Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

TRANSCRIPTOME ANALYSIS OF *SOLANUM VIRGINIANUM* AND *IN SILICO* PREDICTION OF ANTIMICROBIAL PEPTIDES

Megha Gowri Thippeswamy[1], Ravikumar Hemagirigowda[2], Rajeshwara Achur[1], Nagaraju Shivaiah[3*]

1. ***Department of Biochemistry, Kuvempu University, Shivamogga, Indian.***
2. ***Department of Life Sciences, Jnana Bharathi Campus, Bangalore University, Bangalore, India.***
3. ***Department of Studies and Research in Biochemistry, Tumkur University, Tumakuru, India.***

ARTICLE INFO

Received:
29 Nov 2022
Received in revised form:
20 Feb 2023
Accepted:
25 Feb 2023
Available online:
28 Apr 2023

Keywords: *Solanum virginianum*, De *novo* transcriptome analysis, Coding sequence (CDS) prediction, Antimicrobial peptides.

ABSTRACT

Solanum virginianum commonly known as wild eggplant or nightshade plant is a prickly herb that grows throughout Asia including India and Australia. *S. virginianum*, a member of the *Solanaceae* family is used by traditional medicinal practitioners to treat different ailments. Several studies have done to scientifically evaluate the potential pharmacological properties of the plant. However, the lack of genetic data on *S. virginianum* restricts its future research, particularly at the molecular level. The current study aims at transcriptome analysis of the *S. virginianum* fruit. 18.19 million high-quality reads were obtained. Afterthe de *novo* transcriptome analysis, 1.4 million unigenes and 60,487 coding sequences were found using Transcoder v5.3.0. 200 maximal length CDS transcripts were translated to protein using the Expasy translate server. Bioactive peptides were identified by different in silico approaches which revealed 58 antimicrobial peptides. All identified peptides were non-toxic. Among the 58 bioactive peptides, 19 are defensins. Four bioactive peptides SVBP1-CITGTTKTFYVN, SVBP-YGKNIVNRGRPRCS, SVBP3-KKCVCGSPRCRGYIGG, and SVBP4-FKIFGCICYAHV have been synthesized, evaluated for hemolytic activity and molecular docking study have been done to evaluate its antimicrobial activity. The identified new bioactive peptides could potentially be used in the next research on antibacterial, anti-inflammatory, and anti-cancer agents.

To Cite This Article: Thippeswamy MG, Hemagirigowda R, Achur R, Shivaiah N. Transcriptome Analysis of *Solanum Virginianum* and *in Silico* Prediction of Antimicrobial Peptides. Pharmacophore. 2023;14(2):1-10. https://doi.org/10.51847/zS9KWm2dCK

Introduction

Solanum virginianum, a widespread and very prickly undershrub belongs to the Solanaceae family. It is widely dispersed across India and is a common growing plant in many sandy soils throughout the globe. In Kannada, it is known as Nelagulla, in Sanskrit, it is known as Kantakari, and in English, yellow-berried nightshade. It is one of the members of The Dashamula of Ayurveda. The immature fruits are glabrous, spherical berries with green and white lines, while the matured fruits appear yellow., the leaves may be up to 10 cm long. Long pickles and a somewhat uneven base are features of the petioles. Flowers are in few-flowered cymes. Oval or lanceolate lobes are seen on a thorny calyx. Approximately 2 cm in diameter, the corolla is violet in hue. Yellow or white with green spots, globose, 2 cm in diameter, and berry [1, 2].

Numerous phytochemicals from various portions of the plant, including phenolics, flavonoids, alkaloids, amino acids, sterols, glycosides, saponins, tannins, and fatty acids, have been identified. Numerous medical systems, including Ayurveda, make substantial use of this herb. The herb has been used to treat sterility in females, leukoderma, scorpion bites, asthma, and chest discomfort. The seed oil has been used to treat arthritis and a significant decrease in arthritis has been observed. Toothaches may be relieved by using fruit ash [3]. According to studies, the plant has exhibited pharmacological activities such as antibacterial, phytotoxic, antioxidant, hemolytic, anthelmintic, anti-inflammatory, cytotoxic, antidiabetic, immunostimulatory, and hepatoprotective activities [4, 5].

These biological activities and their molecular characterization have been characterized by using bioinformatics tools.

Corresponding Author: Nagaraju Shivaiah; Department of Studies and Research in Biochemistry, Tumkur University, Tumakuru, Indian. E-mail: nagarajubiochem@gmail.com.

Bioinformatics, empirical research, and an integrated methodology are used to find bioactive peptides. The bioinformatic technique provides the information necessary to ascertain if peptides are present in the protein [6]. There are many different bioactivities that bioactive peptides possess, and they have been found in both dietary and non-dietary sources. These peptides are effective drugs due to their great selectivity, stability, bioavailability, effectiveness, safety, and tolerance. Genetic or recombination libraries may be employed as an alternative source of bioactive peptides [7].

The recent technique of next-generation sequencing (NGS) technologies and bioinformatics gives high-throughput molecular data for comprehending the full genomes and transcriptome profile of any organism. The advancement of genomic technologies has aided in the identification of genes that can be used to modify medicinal plants to produce higher-quality physiologically active phytocompounds [8]. Furthermore, transcriptome analysis is now a crucial part of almost all genomic investigations of disease and biological processes due to the ease of genome-wide profiling with sequencing technologies. The transcriptome, however, contains a variety of non-coding RNAs (ncRNAs), including messenger or coding RNAs. As a result, particular library preparation techniques are required, as well as appropriate bioinformatics algorithms for data processing and quantification for functional analysis [9]. RNA-Seq analysis and transcriptome assembly for blackberry (*Rubus sp. Var. Lochness*) fruit showed the new functional genes in Rubus sp [10].

