Plant Omics Journal POJ

POJ 5(3):223-230 (2012) ISSN:1836-3644

Functional annotation of expressed sequence tags of *Papaver somniferum*

Avantika Priya[1], Himanshu Tripathi[1], Dharmendra K.Yadav[1], Feroz Khan[1*], Vikrant Gupta[2], Rakesh K. Shukla[2], M.P. Darokar[2]

[1]Metabolic and Structural Biology Department, CSIR-Central Institute of Medicinal and Aromatic Plants, P.O.-CIMAP, Kukrail Picnic Spot Road, Lucknow-226015 (U.P.), India
[2]Biotechnology Division, CSIR-Central Institute of Medicinal and Aromatic Plants, P.O.-CIMAP, Kukrail Picnic Spot Road, Lucknow-226015 (U.P.), India

***Corresponding author: f.khan@cimap.res.in**

Abstract

Papaver somniferum (opium poppy) is the source for several pharmaceutical benzylisoquinoline alkaloids (BIA) including morphine, codeine and sanguinarine. In an attempt to identify additional biosynthetic steps at the molecular level, a large expressed sequence tag (EST) dataset with a total of 20,532 sequences from opium poppy were assembled and functionally annotated. Sequence cleaning filter showed 20,445 (709 trimmed) valid sequences and 87 trashed sequences at 96% minimum identity parameter for an alignment with a contaminant. Repeat masking filter showed length of valid sequences 1,69,60,151 bp, GC level 41.12% and masked bases 5,01,768 bp (2.96%) by using RepBase (update 8.12) repeats library of Arabidopsis. Elements observed in EST sequences were 1017 retroelements, 55 LINEs, 55 L1/CIN4, 962 LTR elements, 412 Ty1/Copia, 550 Gypsy/DIRS1, 56 DNA transposons, 15 hobo-Activator, 7 Tc1-IS630-Pogo, 7 En-Spm, 11 MuDR-IS905, 6 Tourist/Harbinger, 1 satellite, 920 simple repeats and 1697 low complexity elements. NCBI's vector library UniVec (core) and plastids library of Arabidopsis were used for masking vector and organelle sequences, respectively. For EST sequence assembly, the cutoff for overlap sequence identity was kept 80%. A total of 15,279 assembled unique transcripts were identified, which include 1,408 contigs and 13,871 singletons. Functional annotation was performed through BLASTX using UniProtKB/Swiss-Prot database. EST sequences were retrieved from dbEST (NCBI) and processed, cleaned, clustered and assembled through Gencheck, EGassembler, Phred, RepeatMasker, Cross-Match, Phrap and Cap3 softwares. Mostly hypothetical genes were observed apart from important genes related to secondary metabolism such as enzymes of BIA biosynthesis *e.g*., stylopine synthase, S-norcoclaurine synthase 1, salutaridine synthase, salutaridine reductase, (S)-coclaurine N-methyltransferase, S-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase and salutaridinol 7-O-acetyltransferase. This functionally annotated EST dataset will be a useful resource for further studies such as taxonomy, molecular breeding, genetics, genomics and secondary metabolism.

Keywords: *Papver somniferum*, opium poppy, EST, sequence assembly, contigs, functional annotation, metabolic pathways.

Introduction

Opium poppy (2n = 22, *Papaver somniferum* L.) is an annual medicinal herb which belongs to the family Papaveraceae. The medicinal value of opium poppy is due to the presence of numerous alkaloids out of which morphine, codeine, narcotine, thebaine and papaverine are frequently used as pain killer, sedative, analgesic, antitussive and antispasmodic in modern medicine. Amongst these compounds, narcotic analgesic morphine and the antitussive codeine are the most important physiologically active alkaloids obtained from this plant. India is a leading country in the cultivation of opium poppy and is the world's largest producer of licit opium for the world's pharmaceutical industry. Despite its importance as a medicinal crop, very little genetic information is available because of its genetic variability. Partial cDNA sequences known as expressed sequence tags (ESTs) have become the method of choice for the rapid and cost-effective generation of data on the coding regions of genomes in a wide range of organisms. In plants, this method was initially used for the model species *Arabidopsis thaliana* (Höfte et al., 1993) and rice (Yamamoto and Sasaki, 1997). Since then, generation of EST sequences and their analyses has become an efficient approach for identifying a large number of genes expressed during different developmental stages or in response to a variety of environmental conditions in plants such as apple (Newcomb et al., 2006), coffee (Lin et al., 2005), cocoa (Argout et al., 2008), pepper (Kim et al., 2008), sugarcane (Gupta et al., 2010), and a conifer genomics resource of 200,000 spruce (Picea spp.) ESTs (Ralph et al., 2008), Chinese maize (Wang et al., 2011), wheat (Kumar and Sharma, 2011). Recently, large-scale EST sequencing on the 454-GS-FLX platform and their analyses revealed genes involved in alkaloid biosynthesis in plants (Luo et al., 2010). Enormous plant EST sequences from a large variety of species have been deposited in dbEST database (NCBI; www.ncbi.nlm.nih.gov/dbEST). As of January 2011, only 293 nucleotide sequences, 20,863 ESTs, no data of Unigenes, 166 proteins, 2 conserved protein domains, 32 UniSTS (markers and mapping data), and 14 PopSet (population study data set) were available for *P. somniferum* in NCBI repository, USA. Generation of a large-scale expressed sequence tag (EST) dataset is a useful approach to accelerate the researches in non-model plant species such as *P. somniferum*. As a valuable resource for comparative genomics, functional genomics and biodiversity study, EST

