gene  $\beta$ -actin, amplified with Actin\_F and Actin\_R primers (Table 1), was used as an endogenous control to normalize the expression of *RaALS*. The qRT-PCR data generated was analyzed on the basis of comparative Ct (cycle threshold) method as described earlier (Pandith et al. 2016).

## Promoter Isolation

Genome walking approach was followed to determine the promoter sequence of *RaALS* gene using GenomeWalker Universal Kit (Clontech). Genomic DNA was isolated from fresh and young leaves of *R. australe* using DNeasy Plant mini Kit as per the instructions of the manufacturer (Qiagen, Germany). To construct GenomeWalker DNA libraries, the genomic DNA was further digested with four different blunt-end generating restriction endonucleases (DraI, PvuII, EcoRV, and StuI) in separate tubes each containing 5  $\mu$ g of the extracted DNA. The digested samples were recovered using a PCR purification kit (Qiagen) and ligated individually to the GenomeWalker AP adaptor (provided with kit) to produce four adapter-ligated libraries to be used as templates in PCRs consisting of two walking amplification steps (primary and nested PCR). Two-round PCR reactions were carried out using AP1 and AP2 adapter primers (provided with the kit) and the gene specific primers (Table 1) in light of the kit manual with slight modifications. The final products generated from this exercise were analyzed on agarose gel (1.5-2%), purified and further cloned into pTZ57R/T cloning vector for sequencing to obtain the promoter sequence of *RaALS* gene. The sequence thus generated was processed and then subjected to the PlantCare (Lescot et al. 2002) and New PLACE (Higo et al. 1999) databases to identify various putative *cis*-acting elements in the promoter region of this specific Type III PKS.

## Elicitor Treatment and qRT-PCR Expression Analyses of *RaALS*

*In vitro* cultures of *R. australe* were developed to determine the variation in the accumulation pattern of *RaALS* transcripts upon treatments with different plant-derived endogenous elicitors as discussed in our previous study (Pandith et al. 2016). Briefly, the axenic cultures were adapted for 2 weeks in MS (Murashige and Skoog) liquid medium before subjecting to elicitations by MeJ (methyl jasmonate, 0.1 mM) and SA (salicylic acid, 0.1 mM) and UV-B radiation. The elicitors were chosen in correspondence to the promoter motifs identified in the upstream region of the *RaALS* gene. The control and treated plant tissue samples were harvested for RNA extraction at defined time periods of 0 (in case of control), 12, 24 and 48 h, except for UV-B exposed samples which were harvested at 3, 6 and 9 h. The RNA samples from control and treated samples were then reverse transcribed to synthesise cDNA and the effects of elicitor treatments on the expression analysis of *RaALS* gene were studied using qRT-PCR with the same parameters as used for the tissue- and site- specific transcript study discussed above.

## Statistical Analyses

Data was analysed using GraphPad Prism for Windows Version 5.0 by one-way ANOVA followed by Tukey's test ( $P < 0.05$ ) to determine the differences among means.

## Results And Discussion

### Molecular Cloning of *RaALS*

A homology-based approach was used to isolate and identify the ALS gene from *R. australe* while taking benefit from some of the highly conserved amino acid sequences of Type III plant polyketide synthases. First, a 319 bp core cDNA fragment was obtained by PCR using degenerate primers (Table S1) which was subsequently followed by 5' and 3' RACE generating 496 and 361 bp fragments, respectively. Assemblage of all these fragments resulted in the generation of an 1176 bp full length cds sequence of the *RaALS* gene. The obtained sequence (ORF) was submitted to the NCBI GenBank data base under the accession number KC473812. The nucleic acid sequence alignment showed a reasonable level of sequence similarity to related Type III PKS's using BLASTN/BLASTX analysis tools. Additionally, using the BLASTX/BLASTP algorithm, the *RaALS* amino acid sequence displayed extended sequence similarity (64% - 88%) with orthologous sequences from other plant species belonging to the order Caryophyllales. To cite a few: *Chenopodium quinoa* (GenBank accession no. ATY35195.1), *Drosophyllum lusitanicum* (GenBank accession no. ABQ59603.1), *Fagopyrum tataricum* (GenBank accession no. AHA14502.1), *Fallopia multiflora* (GenBank accession no. ATY35195.1), *Polygonum cuspidatum* (GenBank accession no. ABE68892.1), and *Spinacia oleracea* (GenBank accession no. XP\_021835131.1).

### *In silico* characterization and phylogenetic analysis of deduced *RaALS*

The ORF of *RaALS* was subjected to translation to generate 391 amino acids, corresponding to protein of 43.35 kDa with calculated pI value of 5.74. The secondary structure analysis revealed that *RaALS* are predominantly  $\alpha$ -helical protein with respective percentages for  $\alpha$ -helices (43.99%, 172 residues), random coils (34.27%, 134), extended strands (15.09%, 59 residues) and  $\beta$ -turns (6.65%, 26 residues). The computed instability index (II) was found to be 37.26 confirming the stability of the protein secondary structure. Further, the aliphatic index of 83.81 and grand average of hydropathicity (GRAVY) of -0.157 indicates toward the hydrophobic nature of the protein. Pertinently, and as confirmed by SignalP 4.1 and TMHMM v. 2.0 servers, *RaALS* was also found to lack any signal peptides and, thus, the transmembrane helices normally associated with the hydrophobic proteins. Further, analysis of the evolutionary conservation of *RaALS* amino acid sequence revealed various high-score structural residues to be functional (Fig. S1).

The well-known conservative nature of Type III PKSs was utilized to identify the catalytically important residues in *RaALS* using ClustalW multiple sequence alignment tool (Fig. 3). Importantly, the highly conserved amino acid residues found in nearly all Type III PKSs were also observed to maintain their identity in the primary amino acid sequences of *RaALS*. The multiple sequence alignment revealed that the active site residues in CHS family viz Cys<sup>165</sup>, Phe<sup>216</sup>, His<sup>304</sup> and Asn<sup>337</sup> (marked '\*') are well

conserved in *RaALS*. The characterized enzyme also displayed conserved active site residues, which include Phe<sup>139</sup>, Gly<sup>164</sup>, Gly<sup>168</sup>, Leu<sup>215</sup>, Asp<sup>218</sup>, Gly<sup>264</sup>, Gly<sup>336</sup>, Phe<sup>305</sup>, Gly<sup>306</sup>, Gly<sup>307</sup> (marked '\*') along with 'GVLF $\beta$ GF' (displayed by green block) signature motif as found in CHSs. In the malonyl CoA motif, Glu<sup>315</sup>, Lys<sup>317</sup>, Leu<sup>318</sup>, Leu<sup>320</sup>, Glu<sup>321</sup>, Lys<sup>323</sup> and Arg<sup>330</sup> (displayed by red block) remained conserved in *RaALS*. Additionally, five out of the seven amino acid residues including Thr<sup>133</sup>, Met<sup>138</sup>, Phe<sup>216</sup>, Phe<sup>267</sup> and Pro<sup>376</sup> (marked by '+') which are involved in formation of the cyclization pocket (part of the active site architecture of the enzyme) were found to be conserved in *RaALS* with an exception of Ile<sup>256</sup> and Gly<sup>258</sup> which were non-synonymously changed to Met and Leu, respectively. As discussed in our previous investigation (Pandith et al. 2016), the cyclization pockets' topology guides polyketide folding and thereby product formation. Apparently, any change in the cyclization pocket may shake the stereochemistry of the cyclization reaction and temper with product selectivity. Though subject to further empirical experimental findings, it might be one of the possible reasons for difference in substrate selectivity and product formation for *RaALS* which derails a bit from well-studied Type III PKSs, CHSs, in particular. Furthermore, three of the inert residues (Thr<sup>197</sup>, Gly<sup>256</sup>, and Ser<sup>338</sup>; numbering in *Medicago sativa* CHS) affecting the shape/volume of the initiation/elongation cavity of active site and thus known to hold substantial importance in controlling the growing polyketide product (Ferrer et al. 1999b) are again non-synonymously replaced in *RaALS* with Ala, Leu, and Thr, respectively. This event is also anticipated to play a possible role in altered catalytic activity of the isolated enzyme.

Pertinently, there are no crystal structures of aloesone synthase present in Protein Data Bank (PDB). Therefore, homology model of *RaALS* was built using SWISS-MODEL online web service. The crystal structure of CHS with PDB ID 4YJY (Go et al. 2015) was chosen as template structure by BLASTp search against PDB database. Template structure showed 64% sequence identity and 99% query coverage against the query sequence of *RaALS*. The homology model generated by Swiss-Model is shown in Figure 4a. Moreover, *RaALS* shares same evolutionary family with CHS and CHS-like Type III PKSs. Indeed, homology modeling of *RaALS* shows five-layer  $\alpha\beta\alpha\beta\alpha\beta$  core structure similar to that of CHSs (Weng and Noel 2012). The RMSD between template structure and model structure is 0.11 Å which supports the high similarity between both structures and suggests a good model quality. Additionally, the model was validated using Ramachandran plot predicted by PROCHECK. The plot showed 94.1% of the residues in the core region, 5.3% in allowed region while 0.3% in disallowed region as shown in Table S2 and Figure 4b. Residues in the disallowed region were named and colored red on the plot. This revealed that majority of amino acids were in phi-psi distribution of Ramachandran plot, hence the model is of good quality with overall residue-residue geometry and good stereo chemical quality. Furthermore, the ProSA analysis of model was computed to understand erroneous energies of each amino acid in the protein (Fig 4c). As per the plot, all the amino acids show negative energies. The model was found to be ideal with regions mostly falling in correctly determined region by QMEAN Plot (Fig 4d). The verified 3D score of model was found to be 99.74% (Table S2, Fig. 4e) representing very high agreement of 3D structure to the 1D sequence. Hence, the model proved to be validated in terms of structural geometry and energy profiles suggestive of the perfect starting point for prospective phases of model analyses.