The identification of peptide domains, motifs, and active sites in proteins has been accomplished using bioinformatics methods. Therefore, next-generation sequencing (NGS) is a promising way to identify new AMPs [11]. In a wide range of invertebrate, animal, and plant species, various tissues and cell types generate antimicrobial peptides (AMPs). They may connect to and enter membrane bilayers because of their cationic charge, amino acid makeup, size, and amphipathicity. Encoded within the sequences of natural protein precursors, antimicrobial peptides are typically less than 10 kDa and may also be produced in vitro by enzymatic hydrolysis [6]. As therapeutic medicines for a variety of pathogenic microorganisms, AMPs provide a viable option [12]. There are now more than 140 peptide therapies being tested in clinical studies, and more than 60 peptide medications have been released into the market [13]. Przybylski *et al*. (2016) discovered a haemoglobin fragment 137-141. It is a tiny hydrophilic antimicrobial peptide that can also be used as a meat preservative, lowering lipid oxidation by around 60% and delaying meat rancidity. Additionally, for 14 days while being refrigerated, the peptide 137-141 prevented microbial development. These antibacterial properties were comparable to those of BHT. The cationic antimicrobial peptide family known as defensins is effective against a wide variety of infectious microorganisms, including bacteria, viruses, and fungi. Defensins also serve significant roles as innate effectors and immune modulators in the immunological regulation of microbial infection. Plant defensins are a class of short cationic peptides rich in disulfides that have a wide range of antibacterial properties. From the transcriptome of the plants, numerous antimicrobials or defensin peptide was effectively discovered and described. However, there are still many plants that have therapeutic benefits, but relatively little research has been conducted on them. Despite the importance of eggplants for medicinal value for millions of people, genomics studies in this group have been limited [2, 14].

In this study, we have isolated total RNA and library prepared for the *S. virginianum* fruit. The prepared library was analyzed for transcriptome sequencing, De Novo assembly, and in silico prediction of bioactive peptides. Further, the bioactive peptides were synthesized and evaluated for in silico antimicrobial activity using a molecular docking study, and hemolytic activity was conducted. This is the first report on the transcriptome analysis and identification of bioactive peptides from *S. virginianum* fruit.

Materials and Methods

Total RNA Isolation and Library Preparation

Total RNA was extracted from the fruit samples according to the manufacturer's instructions using ZR plant RNA Miniprep (ZYMO Research). Nanodrop was used to assess the quality and amount of the obtained RNA samples, followed by an Agilent Tape station employing high-sensitivity RNA Screentape. The RNA-Seq paired sequencing library was produced from the QC passed RNA sample using Illumina TruSeq Stranded mRNA sample Prep kit. Briefly, mRNA was isolated from the total RNA using Poly-T connected magnetic beads, followed by enzymatic fragmentation, 1st strand cDNA conversion using superscript II and Act-d mix to enhance RNA-dependent synthesis. The 1st strand cDNA was then synthesized to the second strand utilizing the second strand mix. The dscDNA was then purified using AMPure XP beads followed by A-tailing, adapter ligation, and then enriched by a limited no of PCR cycles. The PCR enriched library was evaluated using a 4200 Tape Station system (Agilent Technologies) utilizing high sensitivity D1000 Screen tape as per manufacturer recommendations.

Transcriptome Sequencing, De Novo Assembly

The sequencing raw data for the fruit sample was processed to extract high-quality concordant reads by removing adapters, ambiguous reads (reads with unknown nucleotides "N" more than 5 percent), and low-quality sequences (reads with more than 10 percent quality threshold (QV) 20 phred score) using Trimmomatic v0.38 [15]. Paired-end readings were utilized for de novo sample assembly. Fruit sample readings of good quality were assembled into transcripts using Trinity de novo assembler (version 2.8.4) and a kmer value of 25 [16]. The assembled transcripts were then clustered using CD-HIT-EST-4.6 to exclude the isoforms created during assembly [17]. Consequently, sequences can no longer be extended. These sequences are classified as unigenes and considered for further study. The above-mentioned unigenes were utilized to predict coding sequences using TransDecoder- v5.3.0 (https://github.com/TransDecoder). TransDecoder finds potential coding areas within the sequences of

unigenes [18].

Translation of Proteins

To find the maximum length range, CDS were sent to the BLAST2Go platform [19]. To translate proteins, transcripts with the longest length range were subjected to a translation by using the Expasy translate tool (https://web.expasy.org/translate) [20].

In Silico Prediction of Bioactive Peptides

Bioactive peptides were predicted by using a modified bioinformatics strategy. DRAMP (Data Repository of Antimicrobial Peptides) now has 22259 entries, with 5891 being general AMPs (including both natural and synthetic AMPs) [21]. 181 stapled antimicrobial peptides belonging to specific AMPs were included in the latest update. The AMPA (Antimicrobial Sequence Scanning System) algorithm generates an antimicrobial profile employing a sliding window system [22]. CAMPR3 was used for AMP prediction using four different algorithms [23]. Version 2 of the server uses a deep neural network to classify peptides as AMPs or Non-AMPs. The Antimicrobial Peptide (AMP) Scanner (https://www.dveltri.com/ascan/v2-ascan.html) [24]. was used to predict if a peptide sequence may be an AMP active against Gram-positive and Gram-negative bacteria.

