

**Prediction and Characterization of micro-RNAs
from *Agave sisalana* and Targeting of Pathogen
Responsive Genes Under Biotic Stress**



Basit Jabbar

National Centre of Excellence in Molecular Biology

University of The Punjab

Lahore, Pakistan

2022

Prediction and Characterization of micro-RNAs from *Agave sisalana* and Targeting of Pathogen Responsive Genes Under Biotic Stress

A Thesis Submitted to

UNIVERSITY OF THE PUNJAB

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

MOLECULAR BIOLOGY

By

Basit Jabbar

Supervisors

Prof. Dr. Bushra Rashid

Dr. Sameera Hassan

National Centre of Excellence in Molecular Biology

University of The Punjab

Lahore, Pakistan

2022

DEDICATED

to

my Parents

and

all Family Members

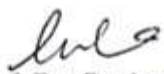


**CENTRE OF EXCELLENCE
IN MOLECULAR BIOLOGY
University of the Punjab**

Ministry of Federal Education
& Professional Training

SUPERVISORS DECLARATION

It is certified that the research work described in this thesis is the original work done by **Mr. Basit Jabbar** and has been carried out under our direct supervision. We have personally gone through all the data reported in the manuscript and certify their correctness/authenticity and found that the thesis has been written in pure academic language and are free from any typos and grammatical errors. It is also certified that the thesis has been prepared according to the prescribed format and we endorse its evaluation for the award of Ph.D degree through the official procedures of the University.


Prof. Dr. Bushra Rashid
(Thesis Supervisor)


Dr. Sameera Hassan
(Thesis Supervisor)

Certificate

It is to certify that the research work described in this thesis is the original work of the author Mr. Basit Jabbar and has been carried out under our direct supervision. We have personally gone through all the data reported in the manuscript and certify their correctness/authenticity and found that the thesis has been written in a pure academic language and is free from any typos and grammatical errors. It is further certified that the material included in this thesis has not been used in part or full in a manuscript already submitted or in the process of submission in partial/complete fulfillment of the award of any other degree from any other institution. It is also to certify that the thesis has been prepared under our supervision according to the prescribed format and we endorse its evaluation for the award of the Ph.D. degree through the official procedures of the University.

In accordance with the rules of the Centre data book #1519 is declared as an un-expendable document that will be kept in the registry of the Centre for a minimum of three years from the date of the thesis defense examination.

Signature: 

Name of Supervisor: Dr. Bushra Rashid
(Professor)

Signature: 

Name of Supervisor: Dr. Sameera Hassan
(Associate Professor)

Author's Declaration

I, Basit Jabbar, hereby state that my Ph.D. thesis titled, "Prediction and Characterization of micro-RNAs from *Agave sisalana* and Targeting of Pathogen Responsive Genes Under Biotic Stress" is my own work and has not been submitted previously by me for taking any other degree from this university, "University of the Punjab" or anywhere else in the country/world. At any time, if my statement is found to be incorrect even after my graduation, the university has the right to withdraw my Ph.D.degree.

Date: 30th September, 2022

Signature:  _____

Plagiarism Undertaking

I solemnly declare that research work presented in the thesis titled, "Prediction and Characterization of micro-RNAs from *Agave sisalana* and Targeting of Pathogen Responsive Genes Under Biotic Stress" is solely my research work with no significant contribution from any other person. Small contributions/help wherever taken, has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and university, "University of the Punjab" towards plagiarism. Therefore, I as an author of the above-titled thesis, declare that no portion of my thesis has been plagiarized and any material used as a reference is properly referred/ cited.

I undertake that if I am found guilty of any formal plagiarism in the above-titled thesis even after the award of the Ph.D. degree, the university reserves the rights to withdraw/ revoke my Ph.D. degree and that HEC and the university has the right to publish my name on the HEC/ university website on which names of students are placed who submitted plagiarized thesis.

Date: 30th September, 2022

Signature:  _____

Acknowledgements

“In the name of ALLAH, the Most Gracious and the Most Merciful”.

All the praises are for the Almighty “**ALLAH**”, the ultimate source of knowledge, WHO has given me the ability to understand and learn. Whosoever I was, I am or I may become is due to HIS blessings upon me. All praises to Holy Prophet “**MUHAMMAD**” (PBUH); the reason for the creation of whole universe, the source of knowledge and guidance for the whole mankind.

I would like to thank **my parents** and all family members for their love, prayers and moral support which gave me confidence and motivation. I am grateful to my Research Supervisors: **Prof. Dr. Bushra Rashid** and **Dr. Sameera Hassan** for the support, guidance, and knowledge that I gained from them. Their encouragement and momentum throughout the course of research has propelled me to achieve the research milestones. It gives me an immense pleasure to thank **Prof. Dr. Kausar Malik**, Director CEMB, for providing me the research facilities and the opportunity to perform research work in this prestigious research institution in Pakistan, and I also express gratitude to **Prof. Dr. Tayyab Husnain** and **Prof. Dr. Ahmad Ali Shahid**- the ex-Directors of CEMB.

I am thankful to the **Director, IAGS** (Institute of Agricultural Sciences), University of the Punjab for providing culture of *Aspergillus niger* from IAGS culture bank. I also owe gratitude to **Dr. Azra Mehmood** for sanctioning the use of microscopy and nanodrop facilities in the Stem Cells Lab at CEMB for this research work. Helping in access to the *Agave sisalana* transcriptome (RNA-seq) assembly by **Bilal Sarwar** is also acknowledged.

I would also like to express gratitude towards my seniors/juniors at CEMB including **Moon Sajid, Zarnab Ahmad, Bilal Sarwar, Fizza Zulfiqar, Iqra Maqsood, Muhammad Umair, Muhammad Tariq, Anicet Batcho, Atif Shahzad, Abdullah Jamil, Saba Nawaz and Hassan Zaman** who were friendly, cooperative, and lent their technical and moral support at times of need.

Basit Jabbar

Table of Contents

Supervisors Declaration	ii
Certificate.....	iii
Author's Declaration.....	iv
Plagiarism Undertaking.....	v
Acknowledgements.....	vi
Table of Contents	vii
List of Tables	xii
List of Figures	xiv
Abbreviations.....	xvii
ABSTRACT.....	xx
1. INTRODUCTION	1
2. REVIEW OF LITERATURE	6
2.1 Micro-RNAs	6
2.2 miRNA Biogenesis	7
2.3 Next-Generation Sequencing (NGS)	10
2.4 RNA Sequencing (RNA-Seq): A Method for Transcriptome Analysis.....	11
2.5 Applications of RNA-Seq.....	12
2.6 Approaches for miRNA Identification.....	13
2.7 miRNA Prediction from RNA-seq Data	14
2.8 Biotic Stress	16
2.9 Pathogen Responsive miRNAs under Biotic Stress.....	17
2.10 Universal Stress Proteins	20
2.11 Universal Stress Proteins (USPs) in Biotic and Abiotic Stresses.....	23
2.12 <i>Agave sisalana</i> and its Applications	26
2.13 Bole Rot Disease Caused by <i>Aspergillus</i> sp.	28