Phylogenetic tree was constructed to understand the evolutionary relationship of *RaALS* (Fig 5) with orthologous Type III PKSs. Deduced protein sequence from cloned *RaALS* gene was employed to construct the phylogenetic tree using Maximum likelihood Method with MEGA X software (Kumar et al. 2018). CHS from Chinese peony (*Paeonia lactiflora*) was used as an outgroup to root the phylogenetic tree. The taxon is a member of the family Paeoniaceae (order, Saxifragales) that is known to share a close evolutionary association with Caryophyllales. The results obtained here indicated that the Type III PKSs from Caryophyllales show an interesting evolutionary history. *RaALS* from *R. australe* is seen grouped with PKS enzymes from members of Polygonaceae and Plumbaginaceae both of which belong to the Caryophyllales order. In general, most of the PKSs from genus *Rheum* form a separate monophyletic group like the ones from *Polygonum*, *Fallopia*, *Fagopyrum* and *Dianthus*. The specific heptaketide producing PKS characterised in the present study exhibited a close relationship with its orthologous member from the congeneric species *R. palmatum* (Fig. 5). This little clad advocates their early divergence. In other words, the two members seem to have diverged before the speciation event as shown for *R. australe* CHS members in our previous investigation (Pandith et al. 2016).

### Molecular docking of *RaALS*

Acetyl-CoA is the starter molecule for *RaALS* and carries out successive six condensation reactions combined with malonyl-CoA in order to produce aloesone (Fig. 1). Docking studies of acetyl-CoA and aloesone helped to gain insights into interactions at molecular level. Acetyl-CoA being a large bulky molecule takes a large cavity for binding. We found that acetyl-CoA shows major interactions with Leu<sup>257</sup> and Thr<sup>339</sup> active site residues, and Ala<sup>198</sup> which helps in extending the buried pocket (Fig 6a). Interestingly, all the docked complexes of acetyl-CoA and *RaALS* show high binding energy which might be due to bulky size of molecule or possibly the limitation of docking algorithm to dock the large ligand. On the other hand, product aloesone was observed to bind deeper inside the pocket with low binding energy of -6.86 Kcal/mol (Fig. 6b).

### Recombinant heterologous expression in alternate microbial hosts

The cds region of *RaALS* was expressed in *E. coli* BL21 (DE3) to establish the identity of isolated PKS. The recombinant protein was expressed under the control of Ptac hybrid-promoter using pGEX-4T-2, an *E. coli* expression vector induced by IPTG. SDS-PAGE analysis demonstrated that the fusion protein showed optimum expression when induced with 0.8 mM IPTG for 6 h at 30 °C (Fig. 7). The molecular weight of the expressed recombinant enzyme was apparently ~ 69 kDa and was in agreement with the predicted molecular mass of *RaALS* (calculated from its deduced amino acid sequence) including that of GST (25.99 kDa). In our previous investigation (Pandith et al. 2016), we had demonstrated the functional expression of recombinant Type III PKS's in soluble fraction with a homogeneous molecular weight of approximately 42 kDa, excluding that of the associated GST tag. However, *RaALS* could not be expressed in soluble fraction in sufficient amount as most part of it remained localized to the inclusion bodies hampering its purification. In fact, the specific Type III PKS was also expressed in *S. pombe* using pDS472a expression vector, but could not be purified for some abstruse reasons.

## **Determination of gene copy number— *RaALS* exists as single copy in *R. australe***

To estimate the number of *RaALS* gene copies in *R. australe* and to validate this Type III plant PKS, genomic southern blot analysis was performed under high stringency conditions as discussed in some of our previous studies (Pandith et al. 2016; Dhar et al. 2014; Rana et al. 2013). Three restriction enzymes used viz *Bam*HI and *Eco*RI— without any restriction site in the *cds* region of the gene, and *Not*I— with a single cutting site, generated the restriction pattern which was consistent with a single-copy gene (Fig. 8). A single band was scored in *Bam*HI and *Eco*RI digested DNA samples and two bands were observed with *Not*I digestion which indicated that PCR synthesized DIG-labeled probe hybridized to the specific gene of interest. The results obtained in southern hybridization thus suggest that *R. australe* genome contains a single copy of *RaALS* gene. The gene family members vary in number from one to the other species owing to the need of attributes they are ascribed to. Interestingly, including two earlier characterized members (Pandith et al. 2016), so far, we have identified three members of the Type III PKS's from *R. australe* which exhibit varied/promiscuous functions in the path of secondary metabolism vis-à-vis dynamic eco-physiological attributes besides environmental agitations of varied nature. Though subject to further empirical experimental investigations, the Type III PKS gene family does not seem to have evolved much (to generate its various homologous members) in this important medicinal herb over the course of evolution.

## ***RaALS* is highly expressed in leaves of *R. australe***

The Type III plant PKS's, in general, exhibit differential expression pattern in different tissues and in response to the environmental perturbations. In our previous study (Pandith et al. 2016) on *R. australe* CHS's— the simplest representatives of Type III PKS's, a comparative analyses of flavonoids and anthraquinones revealed significantly higher content of rutin (the predominant flavonoid) and chrysophanol (the predominant anthraquinone) in leaves as compared to that of stem and the root tissues. The same samples were analysed using relative quantitative real-time PCR to examine the spatial regulation of *RaALS* gene in different tissues (leaf, stem and root) of *R. australe* (Fig. 9a). Though expressed in all tissues, the transcripts of *RaALS* gene exhibited a distinct expression pattern with leaves ( $1.01 \pm 0.17$ ) showing maximum transcript abundance, followed by stem ( $0.85 \pm 0.63$ ) and the root ( $0.19 \pm 0.01$ ). Here, *RaALS* expression corroborates well with the specific metabolite accumulation we had observed in leaves as compared to the other plant tissues (Pandith et al. 2016); and supplemented by another study on *Fagopyrum esculentum*, a Polygonaceous member well studied for rutin biosynthesis (Li et al. 2010). Similar results are reported in mulberry wherein specific Type III PKS's (cluster B genes) were shown to express more in leaves and flowers compared to other tissues (Li et al. 2016). It thus seems that the PKS isolated here is possibly involved in the biosynthesis of specified major metabolites.

## ***RaALS* exhibits differential expression pattern along the altitudinal gradient**

Altitude, one of the significant natural experiments to test the ecological and evolutionary responses of living world to the non-living geo-physical influences, is known to have a prominent effect on accumulation of the secondary chemical constituents in plants. In our previous investigation's, we found

a general increasing trend of secondary metabolite constituents with increasing altitude (Pandith et al. 2016; Pandith et al. 2014), and the ploidy status as well (Farooq et al. 2013). In particular, major flavonoid constituents viz naringenin and rutin were seen to accumulate at greater concentrations at higher altitudes. This observation prompted us to ascertain the levels of *RaALS* transcripts in same tissues (as used in previous investigations) collected from four different geographic regions, and to correlate the findings with the targeted metabolite accumulation. Fortunately, we noticed consonance in *RaALS* expression levels and the altitude wherein its transcript level was nearly six fold in plant samples collected from Nyoma valley, Ladakh ( $11.66 \pm 2.34$ ;  $33^{\circ} 08' 661''$  N,  $78^{\circ} 34' 742''$  E, 4415 m asl) to that in Pen se La top, Ladakh ( $2.04 \pm 0.49$ ;  $33^{\circ} 51' 08''$  N,  $76^{\circ} 21' 57''$  E, 4287 m asl) samples which was followed by two sites of Kashmir viz Yarikhah farm, Gulmarg ( $1.84 \pm 0.97$ ;  $34^{\circ} 04' 797''$  N,  $74^{\circ} 26' 448''$  E; 2119 m asl) and Bonera farm, Pulwama ( $1.05 \pm 0.39$ ;  $33^{\circ} 52' 59''$  N,  $74^{\circ} 55' 00''$  E; 1630 m asl), respectively (Fig. 9b). In general, the linear increasing trend of *RaALS* transcripts with the increasing metabolite accumulation vis-à-vis altitude suggests a possible role of the specific Type III PKS in biosynthesis of the metabolites determined. Further, such observations of this exercise were in conformity with some of the investigations of our own research group (Bhat et al. 2014a; Bhat et al. 2014b; Bhat et al. 2013), besides earlier reports on *F. tataricum* (Guo et al. 2011) and *Arnica montana* (Spitaler et al. 2006).

### Identification of putative *cis*-regulatory elements from isolated promoter sequence

In eukaryotes, the process of transcription typically follows an event of specific recognition between a transcription factor (TF) and its binding site (*cis*-element) which is usually located at the upstream region of a gene. Knowing the particular sequence of this TF binding site holds critical importance in gene regulation (Walhout 2006). Taking this in mind, and to elucidate the transcriptional regulation of *RaALS*, genome walking approach was used to isolate the 5' upstream region of this gene. This strategy led to the isolation of 416 bp promoter sequence which was scanned *in silico* with PlantCare (Lescot et al. 2002) and PLACE (Higo et al. 1999) online tools for the identification of various putative *cis*-regulatory elements. Pertinently, the *RaALS* promoter sequence was found to contain various important *cis*-regulatory elements (Fig. 9c, Table 1), besides having a typically high A + T content (56.25%) usually seen in other plant promoters.