database has been established for model plant species and non-model species (citrus, grape, kiwifruit, etc.). So far, a very limited number of EST sequences from only three Ranunculales species (*Aquilegia Formosa* × *Aquilegia pubescens*, *Eschscholzia californica*, and *Papaver somniferum*) are available in the public databases. Besides, the ESTs can be useful in molecular marker development, construction of genetic and physical maps, comparative mapping, and gene discovery (Wang et al., 2001). Despite the numerous compounds identified from poppy, the biosynthetic pathways and the participating enzymes or cDNAs have been characterized only for a few selected members, whereas the biosynthesis of the majority of the compounds is still unknown (Pienkny et al., 2009). In the present study, an attempt was made to identify the existing and additional biosynthetic steps in Benzylisoquinoline alkaloid (BIA) biosynthesis pathway in opium poppy. Currently available ESTs dataset with a total of 20,532 ESTs from *P. somniferum* were assembled and functionally annotated aiming to provide information for deciphering their role in secondary metabolism. The primary goal of this investigation was to annotate and assign putative functions to 1,408 assembled contig sequences. The functionally annotated ESTs dataset of opium poppy will be a useful resource for further studies such as taxonomy, molecular breeding, genetics, genomics and secondary metabolism.

Results and Discussion

Assembly of EST sequences

Of the 15,279 *Papaver* assembled unique transcripts, 1408 (9.22%) and 13,871 (90.78%) comprised of contigs and singletons, respectively. The assembled 1,408 contigs sequences were compared to the GenBank non-redundant database using BLASTX (cut-off E-value of $\leq 1e^{-5}$) to assign putative function. Of all assembled sequences 61.01% could be assigned a putative identity, 26.49% with hypothetical functions and 12.5% showed no match with existing sequences at an E-value of $1e^{-5}$ (Fig 1). Assembled contig sequences were categorized with respect to functionally annotated genes in *Papaver* and grouped into different categories of biological roles by using sequence similarity. Almost 61.01% of assembled contig sequences were categorized and the largest fraction of the transcripts with a putative identity coded for enzymes catalyzing various metabolic and biosynthetic reactions (29.12%) whereas the functions of 38.99% transcripts were not classified. Fig 2 shows the comparative distribution of functional categories among the classified genes from *Papaver* assembled sequences. The genes with 'housekeeping' roles such as protein and amino acid metabolism were over-represented.

Structural and functional annotation of ESTs

The consensus sequences are blasted against non-redundant protein database and 20,532 ESTs comprising of 1,408 contigs and 13,871 singletons showed significant hits. Annotated contigs were further categorized with respect to their biological roles. After enzymes, ribosomal proteins and retrotransposon-related proteins are the second major contributor to total annotated contigs comprising of 4.33% each (Table 1). Transporters are another large group (3.41%) and cytochrome-related proteins are also present (0.64%) that might participate in various hydroxylation reactions in secondary metabolite synthesis with the primary metabolic reactions. It is followed by the translation-related at 0.28%,

Table 1. Functional classification of *P. somniferum* assembled contigs with known function.

Predicted Function	No. of Contigs
Transcription terminator	1
Senescense associated	3
Antimicrobial peptide	3
Lipid binding protein	4
Translation related	4
Extracellular proteins	4
Proteosome	5
Heat shock	5
Chaperone	6
Polyproteins	14
Pathogenicity related	8
Organelle protein	13
Ubiquitin related	8
Stress associated	8
Cytochrome related	9
Receptors	10
Histone	10
Chromosomal protein	11
Calcium binding	12
Inhibitor proteins	13
Transcription regulators	47
Nucleic acid binding proteins	15
Protein binding	15
NTP binding proteins	15
Membrane protein	36
Transporter	48
Retrotransposon related protein	61
Ribosomal proteins	61
Enzymes	410
Total =	859

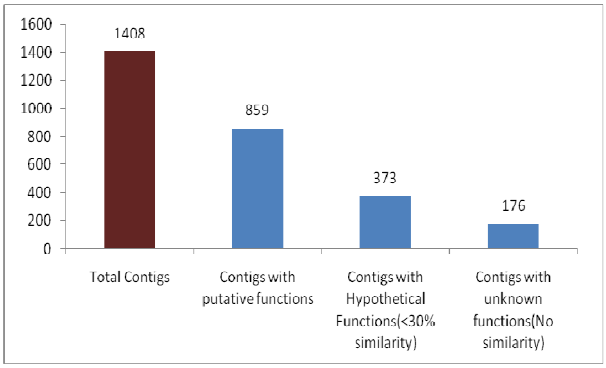


Fig 1. Classification of *P. somniferum* assembled contigs after functional annotation through BLASTX.

hydrolase activity (6.89%), transferase activity (9.23%), and transcription regulator activity (3.34%) and membrane-related proteins are 2.56% (Fig 1-3). Many other unidentified compounds are also present in *Papaver* species. Thus, genes related to secondary metabolism especially for the alkaloid biosynthesis, are studied further in order to decipher the molecular mechanism of natural variation of phytochemical components and facilitate the molecular breeding of medicinal cultivars. In the annotated EST contigs, 21 EST consensus sequences are related to secondary metabolic process, which include genes encoding key enzymes of the BIA biosynthetic pathway such as stylopine synthase

Table 2. Details of functionally annotated Benzylisoquinoline alkaloids pathway enzymes of *P. somniferum* with their contig number, enzyme classification (EC) number and developed 3D protein structure homology models submitted to PMDB (Protein Model DataBase) database.

