AMOGHVARTA

ISSN : 2583-3189

Transcriptome Sequencing and Analysis for Identification of Putative candidate gene encoding Serine Acetyl Transferase gene (*LsSAT)* related Biosynthesis of neurotoxin, β-ODAP in *Lathytus sativus*

ORIGINAL ARTICLE

Authors
Jajati Keshari Nayak,
Miranda Kongbrailatpam
Dept of Plant Molecular Biology and Biotechnology, Indira Gandhi Krishi Vishwavidyalaya, College of Agriculture, CoA, Raipur, Chhattisgarh, INDIA

Kalp Das
Research Scholar
Dept. of Vegetable Science
Punjab Agricultural University
Ludhiana, Punjab, INDIA

Taruna Borule, Supriya Nayak
Shubha Banerjee
Dept of Plant Molecular Biology and Biotechnology, Indira Gandhi Krishi Vishwavidyalaya, College of Agriculture, CoA, Raipur, Chhattisgarh, INDIA

Abstract

Grass pea is a nutritious crop with high seed protein content of good quality yet presence of neurotoxin β-ODAP (beta-oxalyl di-amino propionic acid) has limited its human consumption. Information of gene sequence is prerequisite for application of advanced biotechnological tools like gene editing therefore transcriptome sequencing of Lathyrus sativus genotypes was done to identify the key genes involved in β-ODAP biosynthesis. Alignment of the RNA sequence data from 10 days old leaf tissue with bacterial serine acetyl transferase (SAT), followed by homology modelling, protein domain characterization, and phylogenetic analysis led to the identification of transcript (ID-15184) as putative LsSAT encoding transcript. Pathway mapping and study of conserved domain confirmed its role in cysteine and methionine metabolism as it carried domain like PLN02739, SATase_N, DUF3422, CysE, LbH_SAT, which are involved in transfer of acetyl group to serine, leading to formation of O-acetyl serine (precursor of β-ODAP). Molecular docking also showed interaction with L-serine as ligand at amino acid residues at active site. Semiquantitative RT-PCR was done in 10 days old leaves was done with primers designed from ID-15184. Sequencing of the resultant amplicon (ncbi-Accession no-MW590957) was found to be significantly similar to known SAT genes. However, expression of putative candidate LsSAT gene in both high and low ODAP containing genotypes indicated it does not determine the rate of ODAP biosynthesis, but is involved in synthesis of acetyl serine which is utilized by other enzymes for synthesis of neurotoxin as well as cysteine.

Key Words

β-ODAP, Lathyrus sativus, Neuro-lathyrism, Serine acetyltransferase, Transcriptome.

Introduction

The grass pea (*Lathyrus sativus*) is a major pulse crop produced in Central India, Africa, Ethiopia, and South Asian nations[1,2]. It possesses tolerance to water scarcity [3], cold and common insect pests[4]. It is usually cultivated as a fellow crop after rice on residual soil moisture [5,6,7]. Apart from favorable agronomic traits, grass pea is known to have high biological value seed protein ranging from 24% to 34.6% [8,9,10]. Despite these agronomic and nutritional advantages, grass pea is not cultivated widely. It is frequently overlooked in conventional agriculture because it causes a neurological condition caused by an antinutritional factor, β -N-oxalyl-L- α, β -diamino-propionic acid (β -ODAP)[11,12,13,14]. Excess consumption of the *Lathyrus sativus* seeds for at least 2–3 months causes neurolathyrism[15,16].

β- ODAP is postulated to be synthesized in a two-step reaction addition to the O-acetyl serine metabolic pathway in Chloroplast and mitochondria[17,18,6]. It is made from a heterocyclic -isoxazoline-L-alanine (BIA) by the enzyme -Cyanoalanine synthase (CAS)[18]. B.I.A. is the primary precursor of β-ODAP, made from O-acetyl-L-serine (OAS) and the isoxazoline-5-one ring, potentially by the cysteine synthase (CSase) enzyme. This OAS is synthesized from Serine by the Serine acetyltransferase(SAT) enzyme[19]. So SAT is the main pioneer enzyme of the β- ODAP biosynthesis pathway[20]. This serine O-acetyltransferase (SAT) makes trimer with cysteine synthase and forms a reversible cysteine regulatory complex (CRC) that is critical in regulating sulfur homeostasis[21].

By conventional breeding, somaclonal variation, mutation breeding, and microbiological degradation, large efforts have been made to reduce the seed β-ODAP content in the last two decades. Still, little has been achieved[22]. Silencing the key enzymes coding genes of the ODAP biosynthesis pathway like CoA synthetase, Serine acetyl Transferase, and ODAP synthase[23,24, 20,25] by gene-editing techniques can lead to the development of zero β-ODAP containing genotypes or prevent its synthesis and storage in seeds at least. So identification and characterization of the genes encoding key pathway enzymes are critical to using the gene-editing approach.

To characterize the key enzymes of β-ODAP biosynthesis in *L sativus* and obtain the sequence of putative gene serine acetyltransferase, we made an effort in this study to analyze the leaf tissue transcriptome of two *L sativus* genotypes viz. RLK1950 – high β-ODAP content landrace and Mahateora – low β-ODAP contains landrace[26]. The transcriptome sequence was screened to identify the coding sequence of serine acetyltransferase (Accession no-MW590957). In silico characterization, docking between ligand L-serine and SAT gave a positive output about the identification of SAT. A detailed study on subcellular localization and pathway mapping correlates its link to sulfur metabolism, a key criterion for serine acetyltransferase. Further constitutive expression of genes and transcript confirmation after sequencing have opened a new window for fellow researchers to produce toxin-free grass pea soon.

Materials and Methods

In silico identification of Putative transcripts

Local BLAST was performed using BLAST2GO (https://www.blast2go.com/) with a cut-off e-value of 10 to identify sequences encoding or corresponding to serine-acetyl transferase encoded genes. The nucleotide of a sequence of gene encoding SAT (Serine acetyltransferase) from Staphylococcus *aureus (strain MRSA252)* was retrieved from the *Uniprot id*- Q6GJE0 (CYSE_STAAR). It was used as a query sequence for alignment search with the transcriptome sequence file of RLK-1950 (202 R).

Sequence-based homology characterization of putative transcripts

The sequence homology search using NCBI-BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to compare probable candidate sequences from the Lathyrus sativus transcriptome to other known

serine acetyltransferases.

Sub-cellular localization of putative were analysed using TargetP program (http://www.cbs.dtu.dk/services/TargetP/predictions/pred.html/) and WoLF PSORT (http: //www. genscript.com/tools/wolf-psort/)[27]. Target P[28] retrieved the data from SWISS-PROT[29]. Conserve domain search of putative transcripts was done with CDD data version 3.19. (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Redundancy reduction was performed by maintaining the specificity cut-off at >0.95% by using Hobohm algorithm 2[30], and we eliminated undeserving sequences. Using MEGA X software's pairwise alignment[31,32], we used the Smith-Waterman algorithm and the PAM 250 scoring matrix (https://www.megasoftware.net/) used for evolutionary study through the maximum composite likelihood method[33]. To determine the metabolic pathways involved in candidate transcripts, we performed KEGG pathways mapping (http://genome.jp/kegg/mapper/).

To determine the protein structure of putative transcripts, the nucleotide sequences were converted into their respective peptide sequences using EMBOSS Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/). These peptide sequences were searched for the best protein structure in the PDB data bank through the swiss-model (swissmodel.expasy.org). After that, the resulted protein structures were analyzed in PyMol software. Stereochemical feasibility of the protein model as checked by Ramachandran plot through PROCHECK online server (https://saves.mbi.ucla.edu)[34].

Docking

The docking was performed through the pathdock server, and then the result was analyzed in PYMOL. We took a template of candidate gene SAT as protein and L- serine as a ligand in this docking process. We have downloaded the ligand o-acetyl Serine XML format from the ligand book (https://ligandbook.org).Then we have converted it into PDB format through open bible 2.4 software. Template taken from protein homology modeling and downloaded from PDB database. (https://www.rcsb.org/structure).