The predicted transcription initiation site (TIS, +1) was found to be located at 116 bp upstream of the *RaALS* start codon. SOGO (New PLACE) analyses of the isolated promoter sequence showed the location of putative TATA box (essential for initiating transcription process) at position 98 (+) upstream with respect to TIS. The eukaryotic promoter regulatory consensus *cis*-acting element CAAT box, commonly found in promoter and enhancer regions, was observed at positions 153 (+), 220 (+), 261 (-), 262 (-), 380 (-), 381 (-). The light responsive elements viz Box I, GATA box and GT-1 motif were also found in the *RaALS* promoter region. Further, several other functional elements such as MYB core, CGTCA motif, TC rich repeats, TGACG motif/TCA element and Skn-1\_motif were also found at various positions in the promoter sequence. These elements are reportedly known for various activities including defense and stress responsiveness and tissue specific expression etc. MYB elements, found in numerous plant promoters, are known to serve as binding sites for R2R3 MYB TFs (Duraisamy et al. 2018). Additionally,

the light responsive element, GT-1 motif, is also known to be involved in SA elicited gene expression (Bhat et al. 2014b). Specific regulatory elements (Box I, CGTCA motif, TGACG motif/TCA element, TC rich repeats and GT-1 motif) from *RaALS* promoter region were selected to determine their role in stress responsiveness under *in vitro* conditions to examine their inducible/repressible nature.

### Expression of *RaALS* is inducible by MeJA, SA and UV-B elicitation

Elicitors, in general, are the chemicals or bio-factors from varied sources which can induce the physiological and morphological changes within a living host. In plants, the elicitors, biotic or abiotic, generate a range of defense reactions which mostly include the accumulation of an array of defensive secondary chemical constituents. Indeed, contemporary developments in comprehension of plant signaling pathways have flagged an approach of utilizing elicitor-induced secondary metabolite variation in plants as a projecting device for potential pathway elucidation, regulation and prospective engineering studies. Therefore, to examine and evaluate the constitutive or inducible nature of *RaALS* gene, and to validate the results generated with the *cis*-regulatory elements identified from the promoter region of the target gene, *RaALS* transcripts were assayed in response to the exogenous induction of MeJ (0.1 mM), SA (0.1 mM), and UV-B light (1,500  $\mu\text{J m}^{-2}$ ) exposure.

Jasmonic acid, its related (MeJ) and even conjugated compounds, is reportedly well known transducers and mediators leading to the production and accumulation of various classes of plant secondary metabolites (Zhao et al. 2005; Tamogami et al. 1997; Farmer et al. 2003). In present study, MeJ treatment significantly instigated the expression of *RaALS* gene with a steep increase. The mRNA levels peaked at 24 h with nearly 18 fold increase as compared to that of control. However, the transcript level slightly declined afterwards as shown in the Figure 9d. These observations are in conformity with some of the earlier investigations on Type III plant PKSs (Richard et al. 2000; Yu et al. 2015; Pandith et al. 2016).

Though not a universal agent to induce production of the defensive compounds in plants, SA is known to tempt expression of genes related to the generation of some classes of plant secondary metabolites (Taguchi et al. 2001). In *R. australe*, SA elicited a gradual, but nearly significant, increase in the transcript abundance of *RaALS* gene. About two fold rise was registered in the expression profile at 24 h post induction after which the expression declined to less than the control value (Fig. 9e).

The varied response of plants against irradiation acts by triggering protection/repair mechanisms which mostly involve biosynthesis of the UV absorbing secondary metabolite constituents; in particular, phenolic compounds (Hahlbrock and Scheel 1989; Dao et al. 2011). Earlier studies have revealed that these UV absorbing chemical compounds accumulate distinctly upon UV-B (280–320 nm) exposure which has been further shown to be the consequence of upsurge in transcript levels of various enzymes of phenylpropanoid biosynthetic pathway including Type III PKSs (Jaakola et al. 2004; Davies and Schwinn 2003). To analyze the effects of UV-B irradiation on *RaALS* transcript abundance, the *in vitro* raised plantlets were subjected to UV-B light treatment for 9 h under controlled conditions. As shown in the Figure 9f, the transcripts of *RaALS* gene displayed an irregular, but increasing, trend in their

accumulation. Nearly a 6 and 4 fold increase was observed at 9 and 3 h post UV-B induction, respectively. In general, the results obtained were in agreement with some of the earlier investigations of Type III PKS gene family members on Soybean (Tuteja et al. 2004) and Mulberry (Li et al. 2016).

## Conclusions And Future Perspectives

Owing to the treasury of polyketide-derived bio-active secondary chemical constituents, PKSs have cultivated a growing and vital interest among scientists toward the polyketide biosynthetic machinery. Indeed, the simplicity of Type III PKSs has made them soft and achievable targets for their understanding and rational protein engineering. In this context, the present work is an endeavor to understand the molecular basis of biosynthesis of species-specific bio-active phytoconstituents. This minor effort entails the identification, *in silico* and molecular characterization, copy number determination, heterologous expression and phylogenetic evaluation of one of the important Type III PKS orthologs from *R. australe*. The transcript levels of *RaALS* differed spatially and with altitude, and were also found to be modulated by various cues (exogenous elicitations) based on the identified *cis*-regulatory elements vis-à-vis accumulation of major bio-active chemical constituents. Further, the promiscuous behaviour of *RaALS* in generating a heptaketide intermediate was also highlighted in relation to the amino acid changes in some of the vital (and conserved) positions of this peculiar enzyme. Present work may pave the way toward understanding the role of CHS-like (and other) Type III PKSs in the biosynthesis of pharmacologically important phytoconstituents; anthraquinones, in particular. Moreover, the recent advancements in synthetic and systems biology have opened up several new ways toward understanding and realizing the potential of PKSs and further utilizing them, through rational engineering of their biosynthetic machinery, for the generation of more diverse and novel lead compounds with desirable pharmacological efficacy.

## Declarations

### Funding

Work in the SAP laboratory is supported by Department of Science and Technology (DST), Govt. of India, under the INSPIRE Faculty Scheme [DST/INSPIRE/04/2016/001059]. Earlier, the work was partly supported by financial grant from Council of Scientific and Industrial Research (CSIR)-Indian Institute of Integrative Medicine under Major Lab Project MLP-3012.

### Author contributions

Conceived and designed the experiments: SAP, SL, VB; Performed the experiments: SAP, SR, SB, ND; Analyzed the data: SAP, VB, ND, MAS, SL; Contributed reagents/materials/analysis tools: SAP, SL; Wrote the paper: SAP, SR, SB, VB, ND.

### Conflict of interest

The authors declare that there is no conflict of interest.

## References

- Abe I, Watanabe T, Lou W, Noguchi H (2006) Active site residues governing substrate selectivity and polyketide chain length in aloesone synthase. *FEBS J* 273 (1):208-218
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*:gkq399
- Baharum H, Morita H, Tomitsuka A, Lee F-C, Ng K-Y, Rahim RA, Abe I, Ho C-L (2011) Molecular cloning, modeling, and site-directed mutagenesis of type III polyketide synthase from *Sargassum binderi* (Phaeophyta). *Marine Biotechnol* 13 (5):845-856
- Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27 (3):343-350
- Bhat WW, Dhar N, Razdan S, Rana S, Mehra R, Nargotra A, Dhar RS, Ashraf N, Vishwakarma R, Lattoo SK (2013) Molecular characterization of UGT94F2 and UGT86C4, two glycosyltransferases from *Picrorhiza kurrooa*: comparative structural insight and evaluation of substrate recognition. *PLoS ONE* 8 (9):e73804
- Bhat WW, Rana S, Dhar N, Razdan S, Pandith SA, Vishwakarma R, Lattoo SK (2014a) An inducible NADPH-cytochrome P450 reductase from *Picrorhiza kurrooa*—an imperative redox partner of cytochrome P450 enzymes. *Funct Integr Genomics* 14 (2):381-399
- Bhat WW, Razdan S, Rana S, Dhar N, Wani TA, Qazi P, Vishwakarma R, Lattoo SK (2014b) A phenylalanine ammonia-lyase ortholog (PkPAL1) from *Picrorhiza kurrooa* Royle ex. Benth: Molecular cloning, promoter analysis and response to biotic and abiotic elicitors. *Gene* 547 (2):245-256
- CTAB DNA extraction buffer (2009). *Cold Spring Harbor Protocols* 2009 (3):pdb.rec11718. doi:10.1101/pdb.rec11718
- Dao T, Linthorst H, Verpoorte R (2011) Chalcone synthase and its functions in plant resistance. *Phytochem Rev* 10 (3):397
- Davies KM, Schwinn KE (2003) Transcriptional regulation of secondary metabolism. *Functional Plant Biol* 30 (9):913-925
- Dhar N, Rana S, Razdan S, Bhat WW, Hussain A, Dhar RS, Vaishnavi S, Hamid A, Vishwakarma R, Lattoo SK (2014) Cloning and functional characterization of three branch point oxidosqualene cyclases from *Withania somnifera* (L.) dunal. *J Biol Chem* 289 (24):17249-17267
- Duraisamy G, Mishra A, Kocábek T, Matoušek J (2018) Activation of polyketide synthase gene promoter in *Cannabis sativa* by heterologous transcription factors derived from *Humulus lupulus*. *Biol Plant* 62 (2):250-260