S.No.	Contig No.	Functional annotation (BIA pathway enzymes)	Plants with matching genes	EC No.	PDB ID	PMDB ID
1	contig275	chain A, salutaridine reductase	*P. somniferum*	1.1.1.248	3O26	-
2	contig308	salutaridine synthase	*P. somniferum*	1.14.21.4	-	PM0077913
3	contig860	(S)-adenosyl-L-methionine:norcoclaurine 6-O-methyltransferase	*P. somniferum, Thalictrum flavum subsp. Glaucum*	2.1.1.128	-	PM0077914
4	contig1112					
5	contig879	tyrosine decarboxylase	*P. somniferum*	4.1.1.25, 4.1.1.28	-	PM0077915
6	contig924	cheilanthifoline synthase	*P. somniferum*	1.14.21.2	-	PM0077916
7	contig1059	polyphenol oxidase	*Nelumbo nucifera, Populus tremuloides*	1.10.3.1, 1.14.18.1	-	PM0077917
8	contig1145					
9	contig1208					
10	contig1339					
11	contig1130	reticuline oxidase	*P. somniferum, Ricinus communis*	1.21.3.3	-	PM0077919
12	contig1375					
13	contig1151	aspartate aminotransferase, cytoplasmic	*Daucus carota*	2.6.1.1	-	PM0077920
14	contig1164	(S)-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase 1 (Q7XB11)	*P. somniferum*	2.1.1.116	-	PM0077926
15	contig1183	(S)-adenosyl-L-methionine:coclaurine N-methyltransferase	*P. somniferum*	2.1.1.140	-	PM0077921
16	contig1268	salutaridinol 7-O-acetyltransferase	*P. somniferum*	2.3.1.150	-	PM0077922
17	contig1295	(S)-norcoclaurine synthase 1	*P. somniferum*	4.2.1.78	-	PM0077923
18	contig1304	stylopine synthase	*P. somniferum*	1.14.21.1	-	PM0077924
19	contig1384	NADPH-dependent codeinone reductase-like protein	*P. nudicaule*	1.1.1.247	-	PM0077925
20	contig1385	(S)-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase 2 (Q7XB10)	*P.somniferum*	2.1.1.116	-	PM0077927
21	contig1387	(S)-N-methylcoclaurine 3'-hydroxylase	*P.somniferum*	1.14.13.71	-	PM0077928

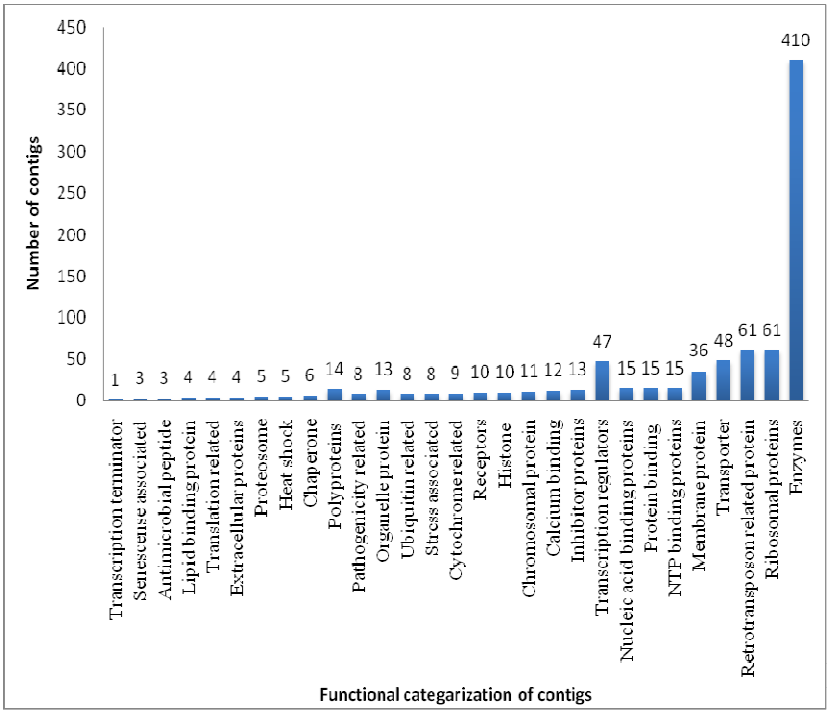


Fig 2. Functional categorization of *P. somniferum* contigs on the basis of their biological roles.