3.	MATERIALS AND METHODS.....	30
3.1	Retrieval of mature miRNA Dataset.....	30
3.2	RNA-Seq Data Resource	31
3.3	Standalone BLAST-N Analysis.....	31
3.4	Removal of Protein Coding Transcripts.....	32
3.5	Prediction of pre-miRNA Stem-Loop (Secondary) Structures	32
3.6	Selection of Stable Pre-miRNA Secondary Structure.....	32
3.7	Growth of <i>Agave sisalana</i> in Greenhouse.....	34
3.8	Tissue Sampling for RNA Extraction	34
3.9	Isolation of Total RNA from <i>A. sisalana</i>	34
3.10	Separation of Small and Large RNA Pool.....	35
3.11	Polyadenylation of Small RNA Pool	36
3.12	Precipitation of Poly(A) RNA.....	37
3.13	cDNA Preparation with Reverse Adaptor.....	37
3.14	Primers Designing for Poly(A)RT PCR.....	38
3.15	Poly(A) RT-PCR for Validation of Predicted miRNAs.....	41
3.16	Target Prediction of miRNAs Against TAIR.....	42
3.17	Gene Ontology Analysis of Predicted Target Genes	42
3.18	Prediction of Candidate Coding Regions from <i>A. sisalana</i> RNA-seq Assembly	42
3.19	Hidden Markov Models (HMM) Search for USP Family Genes.....	42
3.20	cDNA Preparation from Large RNA Pool.....	43
3.21	Primer Design for Amplification of USP Genes.....	43
3.22	Amplification of Universal Stress Protein Family Genes- PF00582	44
3.23	Agarose Gel Electrophoresis.....	45
3.24	Purification of PCR Product	45
3.25	Cloning with TA Cloning® Kit Dual Promoter (pCR®II)	46

3.26	Preparation of <i>Escherichia coli</i> TOP 10 Competent Cells.....	48
3.27	Transformation of Recombinant Vector in Competent <i>E. coli</i> TOP10.....	48
3.28	Blue-White Screening	49
3.29	Plasmid Extraction with GeneJET Plasmid Miniprep	49
3.30	Restriction Digestion with <i>EcoRI</i>	50
3.31	Sanger Sequencing Analysis	50
3.32	Target Prediction of miRNAs Against AsUSPs.....	50
3.32.1	RNAhybrid.....	51
3.32.2	psRNATarget	51
3.32.3	psRobot	51
3.33	miRNA 3D Structure Modelling (MC-Fold)	51
3.34	3D Structure Modelling of AsUSP Genes mRNA (RNA Composer)	52
3.35	miRNA-mRNA Molecular Docking Analysis	52
3.36	Multiple Sequence Alignment of pre-miRNAs and Phylogenetic Analysis.....	52
3.37	Biotic Stress Induction	53
3.37.1	<i>Aspergillus niger</i> Culture	53
3.37.2	Preparation of <i>Aspergillus niger</i> Suspension	53
3.37.3	<i>Aspergillus niger</i> Inoculation in <i>Agave sisalana</i>	53
3.38	Estimation of Chlorophyll Content	53
3.39	Estimation of Proline Content.....	54
3.40	Microscopic Examination of Leaf Tissues.....	55
3.41	Statistical Data Analyses.....	55
4.	RESULTS	56
4.1	Standalone BLAST-N Result.....	56
4.2	Pre-miRNA Secondary Structures	58
4.3	Characteristics of the predicted miRNA and pre-miRNA Candidates.....	63
4.4	Electrophoretic Analysis of Extracted RNA	65

4.5	Quantification and Quality Assessment of RNA via NanoDrop.....	65
4.6	Poly(A)RT PCR for Experimental Validation of miRNAs.....	66
4.7	Predicted Targets of <i>A. sisalana</i> miRNAs against TAIR v.10.....	71
4.8	Gene Ontology Analysis of Predicted Targets in TAIR v.10	71
4.8.1	GO Biological Processes.....	71
4.8.2	GO Molecular Function	72
4.8.3	GO Cellular Component	73
4.9	Targets of <i>A. sisalana</i> miRNAs among USP Genes in TAIR.....	74
4.10	<i>A. sisalana</i> USP Family Genes Predicted by TransDecoder and HMM3 Search	77
4.11	PCR Amplification of Predicted <i>A. sisalana</i> USP Family Genes	77
4.12	Ligation with pCR®II vector, Transformation in Competent <i>E. coli</i> TOP 10 and Blue White Screening	79
4.13	Confirmation of Recombinant Clones by Restriction with <i>EcoRI</i>	79
4.14	Sequencing Analysis and Submission to NCBI.....	80
4.15	miRNA Target Prediction Against AsUSP Genes	91
4.15.1	Predicted Target Sites with RNAhybrid	91
4.15.2	Predicted Target Sites with psRNATarget.....	92
4.15.3	Predicted Target Sites with psRobot.....	93
4.15.4	Consensus of miRNA Target Prediction Algorithms.....	94
4.15.5	Estimations of the miRNA-Gene Hybridization Potential.....	98
4.16	Evaluation of asi-miRNA-AsUSP Gene Complexes through Molecular Docking Analysis.....	99
4.16.1	Docking complex of asi-miR2118l and AsUSP Gene 1	99
4.16.2	Docking complex of asi-miR165a-5p and AsUSP Gene 2	99
4.16.3	Docking complex of asi-miR166b and AsUSP Gene 2	100
4.16.4	Docking complex of asi-miR482a-3p and AsUSP Gene 7	100
4.17	MSA and Phylogenetic Analysis of asi-miRNAs	110
4.18	Microscopic Analysis of <i>Aspergillus niger</i> Culture.....	114
4.19	<i>Agave sisalana</i> Growth under Control and Biotic Stress.....	114
4.20	Estimation of Chlorophyll Content	115

4.21	Estimation of Proline Content.....	117
4.22	Microscopic Examination of Transverse Leaf Sections.....	118
5.	DISCUSSION	120
6.	REFERENCES	128
	APPENDICES	154

List of Tables

Table 3.1: Summary of pathogen responsive miRNAs, fetched from literature, belonging to various miRNA families	30
Table 3.2: Reaction mixture for generation of poly(A)tailed RNA from miRNAs	36
Table 3.3: Reaction mixture for cDNA preparation from poly(A)tailed RNA	37
Table 3.4: Composition of the cDNA synthesis mixture	38
Table 3.5: Details of forward (miRNA specific) primers used in Poly(A)RT PCR	39
Table 3.6: Poly(A)RT PCR reaction mixture for miRNA validation	41
Table 3.7: PCR profile for poly(A)RT PCR for miRNA validation	41
Table 3.8: Primer Sequences designed for PCR amplification of AsUSP genes	43
Table 3.9: Reaction mixture for PCR amplification of AsUSP genes	44
Table 3.10: PCR profile for amplification of AsUSP genes	44
Table 3.11: Reaction mixture for ligation of PCR products with pCR®II vector	46
Table 3.12: Restriction digestion mixture for <i>EcoRI</i> digestion of the recombinant plasmids	50
Table 4.1: List of 74 potential miRNAs predicted in the study including miRNA sequences, homologues, names of transcripts in RNA-seq assembly of <i>A. sisalana</i> giving positive hits and various characteristics of miRNAs and pre-miRNA	60
Table 4.2: Quality assessment and quantification of small and large RNA fraction	66
Table 4.3: Summary of output from target prediction of <i>A. sisalana</i> miRNAs against USP genes in TAIR v.10	75
Table 4.4: Intersection of results of the three miRNA target prediction algorithms (RNAhybrid, psRNATarget and psRobot).....	97
Table 4.5: Molecular docking results of miRNAs and genes from Haddock. miRNA-Gene pairs along with the docking scores, RMSD, van der Waals energies, electrostatic energies, z-scores and other output parameters have been indicated	101
Table 4.6: Intermolecular hydrogen bonding interactions between asi-miR2118l and AsUSP Gene 1 along with bond distances (Å)	103
Table 4.7: Intermolecular hydrogen bonding interactions between asi-miR165a-5p and AsUSP Gene 2 along with bond distances (Å)	105
Table 4.8: Intermolecular hydrogen bonding interactions between asi-miR166b and AsUSP Gene 2 along with bond distances (Å)	107
Table 4.9: Intermolecular hydrogen bonding interactions between asi-miR482a-3p and AsUSP Gene 7 along with bond distances (Å)	109