Primer Designing

Putative candidate transcripts were screened and selected based on the parameters mentioned above. The most suitable candidate sequence belonging to the transferase family was used to design primers for RT-PCR-based validation. All the sequences were subjected to Primer designing through primer quest IDT (Integrated DNA Technology https://eu.idtdna.com/PrimerQuest/Home/Index) (Supplementary File Table 2).

Expression analysis of selected transcripts

(a) RNA extraction and cDNA synthesis

Leaf samples of 11 *Lathyrus* genotypes were collected from 10 days old plants because ODAP Synthesis is higher in 10 days old plants. Healthy and tender leaves were used for RNA extraction using TRIZOL Reagent. RNA quantity was checked using Nanodrop (Thermo Fisher Scientific, USA). Quality was measured through Formaldehyde treated agarose gel, followed by cDNA synthesis using Bio-rad script cDNA Synthesis kit (as per manufacturer's instructions).

(b) Semi-quantitative RT- PCR based gene expression analysis

Semi-quantitative reverse transcriptase PCR was carried out to study the expression of ODAP related genes. The cDNA generated from the total RNA isolated from leaf tissues of 11 different *Lathyrus* genotypes having differential ODAP content were subjected to semi-quantitative expression profiling in a 10µl reaction using candidate gene-specific primers designed from transcriptome sequence. The resultant PCR product was resolved on 1.5 % Agarose gel at 70V. The presence of amplicons and their respective intensity were recorded under a gel documentation system (BioRad make). The expression was analyzed by comparing the relative fluorescent intensities of cDNA amplicons under a gel documentation system.

Semi Q RT-PCR was executed in 10µl reaction solution & *LsActin* primer was used as an internal control for normalization of cDNA conc. Master mix of PCR buffer contain 10mM polymerase buffer, 1mM dNTPs, 0.1 units *Taq polymerase* (Invitrogen) & 1µM primer. PCR thermal profile was 94^{0}C for 5 mins, followed by 35 cycles of 94^{0}C for 45 sec, 55^{0}C for the 30s, 72^{0}C for 1 min, and a final extension of 72^{0}C for

2 mins. As a result of PCR, the amplified product was resolved on 1% TAE agarose gel with EtBr 0.0003 %. Quantity one 4.5.1 chemidoc EQTM software from Bio-Rad was used for imaging.

The amplified fragments of the desired size (~ 500bp) were eluted from gel using a Thermo fisher Gel Extraction kit (as per the manufacturer's instruction). The eluted DNA sample was sent for sequencing to Thermo Fisher Scientific, Gurgaon, for bidirectional sequencing on the Ion Torrent PGM sequencer.

Results

An analysis of local blast searches using bacterial SAT (Serine acetyltransferase) from *Staphylococcus aureus* (strain MRSA252) as query sequence revealed transcripts encoding SAT in Lathyrus. It led to identification of 10 most similar sequences with contig ids as ID-15184, ID-68475, ID-19615, ID-79795, ID-1498, ID-45, ID-44, ID-52098, ID-42842, ID-19615 and ID-8529 (Supplementary File Table 1) . Out of these ten transcripts, the contig, ID-15184, showed a low e value 3.22046e-7 with a high hit length of 1490, aligning the length of 79 and a high similarity of 79.7%. Although, few other transcripts like ID-79795 were similar (similarity 92.6%) but showed a less significant e value of 0.116651 compared to ID-15184. Therefore, considering all the parameters like sequence identity, similarity, and e value, four sequences, i.e., ID-15184, ID-68475, ID-19615, ID-79795, were selected for further analysis

Homology based characterization of Putative gene LsSAT

The selected transcript sequences corresponding to contig ID-15184, ID-79795, ID-68475, and ID-19615 were aligned using BLASTn (NCBI). For alignment, a standard non-reductant database optimized for the mega blast was used as it is most suitable for a closely related sequence with a sequence having more than 95% similarity. BLAST analysis yielded 34 homologous sequences matching with query transcript sequence. Among them, serine acetyltransferase (SAT) of *Medicago truncatula* (Accession no-XM003612271) showed the highest similarity with the putative candidate transcript ID-15184 (score of 1096 and query cover of 93%), followed by SAT mRNA fragment from *Cicer arietinum* (Accession No-XM004512190).

To understand the genetic relatedness and evolutionary development of SAT encoding candidate transcript in L. sativus, phylogenetic analysis was carried out. The 28 homologous nucleotide sequences obtained with BLASTn of ID-15478 transcript were compared by the Neighbour-Joining method[35] with 1000 replicate bootstrap analysis [36]. Codon positions incorporated as 1st +2nd+3rd+ Non coding. After deleting all ambiguous positions from each sequence, 1002 positions were present in the dataset.

The greatest composite likelihood technique was utilized to calculate phylogenetic distance (Figure 1b) and phylogenetic tree (Figure 1b) using MEGA X[37] software, where pairwise distance and related substitutions parameters are correctly determined by maximizing the composite likelihood. Serine Acetyl Transferase referred to as SAT, is reported to be involved in L-cysteine biosynthesis from L-serine. It is found in almost all crop plants in Chloroplast and also mitochondrial genome[38]. The isoform-specific induction of SERAT or SAT has been reported on exposure of the plant to abiotic stresses like cold or soil stresses. The putative SAT candidate transcript (ID-15184) comparison showed the highest similarity to variant forms of the SAT gene in *Cicer ariitinum* and *Medicago truncatula* (Figure 1a) with lesser similarity to *Lupin*. The *L. sativus* SAT transcript (ID-15184) made a different subgroup from the *Arachis* genus. At the same time, the SAT mRNA of *P. Vulgaris*, *V. radiata*, and *V. unguiculate* consisted of an altogether different subgroup which is relatively more similar to that of ornamental plants *Rosa chinensis* and *Ziziphus jujuba*.

Table 1: Conserve domain comparison between our candidate transcript Id-15184 and SAT gene of OS= *Staphylococcus aureus* (strain MRSA252) Q6GJE0

Transcript	ID-15184					Serine acetyltransferase OS=Staphylococcus aureus (strain MRSA252) Q6GJE0				
S. No.	Domain name	Accession	Description	Interval	E-value	Domain name	Accession	Description	Interval	E-value
1	PLN02379	PLN02379	Serine acetyltransferase	1-933	2.28e-133					
2	CysE	COG1045	SAT amino acid transport and metabolism	361-858	1.26e-65	CysE	COG1045	SAT amino acid transport and metabolism	1-195	1.62e-97
3	CysE	TIGR01172	Serine-o-acetyltransferase; Cysteine biosynthesis & AA biosynthesis	364-849	8.34e-61	CysE	TIGR01172	Serine-o-acetyltransferase; Cysteine biosynthesis & AA biosynthesis	7-167	1.07e-93
4	LbH_SAT	Cd03354	S.A.T. and SAT catalyze the Co-A dependent acetylation	547-843	1.17e-36	LbH_SAT	Cd03354	S.A.T. and SAT catalyze the Co-A dependent acetylation	66-166	1.45e-54
5	SATase_N	Pfam06426	S.A.T.; N terminal domain hydroxyl group of L-serine to form O-acetylserine	205-459	7.37e-22	CysE	PRK11132	SAT Provisional	10-173	3.84e-54
6	SATase_N	Smart00971	N-terminal domain of serine acetyltransferase, which is conserved in plants and bacteria.	205-459	9.87e-17	Hexapep	Pfam00132	Bacterial transferase hexapeptide	119-148	2.71e-03
7	DUF3432	Pfam11914	This presumed domain is functionally uncharacterized, which is found in eukaryotes. It is associated with pfam0096 & has two conserved motifs: YPSPV AND PSP	600-845	6.44e-04					

1(a)

1(b)