- Farmer EE, Alm ras E, Krishnamurthy V (2003) Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. *Curr Opin Plant Biol* 6 (4):372-378
- Farooq U, Pandith SA, Saggoo MIS, Lattoo SK (2013) Altitudinal variability in anthraquinone constituents from novel cytotypes of *Rumex nepalensis* Spreng—a high value medicinal herb of North Western Himalayas. *Ind Crop Prod* 50:112-117
- Ferrer J-L, Jez JM, Bowman ME, Dixon RA, Noel JP (1999a) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Biol* 6 (8):775-784
- Ferrer J-L, Jez JM, Bowman ME, Dixon RA, Noel JP (1999b) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Mol Biol* 6 (8):775
- Flores-Sanchez IJ, Verpoorte R (2009) Plant polyketide synthases: a fascinating group of enzymes. *Plant Physiol Biochem* 47 (3):167-174
- Funa N, Ohnishi Y, Fujii I, Shibuya M, Ebizuka Y, Horinouchi S (1999) A new pathway for polyketide synthesis in microorganisms. *Nature* 400 (6747):897-899
- Funabashi M, Funa N, Horinouchi S (2008) Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. *J Biol Chem* 283 (20):13983-13991
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31 (13):3784-3788
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. Springer,
- Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer applications in the biosciences: CABIOS* 11 (6):681-684
- Go MK, Wongsantichon J, Cheung VWN, Chow JY, Robinson RC, Yew WS (2015) Synthetic polyketide enzymology: platform for biosynthesis of antimicrobial polyketides. *ACS Catal* 5 (7):4033-4042
- Gokhale RS, Sankaranarayanan R, Mohanty D (2007) Versatility of polyketide synthases in generating metabolic diversity. *Curr Opin Struct Biol* 17 (6):736-743
- Guo X-D, Ma Y-J, Parry J, Gao J-M, Yu L-L, Wang M (2011) Phenolics content and antioxidant activity of tartary buckwheat from different locations. *Molecules* 16 (12):9850-9867
- Hahlbrock K, Scheel D (1989) Physiology and molecular biology of phenylpropanoid metabolism. *Annu Rev Plant Biol* 40 (1):347-369

- Hashimoto M, Nonaka T, Fujii I (2014) Fungal type III polyketide synthases. *Nat Prod Rep* 31 (10):1306-1317
- Hertweck C (2009) The biosynthetic logic of polyketide diversity. *Angewandte Chemie International Edition* 48 (26):4688-4716
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27 (1):297-300
- Jaakola L, Määttä-Riihinen K, Kärenlampi S, Hohtola A (2004) Activation of flavonoid biosynthesis by solar radiation in bilberry (*Vaccinium myrtillus* L.) leaves. *Planta* 218 (5):721-728
- Jiang W, Yin Q, Wu R, Zheng G, Liu J, Dixon RA, Pang Y (2015) Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *J Exp Bot* 66 (22):7165-7179
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44 (D1):D457-D462
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biol Evol* 35 (6):1547-1549
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J App Cryst* 26 (2):283-291
- Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30 (1):325-327
- Li H, Liang J, Chen H, Ding G, Ma B, He N (2016) Evolutionary and functional analysis of mulberry type III polyketide synthases. *BMC Genomics* 17 (1):540
- Li X, Park NI, Xu H, Woo S-H, Park CH, Park SU (2010) Differential expression of flavonoid biosynthesis genes and accumulation of phenolic compounds in common buckwheat (*Fagopyrum esculentum*). *J Agr Food Chem* 58 (23):12176-12181
- Lim YP, Go MK, Yew WS (2016) Exploiting the biosynthetic potential of type III polyketide synthases. *Molecules* 21 (6):806
- Martinelli L, Redou V, Cochereau B, Delage L, Hymery N, Poirier E, Le Meur CL, Foch L, Le G, Cladiere L (2020) Identification and Characterization of a New Type III Polyketide Synthase from a Marine Yeast, *Naganishia uzbekistanensis*. *Marine Drugs* 18 (12):637
- ME AMSTB, Moore HSKA (2006) Kay RR Noel JP Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat Chem Biol*

- MEGA X (2018) molecular evolutionary genetics analysis across computing platforms; S Kumar, G Stecher, M Li, C Knyaz, K Tamura. *Mol Biol Evol* 35:1547-1549
- Möller S, Croning MD, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17 (7):646-653
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30 (16):2785-2791
- Nair DR, Anand S, Verma P, Mohanty D, Gokhale RS (2012) Genetic, biosynthetic and functional versatility of polyketide synthases. *Curr Sci* 277-287
- NCBI RC (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 41 (Database issue):D8
- Pandith SA, Dar RA, Lattoo SK, Shah MA, Reshi ZA (2018) *Rheum australe*, an endangered high-value medicinal herb of North Western Himalayas: a review of its botany, ethnomedical uses, phytochemistry and pharmacology. *Phytochem Rev* 17 (3):573-609
- Pandith SA, Dhar N, Rana S, Bhat WW, Kushwaha M, Gupta AP, Shah MA, Vishwakarma R, Lattoo SK (2016) Functional promiscuity of two divergent paralogs of type III plant polyketide synthases. *Plant Physiol* 171 (4):2599-2619
- Pandith SA, Hussain A, Bhat WW, Dhar N, Qazi AK, Rana S, Razdan S, Wani TA, Shah MA, Bedi Y (2014) Evaluation of anthraquinones from Himalayan rhubarb (*Rheum emodi* Wall. ex Meissn.) as antiproliferative agents. *S Afr J Bot* 95:1-8
- Pandith SA, Ramazan S, Khan MI, Reshi ZA, Shah MA (2020) Chalcone synthases (CHSs): the symbolic type III polyketide synthases. *Planta* 251 (1):15
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8 (10):785-786
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25 (13):1605-1612
- Rana S, Lattoo SK, Dhar N, Razdan S, Bhat WW, Dhar RS, Vishwakarma R (2013) NADPH-cytochrome P450 reductase: molecular cloning and functional characterization of two paralogs from *Withania somnifera* (L.) Dunal. *PLoS ONE* 8 (2):e57068

- Reimold U, Kröger M, Kreuzaler F, Hahlbrock K (1983) Coding and 3' non-coding nucleotide sequence of chalcone synthase mRNA and assignment of amino acid sequence of the enzyme. *The EMBO J* 2 (10):1801-1805
- Richard S, Lapointe G, Rutledge RG, Séguin A (2000) Induction of chalcone synthase expression in white spruce by wounding and jasmonate. *Plant Cell Physiol* 41 (8):982-987
- Shimizu Y, Ogata H, Goto S (2017) Type III polyketide synthases: functional classification and phylogenomics. *Chembiochem* 18 (1):50-65
- Sievers F, Higgins DG (2014) Clustal omega. *Current protocols in bioinformatics*:3.13. 11-13.13. 16
- Spitaler R, Schlorhauser PD, Ellmerer EP, Merfort I, Bortenschlager S, Stuppner H, Zidorn C (2006) Altitudinal variation of secondary metabolite profiles in flowering heads of *Arnica montana* cv. ARBO. *Phytochem* 67 (4):409-417
- Taguchi G, Yazawa T, Hayashida N, Okazaki M (2001) Molecular cloning and heterologous expression of novel glucosyltransferases from tobacco cultured cells that have broad substrate specificity and are induced by salicylic acid and auxin. *Eur J Biochem* 268 (14):4086-4094
- Tamogami S, Rakwal R, Kodama O (1997) Phytoalexin production elicited by exogenously applied jasmonic acid in rice leaves (*Oryza sativa* L.) is under the control of cytokinins and ascorbic acid. *FEBS letters* 412 (1):61-64
- Tuteja JH, Clough SJ, Chan W-C, Vodkin LO (2004) Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. *Plant Cell* 16 (4):819-835
- Walhout AJ (2006) Unraveling transcription regulatory networks by protein–DNA and protein–protein interaction mapping. *Genome Res* 16 (12):1445-1454
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46 (W1):W296-W303
- Weng J-K, Noel JP (2012) Structure–function analyses of plant type III polyketide synthases. *Method Enzymol* 515:317-335
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35 (suppl\_2):W407-W410
- Yu H-N, Wang L, Sun B, Gao S, Cheng A-X, Lou H-X (2015) Functional characterization of a chalcone synthase from the liverwort *Plagiochasma appendiculatum*. *Plant Cell Rep* 34 (2):233-245

Zeng J, Decker R, Zhan J (2012) Biochemical characterization of a type III polyketide biosynthetic gene cluster from *Streptomyces toxytricini*. *Appl Biochem Biotechnol* 166 (4):1020-1033