Toxicity Prediction of Predicted AMPs by in Silico

Prediction of Peptides toxicity was performed by using webserver *ToxinPred (*http://crdd.osdd.net/raghava/toxinpred*),* a unique in silico method of its kind, which will be useful in predicting the toxicity of peptides/proteins [25].

Identification of Defensins Peptides

Predicted AMPs were analyzed for defensins by employing a server known as defpred (https://webs.iiitd.edu.in/raghava/defpred/predict.php)webserver is an attempt to establish a prediction technique for the identification and optimization of such defensins peptides [3].

Peptide Synthesis

Peptides based mainly on the defensins peptide prediction tools mentioned above were synthesized at Grey matter research foundation pvt ltd in Tamil Nadu, India, using solid-phase peptide synthesis methods. The peptides were then purified to >95% purity using high-performance liquid chromatography, and the purity was confirmed using mass spectrometry. The peptides were dissolved in acidified distilled water (0.01 percent acetic acid) and stored at -20° C until further use.

Hemolytic Assay

The peptides' hemolytic activity was determined by measuring the release of hemoglobin from human erythrocytes at 540 nm [26]. For the hemolytic assay, 20 µL of each peptide solution was mixed with 180 µL of a 2.5 % (v/v) suspension of human erythrocytes in phosphate-buffered saline (PBS). After 30 minutes of incubation at 37°C, 600 µL of PBS was added to each tube. The supernatant was removed after 3 minutes of centrifugation at 10,000 g, and the absorbance at 540 nm was measured. The results of at least three independent experiments, each carried out in triplicate, were used to make the assessments.

Molecular Docking Studies

Evidence on binding conformation, pattern, and affinity can be found in silico molecular docking studies of chemical drug compounds or bioactive peptides that work by interacting with receptors. The protein was obtained from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (http://www.rcsb.org/pdb) and assigned with proper three - dimension orientation. peptides were converted to a PDB file by using the PREPFOLD server [27]. The energy-minimized protein was then used as input for HPEPDOCK SERVER to carry out the docking simulations. Protein *Klebsiella pneumonia* (pdb id: *6CP9*) was obtained from the RCSB PDB they were used as receptor molecules [28]. Before analysis, water molecules and other unwanted residues were removed from all proteins, when necessary, using Discovery studio software. The sequences were then subjected to energy minimization by Swiss-PdbViewer v4.1.0. The docking algorithm provided with HPEPDOCK was used to search for the best-docked conformation between ligand and protein [29]. During the Docking process, a maximum of 15 conformers were considered for ligand. Discovery Studio software was used to deduce the 2D and 3D pretorial representation of the interaction between the peptides and receptors.

Results and Discussion

Transcriptome Sequencing and De Novo Assembly

Using NextSeq500 and 2X 150bp chemistry, 5.31 Gb of high-quality paired-end data was generated, yielding 18,184,076 PE reads and 5,314,463,360 bases. There were 1,60,162 transcripts obtained. During assembly, unigenes were eliminated using CD-HIT-EST-4.6, yielding 1, 40, and 200 unigenes. Transcoder-v5.3.0 software was used to predict coding sequences from unigenes. This yields a total of 60, 487 coding sequences (CDS) **(Tables 1 and 2)**.

Table 1. High-quality read statistics, Transcript summary, Unigenes summary, CDS Statistics

High quality read statistics	
No. of PE Reads	18,184,076

Number of bases	5,314,463,360
Transcript summary	
No. of Transcripts	1,60,162
Total transcript length (bp)	173,343,596
N50 (bp)	1,905
Maximum transcript length (bp)	15,780
Minimum transcript length (bp)	201
Mean transcript length (bp)	1,082
Unigenes summary	
No. of Unigenes	1,40,200
Total unigene length (bp)	136,839,529
N50 (bp)	1,743
Maximum unigene length (bp)	15,780
Minimum unigene length (bp)	201
Mean unigene length (bp)	976
CDS Statistics	
No. of CDS	60,487
Total CDS length (bp)	55,354,977
N50 (bp)	1,176
Maximum CDS length (bp)	15,273
Minimum CDS length (bp)	255
Mean CDS length (bp)	915

Table 2. Data Distribution Statistics

Description	Total no. of CDS	No. of CDS with Blast Hit	No. of CDS without Blast Hit
CDS	60,487	53,317	7,170

Translation of Protein

CDS were sent to the BLAST2Go platform to predict transcripts with a maximum length. Blast2GO is a bioinformatics platform for the functional analysis of genomic datasets. The length distribution statistics of the CDS are shown in **Table 3**. In this study, we have used the Expasy translate tool to translate transcripts with lengths ranging greater or equal to 5000 base pair CDS to find full-length ORF.

Table 3. Length distribution statistics of the CDS

Length range (bp) of CDSs	No. of CDS
CDS ≤ 500	20,463
500 ≤ CDS ≤ 1000	21,059
1000 ≤ CDS ≤ 2000	14,669
2000 ≤ CDS ≤ 3000	3,119
3000 ≤ CDS ≤ 4000	780
4000 ≤ CDS ≤ 5000	200
>= 5000	197

In silico Prediction of Bioactive Peptides

By adopting DRAMP IDs with E-values under 5, peptide predictions were made using DRAMP. A cutoff of 0.5 prediction probability score was used in AMPA to predict AMPs. To achieve peptides with a projected probability score of 0.5, a Random Forest-based prediction method was ultimately applied for the final predicted peptides. A peptide with a prediction probability >0.5 is considered an AMP by the AMP scanner algorithm. Therefore, peptides having a probability of more than 0.5 were considered. We assembled all the data for the final peptides and displayed those peptides that correctly predicted AMPs using all four databases and methodologies. As stated in **Table 4**, in this study, we have used AMPA as the strategy for anticipated peptides.