Table 4.10: Analysis of variance (ANOVA) and mean performance analysis for Chlorophyll-a, Chlorophyll-b and total Chlorophyll contents in control and biotic stressed (infected) <i>A. sisalana</i>	117
Table 4.11: Analysis of variance (ANOVA) and mean performance analysis for total proline content in control and biotic stressed (infected) <i>A. sisalana</i>	118

List of Figures

Figure 2.1: An overview of various major processes involved in biogenesis and action of miRNAs in plants (Zhang et al., 2022)	9
Figure 3.1: Overview of pipeline for prediction of <i>A. sisalana</i> miRNAs from RNA-seq assembly	33
Figure 3.2: Plasmid map of pCR®II vector including restriction sites, origin of replication and other details	47
Figure 4.1: Representation of pair-wise nucleotide sequence alignments of selective query miRNAs (one from each miRNA family), with <i>A. sisalana</i> transcriptome assembly unigenes, obtained from the results of standalone BLAST-N	58
Figure 4.2: Secondary structures of selective pre-miRNAs with high stability of structures as indicated by the minimum free energy (ΔG) values	59
Figure 4.3: Box and whisker plots showing distribution of values for miRNA length and various parameters analyzed for pre-miRNAs.....	63
Figure 4.4: Bar chart representing the number of miRNAs predicted in each miRNA family	64
Figure 4.5: Agarose Gel Electrophoresis (2%) of the large RNA fraction. Lane 1: Gene ruler (100 bp), Lane 2-3: RNA sample loaded.....	65
Figure 4.6 (A-F): Poly(A)RT PCR analysis for experimental validation of miRNAs.....	67
Figure 4.7: Gene Ontology classification into biological processes of the identified miRNA targets in TAIR v.10	72
Figure 4.8: Gene Ontology classification into molecular functions of the identified miRNA targets in TAIR v.10	73
Figure 4.9: Gene Ontology classification into cellular components of the identified miRNA targets in TAIR v.10	74
Figure 4.10: 1.2% Agarose Gel Electrophoresis for analysis of PCR products from amplification of AsUSP Gene 5, AsUSP Gene 6, AsUSP Gene 7, AsUSP Gene 9, AsUSP Gene 10.....	77
Figure 4.11: 1.2% Agarose Gel Electrophoresis for analysis of PCR products from amplification of AsUSP Gene 1 and AsUSP Gene 3	78
Figure 4.12: 1.2% Agarose Gel Electrophoresis for analysis of PCR products from amplification of AsUSP Gene 2, AsUSP Gene 4 and AsUSP Gene 8	78
Figure 4.13: Transformation results of pCR®II vector ligated with AsUSP Gene 3 (A) and pCR®II vector ligated with AsUSP Gene 9 (B). Blue and white colonies can be observed in the respective figures.....	79
Figure 4.14: 1% Agarose Gel Electrophoresis for analysis of restriction products with <i>EcoRI</i> digestion .	80

Figure 4.15 A-I: Nucleotide sequences of AsUSP Gene 1-9 along with relevant information submitted to NCBI Genbank.....	81
Figure 4.16: Plot of predicted miRNA target sites in AsUSP genes using RNAhybrid	91
Figure 4.17: Plot of predicted miRNA target sites in AsUSP genes using psRNATarget.....	92
Figure 4.18: Plot of predicted miRNA target sites in AsUSP genes using psRobot.....	93
Figure 4.19: Plot of predicted miRNA target sites in AsUSP genes using consensus of three miRNA target prediction algorithms	95
Figure 4.20: Venn Diagram representing union and intersection of the miRNA target sites predicted in AsUSP genes.....	96
Figure 4.21: Bar chart representing miRNAs and their respective mfe ratio as a measure of miRNA-gene hybridization potential	98
Figure 4.22: A) Complex generated from molecular docking of asi-miR2118l and AsUSP Gene 1; B) View of selective hydrogen bonding interactions between nucleotides of the complex	102
Figure 4.23: A) Complex generated from molecular docking of asi-miR165a-5p and AsUSP Gene 2; B) View of selective hydrogen bonding interactions between nucleotides of the complex	104
Figure 4.24: A) Complex generated from molecular docking of asi-miR166b and AsUSP Gene 2; B) View of selective hydrogen bonding interactions between nucleotides of the complex	106
Figure 4.25: A) Complex generated from molecular docking of asi-miR482a-3p and AsUSP Gene 7; B) View of selective hydrogen bonding interactions between nucleotides of the complex	108
Figure 4.26: Multiple sequence alignment (A) and phylogenetic relationship (B) between asi-miR2118l and homologues in miR-2118 family.....	111
Figure 4.27: Multiple sequence alignment (A) and phylogenetic relationship (B) between asi-miR166b and homologues in miR-166 family.....	112
Figure 4.28: Multiple sequence alignment (A) and phylogenetic relationship (B) between asi-miR482a-3p and homologues in miR-482 family.....	113
Figure 4.29: Microscopic examination of <i>Aspergillus niger</i> growth as observed under optical microscope (with magnification of 100x). Conidiophore (A) and various conidia (B) could be observed under microscope.....	114
Figure 4.30: (A) <i>A. sisalana</i> Control and (B) infected with <i>A. niger</i> suspension after 14 days of infection	115
Figure 4.31: (A) Statistical comparison of Chlorophyll-a, (B) Chlorophyll-b and (C) total Chlorophyll contents in <i>A. sisalana</i> control and infected	116
Figure 4.32: Statistical comparison of proline contents in <i>A. sisalana</i> control and infected	117

Figure 4.33: Microscopic view of cross section produced from microtomy of <i>A. sisalana</i> control (A) and infected (B) leaf samples at 10x magnification (infected regions marked with blue arrows).....	118
Figure 4.34: Microscopic view of leaf cross section produced from microtomy of <i>A. sisalana</i> control (A) and infected (B) leaf samples at 20x magnification (infected regions marked with blue arrows).....	119

Abbreviations

%	percent
°C	Degree centigrade
µg	micro gram
µl	micro liter
µM	micro Molar
3D	Three dimensional
Å	Angstrom
A260/A280	Ratio of absorbance at 260 nm to absorbance at 280 nm
AFME	Adjusted minimum free energy
ANOVA	Analysis of variance
approx.	approximately
asi-miRNA	<i>Agave sisalana</i> micro-RNA
AsUSP	<i>Agave sisalana</i> Universal Stress Protein
ATP	Adenosine triphosphate
BLAST	Basic local alignment and search tool
CaCl ₂	Calcium chloride
cDNA	Complementary Deoxyribonucleic Acid
CTAB	Cetyltrimethylammonium bromide
DEPC	Diethyl pyro-carbonate
DNA	Deoxyribonucleic Acid
dNTPs	Dinucleotide Triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetic acid
ESTs	Expressed sequence tags
FW	Fresh weight

gm	gram
GO	Gene ontology
H&E	Hematoxylin and eosin
HMM	Hidden Markov Model
i.e	That is
kb	kilobase
LQ	Lower quartile
LSD	Least significance difference
M	Molar
Mfe	Minimum free energy
MFEI	Minimum free energy index
mg	milligram
MgCl ₂	Magnesium chloride
miRNA	micro-RNA
ml	millilitre
mm	millimeter
mM	millimolar
MSA	Multiple Sequence Alignment
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
ng	nano gram
NGS	Next Generation Sequencing
nm	nano meter (wavelength)
ns	non-significant
ORF	Open Reading Frame
PANTHER	Protein annotation through evolutionary relationship
PCR	Polymerase chain reaction