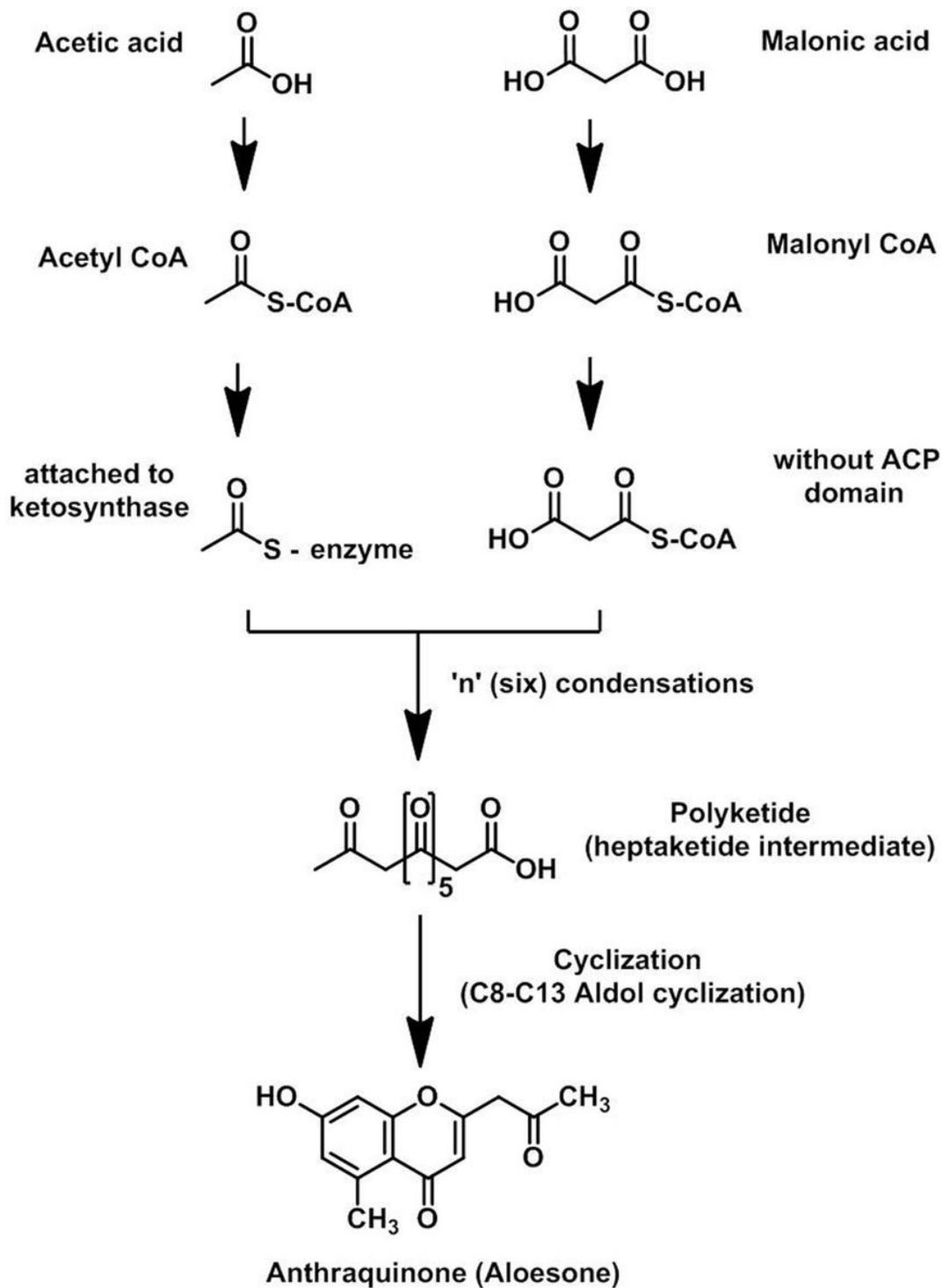
Zhao J, Davis LC, Verpoorte R (2005) Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnol Adv* 23 (4):283-333

## Tables

**Table 1: Putative *cis*-acting regulatory elements identified in the promoter of *RaALS* using PLANTCARE database**

<b><i>Cis</i>-element</b>	<b>Position (strand)</b>	<b>Signal sequence</b>	<b>Putative function</b>
ARE	33 (-)	AAACCA	<i>Cis</i> -acting regulatory element essential for the anaerobic induction
Box I	58 (+)	TTTCAAA	Light responsive element
CAAT-box	153 (+), 220 (+), 261 (-), 262 (-), 380 (-), 381 (-)	CAAT, CAATT	Common <i>cis</i> -acting element in promoter and enhancer regions
CCGTCC-box	125 (+)	CCGTCC	<i>Cis</i> -acting regulatory element related to meristem specific activation
CGTCA-motif	217 (+)	CGTCA	<i>Cis</i> -acting regulatory element involved in the MeJA-responsiveness
GATA-box	2 (+), 186 (-), 190 (-), 320 (-), 340 (-), 394 (-)	GATA	Required for high level, light regulated, and tissue specific expression
TATA-box	254 (+), 256 (+), 313 (-), 312 (+), 314 (+), 255 (-), 258 (-)	TATA, TATAA, ATTATA, TATAAA, TTTTA	Core promoter element around -30 of transcription start
TC-rich repeats	71 (+)	ATTCTCTAAC	<i>Cis</i> -acting element involved in defense and stress responsiveness
TGACG-motif	217 (-)	TGACG	<i>Cis</i> -acting element involved in transcriptional activation of several genes by auxin and/or salicylic acid
MYB-core	25 (+), 77 (-), 95 (-), 99 (+), 358 (-)	CNGTTR	Binding site for all animal MYB and at least two plant MYB proteins; ATMYB1 and ATMYB2, both isolated from Arabidopsis
Skn-1_motif	83 (+), 119 (-)	GTCAT	<i>Cis</i> -acting regulatory element required for endosperm expression
GT1-motif	357 (-)	GGTTAA	Light responsive element

## Figures

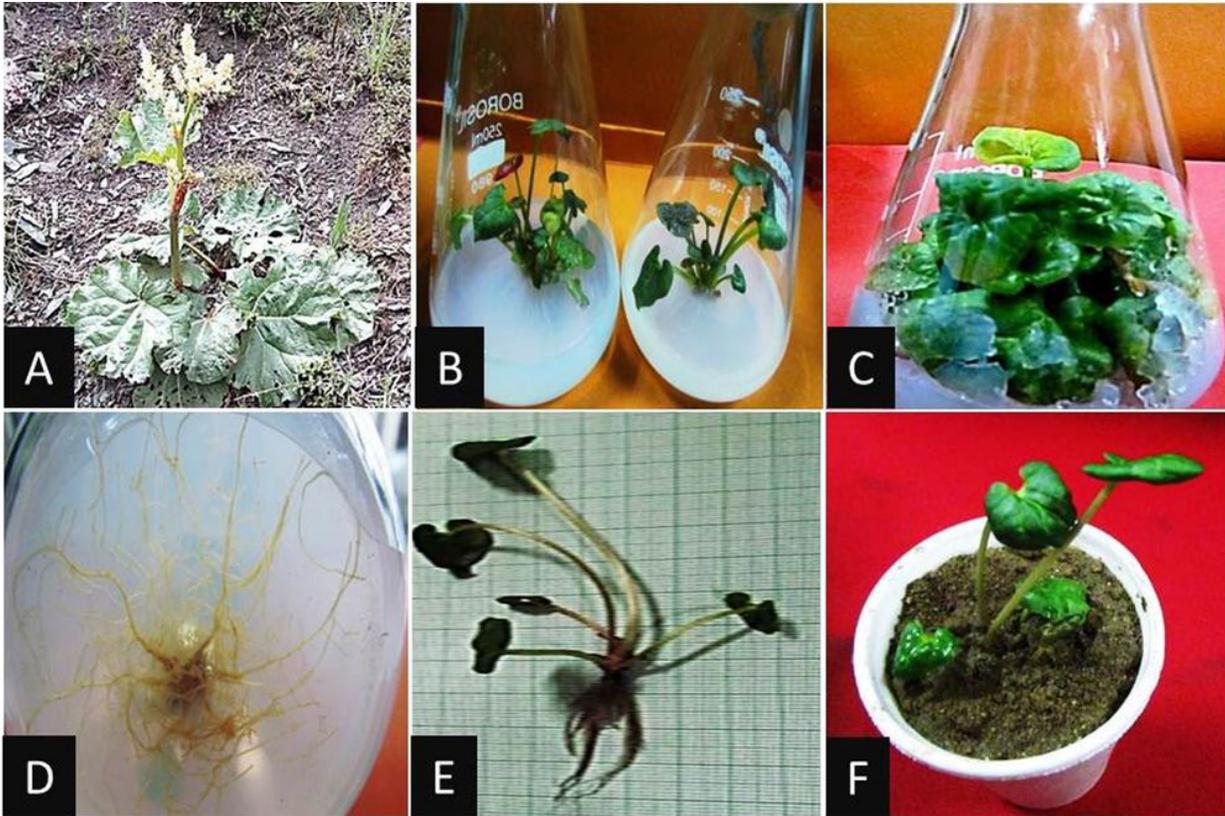


**Fig. 1**

**Figure 1**

The chemistry of polyketide chain assembly; Acetic acid and malonic acid are converted to their coenzyme A esters and then attached, by specific acyl transferases, to components of the polyketide synthase: acetyl-CoA is attached to the active site of the ketosynthase, and malonyl-CoA to a structural component of the PKS called the acyl carrier protein (ACP), usually absent in Type III PKSs. Condensation of the two units by the ketosynthase, with loss of one carbon from malonyl-CoA as carbon dioxide, produces a four-carbon chain. This is transferred back to the ketosynthase, and further rounds (total 6) of

condensation with malonyl-CoA or other chain extender units produce a heptaketide chain which then cyclises to form the aloesone.



**Fig. 2**

**Figure 2**

*Rheum australe*; a) In natural habitat (Pissu top, Pahalgam, Kashmir; 33° 51'08"N, 76° 21'57"E; 4287 m asl), b-c) in vitro raised cultures, d) rooting, e) regenerated plantlet and f) acclimatization.