Table 4. Amino acids predicted stretch for AMP from the ORF of transcripts with lengths ranging from 4000 to 5000 base pair CDS, prediction of defensin peptides

Sl. No	Seq id	Seq	Score	Prediction
1	CDS_7295	VFRTRRKDIKTNWP	0.11	Non-Defensins
2	CDS_13093	LVTCRRTFKNLLV	0.02	Non-Defensins

3	CDS_15151	NNKQGKAHGVWRQRGS	0.95	Defensins
4	CDS_16378	AYNIHTYAVHYTLQ	0.89	Defensins
5	CDS_17473	CRRPKTRQTRHQRAS	0.08	Non-Defensins
6	CDS_19305	NIRIMPWGHQHRN	0.8	Defensins
7	CDS_20091	VVHRYIGRQTQVM	0.09	Non-Defensins
8	CDS_20804	VRSYVQSRGRARQT	0.14	Non-Defensins
9	CDS_21079	CITGTTKTFYVN	0.99	Defensins
10	CDS_21374	ITRHHHPRFLSKL	0	Non-Defensins
11	CDS_21704	KKKSSSRQKGGRNSG	0.31	Non-Defensins
12	CDS_22055	FRWTNTHQRSKG	0.62	Defensins
13	CDS_22900	YRMTLIARRQNSP	0.16	Non-Defensins
14	CDS_22961	KIAHHVNTSKICHVLS	0.66	Defensins
15	CDS_24635	CTITKFFSKTVAL	0.61	Defensins
16	CDS_24876	GTRCSVCFIVVAC	0.85	Defensins
17	CDS_24924	VKQIYRGVVFLY	0.07	Non-Defensins
18	CDS_25611	KLQPRGIWFLTVL	0	Non-Defensins
19	CDS_25613	GLRSGLRHRIYDS	0.02	Non-Defensins
20	CDS_26176	STRNVVGNVKIPLLF	0.05	Non-Defensins
21	CDS_26984	VTIKRANNLKQVM	0.03	Non-Defensins
22	CDS_27112	IYKLVKQLQTVS	0.02	Non-Defensins
23	CDS_29100	IHRVQGTVCVKVASII	0.1	Non-Defensins
24	CDS_29910	NKWRISCVHTQIL	0.06	Non-Defensins
25	CDS_30112	REIKQLKQLRGQ	0.02	Non-Defensins
26	CDS_32070	YAHHNKLLTIQVRCLP	0.06	Non-Defensins
27	CDS_32073	KKCVCGSPRCRGYIGG	0.97	Defensins
28	CDS_32581	KRLNVQKFHFGG	0.45	Non-Defensins
29	CDS_32867	SHKYALVHQRVH	0.03	Non-Defensins
30	CDS_32877	RVHFHWSKIHMG	0.03	Non-Defensins
31	CDS_33424	GRYTNLIGRVNINNKGS	0.83	Defensins
32	CDS_33515	VFYGQIIYVCFFVGQR	0.12	Non-Defensins
33	CDS_33723	LRVSRLRAMGVRMT	0.01	Non-Defensins
34	CDS_34029	YGKNIVNRGRPRCS	0.98	Defensins
35	CDS_34557	YLGTGCGKTHIA	0.68	Defensins
36	CDS_35542	HLKVLSSWKCGFLVG	0.01	Non-Defensins
37	CDS_35543	KTIRSKPSNKYS	0.98	Defensins
38	CDS_36903	KPRLTCWVLPKL	0.04	Non-Defensins
39	CDS_38476	IYGSLRMSVKIQLL	0.01	Non-Defensins
40	CDS_40195	RVKLEIYKTERK	0	Non-Defensins
41	CDS_40946	GRLQVQLSYSKVVTL	0.02	Non-Defensins
42	CDS_41570	FMRRWMRAHILLL	0	Non-Defensins
43	CDS_43309	FNLKNNYSGLKACHTHCHL	0.95	Defensins
44	CDS_43777	LFKLVVITVLVI	0.12	Non-Defensins
45	CDS_44135	YRRYKANVAVCKA	0.52	Defensins
46	CDS_44840	MSKLLHHLRLSY	0.01	Non-Defensins
47	CDS_46293	WKSHFRHSFLRNVRHVRNSSV	0.01	Non-Defensins
48	CDS_46391	GRNCFRIHQCIKAF	0.92	Defensins
49	CDS_46393	WKSHFRHSFLRNVRHVRNSSV	0.01	Non-Defensins
50	CDS_46506	VKRARVRMGRSA	0.02	Non-Defensins
51	CDS_47785	IVRRAVALGRYL	0.01	Non-Defensins
52	CDS_49566	VPKKPLTWHRTG	0.01	Non-Defensins
53	CDS_49569	KQRAATTKNIVPF	0.77	Defensins
54	CDS_49852	RHKCLSVIGKLMYFS	0.13	Non-Defensins
55	CDS_51057	EKRHKDYLKKSK	0.01	Non-Defensins
56	CDS_56439	RMRLVLGNRTFSQW	0.02	Non-Defensins
57	CDS_57030	FKIFGCICYAHV	0.95	Defensins
58	CDS_57422	YVKNVTPKGCFVILSRK	0.53	Defensins

Toxicity Prediction of Predicted AMPs by in Silico

The toxicity of the predicted AMPs was detected using ToxinPred. ToxinPred is a web server that can predict the toxicity or non-toxicity of the AMPs. the minimum mutations in peptides for increasing or decreasing their toxicity and the toxic areas of proteins. ToxinPred is a first-of-its-kind in silico approach for predicting the toxicity of peptides and proteins. In addition, it will be useful for developing the least toxic peptides and identifying toxic protein areas [25]. Predicted AMPs were subjected to toxicity prediction. As a result, Predicted AMPs are non-toxic as shown in **Table 4**.