Gene.15478::202R_contig_15184	0																			
XM_003612271Medicago_truncatula_serine_acetyltransferase_2_(LOC11409367)	0.135																			
XM_013598127.Medicago_truncatula_serine_acetyltransferase_2_(LOC11409367)	0.126	0.000																		
XM_004512190 Cicerarietinum_serine_acetyltransferase_2_(LOC101491313)	0.142	0.099	0.100																	
XM_016345907.Arachis_ipaensis_serine_acetyltransferase_2_(LOC107642518)	0.187	0.158	0.156	0.157																
XM_016345906.Arachis_ipaensis_serine_acetyltransferase_2_(LOC107642518)	0.187	0.158	0.156	0.157	0.000															
XM_025842463.Arachis_hypogaea_serine_acetyltransferase_2(LOC112800272)	0.190	0.159	0.158	0.157	0.005	0.005														
XM_029298196.Arachis_hypogaea_serine_acetyltransferase(LOC112800272)	0.190	0.159	0.158	0.157	0.005	0.005	0.000													
XM_016108055.Arachis_duranensis_serine_acetyltransferase(LOC107487415)	0.181	0.155	0.158	0.156	0.005	0.005	0.000	0.000												
XM_016108054.Arachis_duranensis_serine_acetyltransferase_2_(LOC107487415)	0.181	0.155	0.158	0.156	0.005	0.005	0.000	0.000	0.000											
XM_007157922.(PHAVU_002G114700g)	0.183	0.174	0.173	0.173	0.177	0.177	0.175	0.175	0.175	0.175										
XM_007157921.Phaseolus_vulgaris(PHAVU_002G114700g)	0.183	0.174	0.173	0.173	0.177	0.177	0.175	0.175	0.175	0.175	0.000									
XM_017570563.Vigna_angularis_serine_acetyltransferase_2-(LOC108334667)	0.205	0.181	0.180	0.179	0.184	0.184	0.182	0.182	0.185	0.185	0.029	0.029								
XM_028056859.Vigna_unguiculata_serine_acetyltransferase_2-(LOC114172263)	0.207	0.188	0.188	0.181	0.181	0.181	0.180	0.180	0.182	0.182	0.030	0.030	0.022							
XM_028056854.Vigna_unguiculata_serine_acetyltransferase_2(LOC114172263)	0.207	0.188	0.188	0.181	0.181	0.181	0.180	0.180	0.182	0.182	0.030	0.030	0.022	0.000						
XM_014663911.Vigna_radiata_var._radiata_serine_acetyltransferase_2(LOC106776453)	0.199	0.180	0.180	0.177	0.184	0.184	0.182	0.182	0.185	0.185	0.034	0.034	0.012	0.028	0.028					
XM_022775644.1Vigna_radiata_var._radiata_serine_acetyltransferase_2_(LOC106776453)	0.199	0.180	0.180	0.177	0.184	0.184	0.182	0.182	0.185	0.185	0.034	0.034	0.012	0.028	0.028	0.000				
XM_019589334._Lupinus_angustifolius_serine_acetyltransferase_2-like_(LOC109348784)	0.224	0.184	0.185	0.185	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.169	0.172	0.172	0.169	0.169			
XM_025076710.Zziziphus_jujuba_serine_acetyltransferase_2_(LOC107424304)	0.279	0.264	0.254	0.250	0.226	0.226	0.231	0.231	0.226	0.226	0.226	0.226	0.243	0.235	0.235	0.246	0.246	0.232		

Figure. 1: (a) Phylogenetic tree of an aligned homologous sequence of candidate contig id-15184 from MEGAX (b) Phylogenetic distance of ID-15184 with other homologous sequences obtained from MEGA X

A conserved domain search of putative transcript ID-15184 of *Lathyrus sativus* with SAT of *Staphylococcus aureus* (strain MRSA252, Q6GJE0) by NCBI protein BLAST showed that it belongs to the PLN superfamily having a conserved domain like of PLN02379, COG1045, TIGR01172, Cd03354, Pfam06426, Smart00971, Pfam11914(Table 1). There are three domains specific to CysE in the protein domain that corresponds to ID-15184: COG1045 for Serine acetyltransferase [Amino acid transport and metabolism], TIGR01172 for Cysteine biosynthesis [Amino acid biosynthesis, Serine family]., and LbH_SAT with Cd03354 domain (Table 1). All three domains are related to the enzyme Serine acetyltransferase (SAT), which catalyzes co-enzyme A (CoA) dependent acetylation of L-serine in its hydroxyl (-OH) group side. This is the 1st step to forming L-cysteine in bacteria & plants. The O-acetylserine is also a precursor of β-ODAP biosynthesis wherein the presence of cyanolalanine synthase (CAS) enzyme reacts to isoxazoline-5-one leading to the formation of a short-lived intermediate, which is finally converted to β-ODAP by an uncharacterized enzyme. Thus, the putative SAT transcript identified in L. sativus can help study and identify the branching point in the β-ODAP biosynthesis pathway.

Sub Cellular Localisation of Putative SAT transcript

SAT and its isoforms are known to be located in Chloroplast and mitochondria. Thus, the putative SAT encoding transcript analysis was conducted to understand its cellular or subcellular localization. Various online servers are available to analyze the possible sub-cellular location of proteins or enzymes[39]. Possible cellular localization of *LsSAT* was predicted through Target P & WoLF-PSORT online server. Scanning the LsSAT suggested that it possibly be located in Chloroplast & mitochondria (Table 2). As per the Target P result section results, cTP-a chloroplast transit peptide; SP-signal peptide; mTP- mitochondrial targeting peptide were found in the transcript ID-15184; with reliability class value (RC) as 0.012 for Chloroplast and 0.003 for signal peptide. A lower RC value, i.e., less than 0.2, corresponds to a more probable and significant localization prediction [28].

Wolf P Sort's alternate method of analysis of subcellular location resulted in 14 hits for transcript ID-15184. Out of these 14 hits, five hits corresponded to the Chloroplast, five corresponded to mitochondria, one hit to the vacuole, and one hit to mitochondria chloroplast, both mitochondria (supplementary file Figure

1). Thus, the average value in Chloroplast and mitochondria is 6, suggesting localization of the encoded SAT protein in these subcellular compartments in *L. sativus* (Table 2).

Table 2: Subcellular localization of putative contig ID-15184 based on Target P and Wolf P Sort analysis

Contig name	Cellular localization	
	Target P	WoLF PSORT
ID-15184	Sequence 331 cTP 0.012; mTP 0.618; SP 0.030; other 0.631; Loc RC-2(cTP- Chloroplast transit peptide; SP-Signal Peptide; mTP-Mitochondrial targeting peptide; Loc RC: Localization reliability class)	prediction chlo: 5.5, mito: 5.5 (chlo: chloroplast; cyto: Cytoplasm)

Table 3: β-ODAP content in 11 diverse genotypes. *Lathyrus sativus*[40]

S. No.	Genotype	B-ODAP content (%)
1	No-2208	0.59
2	SEL-519	1.49
3	IL-143099	1.41
4	IL-527	1.51
5	AKL-19	1.54
6	MAHATEORA	0.32
7	PUSA-24	0.84
8	BIO-L-202	0.50
9	RLK-498	1.00
10	SEL-522	1.42
11	RLK-1950	1.90

Semi-Quantitative RT PCR based validation of candidate transcript ID-15184

Out of all the ten putative SAT-related transcripts, four candidate transcripts(i.e., ID-15184, ID-68475, ID-19615, ID-79795) were selected based on high similarity and low e value (local blast data) for validation in the high and low β-ODAP containing *L. sativus* genotypes. By IDT-primer quest, primers were designed for semi-quantitative PCR analysis (PaaY1, PaaY1a, UNK1, UNK2 & NTA) (Supplementary Table 2). Since β-ODAP biosynthesis occurs maximum in 10 days old seedlings and then at the pod formation stage, ten days old seedlings were used for RNA isolation.