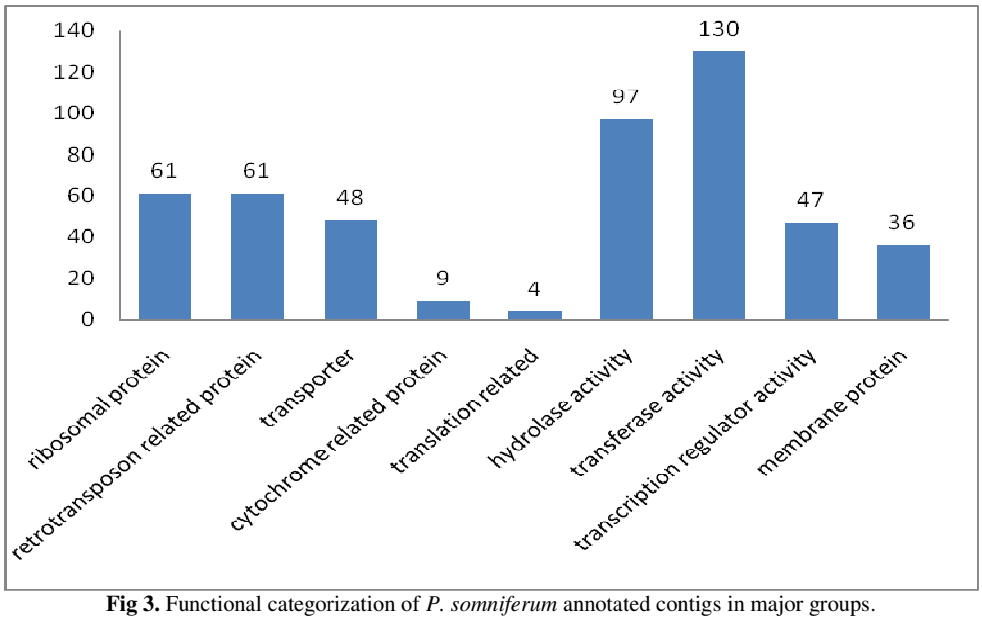


Fig 3. Functional categorization of *P. somniferum* annotated contigs in major groups.

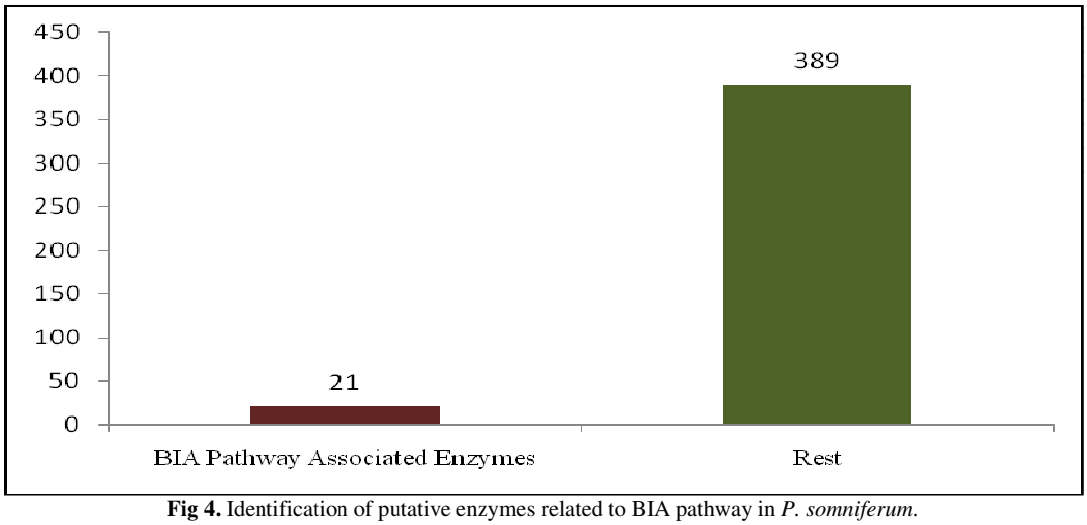


Fig 4. Identification of putative enzymes related to BIA pathway in *P. somniferum*.

(Desgagné-Penix et al., 2010), (S)-norcoclaurine synthase (Liscombe et al., 2005), salutaridine synthase (Gesell et al., 2009), salutaridine reductase (Ziegler et al., 2006), (S)-coclaurine N-methyltransferase, S-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase and salutaridinol 7-O-acetyltransferase (Samanani et al., 2006) (Fig 4, Table 2). Comparison of presently available reference known genes of BIA pathway in *P. somniferum* at KEGG database (http://www.genome.jp/kegg/) and unigenes predicted through ESTs assembly (contigs) in our study are shown in Fig. 5. Experientally identified enzymes represented in standard Enzymes Classification (EC) numbers inside the box, while circle mark represents the pathway metabolites. EC numbers indicated in black were detected in our study, while those indicated in red were not detected through present EST dataset of *P. somniferum* at dbEST (NCBI, USA; www.ncbi.nlm.nih.gov/dbEST/) database. Only contig-275 for chain A, Salutaridine reductase of BIA pathway showed cystal structure data in RCSB Protein DataBank (PDB; www.rcsb.org/) i.e., PDB ID: 3O26, rest predicted contigs (unigenes) were modeled based on homology method and submitted to PMDB (Protein Model DabaBase;http://mi.caspur.it/PMDB/) (PDB; www.rcsb.org/). A diagrammatic representation of BIA pathway showing metabolite-enzyme reaction series from (R)-Reticuline to Morphine with predicted unigenes protein 3D structure models and 2D chemical structures of metabolites are shown in Fig. 6. Similarly, structural representation of BIA pathway showing metabolite-enzyme reaction series from (S)-Reticuline to (S)-Stylopine with predicted unigenes 3D protein structure models and 2D chemical structures of metabolites are shown in Fig. 7. In addition, more than 47 EST consensus sequences (contigs) were annotated relating to transcription regulator activity (Table 1), which seems to play important roles in regulating BIA biosynthesis. For example, regulatory genes are annotated as members of GntR family (Rigali et al., 2002), MYB family (Yanhui et al., 2006), and AraC family (Gallegos et al., 1997) that can form a triplex compound to regulate the specific expression patterns of alkaloids structural genes. Some of these structural and regulatory genes have been already cloned and reported and their functional studies are being carried out