Identification of Defensins Peptides

Defensins primarily belong to the Brassicaceae, Fabaceae, and Solanaceae families. Defensins are short (12–45 amino acids), extremely basic, and include 8–10 cysteines that are involved in disulfide bridges that serve to stabilize these molecules [30]. Defensin peptides prediction among the predicted AMPs was conducted to the predicted AMPs. As a result, 19 AMPs are Defensins peptides as shown in **Table 4**.

Experimental Validation of Bioactive Peptides

Two bioactive peptides with good activity were ultimately chosen based on the in-silico prediction. According to the results of the Ramachandran plot and secondary structure prediction [31], peptides were synthesized. Due to the effectiveness and affordability of peptide synthesis for the creation of bioactive peptides, we selected the α-helix regions based on secondary structure. The two synthesized peptides were studied for hemolytic activity, four peptides were devoid of activity up to a concentration of 1mg/ml **(Figure 1)**.

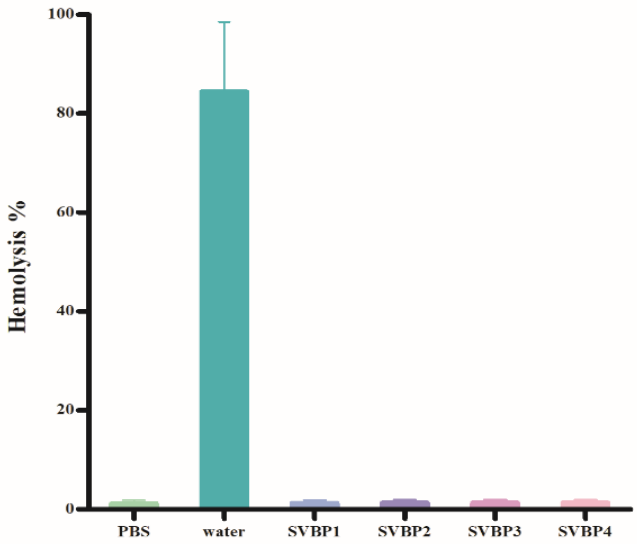

Figure 1. Hemolytic activity in human red blood cells. Data are the average of three independent experiments of water, PBS, SVBP1, SVBP2, SVBP3 and SVBP4. Error bars represent the standard deviations.

Molecular Docking Studies

The HPEPDOCK server was used to perform a molecular docking simulation between the *Klebsiella pneumonia* bacterial protein (pdb id: *6CP9*) and the eleven bioactive peptides. The best-weighted scores for *Klebsiella pneumonia* were SVBP2, SVBP3, and SVBP4 **(Table 5)**. The poses obtained from the HPEPDOCK server were analyzed for their receptor binding domain as well as the interacting bonds between the receptor and the ligand in Discovery studio **(Figure 2)**.

Table 5. Docking score and interactions for peptide-protein complexes of peptides and *Klebsiella pneumoniae*

Seq id	Bioactive peptides	Docking score(kcal/mol)
SVBP1	CITGTTKTFYVN	-162.739
SVBP2	YGKNIVNRGRPRCS	-171.109
SVBP3	KKCVCGSPRCRGYIGG	-173.049
SVBP4	FKIFGCICYAHV	-171.21
Synthetic AMP	MRFRRLRKKW RKRLKKI	-183.180