The amplification obtained in the primer UNK1, UNK2 & NTA irregular banding pattern gives no correlation with the seed ODAP content (Figure 2). The Primer PaaY designed from ID-15184 amplified at almost a similar level in all the 11 genotypes, suggesting a similar expression level with no specific correlation to seed ODAP content (Figure 2b).

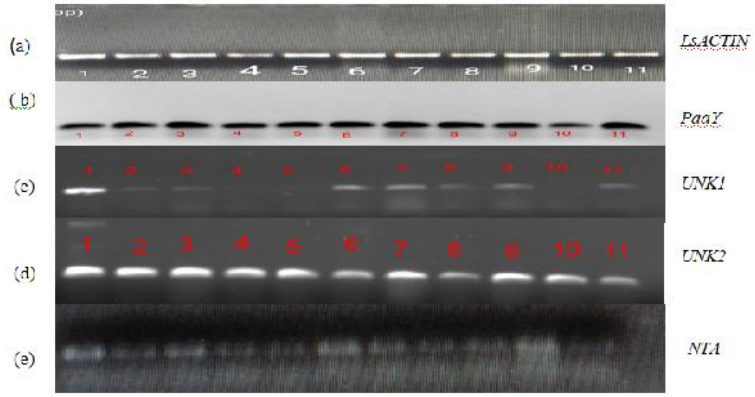

Figure 2: Semi-quantitative RT- PCR profile of 10 days old plant of *Lathyrus sativus* in 11 diverse genotypes. (a) Expression of *LsACTIN* for normalization of cDNA obtained from the leaf of 10 days old seedlings (b) Expression of PaaY primer constructed from ID-15184, all the diverse genotypes showing equal expression as of *LsACTIN*

(c) Expression of UNK1 designed from ID-68475. Only seedling no 1 shows intense expression & the rest others show a faint banding pattern which does not correlate with ODAP content
(d) Expression of UNK2 designed ID-19615 All seedlings except 6, 8, 10 & 11 show intense expression, which means no relation to ODAP content
(e) Expression of NTA primer designed from ID-79795. All seedlings show a faint expression

Sequencing of amplicons obtained with PaaY1 & PaaY1b Primer

The putative SAT gene amplified in selected *L. sativus* genotypes, with primer PaaY1 (Amplicon size-112), was sequenced to confirm the transcript sequence of ID-15184. BLASTn result of the sequenced amplicon of PaaY showed a resemblance with Serine acetyltransferase. PaaY1 has less amplicon size so it may give a false-positive result; therefore, PaaY1b was constructed with more amplicon size around 516 bp from the same contig ID 15184. The amplified fragments were eluted from the agarose gel and sequenced. The DNA sequence data obtained chromatogram in the ABI-SEQ file format was uploaded to the FINCHTV software for quality analysis and trimming. The sequences showing Q values less than 20 (Q<20) were trimmed from both sides. The sequence (Pa7) of the amplicons of Primer PaaY1b corresponding to candidate transcript ID-15184 was then converted into FASTA format for further analysis.

Homology Modelling

A Homology modeling of the candidate transcript ID-15184 was performed through the SWISS-MODEL workspace (https://swissmodel.expasy.org/interactive) and a mod base online data server (https://modbase.compbio.ucsf.edu/modweb/). It is matched with the similar template in the SWISS-MODEL template library (SMTL[41]), a vast structural database derived from the Protein Data Bank containing experimentally established protein structures PDB database. It uses BLAST and HHblits programs[42]. A total of 41 templates were found corresponding to transcript ID-15184. It showed significant matches to the template id such as 1t3d, 6jvu, 3gvd, 6wye, 4h7o, 1ssq, 1ssm (Figure 3) that supported its identity with serine acetyltransferase. Significant matching with the PDB templates 4hzc and 4hzd also indicates the identity of the transcript similar to that of CysE, involved in Cystine biosynthesis of *LsSAT.*

Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Coverage	Description
1t3d.1.A	48.45	homo-hexamer	0.57	HHblits	X-ray	2.20Å	0.44	0.78	Serine acetyltransferase
6jvu.1.A	49.22	homo-hexamer	0.56	HHblits	X-ray	2.80Å	0.44	0.77	Serine acetyltransferase
6jvu.1.A	50.40	homo-hexamer	0.56	BLAST	X-ray	2.80Å	0.44	0.76	Serine acetyltransferase
3gvd.1.A	47.10	homo-hexamer	0.53	HHblits	X-ray	2.40Å	0.43	0.78	Serine acetyltransferase
6wye.1.A	44.32	homo-hexamer	0.54	HHblits	X-ray	2.01Å	0.42	0.80	Serine acetyltransferase
6wye.1.B	44.32	homo-hexamer	0.54	HHblits	X-ray	2.01Å	0.42	0.80	Serine acetyltransferase
4h7o.1.A	46.00	homo-hexamer	0.52	HHblits	X-ray	2.17Å	0.43	0.79	Serine acetyltransferase
4hzd.1.A	46.18	homo-hexamer	0.53	HHblits	X-ray	1.87Å	0.42	0.79	CysE, serine acetyltransferase
1ssq.1.A	45.45	homo-hexamer	0.55	HHblits	X-ray	1.85Å	0.43	0.76	Serine acetyltransferase
4h7o.1.A	46.80	homo-hexamer	0.51	BLAST	X-ray	2.17Å	0.44	0.76	Serine acetyltransferase
4hzd.1.A	47.64	homo-hexamer	0.52	BLAST	X-ray	1.87Å	0.43	0.77	CysE, serine acetyltransferase
4hzc.1.A	46.36	homo-hexamer	0.51	HHblits	X-ray	1.97Å	0.42	0.79	CysE, serine acetyltransferase
1ssm.1.A	46.67	homo-hexamer	0.53	HHblits	X-ray	2.15Å	0.43	0.73	Serine acetyltransferase
4hzc.1.A	47.64	homo-hexamer	0.50	BLAST	X-ray	1.97Å	0.43	0.77	CysE, serine acetyltransferase
4ni69.1.A	55.02	homo-trimer	0.45	HHblits	X-ray	1.80Å	0.46	0.75	Serine Acetyltransferase Apoenzyme
3gvd.1.A	49.34	homo-hexamer	0.51	BLAST	X-ray	2.40Å	0.44	0.69	Serine acetyltransferase
1sst.1.A	45.10	homo	0.50	HHblits	X-ray	2.00Å	0.43	0.77	Serine acetyltransferase

Figure 3: Result of homology modeling report of candidate transcript in ID-15184 obtained from swiss model (https://swissmodel.expasy.org/interactive/5zAYVX/templates/)

Template IT3D has a higher sequence identity of 48.45%. For docking, We selected a template with a higher QSQE (quaternary structure quality estimate)score. The QSQE score is a value between 0 to 1, representing the expected accuracy of inter-chain interactions for a model created using a specific alignment and template. In general, a greater QSQE is "better" when modeling the projected quaternary structure.

Ramachandran plots were constructed to determine whether template 1T3D is stereochemically stable[43] through PROCHECK online server (https://saves.mbi.ucla.edu).

The PDB file containing the coordinates of the structure of interest is the only input required for PROCHECK[44]. This PROCHECK analyzes residue-by-residue geometry and overall structural geometry to determine the stereo-chemical quality of a protein structure. According to the standard stereochemical parameters described by Morris[45], our template 1t3d has the most favored region > 90%, which means our template is stable enough to perform the docking procedure. Moreover, it has an additional allowed region of 9% and only a 0.5% disallowed region (Table 4). From the summary of the Ramachandran plot, out of 251 residues, there are five labeled residues, and in the chi1-chi2 plot, out of 134 residues, one is labeled residue. Labeled residues are those residues that have unfavorable conformations (score < -3.00) (Figure 4b). Other properties are like maximum deviation is 5.3; Bond length/angle is 4.7; a bad contact is 0; 1 cis peptide G-factors Dihedrals is -0.12; covalent 0.23; Moris classes is 1 2 2; overall: 0.02, Planar groups: 98.8% within limits 1.2% highlighted.