(Liscombe and Facchini, 2008; Guirimand et al., 2010). Further studies such as molecular modification (or gene transformation) and metabolic engineering of enzymes are also in progress (Hawkins and Smolke 2008; Ziegler and Facchini, 2008).

Materials and Methods

Retrieval of EST sequences

A total of 20,532 ESTs of *P. somniferum* were retrieved from dbEST database of NCBI, USA (www.ncbi.nlm.nih.gov/dbEST). Complete cDNA sequences were utilized where full-length cDNA sequences were available for mining purpose. Thus, all entries in the GenBank database that belong to EST or cDNA categories were included in the training datasets.

EST sequence cleaning

EST sequences retrieved from dbEST were cleaned, clustered and assembled through PHRED (Ewing and Green, 1998), RepeatMasker, Cross-Match, PHRAP and CAP3 (Huang and Madan, 1999) softwares with the help of Gencheck software (Ocimum Biosolutions, USA; www.ocimumbio.com) and compared with the assembly results of EGassembler (Masoudi-Nejad et al., 2006). After sequence cleaning 20,445 (709 trimmed) valid EST sequences and 87 trashed sequences out of total 20,532 ESTs were resulted at 96% minimum identity parameter for an alignment with a contaminant.

Repeat masking filter

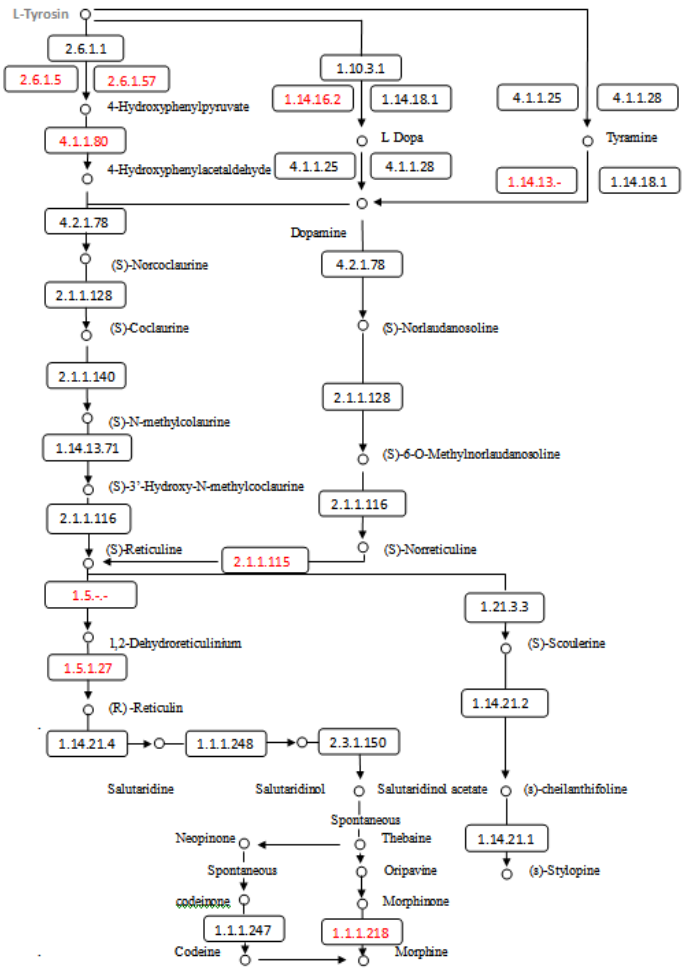
Repeat masking filter study showed length of valid sequences 1,69,60,151 bp, GC level 41.12% and masked bases 5,01,768 bp (2.96%) by using RepBase (update 8.12) repeats library of *Arabidopsis*. Elements observed in EST sequences were 1017 retroelements, 55 LINEs, 55 L1/CIN4, 962 LTR elements, 412 Ty1/Copia, 550 Gypsy/DIRS1, 56 DNA transposons, 15 hobo-Activator, 7 Tc1-IS630-Pogo, 7 En-Spm, 11 MuDR-IS905, 6 Tourist/Harbinger, 1 satellite, 920 simple repeats and 1697 low complexity elements.

Vector and Organelle sequence masking

NCBI's vector library UniVec (core) was used for masking vector sequences. Similarly, the plastids library of *Arabidopsis thaliana* was used for masking sequences from organelle origin.