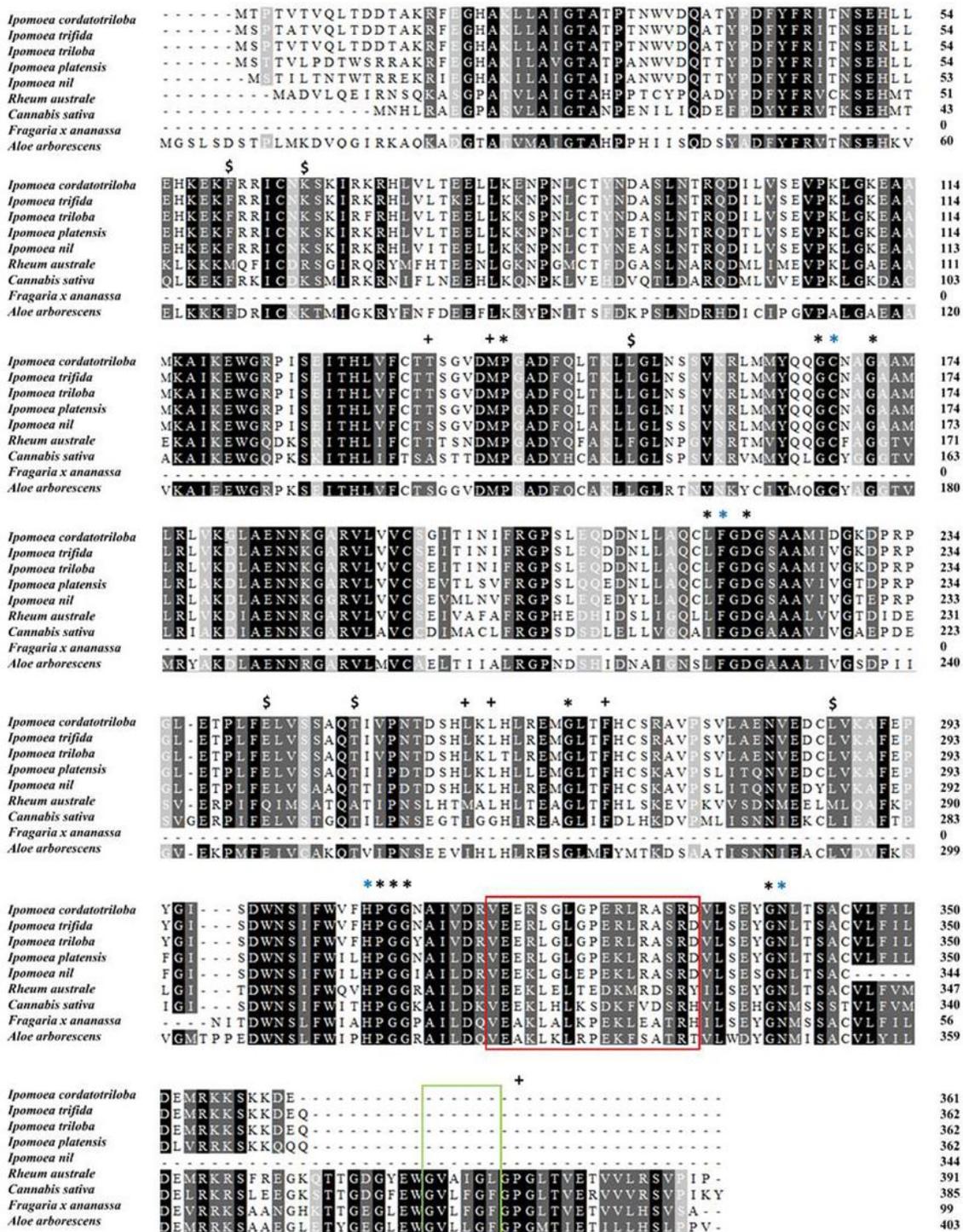
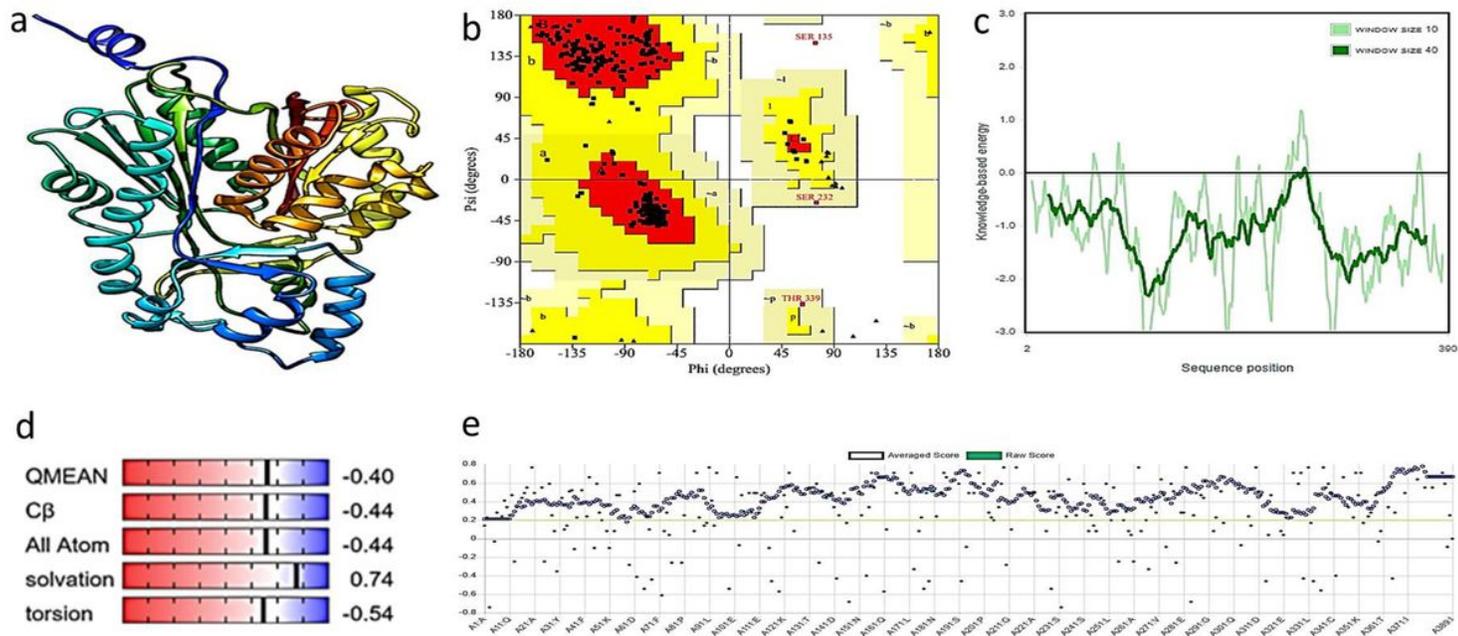


Fig. 3

Figure 3

Multiple sequence alignment; Sequence alignment of RaALS with chalcone synthase proteins from *Medicago sativa*, *Gerbera hybrid* cultivar, *Cannabis sativa*, *Fragaria x ananassa* and *Aloe arborescens*. Aligned residues are highlighted from black to grey to white background according to their identity. The catalytic triad residues are marked with '\*', '+' sign shows active site residues, '+' sign displays amino acid cyclization pocket residues whereas '\$' signifies the variation observed in RaALS with respect to other

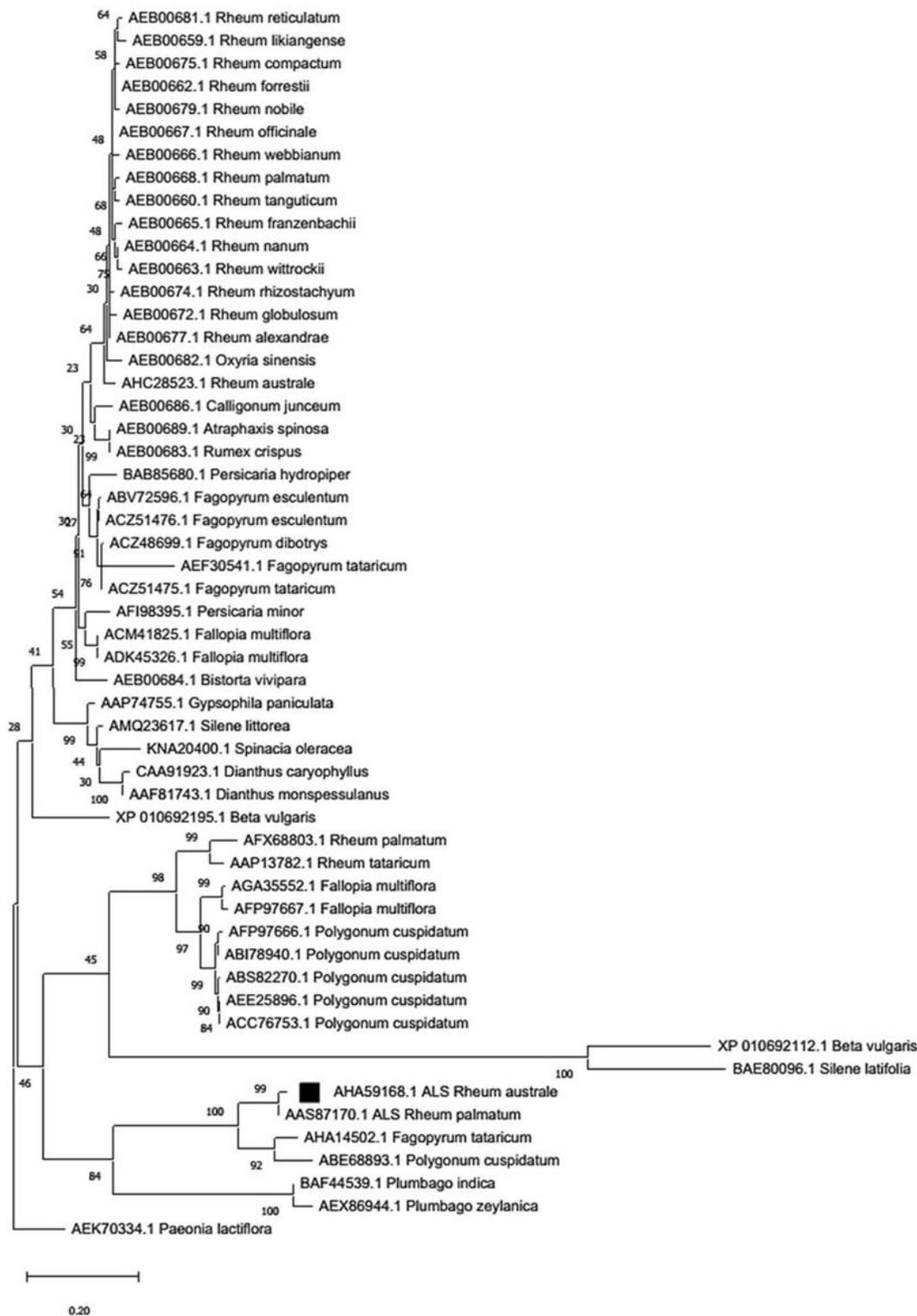
sequences. The red block surrounds the malonyl CoA binding motif and green block shows signature motif in CHS.