To further analyze the stability of template 1t3d, the WHAT CHECK program was run (https://saves.mbi.ucla.edu/). It checks various stereochemical properties of the residues in the model and is derived from the WHAT IF program[46], a subset of protein verification tools. These results showed its 1st generation packing quality as: -0.143, Ramachandran plot appearance is -0.327, chi-1/chi-2 rotator normality is -0.150; backbone conformation is -31.972; bond length & bond angle is 0.789, 0.959 respectively; Side chain planarity is 1.326, and omega angle restraints are 1.149.

The predicted protein contained 261 amino acid residues, out of which 24 residues were Glycine, 14 residues were proline. The 3D protein model of this protein was developed from the Swiss Model (Figure 4a). The Ramachandran plot (Figure 4b) was also produced, and the data showed (Table 4) the likelihood of the predicted SAT domain corresponding to ID-15184.

Table 4: Ramachandran Plot statistics for the predicted protein from SAT candidate transcript ID- 15184

Regions	No. of residues	Percentage
Most favored regions [A, B,L]	200	90.1%
Additional allowed regions[a,b,l,p]	20	9.0%
Generously allowed region [~a,~b,~l,~p]	1	0.5%
Disallowed regions [XX]	1	0.5%*
Non-glycine & Non-proline regions	222	100.0%

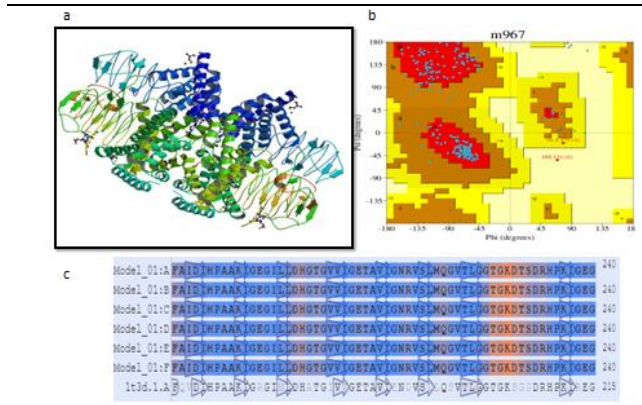

Figure 4: (a) 3D-View of 1T3D protein obtained from (https://swissmodel.expasy.org/templates/1t3d.1) source author [47] (b) Ramachandran plot of 1T3D produced using PROCHECK; (c) Sequence of Amino Acid of template IT3D obtained from the Swiss model.

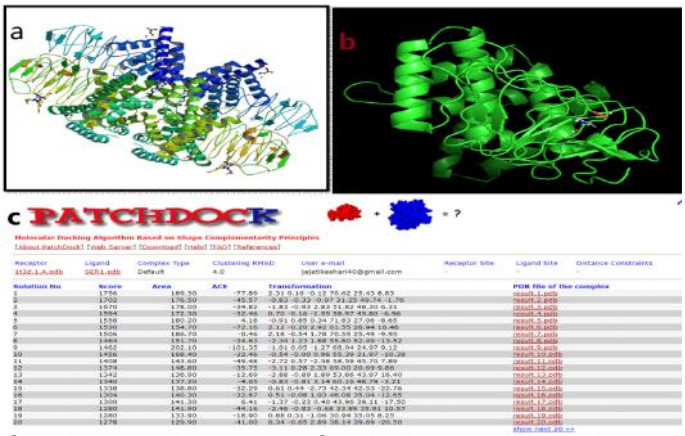

Figure 5(a): Protein before docking (b) Protein after docking with ligand L-serine (c) Results obtained from Patch dock server.

Pathway mapping results depicted that putative transcripts have a role in sulfur metabolism, Glycine, Serine, and threonine metabolism. Figure 6 depicts putative transcripts sharing similarities with EC-2.3.1.30. As we can see, the putative transcript (ID-15184) produces O-acetyl-l-serine by taking L-serine as a substrate. Finally, O-acetyl-l-serine has been used as a substrate in the cysteine and methionine metabolism pathways.

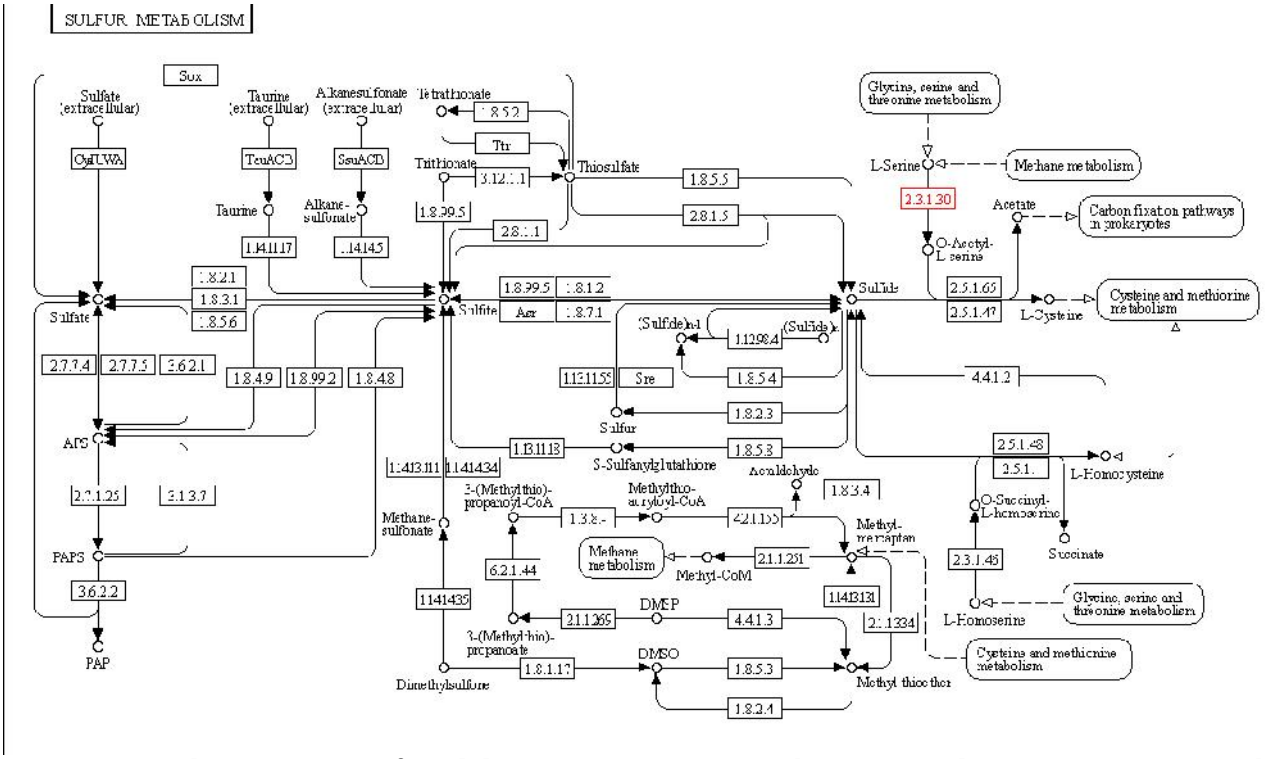

Figure 6: Pathway mapping of candidate transcript ID-15184, the corresponding enzyme serine acetyl transferase (E-2.3.1.30) shown in red colour

3.6 Molecular Docking

The predicted enzyme was docked with L-Serine as Ligand, the SAT enzyme's reported substrate, to confirm its catalytic activity. L-serine was used as a ligand and template IT3D, obtained from homology modeling of candidate transcript ID-15184, to study ligand-protein interactions. Docking simulation was performed in the online platform PatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/).