Clustering and assembly of ESTs

All the 20,532 redundant ESTs of *P. somniferum* retrieved from NCBI were used to produce the non-redundant dataset and for clustering and assembly analysis. It was done through GenChek software (Ocimum Biosolutions, USA) and then cross checked through the results of EGassembler server (Masoudi-Nejad et al., 2006). An automated trimming and screening for various contaminants, low quality and low-complexity sequences was done. The masking of DNA sequences for repetitive elements including small RNA pseudo genes, LINEs, SINEs, LTR elements, vector sequences, organelle and other interspersed repeats was carried out. The above softwares clustered and assembled the sequences into contigs and singletons using CAP3 (Huang



Note: Aspartate aminotransferase, cytoplasmic = 2.6.1.1, tyrosine aminotransferase = 2.6.1.5, aromatic-amino-acid transaminase = 2.6.1.57, 4-hydroxyphenylpyruvate decarboxylase = 4.1.1.80, (S)-norcoclaurine synthase 1 = 4.2.1.78, (S)-adenosyl-L-methionine:norcoclaurine 6-O-methyltransferase = 2.1.1.128, (S)-adenosyl-L-methionine:coclaurine N-methyltransferase = 2.1.1.140, (S)-N-methylcoclaurine 3'-hydroxylase = 1.14.13.71, (S)-adenosyl-L-methionine:3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase = 2.1.1.116, Oxidoreductases = 1.5.-.-, 1,2-dehydroreticulinium reductase (NADPH) = 1.5.1.27, salutaridine synthase = 1.14.21.4, salutaridine reductase = 1.1.1.248, salutaridinol 7-O-acetyltransferase = 2.3.1.150, NADPH-dependent codeinone reductase-like protein = 1.1.1.247, morphine 6-dehydrogenase = 1.1.1.218, reticuline oxidase = 1.21.3.3, cheilanthifoline synthase = 1.14.21.2, stylopine synthase = 1.14.21.1, polyphenol oxidase = 1.10.3.1, 1.14.18.1, tyrosine 3-monooxygenase = 1.14.16.2, tyrosine decarboxylase = 4.1.1.25, 4.1.1.28, (S)-norcoclaurine synthase 1 = 4.2.1.78, (S)-adenosyl-L-methionine:norcoclaurine 6-O-methyltransferase = 2.1.1.128, (RS)-1-benzyl-1,2,3,4-tetrahydroisoquinoline N-methyltransferase = 2.1.1.115 and Oxidoreductases = 1.14.13.-.

Fig 5. Comparison of present status of known genes of Benzylisoquinoline alkaloids (BIA) pathway in *P. somniferum* and unigenes identified through ESTs assembly in our study. Identified enzymes represented in standard Enzymes classification (EC) numbers inside the box, while circle represents metabolites. EC no. indicated in black were identified in our study while those indicated in red were not detected in the present EST dataset of *P. somniferum* at dbEST database.

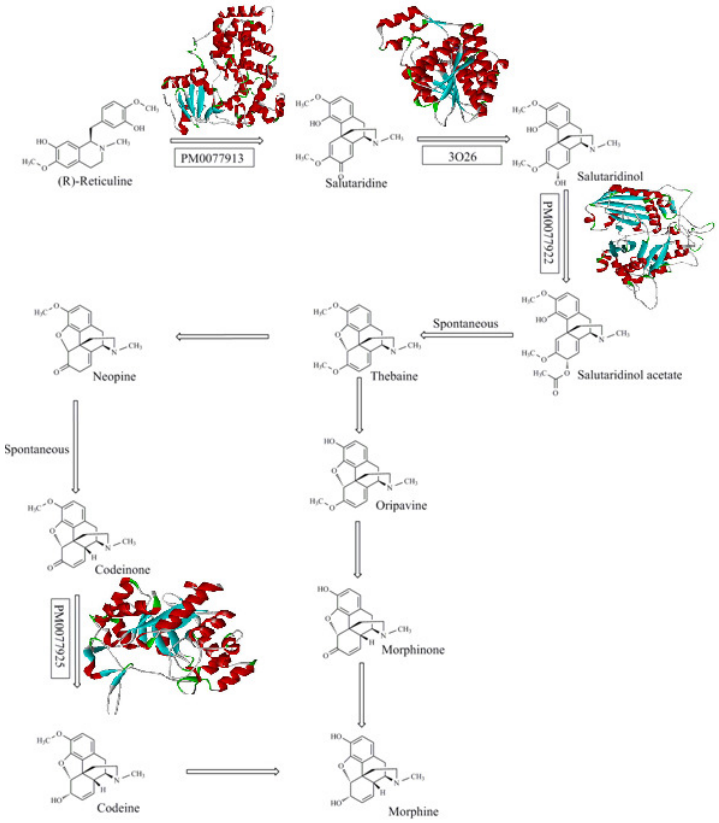


Fig 6. Structural flowchart of *P. somniferum* BIA pathway showing metabolite-enzyme reaction series from (R)-Reticuline to Morphine with identified unigenes 3D protein structure models and 2D chemical structures of metabolites.

and Madan, 1999) with the criterion of 80% overlap identity between the assembled reads. After trimming low-quality (through PHRED) and vector sequences and removing contaminant sequences, the resulting data set contained high-quality, non-redundant ESTs. The 20,532 ESTs were assembled into 1,408 contigs and 13,871 singletons using the PHRAP and CAP3 programs. After the assembly, redundant data sets of 20,532 sequences were reduced to only 15,279 non-redundant sequences. A total of 15,279 assembled unique transcripts were identified. The non-redundant dataset of 1,408 contigs was further used for functional annotation based on conserved domain, motif and gene ontology.