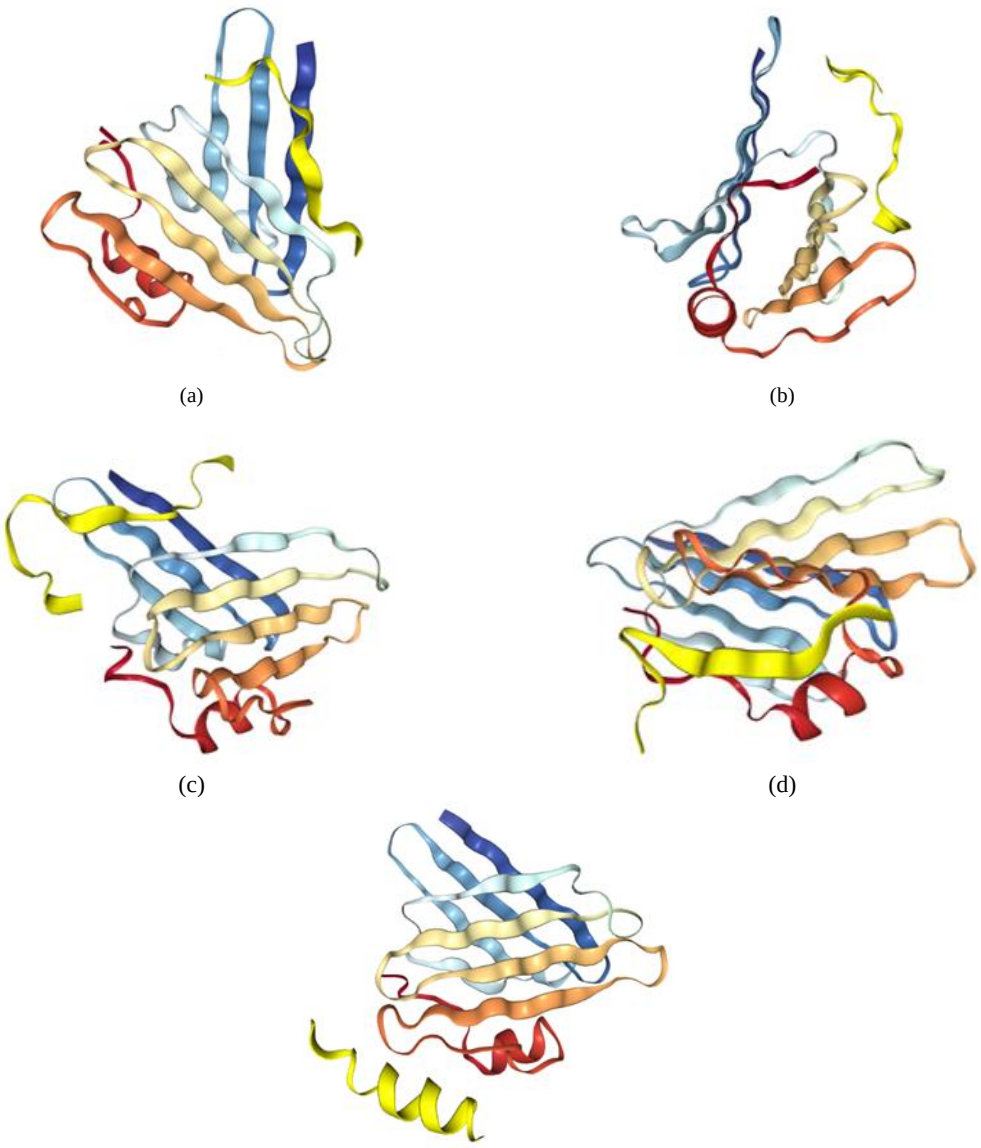

Figure 2. Docking poses of peptides (yellow) obtained using HPEPDOCK are compared with 3D structures (rainbow)

The transcriptome of *S. virginianum* is currently not available for analysis, so in this study, an attempt has been made to create a transcriptome of *S. virginianum* from a fruit sample [7]. A NextSeq500 and 2X 150bp chemistry was used to sequence the two distinct libraries, one for each species.

S. virginianum yielded a total of 5,314,463,360 bases and 18,184,076 PE reads. The assembled transcripts have lengths of roughly 5.31 Gb, with averages of 5000 bp. Gramazio *et al*., 2016 used Bwa, a very quick and memory-efficient mapper that excels at matching reads between 50 and 100 bp, to map the clean reads to the transcriptomes to verify the overall assembly quality. The assembled transcripts for *S. aethiopicum* and *S. incanum* have lengths of roughly 102 and 92 Mbp, respectively, with averages of 946 and 868 bp. The excellent caliber of Trinity assembly was validated by the vast amount of reads that were correctly mapped. The advancements in sequencing technologies, particularly Illumina, have led to steadily bettering assemblies in recent years [32-36].

Transcripts are distinguished from newly duplicated and recognized allelic variations using Trinity software, which identifies splice variants (isoforms) [37]. The RSEM program (RNA-Seq by Expectation-Maximization) was used to select just the highly expressed transcript from each locus' isoforms to create a collection of single-copy gene loci (unigene) [38]. *S. virginianum* had 60,487 unigenes in all, indicating that 22.5% of its transcripts were splice variants.

An understanding of the structural, biochemical, and functional aspects of assembled unigenes is provided by transcriptome annotation [39]. Additionally, the NCBI's non-redundant (NR) protein database was analyzed using BlastX [32] and the Blast2GO program [40] was used to assign GO words (Gene Ontology) and EC numbers (Enzyme Commission) to the proteins.

The Swiss-Prot database, which has been carefully reviewed, was used for the annotation of the vast majority of unigenes in the study by Gramazio *et al*. (2016). They reported 30,630 and 34,231 unigenes, which is comparable to the protein-coding genes reported number for the tomato [36]. (The Tomato Genome Consortium 2012), and other plant species in earlier research.

For example, "Watt [41] and 34,368 of 82,036 unigenes discovered in litchi (Litchi chinesis Sonn.) [42] were annotated in protein databases, as were 32,410 out of 68,132 unigenes in Oryza officinalis Wall and 24,003 out of 31,196 unigenes in the pepper transcriptome (Capsicum annum L.) [43]. Like this, from *S. torvum* and *S. melongena* 28,016 and 29,845 unigenes, were annotated" [44].

Biological processes accounted for the bulk of the GO words. Most of them had GO annotation levels between 4 and 10. Biological activities including protein phosphorylation, metabolic processes, oxidation-reduction, and transcriptional control are often unique to tissues at a stage of development. Molecular functions have been attributed to 30.7% and 35.4% of ontologies, with binding activities being the most prevalent and most of them displaying a GO annotation level of 3 to 9. A cellular component GO was present in the remaining 25.3% and 18.1% of annotated unigenes, primarily concerning the plasma membrane, nucleus, cytosol, mitochondria, and chloroplast. Apart from levels 5 and 8, the distribution of GO levels for this category is rather consistent [45].