Pfam	Protein families database
Poly(A)RNA	Polyadenylated RNA
pre-miRNA	micro-RNA precursor
PVPP	polyvinylpolypyrrolidone
RISC	RNA-induced silencing complex
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
RNA-Seq	RNA Sequencing
Rpm	revolutions per minute
rRNA	Ribosomal RNA
s	seconds
sncRNA	Small non-coding RNA
sp.	Species
SRA	Sequence Read Archive
sRNA	Small RNA
TAIR	The Arabidopsis Information Resource
Taq	Thermus aquaticus
UQ	Upper quartile
USP	Universal Stress Protein
UV	Ultra-violet
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside

ABSTRACT

Micro-RNAs (miRNAs) are non-coding RNA molecules (18-25 nucleotides in length) encoded by genome that escort the associated proteins towards hybridization sites located in mRNAs, resulting in repression of gene expression. miRNAs function as key players in the regulations of genes involved in various biological processes through post-transcriptional regulation of target genes. miRNAs are generated from RNA precursors that have a stem-loop structure with imperfect complementarity. *Agave* species exhibit exceptional tolerance to environmental stresses due to CAM photosynthesis, wax deposition, and protective leaf morphology. Such characteristics makes this species an excellent candidate for the exploration of genomic information governing mechanisms underlying such extra-ordinary features. Research in the area of miRNAome and gene silencing in *Agave sisalana* is needed for getting insights into the stress tolerance mechanisms.

The current study was ambitioned to explore the array of miRNAs that could be identified with the use of computational tools from RNA-seq (transcriptome assembly) of *A. sisalana* and to explore miRNA targeting of *A. sisalana* Universal Stress Protein (AsUSP) family genes under biotic stress. Using a total of 6,880 non-redundant plant miRNAs (fetched from miRbase repository) including pathogen responsive miRNAs, a total of 74 potential candidate miRNAs were predicted from 67,328 unigenes in transcriptome assembly of *A. sisalana* using miRNA prediction pipeline based on comparative genomics approach and secondary structure analysis of miRNA precursors. Small and large RNA fractions were extracted from the leaf tissues of *A. sisalana* under bole rot disease and cDNAs were generated. All the 74 predicted miRNAs were tested for experimental validation and 48 were validated using poly(A)RT PCR for miRNA

amplification, from the small RNA fraction extracted from *A. sisalana* under infection with *Aspergillus niger*, and found to be targeting 5,680 genes in TAIR v.10 including 23 USP family genes (PF00582).

A. sisalana USP family genes were identified among 67,328 unigenes using TransDecoder and HMM3 search that predicted of 9 (nine) full length open reading frames (ORFs) belonging to the USP family. These nine genes were amplified to full length, cloned in *E. coli* TOP 10 using pCR®II vector, confirmed through restriction analysis with *EcoRI* and sequenced. Target prediction analysis of asi-miRNAs against AsUSP genes using three miRNA target prediction algorithms i.e. RNAhybrid, psRNATarget and psRobot, revealed a total of 406 unique miRNA target sites in AsUSP genes. From the consensus (intersection) of the three algorithms, 18 common miRNA target sites were found with minimum free energy (mfe) ratios ranging from 0.55-0.74. 3D structures of asi-miRNAs and AsUSP genes were modelled and the intermolecular interactions between miRNAs and target genes were explored through Molecular Docking analysis using Haddock server. The complexes asi-miR2118l-AsUSP Gene 1, asi-miR165a-5p-AsUSP Gene 2, asi-miR166b-AsUSP Gene 2, and asi-miR482a-3p-AsUSP Gene 7 showed stronger interactions and with a greater reliability based on docking scores and RMSD values, respectively. Considering both hybridization potential (i.e. mfe ratio) and docking scores, it was concluded that asi-miR2118l-AsUSP Gene 1 (mfe ratio 0.72, Haddock score -24.5 +/- 10.2), asi-miR166b-AsUSP Gene 2 (mfe ratio 0.63, Haddock score -37.2 +/- 11.6) and asi-miR482a-3p-AsUSP Gene 7 (mfe ratio 0.61, Haddock score -31.8 +/- 7.4) are more likely to lead to the predicted effect i.e. target mRNA cleavage or translational inhibition.

Biotic stress was induced in *A. sisalana* plants through inoculation of *Aspergillus niger* (a pathogen that causes bole rot disease in *A. sisalana*) suspension into the bole region. Mean total

Chlorophyll content was reduced to 0.77 mg/g FW upon infection development as compared to control, in which the Chlorophyll content was measured to be 1.75 mg/g FW ($P < 0.01$). Microscopic examination of transverse sections of leaves indicated differences between leaf tissues of infected and uninfected plants.

Through this study, the possible interactions between As-miRNAs and the USP genes of *A. sisalana* have been predicted, further confirmation of these miRNA-USP gene interactions and functional investigations of these miRNAs can be performed through in-vitro, wet-lab experimentation to further uncover their roles and function in biotic stress in *A. sisalana*. Results of this study will pave way towards understanding the mechanism of miRNA-mediated gene regulations under biotic stresses in *A. sisalana*. Moreover, small RNA-seq analysis of *A. sisalana* under control and biotic stress can be performed which can reveal novel miRNA sequences in *A. sisalana* and their differential expression under control and biotic stress can unveil miRNAs which are differentially expressed and play role in gene regulation of *A. sisalana* under biotic stress. Differential expression profiling of As-USP genes identified in this study, under control and biotic stress, can be conducted to further investigate their role in defense to infection by *A. niger*. Results of this study will also pave the way towards development of molecular based strategies to improve tolerance to *A. sisalana* to the bole rot disease caused by *A. niger* through over-expression of USP genes after confirmation of their role in pathogen defense.

1. INTRODUCTION

The foundation of molecular biology lies in the genetic code located within the DNA of an organism which is transcribed to RNA for protein synthesis. The discovery of double helical structure of DNA in 1953 was a breakthrough after which many advancements were made in the field of Genomics and Molecular Biology. RNA, collectively referred to as transcriptome, has complex genomic structures with coding and non-coding regions and intermediate molecules between genes and proteins. Analysis of RNA is important for understanding the processes occurring in genome and, origin and development of diseases in an organism (Ergin et al., 2022).

Among the non-coding repertoire of genome, micro-RNAs (miRNAs) act as key players in the regulations of genes as observed initially in *Caenorhabditis elegans*, where they were identified to be functioning in determination of various timings of the development events (Sun et al., 2020). Developments in DNA and RNA sequencing generates vast amounts of sequencing information that can be utilized for the purpose of discovery of miRNAs and their targets (Pritchard et al., 2012). High throughput Next Generation Sequencing (NGS) and comparative genomics based computational methods have been shown to be fruitful in discovery of miRNAs. With the data source accessible from NCBI (National Center for Biotechnology Information), bioinformatic approaches can be applied on the NGS data for prediction of miRNAs. Homology based approach on transcriptomic database using conserved miRNA from other species as query sequences, based on the principle of searching for conserved sequences that fold into extended hairpin structure, can help in identification of orthologs and paralogs of miRNAs (Gomes et al., 2013; Zhang et al., 2006).

As per Food and Agriculture Organization of the United Nations (FAO), plants make up about 80% of the human diet. Productivity of plants with efficiency is rudimental for ensuring

nutritious, safe, affordable, and sufficient food supply to ever-expanding human population (FAO, 2017). Crops are exposed to various biotic and abiotic stresses (such as extreme temperatures, salinity, draught and pathogens) that compromise the yield by reducing production (Pandey et al., 2017). Stress factors coupled with fluctuating climatic conditions around the world act as a major challenge that influences crop production (Kar & Raichaudhuri, 2021). On exposure to stress, post-regulation of gene expression comes into play and try to restore the cellular homeostasis and alleviate the stress (Kawaguchi et al., 2004).