**Fig. 4**

**Figure 4**

Homology model and its validation; a) Homology model of RaALS rendered in cartoon coloured in rainbow theme, b) Ramachandran plot of model, c) Knowledge base energy displaying local model quality, d) QMEAN plot predicted by Swiss Model Assessment tool, and e) Verify 3D score.

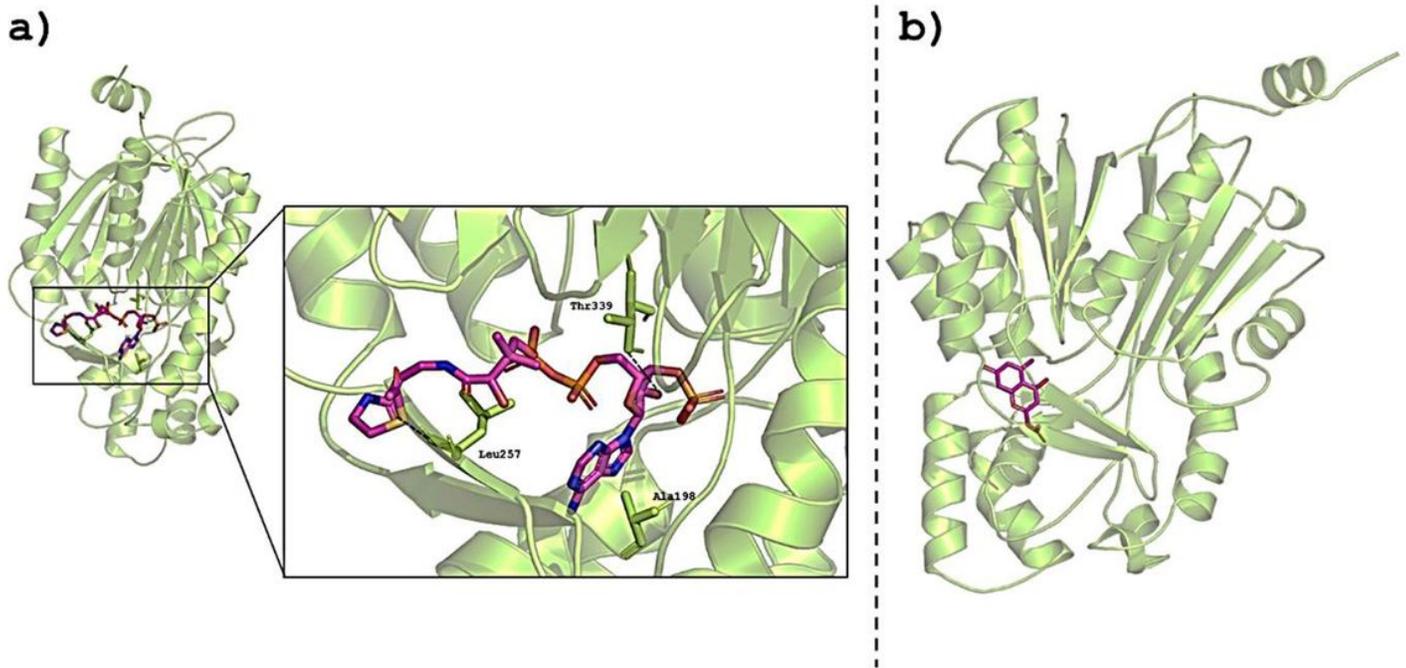


**Fig. 5**

**Figure 5**

Evolutionary relatedness of orthologous Type III plant PKSs: The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The tree with the highest log likelihood (-7803.79) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The numbers on the nodes indicate the bootstrap values after 500 replicates.

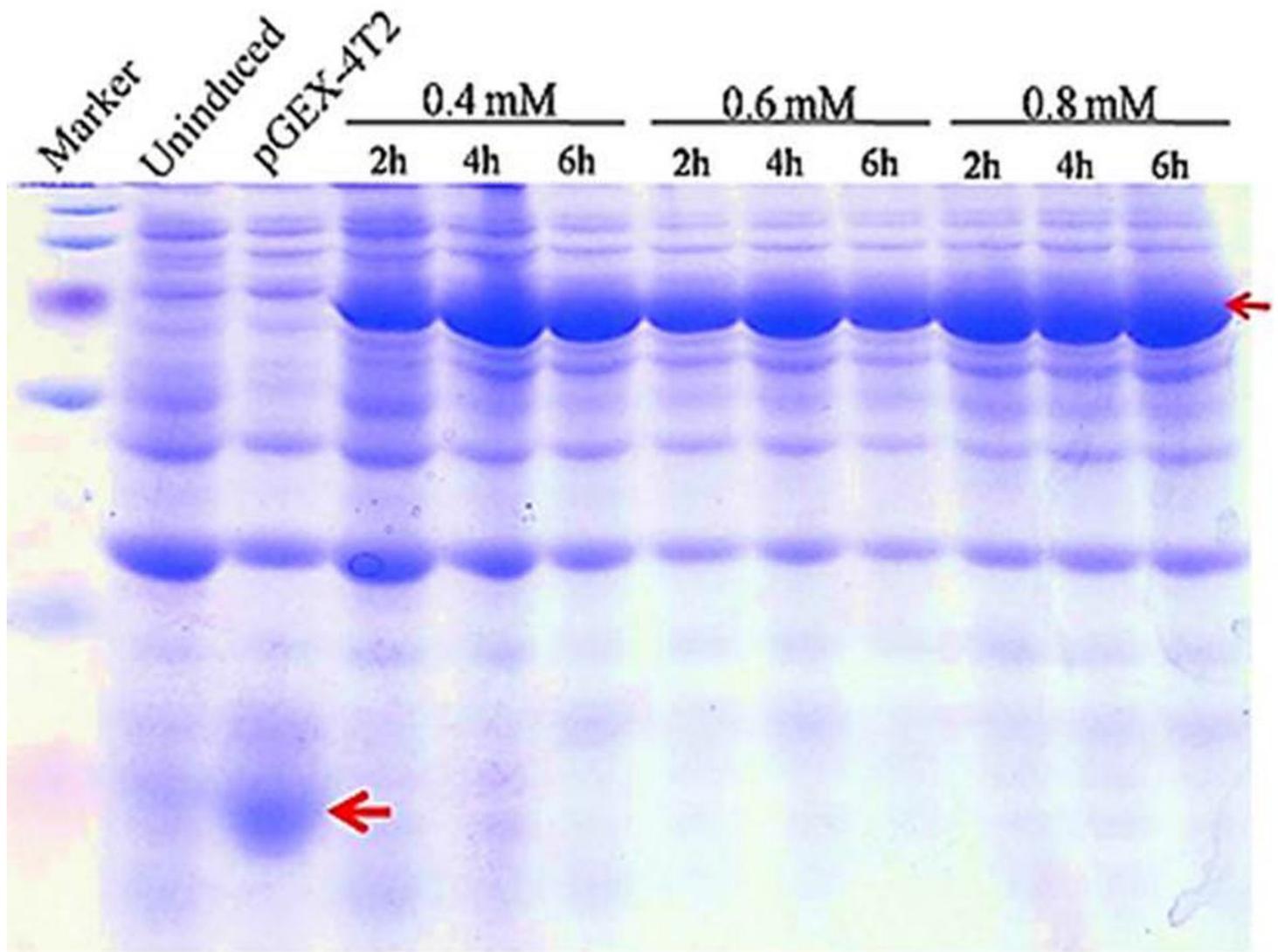
The bar indicates an evolutionary distance of 0.2%. This analysis involved 54 amino acid sequences. There were a total of 406 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.