Based on the in-silico prediction, four antimicrobial peptides which showed SVM based model achieved a maximum MCC of 0.96 with an AUC of 0.99. The synthesized peptides were evaluated following the findings of the Ramachandran plot and secondary structure prediction. The *Klebsiella pneumonia* bacterial protein (pdb id: *6CP9*) and the eleven bioactive peptides were molecularly docked using the HPEPDOCK server. The SVBP2, SVBP3 and SVBP4 values for Klebsiella pneumonia were the highest weighted scores. Additionally, 19 defensin peptides, mostly from the Brassicaceae, Fabaceae, and Solanaceae families, were also discovered.

These antimicrobial peptides or compounds may be generated naturally by the plant or because of an infection, and they may be poisonous to or inhibitive of bacteria, fungi, and/or pests [46, 47]. Plants have developed a complicated and elaborate array of defense mechanisms throughout their lengthy interaction with pathogens, including secondary metabolites, antifungal proteins, and pathogenesis-related proteins. The accessibility of chemicals produced from plants with antifungal properties strong enough to make them useful for agronomic application in disease management is still insufficient to meet the rising demands of the environment [48].

Conclusion

Incredible developments in plant genomics and transcriptomics provide innovative opportunities for understanding the molecular cascade and producing high-value bioactive compounds from medicinally important plants. To the best of our knowledge, this is the first time the transcriptome of *S. virginianum* from fruit samples has been done using Illumina sequencing technology. In addition, antimicrobial peptides for *Solanum virginianum* were generated by support vector machine tools. This is the first illustration of de *novo* sequencing and transcriptome analysis from the fruit of the *S. virginianum* plant.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Morankar P, Jain AP. Extraction, Qualitative and Quantitative Determination of Secondary Metabolites of Aerial Parts of Clematis heynei and Solanum virginianum. J Drug Delivery Ther. 2019; 9(1-s):260-4. doi:10.22270/jddt.v9i1-s.2346
2. Prashith Kekuda TR, Raghavendra HL, Rajesh MR, Avinash HC, Ankith GN, Karthik KN. Antimicrobial, insecticidal, and antiradical activity of Solanum virginianum l.(solanaceae). Asian J Pharm Clin Res. 2017;10(11):163-7. doi:10.22159/ajpcr.2017.v10i11.20180
3. Kaur J, Kumar V, Sharma K, Kaur S, Gat Y, Goyal A, et al. Opioid peptides: an overview of functional significance. Int J Pept Res Ther. 2020;26(1):33-41. doi:10.1007/s10989-019-09813-7
4. Shah MA, Khan H, Khan S, Muhammad N, Ullah Khan F, Shahnaz MA, et al. Cytotoxic, anti-oxidant and phytotoxic effect of Solanum surattense Burm F fruit extracts. Int J Pharmacogn Phytochem. 2013;28(2):1154-8.
5. Kumar SR, Hariprasanth RJ, Siddharth PM, Gobinath M, Rajukutty C. Evaluation of the antioxidant, antimicrobial, antidiabetic and hemolytic activity of organically grown Solanum nigrum and Solanum xanthocarpum. Int J Curr Pharm Rev Res. 2016;7(5):296-9.
6. Mustățea G, Ungureanu EL, Iorga E. Protein acidic hydrolysis for amino acids analysis in food - progress over time: A short review. J Hyg Eng Des. 2019;26:81-7.
7. Daliri EB, Lee BH, Oh DH. Current trends and perspectives of bioactive peptides. Crit Rev Food Sci Nutr. 2018;58(13):2273-84. doi:10.1080/10408398.2017.1319795
8. Kaur D, Patiyal S, Arora C, Singh R, Lodhi G, Raghava GPS. *In-Silico* Tool for Predicting, Scanning, and Designing Defensins. Front Immunol. 2021;12:780610. doi:10.3389/fimmu.2021.780610