The PatchDock algorithm is based on computer vision algorithms for object detection and image segmentation. It consists of 3 important stages: surface patch making, molecular shape representation, and

filtering and scoring. Generally, it employs segmentation techniques to ascertain geometric patches like convex, concave and flat surfaces. Geometric hashing & Pose-clustering algorithms are utilized in surface patch making. Concave patches & flat patches can proceed with any type of patch & eventually, in the filtering and scoring stage, candidate's complexes are screened. Complexes having undesirable penetrations of the receptor's atoms into the ligand's atom are jettisoned. Ultimately, the remaining complexes are assigned a score according to their geometric shape complementarity score and ranked[48,49,50]. Patchdock generates a series of potential complexes between a ligand molecule and a user-specific receptor. These complexes are shown in Figure 5c. The rank of the different complexes depends upon various factors like interface shape, size of the molecule, and interface shape[50]. In the ligand-receptor complex, steric classes give more noise and reduce their score and rank. In interface shape, convex/concave shapes are easier to find out than flat surfaces, and finally, the larger the molecule, the higher the score and rank[50].

From the patch dock results, we have taken solution no 1 (i.e., the proposed docking model in solution no.1) because it has a higher score of 1756 and an area of 189.30, which has the highest among all solutions. In addition to that, it has an ACE of -77.89. It has three rotational angles of 2.31, 0.16,-0.12, and 3 translational parameters of 70.62, 25.43, and 8.83. Clustering RMSD was taken as four standards for docking interaction (The higher the value lesser will be the solution no). Due to its higher score and higher interface area, it was taken as the best ligand-enzyme complex of L-serine and SAT enzyme. As seen in Figure 5a, it is the 3D structure of template 1t3d retrieved from the Protein data bank (https://www.rcsb.org/structure/1T3D), and Figure 5b is the complex of 1t3d-L-serine, the protein-ligand interaction is showing one ligand-binding pocket in the predicted protein. The PYMOL software results showed that the active site consisted of 8 amino acids. GLY 278, ALA 252, PRO 245, ILE 186, LYS 165, ALA 88, VAL230, and TYR-87.

Discussion

Endosymbiotic or endo-symbiosis theory was first forwarded by Russian botanist Konstantin Mereschkowsk[51], which suggested that eukaryotic organelles are derived from prokaryotic cells. The theory holds for mitochondria, plastids such as chloroplasts, based on the evidence such as binary fission of mitochondria and plastids; the presence of porin and cardiolipin proteins found in bacterial cell membranes, and most importantly, circular DNA molecules. Plants use many genes of bacterial origin for the biosynthesis of amino acids, flavonoids, etc., and stress mitigation. β-ODAP is a D-amino acid synthesized as a part of cyanide scavenging pathways in cyanogenic plants like *Lathyrus*[21]. Similar pathways and d-amino acids have been reported in many bacteria [39]. So in this study, to screen *L. sativus* leaf transcriptome data to identify candidate SAT gene involved in ODAP biosynthesis, the sequence of Serine acetyltransferase from *Staphylococcus aureus* (strain MRSA252) was used as a query. Since the ODAP biosynthesis and sulfur metabolism are known to occur in Chloroplast or mitochondria [21,52,53,59], we checked the candidate gene SAT (Contig ID-15184). The presence of the signal peptides in the putative protein model corresponding to Contig ID-15184 showed that the putative protein might be located in the Chloroplast and mitochondria.

Further, the conserved domain search analysis of candidate transcript (ID-15184) indicated the presence of domains like PLN02739, CysE, LbH_SAT, SATase_N, and DUF3432. This domain shows significant similarity to the domain of SAT of Staphylococcus aureus. (Table 1). Downie, in 1989, reported that SAT is a bacterial enzyme belonging to the family o-acetyltransferase[54]. Its SATase_N domain, the N-terminal domain, is conserved both in plants and bacteria. The SAT case encoding transcript ID-15184 also possessed the LbH_SAT domain that catalyzes the CoA-dependent acetylation of the side chain hydroxyl group of L-serine to form O-acetylserine. The formation of the O-acetylserine is the first step of a biosynthetic pathway leading to the formation of L-cysteine in bacteria and plants[19]. This reaction represents a key metabolic point of regulation for the cysteine biosynthetic pathway due to its feedback inhibition by cysteine. The enzyme is a 175 kDa homo-hexamer composed of a dimer of homo-trimers. Each subunit contains an N-terminal alpha-helical region and a C-terminal left-handed beta-helix (LbH) subdomain with five turns. Each contains a

hexapeptide repeat motif characteristic of the acyltransferase superfamily (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?RID=KFMUJKRM014&mode=all).

The *L. sativus LsSAT* encoding transcript (ID-15184) also showed phylogenetic similarity to the SAT of *Cicer arietinum* (accession no XM00451290, Locus id LOC101491313) and serine acetyltransferase two belonging to *Medicago trancatula* (XM003612271) as depicted in a phylogenetic tree (Figure 1a). Both the plant belongs to the *Leguminosae* family-like *L. sativus.* The genetic distance (Figure 1b) calculated through maximum composite likelihood correlates with the phylogenetic tree results. Putative transcript encoding LsSAT also showed a closer genetic distance of 0.126, with *Medicago truncatula* serine SAT (LOC11409367, Accession no-XM013598127) followed by the accession no-XM00361226. The transcript sequence ID15184 was distantly related to *Ziziphus jujube* SAT 2 mRNA fragment (LOC107424304, Accession No-XM 025076710) hibiscus plant is ornamental.

Homology models of candidate transcript ID-15184 encoding *LsSAT* produced from SWISS-MODEL template library P was searched. Out of 41 templates, 1t3d has a higher QSQE score, i.e., 0.57, and high coverage of 0.78 (Figure 3). It is a template of SAT gene from *E. coli*. Due to its high QSQE score and high coverage, this template was screened to check whether this template is stereochemically stable or not. In addition, PROCHECK (https://saves.mbi.ucla.edu/) and the WHAT CHECK program was used to study the stereochemical properties of template 1t3d.

In the pro check program, out of 261 amino acids, 200 amino acids represent the allowed region which is > 90 %(Table 4), indicating the quality of the predicted template . The stability of the template is important to conduct further analyses like docking with a suitable ligand. Molecular docking is a useful tool to check the enzyme-substrate interaction and subsequently provide evidence of the functional properties of the enzyme. The docking of putative SAT enzyme encoding transcript identified from *L. sativus* showed a single binding pocket in the protein corresponding to ID-15184, which had an affinity site for binding L-serine as ligand. The PYMOL view of docked template showed that eight amino acids *viz.*, GLY 278, ALA 252, PRO 245, ILE 186, LYS 165, ALA 88, VAL230, and TYR-8 have electrostatic interaction with the ligand L-serine. The finding is per the beta ODAP biosynthesis pathway proposed by Malathi et al. wherein a single ligand L-serine binds to enzyme serine O-acetyl transferase (EC 2.3.1.30). Interaction of serine O-acetyl transferase with L-serine leads to the formation of O-acetyl l-serine[19]. The docking analysis of putative *LsSAT* interacting with L-serine is correlated with the ODAP biosynthesis pathway in *L. sativus*. Pathway mapping (Figure 6) also revealed that putative transcript has a role in amino acid metabolism[55] and sulfur metabolism[56]. Thus putative transcript (ID-15184) has a major role in sulfur assimilation and cysteine synthase complex [57,58].