Agave sisalana is an herbaceous monocotyledonous plant. Originally from Central America and Mexico, *A. sisalana* is naturalized in many other countries of the tropics, subtropics, and warm and temperate areas (El Mokni & Verloove, 2021). The plants have a large rosette of thick fleshy leaves, each ending generally in a sharp point, and are indigenous to both arid and semi-arid regions (Escamilla-Treviño, 2012). It has many uses owing to properties of its fiber and also due to medicinal properties of its extracts (such as antimicrobial, anti-inflammatory and anthelmintic activities) (Domingues et al., 2010). *A. sisalana* exhibits exceptional tolerance to their xeric environments, and employ various strategies to cope with environmental constraints (Sarwar et al., 2019).

Habitats of *Agave* sp. vary from a range of dry to humid environments and they have developed biological mechanisms to endure and survive under changing climatic conditions (Batcho et al., 2022). Many species of *Agave* have been used in producing spirits (e.g., tequila), food, forage, drinks, drugs, weaving, paper and construction materials. *Agave* are known to accumulate sugars and other commercially important substances in large quantities and therefore have proved to hold immense potential for biofuel production. (Mousumi et al., 2010).

Aspergillus niger belongs to the genus *Aspergillus* which includes a set of filamentous fungi that generally reproduce asexually; however, perfect forms reproducing sexually have also been seen (Pöggeler, 2002). *A. niger* has the ability of growth in the temperature range of 6–47°C with a relatively high temperature optimum at 35–37°C, and over a wide pH range i.e., 1.4–9.8. The water activity limit for *A. niger* (0.88) growth is relatively high compared with other *Aspergillus* species (water activity is the ratio of vapor pressure of water over a sample to the vapor pressure of pure water at the same temperature and is ranged from 0.0-1.0; as most of the foods have water activity >0.95, greater is the likelihood of fungal spoilage). Such abilities favor plentiful production of conidiospores, which are spread through the air and tend to inhabit warm and humid areas (Schuster et al., 2002). Aspergilli, being ubiquitous in nature, have a wide geographical distribution and exist in an array of habitats due to their ability to grow on a broad range of substrates (Gautam et al., 2011). Usually, *A. niger* is a common fungus growing as saprophyte on dead matter (e.g. leaves), compost piles, stored grain and other vegetation under decay process (Sharma, 2012). Diseases caused in plant species include but not limited to black rot of onions, crown rot of peanuts, tuber rot of yam, black mold rot of cherry, fruit rot of grapes and banana, rot of tomatoes, stem rot of *Dracaena* (Gautam et al., 2011). *A. niger* falls in the category of black aspergilli and is one of the pathogenic agent causing bole rot disease of *Agave sisalana* (Coutinho et al., 2006).

Microscopic observation of *Aspergillus niger* has indicated its morphological features. Conidiophores- (a conidia-bearing hyphae or filaments) are smooth walled, translucent, and turning dark towards the vesicle. Head of conidia (asexual reproductive spores) are biserial (arranged in two rows) with the phialides (dilated part of the top of conidiophore) existing on often septate metulae (brown appearance). Conidia show globose (spherical) to subglobose (imperfectly

spherical) appearance and dark brown to black color with diameter of 3.5-5.0 μm , and rough walls (Andge et al., 2017).

Universal Stress Proteins (USPs) are 15-30 kDa (molecular weight) which accumulate in organisms under unfavorable conditions. USP domain (PF00582) contains a conserved motif G-(2X)-G-(9X)-G-(S/T) in the structure and nucleotide-binding/hydrolyzing activity is a key feature of USPs. Under stressed circumstances, USP genes are upregulated and are expressed in substantial amounts by cells. Over-expression of these proteins allows organism to withstand stress conditions; however, the underlying mechanisms remain elusive. A great structural diversity has been observed in USPs (Tkaczuk et al., 2013). USP family genes in *A. sisalana* have not been identified/explored hitherto.

A. sisalana has developed a series of biological mechanisms to cope with and thrive in a range of conditions. To date, very little of the research related to non-coding RNA based gene regulation has been carried out in *A. sisalana* despite it being exceptionally tolerant to various environmental stresses- one of the reasons being a lack of reference genome. Studies on disease response genes in relation to miRNAs in *Agave* sp. is an area yet to be explored to understand defense mechanisms in *A. sisalana*. Recently the transcriptome of *A. sisalana* has been sequenced through NGS (RNA-Seq) (Sarwar et al., 2019) which can be utilized for prediction of miRNAs which act as key regulators of various biological processes including adaptation under extreme conditions. Moreover, there is a lack of research related to pathogen responsive genes/ genes of the defense system such USP genes and their expression regulation under biotic stress (a usual form of which is the bole rot disease of *A. sisalana* caused by black aspergilli including *A. niger*).

The aims and objectives of the current study were prediction of miRNAs in *A. sisalana* from the transcriptome (RNA-seq data), based on comparative genomics approach and analyzing

various characteristics of miRNA precursor sequences. Moreover, the study was directed towards experimentally validating the expression of predicted miRNAs under biotic stress (infection with *A. niger*) and analyzing target genes in The Arabidopsis Information Resource (TAIR) v.10 as reference and further, to identify the potential pathogen responsive genes specifically USP genes in *A. sisalana* from the transcriptome and confirmation of genes through amplification and sequencing analysis. Further, the research was aimed at exploring the targeting and interactions between the experimentally verified miRNAs and USP genes, which potentially play a role in pathogen response in *A. sisalana*.

2. REVIEW OF LITERATURE

2.1 Micro-RNAs

Micro-RNAs (miRNAs) are non-coding RNA molecules (18-25 nucleotides in length) encoded by genome that escort the associated proteins towards hybridization sites of mRNAs resulting in interference in transcription, translation, or epigenetic processes in eukaryotes and viruses (Eldem et al., 2013; Han et al., 2013). The importance of miRNAs was not acknowledged due to technological and methodological limitations until the discovery of two miRNAs, *lin-4* and *let-7* in *Caenorhabditis elegans* (nematode) in 1993 and a kairotic moment in the field of miRNAs and 3'UTRs was the identification of homologs of *let-7* in other species (such as *Homo sapiens*). In the year 2002, a few researchers discovered numerous miRNAs in *Arabidopsis* by Northern blot which was the first evidence of miRNA existence in plants and further research on *Arabidopsis* miRNAs laid down a framework of comprehending miRNA biogenesis and function (Llave et al., 2002; Mette et al., 2002; Reinhart et al., 2002).

Key advancements in sequencing technologies occurred parallel and galvanized a collaborative effort of sequencing that resulted in generation of a plenty of data for prediction of miRNAs; this lead to discovery of thousands of new miRNA sequences (Lau et al., 2001; Nobuta et al., 2010). As the cross-species conservation of miRNAs has been found, miRNAs were indicated to be involved in a multitude of functional processes in metazoans including development, metabolism, signaling and immune system (Ameres & Zamore, 2013).

Advancements in bioinformatics and development of tools for miRNA research have speeded up the discovery, prediction and analysis of miRNAs including their structures and potential targets (L. Chen et al., 2019; Eldem et al., 2013). miRNA regulation of gene expression is not only restricted to controlling developmental timing, cell differentiation, proliferation and

apoptosis, but they are also involved in interactions of host with pathogen (Kulshrestha et al., 2020; Liu, 2008; Šečić et al., 2021).