**Fig. 6**

## Figure 6

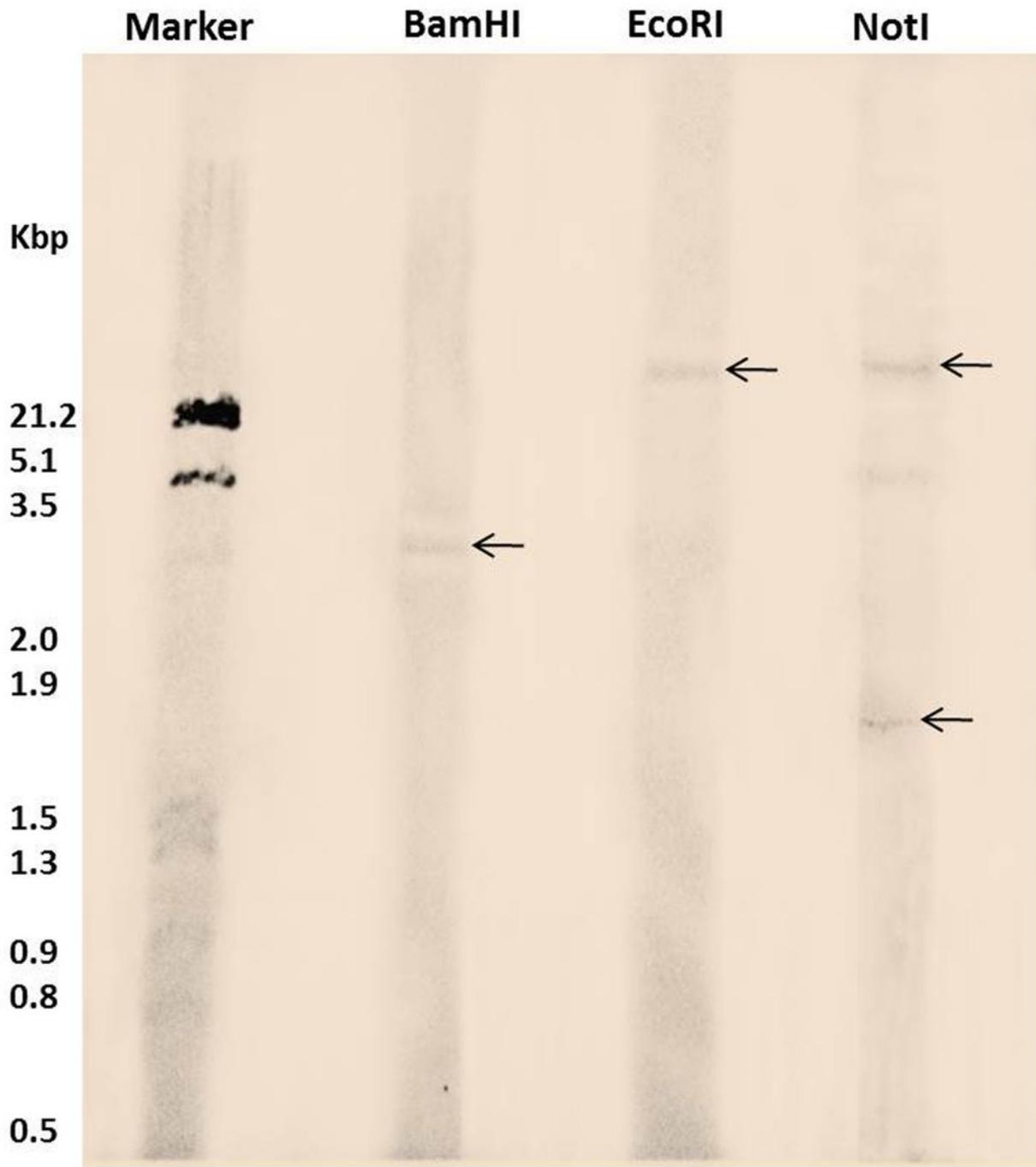
Molecular Docking of RaALS; a) Starter molecule Acetyl-CoA, (b) product aloesone. Protein is rendered in cartoon coloured green while ligand is shown in licorice in magenta colour. H-bond interactions are shown with dotted black line.



**Fig. 7**

**Figure 7**

SDS-PAGE profile of expressed recombinant RaALS protein; SDS-PAGE (10%) of recombinant protein expressed in *E. coli* BL21 (DE3) cells transformed with pGEX-RaALS expression cassettes. Lane 1) standard protein marker, Lane 2) uninduced culture, Lane 3) GST protein, Lane 4-12) Time-course expression of recombinant GST-fused RaALS protein with different concentrations of IPTG (0.4 mM, 0.6 mM and 0.8 mM). The cultures were harvested at different time intervals (2h, 4h, and 6h). Arrows indicate the size of proteins; GST protein (~26 kDa) in L3, and fused RaALS protein (~69 kDa) in L4-12.



**Fig. 8**

**Figure 8**

Southern blot analysis of RaALS: The genomic DNA (>20  $\mu$ g) isolated from *Rheum australe* was digested with BamHI, EcoRI (both non-cutters) and NotI (single-cutter). The digested samples were separated on 0.8% agarose gel, blotted onto a nylon membrane and hybridized with DIG-labeled ORFs of RaALS as probe.

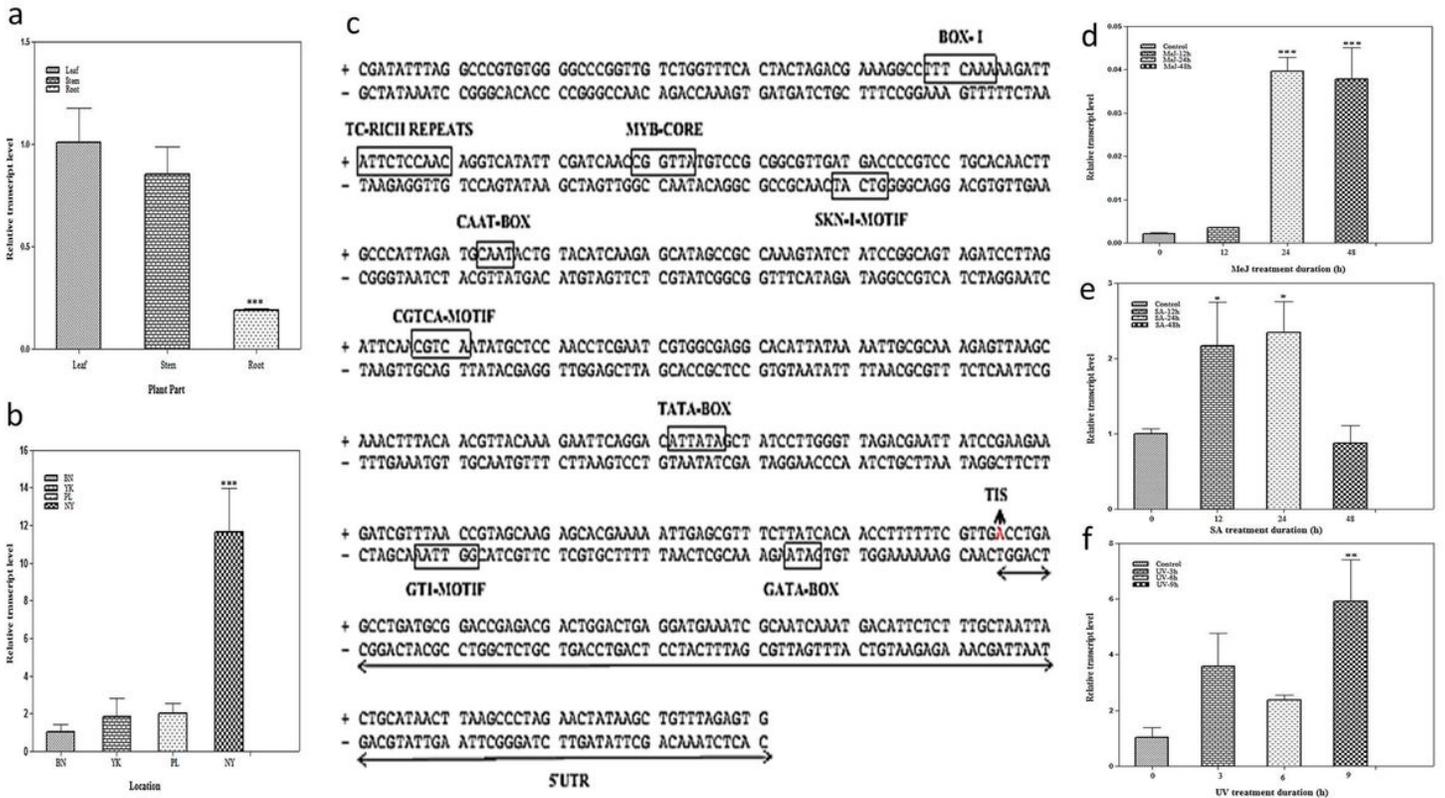


Fig. 9

Figure 9

Real-Time expression analyses of RaALS from different tissues and altitudes, its promoter sequence and transcript abundance vis-à-vis elicitations with MeJ, SA and UV-B radiation: a) Quantitative estimation of the relative expression of RaALS in leaf, stem and root tissues of *R. australe*, b) differential expression pattern of RaALS in plant samples collected from four different geographic locations of NW Himalayas (1600-4500 m asl), c) nucleotide sequences of RaALS gene promoter. The putative core promoter consensus sequences and the motifs with significant similarity to the previously identified cis-acting elements are highlighted and the names are given. TIS represents the transcription initiation site and 5' untranslated region (UTR) is depicted with an arrow, d-f) Time course expression characteristics of RaALS gene in micro-propagated *R. australe* in response to the elicitations by 0.1 mM MeJ (d), 0.1 mM SA (e), and 1500  $\mu\text{j}/\text{m}^2$  UV-B radiation (f). The in vitro raised cultures were precultured in MS liquid medium for about 2 w, elicited and further harvested after different time intervals and the concentration of MeJ and SA for plant treatments was kept as 0.1 mM.

## Supplementary Files

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

- [SupplementaryfileALS.pdf](#)