9. Srivastava A, George J, Karuturi RKM. Transcriptome Analysis, In: Encyclopedia of bioinformatics and computational biology: abc of bioinformatics. Elsevier; 2018:792-805. doi:10.1016/b978-0-12-809633-8.20161-1
10. Garcia-Seco D, Zhang Y, Gutierrez-Mañero FJ, Martin C, Ramos-Solano B. RNA-Seq analysis and transcriptome assembly for blackberry (Rubus sp. Var. Lochness) fruit. BMC Genom. 2015;16(1):1-2. doi:10.1186/s12864-014-1198-1
11. Tavares LS, de Souza VC, Schmitz Nunes V, Nascimento Silva O, de Souza GT, Farinazzo Marques L, et al. Antimicrobial peptide selection from Lippia spp leaf transcriptomes. Peptides. 2020;129:170317. doi:10.1016/j.peptides.2020.170317
12. Cruz J, Ortiz C, Guzmán F, Fernández-Lafuente R, Torres R. Antimicrobial peptides: promising compounds against pathogenic microorganisms. Curr Med Chem. 2014;21(20):2299-321. doi:10.2174/0929867321666140217110155
13. Jois SD. Basic Concepts of Design of Peptide-Based Therapeutics. InPeptide Therapeutics 2022 (pp. 1-50). Springer, Cham.
14. Przybylski R, Firdaous L, Châtaigné G, Dhulster P, Nedjar N. Production of an antimicrobial peptide derived from slaughterhouse by-product and its potential application on meat as preservative. Food Chem. 2016;211:306-13. doi:10.1016/j.foodchem.2016.05.074
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi:10.1093/bioinformatics/btu170
16. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494-512. doi:10.1038/nprot.2013.084
17. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150-2. doi:10.1093/bioinformatics/bts565
18. Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. 2015;43(12):e78. doi:10.1093/nar/gkv227
19. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6. doi:10.1093/bioinformatics/bti610
20. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784-8. doi:10.1093/nar/gkg563
21. Shi G, Kang X, Dong F, Liu Y, Zhu N, Hu Y, et al. DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. Nucleic Acids Res. 2022;50(D1):D488-96. doi:10.1093/nar/gkab651
22. Torrent M, Di Tommaso P, Pulido D, Nogués MV, Notredame C, Boix E, et al. AMPA: an automated web server for prediction of protein antimicrobial regions. Bioinformatics. 2012;28(1):130-1. doi:10.1093/bioinformatics/btr604
23. Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. Nucleic Acids Res. 2016;44(D1):D1094-7. doi:10.1093/nar/gkv1051
24. Veltri DP. A Computational and Statistical Framework for Screening Novel Antimicrobial Peptides. 2015.
25. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R. Open Source Drug Discovery Consortium, et al. In silico approach for predicting toxicity of peptides and proteins. PLoS One. 2013;8(9):e73957. doi:10.1371/journal.pone.0073957
26. Lee JH, Chung H, Shin YP, Kim MA, Natarajan S, Veerappan K, et al. Deciphering Novel Antimicrobial Peptides from the Transcriptome of *Papilio xuthus*. Insects. 2020;11(11):1-10. doi:10.3390/insects11110776
27. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res. 2012;40(Web Server issue):W288-93. doi:10.1093/nar/gks419
28. Gouda H, Torigoe H, Saito A, Sato M, Arata Y, Shimada I. Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. Biochemistry. 1992;31(40):9665-72.
29. Zhou P, Jin B, Li H, Huang SY. HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. Nucleic Acids Res. 2018;46(W1):W443-50. doi:10.1093/nar/gky357
30. Meneguetti BT, Machado LD, Oshiro KG, Nogueira ML, Carvalho CM, Franco OL. Antimicrobial Peptides from Fruits and Their Potential Use as Biotechnological Tools-A Review and Outlook. Front Microbiol. 2017;7:2136. doi:10.3389/fmicb.2016.02136
31. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. Protein Sci. 2018;27(1):129-34. doi:10.1002/pro.3289
32. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. 2013. doi:10.48550/ARXIV.1303.3997
33. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418-26. doi:10.1016/j.tig.2014.07.001
34. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95(1):5-23. doi:10.1016/j.ajhg.2014.06.009

35. Faure D, Joly D. Next-generation sequencing as a powerful motor for advances in the biological and environmental sciences. Genetica. 2015;143(2):129-32. doi:10.1007/s10709-015-9831-8
36. Gramazio P, Blanca J, Ziarsolo P, Herraiz FJ, Plazas M, Prohens J, et al. Transcriptome analysis and molecular marker discovery in Solanum incanum and S. aethiopicum, two close relatives of the common eggplant (Solanum melongena) with interest for breeding. BMC Genom. 2016;17(1):300. doi:10.1186/s12864-016-2631-4
37. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-52. doi:10.1038/nbt.1883
38. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 2011;12(1):323. doi:10.1186/1471-2105-12-323
39. Mutz KO, Heilkenbrinker A, Lönne M, Walter JG, Stahl F. Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol. 2013;24(1):22-30. doi:10.1016/j.copbio.2012.09.004
40. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008;2008:1-12. doi:10.1155/2008/619832
41. Bao Y, Xu S, Jing X, Meng L, Qin Z. De novo assembly and characterization of Oryza officinalis leaf transcriptome by using RNA-seq. Biomed Res Int. 2015;2015:1-7. doi:10.1155/2015/982065
42. Lu X, Kim H, Zhong S, Chen H, Hu Z, Zhou B. De novo transcriptome assembly for rudimentary leaves in Litchi chinesis Sonn. and identification of differentially expressed genes in response to reactive oxygen species. BMC Genom. 2014;15(1):805. doi:10.1186/1471-2164-15-805
43. Ashrafi H, Hill T, Stoffel K, Kozik A, Yao J, Chin-Wo SR, et al. De novo assembly of the pepper transcriptome (Capsicum annuum): a benchmark for in silico discovery of SNPs, SSRs and candidate genes. BMC Genomics. 2012;13(1):571. doi:10.1186/1471-2164-13-571
44. Yang X, Cheng YF, Deng C, Ma Y, Wang ZW, Chen XH, et al. Comparative transcriptome analysis of eggplant (Solanum melongena L.) and turkey berry (Solanum torvum Sw.): phylogenomics and disease resistance analysis. BMC Genomics. 2014;15(1):412. doi:10.1186/1471-2164-15-412
45. Niederhuth CE, Patharkar OR, Walker JC. Transcriptional profiling of the Arabidopsis abscission mutant hae hsl2 by RNA-Seq. BMC Genom. 2013;14(1):37. doi:10.1186/1471-2164-14-37
46. Punja ZK. Genetic engineering of plants to enhance resistance to fungal pathogens—a review of progress and future prospects. Can J Plant Pathol. 2001;23(3):216-35. doi:10.1080/07060660109506935
47. Tarchevsky IA. Pathogen-induced plant proteins. Appl Biochem Microbiol. 2001;37(5):441-55. doi:10.1023/A:1010267704445
48. Balconi C, Stevanato P, Motto M, Biancardi E. Biancardi E. Breeding for biotic stress resistance/tolerance in plants. InCrop production for agricultural improvement 2012 (pp. 57-114). Springer, Dordrecht.