2.2 miRNA Biogenesis

miRNAs are generated from RNA precursors that have a stem-loop structure with imperfect complementarity (Meng et al., 2013). miRNAs are encoded by *MIRNA* genes and the length of such genes is uncertain (Zhang et al., 2007). While the plant *MIRNA* genes mainly harbor at intergenic regions, genes encoding animal *MIRNA* can be located at any position in the genome including the coding repertoire (Millar & Waterhouse, 2005; Zhang et al., 2007).

Similar to protein coding genes, RNA polymerase II is involved in transcribing *MIRNA* genes, albeit certain miRNAs are transcribed by another type of RNA Polymerase i.e. RNA Polymerase III (Faller & Guo, 2008). The resulting miRNA transcripts produced initially as a result of transcription are referred to as pri-miRNAs (primary miRNAs). These pri-miRNAs undergo the process of capping and polyadenylation in both plants and animals (Z. Gao et al., 2021).

The process involved in converting pri-miRNAs to functionally mature miRNAs is complicated involving multiple steps and many enzymes are associated with biogenesis of miRNAs (Sun, 2012). As of now, the major genes coding for miRNA biogenesis enzymes are homologous in animals and plants with the exception of Drosha (which is RNase III endonuclease and its homologue has not been found in plants) (Sun, 2012).

pri-miRNAs have one or more long hairpin structures in addition to capping and polyadenylation features (Xie et al., 2005), and are able to undergo cleavage by multiple protein complexes e.g. DCL1 (dicer-like1), double-stranded RNA-binding protein Hyponastic Leaves (HYL1), C2H2-zinc finger structural protein Serrate (SE), and G-patch structural protein TGH

from either “base-to-loop” or “loop-to-base” direction to produce pre-miRNA (precursor miRNA) (Wang et al., 2019). This process is occurring in the D-body or SmD3/SmB-body which is a nuclear processing center (Z. Gao et al., 2021; Sun, 2012).

Pre-miRNAs are processed into miRNA:miRNA* duplexes, which have 2 nucleotides overhang at 3' end of both strands, and these duplexes can undergo convenient degradation by Small RNA Degrading Nuclease (SDN) -a category of exonucleases (Ramachandran & Chen, 2008). Hua Enhancer 1 (HEN1), a small RNA methyl transferase, causes immediate methylation of the 3' end nucleotides of both strands in the miRNA:miRNA* duplexes and enhances their stability, and thus uridylation and degradation of the duplexes is prevented (Yu et al., 2005).

miRNA/miRNA* duplex is then transported to cytoplasm and the process is aided by methylation as well as by a plant homologue of exportin-5 i.e. HASTY (exportin-5 in animals is involved in shipment of pre-miRNA from nucleus to cytoplasm) (Axtell et al., 2011).

In the cytoplasm, miRNA/miRNA* duplex is dis-united and the miRNA strand is loaded in the RNA-induced silencing complex (RISC) which leads to synthesis of a miR-RISC complex (Calviño et al., 2011; Chen, 2005). ARGONAUTE 1 (AGO1) which is an intrinsic component of RISC causes cleavage of the target mRNA at the center if miRNA-mRNA duplex. Discordant with animal miRNAs, most of the plant miRNAs bring about cleavage of target mRNAs instead of

inhibiting translation (Zhang et al., 2007). An overview of various major processes involved in biogenesis and action of miRNAs in plants are presented in Figure 2.1.

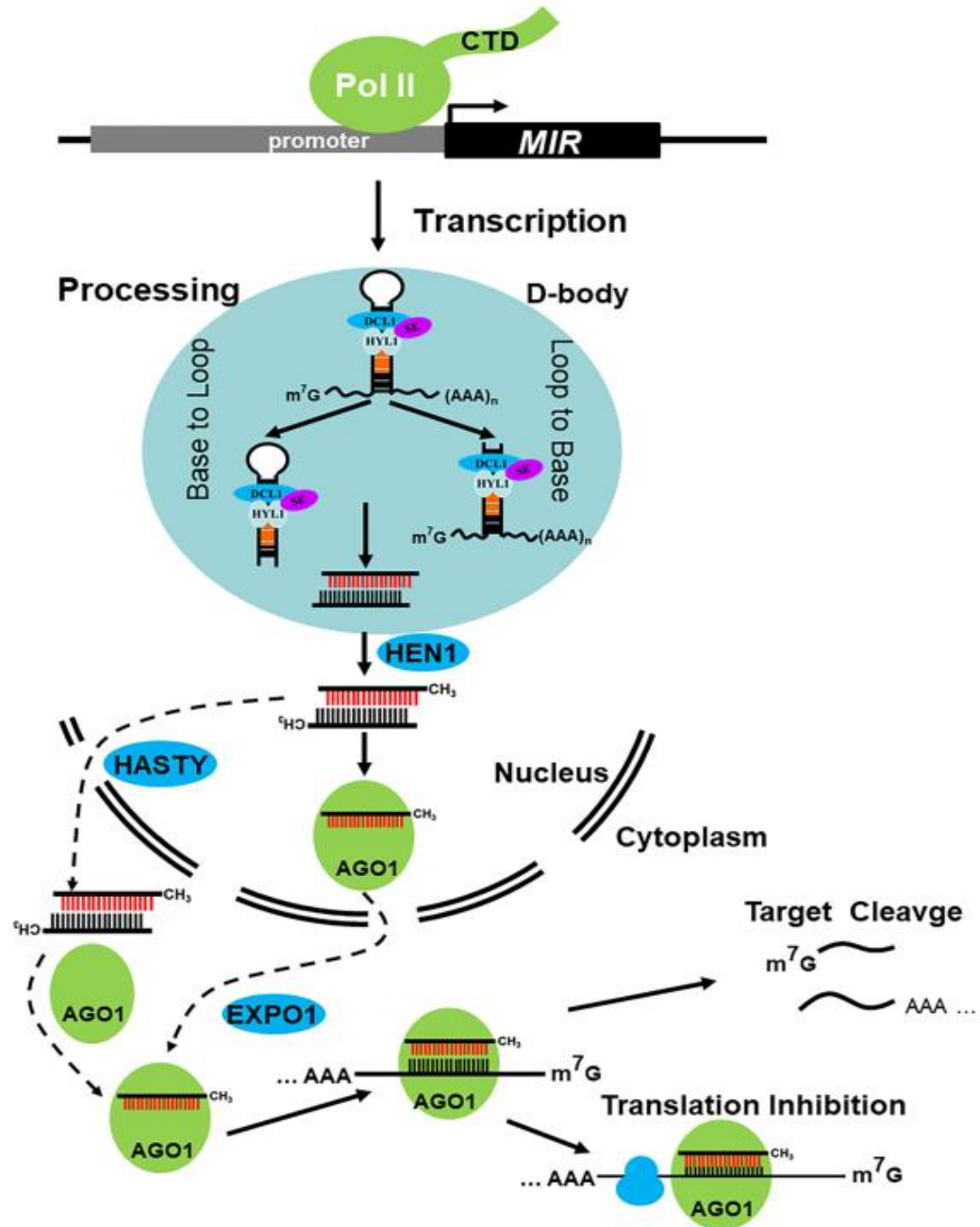


Figure 2.1: An overview of various major processes involved in biogenesis and action of miRNAs in plants (Zhang et al., 2022)

2.3 Next-Generation Sequencing (NGS)

Next-generation sequencing (NGS) involves high-throughput, deep, and in-parallel DNA sequencing techniques that were established a few decades after Sanger DNA sequencing technique was first launched in 1977 (Mardis, 2008). Unlike the Sanger method, NGS has the possibility to perform enormously parallel analysis as it proved to be exceptionally high throughput using a large number of samples at a reduced cost comparatively (Mardis, 2011).