Semi-Quantitative RT PCR Validation of the selected transcript

The semi-Quantitative RT PCR results of primer PaaY1 (designed from putative candidate *LsSAT* encoding transcript ID-15184) were performed in 11 diverse *L. sativus* genotypes with varied seed ODAP content (Table 3). In all genotypes, an amplicon of about 112 bp was obtained with a similar expression level. Although 11 genotypes had varied seed ODAP content (Table 3), the amplicon intensity on the gel did not show differential expression. The enzyme SAT was found to express uniformly in *L. sativus*, which suggests that the formation of substrate o-acetyl-l Serine is a common step shared by the ODAP biosynthesis and *L-cysteine* biosynthesis pathways. The pathway is bifurcated after the formation of o-acetyl-l Serine, which is used by the enzyme *β-(isoxazoline-5-on-2-yl) -L-alanine synthase* (EC 2.5.1.119), *β-(isoxazoline-5-on-4-yl)-L-alanine synthase* (EC 2.5.1.47), *Cysteine synthase*; (EC 2.5.1.51), *β-pyrazolyl alanine synthase* (EC 2.5.1.52) and *L-mimosine synthase* as per the proposed pathway by Malathi et al. and Song et al. [20,21]

Amplicon sequencing of putative LsSAT encoding amplicon and Phylogenetic analysis

To further confirm the results obtained from in-silico analysis, amplicon resequencing was done. The sequencing of the gel eluted fragments matched with the sequence of the known serine-o-acetyl transferase

gene. Analysis of the eluted sequence also consisted of domains like PLN02739, SATase_N, CySE with accession no PLN02739, pfam06246smart00971, TIGR01172, and COG1045 (Supplementary Figure 3). These results indicate that the candidate gene corresponding to ID-15184 mRNA is SAT encoding gene, having a role in the ODAP biosynthesis pathway in the *L. sativus* gene. The phylogenetic tree of eluted sequences Pa7 (Supplementary Figure 2) also showed significant homology with the transcripts of serine acetyltransferase of *Cicer arietinum* and *Medicago truncatula*.

Conclusion

The study conducted to screen the *L. sativus* leaf transcriptome of genotype RLK1950 to identify the genes related to β ODAP biosynthesis and seed ODAP content in *Lathyrus sativus* led to the identification transcript ID-15184, a transcript encoding Serine acetyltransferase in *L. sativus* (*LsSAT*) (Accession No-MW590957). Protein domain search, modelling, and docking results of putative *LsSAT* showed that it carries the reaction of acetylation of Serine to form O-acetyl serine and might be located in Chloroplast or mitochondria. The expression analysis showed a similar expression in the leaves of 10 days old seedlings of both low and high ODAP containing genotypes suggesting that this is the primary step in the β-ODAP synthesis. Further confirmation was obtained by amplicon sequencing of *LsSAT* re-established the structural and functional properties. The transcript ID-15184 is thus considered encoded by the *LsSAT* gene in *L. sativus*. The results form the basis for further screening of the transcriptome to discover other genes corresponding to *alanine synthase*, *cysteine synthase*, and *oxalyl transferase*, which may directly correlate with seeding and leafing ODAP content. It will also pave the way for further identification of key regulatory genes and subsequent genetic modifications to develop toxin-free, safe *Lathryus* genotypes.

Funding: This research received no external funding, the work was funded by the research grant of the IGKV, Raipur for in-house research project (RPF-PMBB13)

Conflicts of interest: The authors declare that they have no conflicts of interest.

Availability of data and material: Not applicable

Code availability: BLAST2GO software used with key B2G-NAYAJAJAm 31B1BCB02261964C6E8BF66D45DB232D

Ethics approval: Not applicable

Informed Consent Statement: Not applicable

References

1. Yan ZY, Spencer PS, Li ZX, Liang YM, Wang YF, Wang CY & Li FM, *Lathyrus sativus* (grass pea) and its neurotoxin ODAP. *Phytochemistry* , 67(2006) 107.
2. Kumar S, Bejiga G, Ahmed S, Nakkoul H & Sarker A, Genetic improvement of grass pea for low neurotoxin (β-ODAP) content. Food Chem Toxicol, 49 (2011) 589.
3. Gusmao M, Siddique KHM, Flower K, Nesbitt H & Veneklaas EJ, Water deficit during the reproductive period of grass pea (*Lathyrus sativus* L.) reduced grain yield but maintained seed size. *J Agron Crop Sci*, 198 (2012) 430.
4. Campbell CG, Grass pea (*Lathyrus sativus* L.) Promoting the Conservation and Use of Underutilized and Neglected Crops. Institute of Plant Genetics and Crop Plant Research, Gatersleben & International Plant Genetic Resources Institute, Rome, Italy, 18 (1997) 1.
5. Campbell CG, Mehra RB, Agrawal SK, Chen YZ, Abd-El-Moneim AM, Khawaja HIT, Yadav CR, Tay JU & Araya WA, Current status and future strategy in breeding grass pea (*Lathyrus sativus*). *Euphytica*, 73 (1994) 167.

6. Yan ZY, Spencer PS, Li ZX, Liang YM, Wang YF, Wang CY & Li FM, *Lathyrus sativus* (grass pea) and its neurotoxin ODAP. *Phytochemistry*, 67 (2006) 107.
7. Xiong JL, Xiong Y, Bai X, Kong H, Tan R, Zhu H, Siddique K, Wang J & Turner NC, Genotypic variation in the concentration of β-N-Oxalyl-L-α,β-diaminopropionic acid (β-ODAP) in Grass Pea (*Lathyrus sativus* L.) seeds is associated with an accumulation of leaf and pod β-ODAP during vegetative and reproductive stages at three levels of water Stress. *J Agric Food Chem*, 63 (2015) 6133.
8. Williams PC, Bhatty RS, Deshpande SS, Hussein LA & Savage GP, Improving the nutritional quality of cool-season food legumes. In : Muehlbauer FJ, Kaiser WJ eds, Expanding the Production and Use of Cool Season Food Legumes,(Kluwer Academic Publishers, Dordrecht Netherlands), 1994, 113.
9. Akalu G, Tufvesson F, Jönsson C & Nai BM, Physico-chemical characteristics and functional properties of starch and dietary fiber in grass pea seeds. *Starch*, 50 (1998) 374.
10. Hanbury C, White C, Mullan B & Siddique K, A review of the potential of *Lathyrus sativus* L. and *L. cicera* L. grain for use as animal feed. *Anim Feed Sci Technol*, 87 (2000) 1.
11. Jiao CJ, Xu QL, Wang CY, Li FM, Li ZX & Wang YF, Accumulation pattern of toxin β -ODAP during lifespan and nutrient elements on β-ODAP content in *Lathyrus sativus* seedlings. *J Agric Sci*, 144 (2006) 369.
12. Prakash S, Mishra BK, Adsule RN & Barat GK, Distribution of β-*N*-oxalyl-L-α, β-diaminopropionic acid in different tissues of aging *Lathyrus sativus* plants. *Biochemie und Physiologie der Pflanzen*, 171 (1977) 369.
13. Rao SLN, Adiga PR & Sarma PS, The isolation and characterization of β-Noxalyl-L-α,_β-diaminopropionic acid, a neurotoxin from the seeds of Lathyrus sativus. *Biochemistry*, 3 (1964) 432.
14. Addis G & Narayan RKJ, Developmental variation of the neurotoxin, β-Noxalyl- L-α,β-diaminopropionic acid (ODAP) in Lathyrus sativus. *Annals of Botany*,74 (1994) 209.
15. Roy DN & Spencer PS, Lathyrogens: Toxicants of Plant Origin, In Cheeke PR Eds, Proteins and Amino Acids, CRC Press, Boca Raton, FL, USA, 3 (1991)170.
16. Jiao CJ, Studies on accumulation and biological significance of β-ODAP in *Lathyrus sativus* L. (grass pea), Master Dissertation, Lanzhou University, Lanzhou, China, 2005.
17. Lambein F, Godelieve O & Haey KY, β-isoxazolinone-alanine is involved in the biosynthesis of the neurotoxin β-N-oxalyl-l-α,β-diaminopropionic acid. *Phytochemistry*, 29 (1990) 3793.
18. Ikegami F, Ongena G, Sakai R, Itagaki S, Kobori M, Ishikawa T, Kuo YH, Lambein F & Murakoshi I, Biosynthesis of β-(isoxazolin-5-on-2-yl)-alanine, the precursor of the neurotoxin β-N-oxalyl-L-α, β diaminopropionic acid, by cysteine synthase in Lathyrus sativus. *Phytochemistry*, 33(1993) 93.
19. Xu Q, Liu F, Chen P, Jez JP & Krishnan HB, β-N-Oxalyl-l-α,β-diaminopropionic Acid (β-ODAP) Content in Lathyrus sativus: The Integration of Nitrogen and Sulfur Metabolism through β-Cyanoalanine Synthase .*Int J Mol Sci*, 18(2017) 526.
20. Malathi K, Padmanaban G & Sarma PS, Biosynthesis of β-N-oxalyl-L-α_,β_-diamino-propionic acid, the *Lathyrus sativus* neurotoxin. *Phytochemistry*, 9(1970) 1603.
21. Song Y, Wang L, Liu F, Jiao C, Nan H, Shen X, Chen H, Li Y, Lei B, Jiang J, Chen P & Xu Q, β-Cyanoalanine Synthase Regulates the Accumulation of β-ODAP via Interaction with Serine Acetyltransferase in *Lathyrus sativus*. J. *Agric Food Chem*, 69 (2021) 1953.
22. Dixit GP, Parihar AK, Bohra A & Singh NP, Achievements and prospects of grass pea (*Lathyrus sativus* L.) improvement for sustainable food production. Crop J, 4 (2016) 407.