Functional annotation of contig sequences

Gene ontology based functional annotation of contig sequences was performed through BLASTX using UniProtKB/Swiss-Prot database. BLAST hits were selected which met the following criteria: E-value $<1e^{-4}$, and similarity >80%. The most significant matches were considered. However, gene ontology descriptions were assigned to contig sequences (unigenes) on the basis of Protein knowledgebase (UniProtKB: Swiss-Prot, TrEMBL) (http://www.uniprot.org/) protein sequence matches.

Homology Modeling

The three-dimensional (3D) protein structures of identified unignes of BIA pathway were modeled through Swiss-

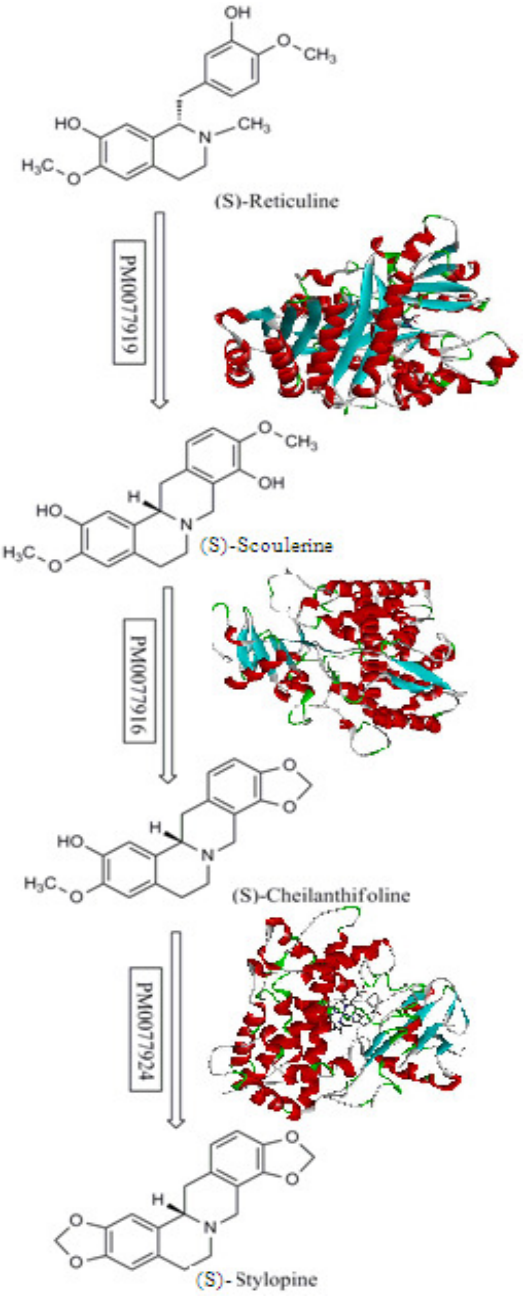


Fig 7. Structural flowchart of *P. somniferum* BIA pathway showing metabolite-enzyme reaction series from (S)-Reticuline to (S)-Stylopine with identified unigenes 3D protein structure models and 2D structures of metabolites.

Model, an automated protein homology modeling server (http://swissmodel.expasy.org).

Conclusion

A large-scale assembled EST dataset with 15,279 consensuses sequences derived from *P. somniferum* is reported in this study. A total of 1,408 assembled contigs have been successfully annotated based on the known sequences, and a fraction of these unique sequences are

predicted to be involved in the alkaloid metabolic pathway. This information would facilitate deciphering the molecular mechanism of secondary metabolism in *Papaver* species. In the studied work a large scale ESTs have been analyzed; i) a total of 20,532 high-quality *Papaver* EST sequences were retrieved and these ESTs were assembled into 15,279 putative unique transcripts, ii) of the 15,279 assembled sequences, 5.62% were assigned a putative function based on sequence identity in public databases, iii) annotation of the assembled 1,408 contig sequences showed categorization of 4.69% *Papaver* genes with pathogenicity related, ubiquitin related, stress associated, calcium binding, nucleic acid binding and nucleotide binding. The EST dataset developed by this effort will provide a fundamental basic resource for understanding secondary metabolism in *P. somniferum* and its both genetic and crop improvement.

Acknowledgements

We acknowledge the Council of Scientific & Industrial Research, New Delhi, for financial support at Central Institute of Medicinal & Aromatic Plants, Lucknow, India.

References

Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci AM, Da Silva C, Cascardo J, Allegre M, Kuhn D, Verica J, Courtois B, Loor G, Babin R, Sounigo O, Ducamp M, Guiltinan MJ, Ruiz M, Alemanno L, Machado R, Phillips W, Schnell R, Gilmour M, Rosenquist E, Butler D, Maximova S, Lanaud C (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. BMC Genomics 9:512.

Desgagné-Penix I, Khan MF, Schriemer DC, Cram D, Nowak J, Facchini PJ (2010) Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. BMC Plant Biol 10:252.