NGS has the capability to sequence billions of DNA nucleotides in parallel analysis, thereby resulting in substantial output and minimizing the requirement of other fragment-cloning techniques which were needed with the Sanger sequencing method (Metzker, 2010). Establishment of NGS platform has brought a revolution in the field of genomics and modernization of the whole field of molecular biology. Nowadays, the major suppliers of high throughput instruments for NGS include the Illumina/HiSeq Genome Analyzer, Life Technologies, Helicos Biosciences, ABI SOLID system (Applied Biosystems), and Roche 454 Genome Sequencer.

- i) On an NGS platform, input DNA fragmentation occurs, and the DNA library is produced either by the amplification or the ligation with other custom adapter sequences. The adapters let the hybridization of library to sequencing chips and play its role as a general priming site for the (other) sequencing primers (Meera Krishna et al., 2019).
- ii) Afterwards, each library fragment is amplified on some solid surface (beads or any flat silicon derivative surface) which covalently linked DNA linkers that bind with the library adapters. The amplification leads to the creation of DNA clusters (each produced from a single library fragment). Each cluster then sequenced as a unique sequencing reaction (Üner, 2020).

- iii) Sequence of each cluster is then optically read through light or fluorescent signals by repeated nucleotide incorporation cycles (McCombie et al., 2019). Cycling condition is dependent on the instrument being used. Reversibly fluorescent and terminated nucleotides are incorporated with repeated cycles followed by signal accession and removal of the terminator groups and fluorescent groups (Olliff et al., 2022).
- iv) At the conclusion of the sequencing reaction, raw data is generated which is a collection of the DNA sequences generated at each cluster. To generate useful results, this data is further analyzed. An NGS platform may differ from another mainly with regards to the technical details accompanied with sequencing reactions (Meera Krishna et al., 2019; Üner, 2020).

2.4 RNA Sequencing (RNA-Seq): A Method for Transcriptome Analysis

RNA sequencing (RNA-seq) emerged in 2008 for a comprehensive study of transcriptome and it is also a high-throughput technology (Wang et al., 2009). RNA-seq makes use of a set of experimental techniques to synthesize cDNA sequences from a repertoire of all RNA molecules. This is followed by library creation and enormously parallel deep sequencing (Han et al., 2015).

RNA is separated from the biological material of interest at the beginning of the method. Then, specific protocols are used to isolate certain subsets of RNA molecules, such as the poly-A selection technique to enrich for polyadenylated transcripts or the ribo-depletion protocol to remove ribosomal RNAs. Reverse transcription is used to create cDNA from RNA, and sequencing connectors (adaptors) are attached to the ends of the cDNA fragments, followed by amplification and sequencing of RNA-Seq library on an NGS platform (Kukurba & Montgomery, 2015).

2.5 Applications of RNA-Seq

RNA-Seq has application in food and agriculture as well as in health sciences, ranging from differential gene expression, alternative splicing, allele-specific expression, variants calling, alternative polyadenylation site usage and detection, precisely locating regulatory elements, circular RNAs, microRNA profiling, meta-transcriptomics and systems biology (Arnold et al., 2013; Conde et al., 2013; Sripathi et al., 2021). In stem cell biology and neuroscience, single-cell RNA-seq analysis has been extensively employed to examine cellular heterogeneity and diversity (Wilson et al., 2015).

Utilizing RNA-seq, it is possible to quantify the relative changes in each transcript abundance level over the course of development or in response to various treatments. With a single base pair resolution, an enormous dynamic detection range, and little background signals, RNA-seq enables unbiased investigation of the transcriptome. Unlike the techniques based on hybridization, it is not confined to probing of particular probes on an array, and it is applicable to the species for which complete reference genome has not been known yet (Han et al., 2015)

The use of RNA-seq has been notable for discovering various novel transcript types such as long non-coding RNA, miRNA, siRNA, and other classes of small RNAs (such as snRNA and piRNA) that are involved in the regulation of RNA stability, protein translation, or the transition between chromatin states (Robertson et al., 2010). The use of alternative promoters and premature transcription termination at the 3' end, which is important for mRNA stability, can also be revealed by RNA-seq (Wilhelm et al., 2011).

2.6 Approaches for miRNA Identification

For identification of plant miRNAs, both experimental methods and computational (bioinformatic) approaches have been implemented by researchers. Initial approaches towards miRNA identification were focused on direct cloning techniques and it has been utilized for identification of miRNAs in various species of plants (Reinhart et al., 2002). However, direct cloning approach being more time requiring, needing extensive lab methods and a high cost has been superseded by other approaches for miRNA identification (Sun, 2012).

Computational approaches have emerged as the understanding about plant miRNAs improved specifically about the properties of conservation and secondary structure of miRNAs, and this resulted in identification of additional miRNAs from a diversity of plants (Jones-Rhoades & Bartel, 2004; Zhang et al., 2005). Lately, as deep sequencing technologies have come to be the point of focus and have become easily accessible, even more miRNAs have been discovered and functionally characterized in plants (Creighton et al., 2009; Fahlgren et al., 2007).

Ranging from moss to land plants, evidence has been gathered regarding evolutionary conservation of miRNAs in plants (Arazi et al., 2005). Thus, a BLAST-N search using confirmed miRNAs from other plants has been a powerful approach towards discovering miRNA homologues miRNAs in other plants species. However, relying merely on homology analysis remains speculative as miRNAs, being only about 20-22 nucleotide in length, getting a high percentage of identity in result of pairwise alignment may not be a powerful evidence of miRNA existence (Zhang et al., 2006).

In order to have a better identification and reducing the chances of false positives, homology searches are coupled with the various major characteristics/properties of miRNAs, specifically the stem-loop hairpin structures (Zhang et al., 2006). On the basis of the sequence

conservation and characteristics of miRNAs, Zhang et al. (2005) established an EST analysis method for identification of miRNAs in a large diversity of plant species. Through implementation of this approach, 481 miRNAs in total, related to 37 families of miRNAs in 71 disparate plants species using greater than 6M EST sequences have been identified.

EST analysis was also extended to other genomic sequences, with the inclusion of the genome survey sequence (GSS) database (Pan et al., 2007). This strategy is being currently used by numerous scientists across the world and miRNAs have been identified in important crops such as wheat, cotton, switchgrass, soybean, corn, oilseed, apple and tobacco (Dezulian et al., 2006; Frazier et al., 2010; Matts et al., 2010).

2.7 miRNA Prediction from RNA-seq Data

Analysis of transcriptome (RNA-seq) data to computationally predict miRNAs in different plants has been reported by many scientists. Bano et al. (2022) used the leaf transcriptome (RNA-Seq) data from *Cestrum nocturnum* L. (147,707 contigs) and *Cestrum diurnum* L. (145,483 contigs) and identified miRNAs and their targets in corresponding species in stress responses through homology-based approaches and secondary structure evaluation. A total of 32 miRNA candidates (harbored in 26 miRNA families and regulating 1024 target genes) were discovered in *C. nocturnum* while a total of 12 miRNA candidates (harbored in 10 miRNA families and regulating 1007 target genes) were found in *C. diurnum*.

Jabbar et al. (2021) applied a comparative genomics approach on RNA-seq data of *Agave deserti* consisting of 128,869 *Agave* contigs. On the basis of nucleotide homology analysis 6,028 non-redundant plant miRNAs and secondary structure analysis of potential pre-miRNAs, 30 unique miRNAs belonging to 14 different categories of precursors. Target genes of miRNAs were predicted subsequently, and possible functions were defined by functional annotation analysis.