23. Lambein F, Haque R, Khan JK, Kebede N & Kuo YH, From soil to brain: zinc deficiency increases the neurotoxicity of *Lathyrus sativus* and may affect the susceptibility for the motorneurone disease neurolathyrism. *Toxicon,32(* 1994) 32(4) , 461–466.
24. Malathi K, Padmanaban G, Rao SLN & Sarma PS, Studies on the biosynthesis of β-N-oxalyl-L-α, β-diaminopropionic acid, the *Lathyrus sativus* neurotoxin. *Biochim Biophys Acta*, 141 (1967) 71.
25. Lambein F, Kuo YH, Kusama-Eguchi K & Ikegami F, β-N-oxalyll-α,β-diaminopropionic acid, a multifunctional plant metabolite of toxic reputation. *Arkivoc* 9 (2007) 45.
26. Bhariya S, Kakade D, Wallalwar M & Banerjee S, Alternative Protocol for Isolation of Total RNA. From *Lathyrus sativus* Without Using Liquid N2 for Molecular Analysis. *Bulletin of Environment, Pharmacology and Life Sciences* 6 (2017) 495-498.
27. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ & Nakai K, WoLF PSORT: protein localization predictor. *Nucleic Acids Res*, 35 (2007) 585.
28. Emanuelsson O, Nielsen H, Brunak S & Heijne G, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*, 300 (2000) 1005.
29. Bairoch A & Apweiler R, The SWISS-PROT protein sequence database and its supplement TrEMBL. *Nucl Acids Res*, 28 (2000) 45.
30. Hobohm U, Scharf M, Schneider R & Sander C, Selection of representative protein data sets. *Protein Sci*, 1 (1992) 409.
31. Smith TF & Waterman MS, Identifcation of common molecular subsequences. J Mol Biol. 147 (1981) 195.
32. Pearson WR, Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol*, 183 (1990) 63.
33. Tamura K, Dudley J, Nei M & Kumar S, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*, 24 (2007) 1596.
34. Laskowski RA, MacArthur MW & Moss DS, Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*, 26 (1993) 283.
35. Saitou N & Nei M, The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4 (1987) 406.
36. Felsenstein J, Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39 (1985) 783.
37. Kumar S, Stecher G, Li M, Knyaz C & Tamura K, MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* , 35 (2018) 1547.
38. Noji M, Inoue K, Kimura N, Gouda A & Saito K, Isoform-dependent differences in feedback regulation and subcellular localization of serine acetyltransferase involved in cysteine biosynthesis from Arabidopsis thaliana. *J Biol Chem*, 273 (1998) 32739.
39. Chakraborty S, Mitra J, Samanta MK, Sikdar N, Bhattacharyya J, Manna A, Pradhan, S, Chakraborty A & Pati BR, Tissue specific expression and in-*silico* characterization of a putative *Cysteine synthase* gene from *Lathyrus sativus* L. *Gene Expr. Patterns*, 27 (2018) 128.
40. Nayak JK & Banerjee S, Analysis of de novo transcriptome sequence for identification of putative candidate genes related to β- ODAP biosynthesis pathway in *lathyrus sativus*, M.sc(Ag.) thesis, Indira Gandhi Krishi Vishwavidyalaya, Raipur,2109.
41. Camach OC, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL, BLAST+: architecture and applications. *BMC Bioinform*, 10 (2009) 421.

42. Remmert M, Biegert A, Hauser A & Söding J, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*, 9 (2012) 173.
43. Ramachandran GN, Ramakrishnan C & Sasisekharan V, Stereochemistry of Polypeptide Chain Conformations. *J Mo Bio*, 7 (1963) 95.
44. Bernstein FC, Koetzle TF, Williams GJB, Meyer EFJr, Brice MD, Rogers JR, Kennard O, Shimanouchi T & Tasumi M, The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mo Bio*, 112 (1977) 535.
45. Morris AL, MacArthur MW, Hutchinson EG & Thornton JM, Stereochemical quality of protein structure coordinates. *Proteins*, 12 (1992) 345.
46. Vriend G, WHAT IF: a molecular modelling and drug design program. *J Mol Graph*, 8 (1990) 52.
47. Pye VE, Tingey AP, Robson RL & Moody PC, The structure and mechanism of serine acetyltransferase from Escherichia coli. *J Biol Chem*, 279 (2004) 40729.
48. Schneidman-Duhovny D, Inbar Y, Nussinov R & Wolfson HJ, PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucl Acids Res*, 33 (2005) 363.
49. Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, Barzilai A, Dror O, Haspel N, Nussinov R & Wolfson HJ, Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins*, 52 (2003) 107.
50. Duhovny D, Nussinov R & Wolfson HJ, Efficient Unbound Docking of Rigid Molecules, in Gusfield et al, Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science, 2452 (2002) 185.
51. Mereschkowsky C, Über Natur und Ursprung der Chromatophoren im Pflanzenreiche.*Biol. Centralbl*, 25 (1905) 593.
52. Hendrickson HR & Conn EE, Cyanide metabolism in higher plants. IV. Purification and properties of the beta-cyanolanine synthase of blue lupine. *J Biol Chem*, 244 (1969) 2632.
53. Ikegami F, Takayama K & Murakoshi I, Purification and properties of β-cyano-l-alanine synthase from *Lathyrus latifolius*. *Phytochemistry*, 27 (1988) 3385.
54. Downie JA, The nodL gene from Rhizobium leguminosarum is homologous to the acetyl transferases encoded by lacA and cysE. Mol Microbiol, 3 (1989) 1649.
55. Liu FJ, Jiao CJ, Bi CX, Xu QL, Chen P, Heuberger AL & Krishnan H, Metabolomics research related to ODAP metabolism in *Lathyrus sativus* L. *J Agric Food Chem* 65 (2017) 10206.
56. Leustek T & Saito K, Sulfate transport and assimilation in plants. *Plant Physiol*, 120 (1999) 637
57. Francois JA, Kumaran S & Jez JM, Structural basis for interaction of O-acetylserine sulfhydrylase and serine acetyltransferase in the Arabidopsis cysteine synthase complex. *Plant Cell*, 18 (2006) 3647.
58. Feldman-Salit A, Wirtz M, Hell R & Wade RCA, Mechanistic model of the cysteine synthase complex. *J Mo Biol*, 386 (2009) 37.
59. Wirtz M, Beard KFM, Lee CP, Boltz A, Schwarzländer M, Fuchs C, Meyer AJ, Heeg C, Sweetlove LJ, Ratcliffe RG & Hell R, Mitochondrial cysteine synthase complex regulates Oacetylserine biosynthesis in plants. J Biol Chem, 287 (2012) 27941.

——==00==——