Ewing B, Green P (1998) Base-Calling of automated sequencer traces using Phred. II. Error Probabilities. Genome Res. 8 (3):186-94.

Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL (1997) AraC/XylS family of transcriptional regulaors. Microbiol Mol Biol Rev 61 (4):393-410.

Gesell A, Rolf M, Ziegler J, Chavez MLD, Huang FC, Kutchan TM (2009) CYP719B1 is salutaridine synthase, the C-C phenol-coupling enzyme of morphine biosynthesis in opium poppy. J Biol Chem 284:24432.

Guirimand G, Courdavault V, St-Pierre B, Burlat V (2010) Biosynthesis and Regulation of Alkaloids. Plant Developmental Biology - Biotechnological Perspectives 2:139-160.

Gupta V, Raghuvanshi S, Gupta A, Saini N, Gaur A, Khan MS, Gupta RS, Singh J, Duttamajumder SK, Srivastava S, Suman A, Khurana JP, Kapur R, Tyagi AK (2010) The water-deficit stress- and red rot-related genes in sugarcane. Funct Integr Genomics 10 (2):207-14.

Hawkins KM, Smolke CD (2008) Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. Nat Chem Biol 4(9):564–573.

Höfte H, Desprez T, Amselem J, Chiapello H, Caboche M, Moisan A, Jourjon MF, Charpenteau JL, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu DY, Mache R, Raynal M, Cooke R, Grellet F, Delseny M, Parmentier Y, Marcillac GD, Gigot C, Fleck J, Philipps G, Axelos M, Bardet C, Tremousaygue D, Lescure B (1993) An inventry of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. Plant J 4 (6):1051-61.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9 (9):868-77.

Kim HJ, Baek KH, Lee SW, Kim J, Lee BW, Cho HS, Kim WT, Choi D, Hur CG (2008) Pepper EST database: comprehensive *in silico* tool for analyzing the chili pepper (*Capsicum annuum*) transcriptome. BMC Plant Biol 8:101.

Kumar A, Sharma M (2011) Wheat genome phylogeny and improvement. Aust J Crop Sci. 5 (9):1120-1126.

Lin C, Mueller LA, Mc Carthy J, Crouzillat D, Pétiard V, Tanksley SD (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor Appl Genet 112 (1):114-130.

Liscombe DK, Facchini PJ (2008) Evolutionary and cellular webs in benzylisoquinoline alkaloid biosynthesis. Curr Opin Biotechnol. 19 (2):173–180.

Liscombe DK, MacLeod BP, Loukanina N, Nandi OI, Facchini PJ (2005) Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. Phytochemistry 66:1374.

Luo H, Li Y, Sun C, Wu Q, Song J, Sun Y, Steinmetz A, Chen S (2010) Comparison of 454-ESTs from *Huperzia serrata* and *Phlegmariurus carinatus* reveals putative genes involved in lycopodium alkaloid biosynthesis and developmental regulation. BMC Plant Biol 10:209.

Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S (2006) EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res 34 (Web Server issue):W459-62.

Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EH, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, Laing WA, McArtney S, Nain B, Ross GS, Snowden KC, Souleyre EJ, Walton EF, Yauk YK (2006) Analyses of expressed sequence tags from apple. Plant Physiol 141(1):147-66.

Pienkny S, Brandt W, Schmidt J, Kramell R, Ziegler J (2009) Functional characterization of a novel benzylisoquinoline O-methyltransferase suggests its involvement in papaverine biosynthesis in opium poppy (*Papaver somniferum* L). Plant J 60:56.

Ralph SG, Chun HJ, Kolosova N, Cooper D, Oddy C, Ritland CE, Kirkpatrick R, Moore R, Barber S, Holt RA, Jones SJ, Marra MA, Douglas CJ, Ritland K, Bohlmann J (2008) A conifer genomics resource of 200,000 spruce (Picea spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). BMC Genomics 9:484.

Rigali S, Derouaux A, Giannotta F, Dusart J (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J Biol Chem 277:12507.

Samanani N, Alcantara J, Bourgault R, Zulak KG, Facchini PJ (2006) The role of phloem sieve elements and laticifers in the biosynthesis and accumulation of alkaloids in opium poppy. Plant J 47:547.

Wang M, Zhang X, Zhao J, Song W, Zheng Y (2011) Evaluation of the genetic diversity and genome-wide linkage disequilibrium of Chinese maize inbred lines. Aust J Crop Sci. 5 (13):1790-1795.

Wang Z, Taramino G, Yang D, Liu G, Tingey SV, Miao GH, Wang GL (2001) Rice ESTs with disease-resistant gene- or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol Genet Genomics 265:302.

Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35 (1-2):135-44.

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q, (2006) The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparision with the rice MYB family. Plant Mol Biol 60 (1):107-24.

Ziegler J, Facchini PJ (2008) Alkaloid Biosynthesis: Metabolism and Trafficking. Annu Rev Plant Biol. 59:735–69.

Ziegler J, Voigtländer S, Schmidt J, Kramell R, Miersch O, Ammer C, Gesell A, Kutchan TM (2006) Comparative transcript and alkaloid profiling in *Papaver* species identifies a short chain dehydrogenase/reductase involved in morphine biosynthesis. Plant J 48:177.