Gadhavi et al. (2020) performed transcriptome-wide identification of miRNAs from *Bacopa monnieri* (a well-known medicinal plant for its nootropic properties) to identify a total of 12 miRNAs relating to 11 miRNA families and found 68 target genes (human) associated with signaling pathways such as NF-kB and MAPK (give full forms).

Kalariya et al. (2019) performed screening of 6028 unique plant miRNAs against 272,161 unigene sequences belonging to *Gymnema sylvestre* for identification of 16 candidate miRNAs that belonged to 12 miRNA families.

Najafabadi and Naghavi (2018) used RNA-seq data of *Ferula gummosa* (from root and flower samples) to predict 220 miRNAs belonging to 94 families and their targets using computational approaches. Regulatory effects of miRNAs (direct or indirect) on target genes associated with terpene biosynthesis were predicted through biological network analysis. miR5658, miR1533, miR5021, miR414, and miR1436 were among the highly abundant miRNAs and were found to be associated with terpene biosynthesis pathway. Differential gene expression analysis showed that these miRNAs had a negative correlation with expression of TFs and their co-expressed terpene biosynthesis genes.

Singh and Sharma (2017) identified miRNAs, using ESTs, nucleotide and Sequence Read Archive (SRA) data of *Curcuma longa*. 18 miRNA families were identified, and 16 families were found to be associated with regulation of 238 target transcripts. Annotation of target genes and analysis of pathway showed the involvement of miR156 and miR5015 in rhizome development. miR5021 indicated regulation for biosynthesis of terpenoid backbone and pathways of isoquinoline alkaloid biosynthesis, while miR2919 was found to be regulating flavonoid biosynthesis pathway. miR1168.2, miR156b and miR1858 were likely involved in biosynthesis of curcumin.

Prakash et al. (2016) used 7201 sequences of mature miRNAs from 53 species of plants for BLAST-N analysis against 99,763 nucleotides entries including TSA dataset (contigs) of *Rauvolfia serpentina*. They identified 15 conserved miRNAs harbored in 13 families. Predicted targets of miRNAs included transcription factors as well as targets involved in various processes including metabolism, stress response, growth and development.

Naganeeswaran et al. (2015) used RNA-seq (cotton leaf transcriptome data) of the cultivar Chowghat Green Dwarf (SRX436961) .e. 130942 coconut transcriptome unigenes generated using an Illumina HiSeq 2000 platform and predicted 16 miRNAs (using homology and structural features based approach), which belonged to 11 miRNA families. Targets were identified for 7 miRNAs and most of the target genes seemed to code for transcription factors.

2.8 Biotic Stress

Environmental stress, such as biotic and abiotic stress, limit crop yield and quality by limiting plant growth and development. Results of different investigations on model species have indicated that thousands of protein coding genes are differentially expressed and such differential expression is induced by various environmental stresses (Kar & Raichaudhuri, 2021; Khraiwesh et al., 2012). Among the stress-induced genes, some of them have been identified and harnessed for improving enhancing tolerance of crops to stress through implementation of transgenic technology. Nonetheless, the mechanism involved in regulation of such protein coding genes remains largely obscure (Sun, 2012).

Biotic stress is a state in plant which is induced by biotic agents such as bacteria, fungi, insects, viruses, nematodes. To withstand such stressed circumstances, gene expression in plant is either up-regulated or down-regulated. Role of miRNAs in regulatory pathways has been ascertained through sequencing of small RNAs under stress condition. Zuo et al. (2011) performed

sequencing in tomato under biotic stress to reveal several conserved and novel miRNAs (Zuo et al., 2011).

Pathogenic viruses contain DNA/RNA as genetic material which cause infection in plants cells by making use of the cytoplasmic machinery or by integration into the host genome. Upon viral infection, expression of viral genes is made possible due to the presence of reverse transcriptase and RNA-dependent RNA polymerase. miRNAs expression can be changed during viral infection that influences the gene expression (Bazzini et al., 2009; Pradhan et al., 2015).

2.9 Pathogen Responsive miRNAs under Biotic Stress

In the recent past, miRNAs playing key role in plant response to environmental stress have been identified. These miRNAs control the expression of genes coding for proteins under the conditions of biotic and abiotic stress (Kulcheski et al., 2011). While understanding regarding the function of miRNAs in regulating uptake of nutrients, circadian rhythms and responsiveness to abiotic stresses is well-established, a thorough comprehension in immune-regulatory roles of miRNAs in biotic stress response still remain insufficient (Šečić et al., 2021).

Both miRNAs and siRNAs can bring about induction of post-transcriptional gene silencing (PTGS) by the way of cleaving/degrading mRNA or inhibition of translation. On the other hand, transcriptional gene silencing (TGS) resulting in DNA methylation, modification of histone or chromatin is normally regulated by siRNAs and some particular miRNAs (Baulcombe, 2004; Chellappan et al., 2004; Cui & Cao, 2014; Vaucheret, 2006; Wu et al., 2015).

Many small RNAs (sRNAs) and miRNAs have been associated with responses to biotic stress. RNA silencing machinery in plants enables pathogen-associated molecular pattern-triggered immunity and effector-triggered immunity that is helpful in defense against pathogen attack or insect herbivores. On the other hand, pathogens also produce effectors and sRNA

molecules that assist them in countering the host immune response. Numerous studies have shown the roles of sRNAs in plant defense and application of sRNAs for defense against pathogenic attack (Iqbal et al., 2017; Jabbar et al., 2019). As a solution for pathogen or insect herbivore attack, transgenic plants have been produced with enhanced ability of pathogen resistance (Huang et al., 2016; Murtaza et al., 2022).

It is important to have improved food production to confront the challenges of ever-rising population of the world and loss of fertile lands. Biotic threats (such as bacteria, oomycetes, viruses, fungi and insect pests) limiting food production altogether account for up to 30% drop in crop content (both pre- and post-harvest) (Bebber & Gurr, 2015; Flood, 2010). Therefore, it is crucial to reveal responses to biotic stress in plants for developing innovative tools of crop protection from various pathogens and pests. Activation or suppression of an array of genes in plants helps in countering the pathogen infection (Jones & Dangl, 2006).

Xu et al. (2021) identified miRNA-target gene pairs responsive to fusarium wilt of cucumber (*Cucumis sativus* L.) that is caused by *Fusarium oxysporum* f. sp. *cucumerinum*, and discovered that miR319a-JRL3, miR6300-BEE1, miR6300-DAHP1 and miR6300-PERK2 were involved in regulation of cucumber defense response against fusarium wilt.

Wu et al. (2020) studied positive regulation of tomato (*Solanum lycopersicum*) resistance against infection caused by *Botrytis cinerea* (a necrotrophic fungus) by miR319c targeting of TCP29, which acts in negative regulation to *B. cinerea* infection. Transgenic plants over-expressing miR319c indicated improvement in their resistance to *B. cinerea*.

Upregulation of several miRNAs, including miR156, miR167 and miR168, was shown in tomato (*Solanum lycopersicum*) roots colonized by the fungus *Pochonia chlamydosporia*. The predicted targets were associated with pathways controlling defense and regulation of apoptosis

(Pentimone et al., 2018). Expression of two miRNAs (miR2911 and miR1030) have been observed upon infection of *Xanthomonas axonopodis* in *Populus* species. Resistance of *Malus hupehensis* against *Botryosphaeria dothidea* (a fungal pathogen) is achieved through targeting of AGO1 by miR168 (Yu et al., 2017).

An elevated expression of miR811 and miR845 has been observed in intolerant maize plant infected with *Exserohilum turcicum* (Wu et al., 2014). In cotton infected with *Verticillium dahliae*, an upregulation of miR159 was noted, as well as in Arabidopsis and, like miR166, thereby bringing about silencing of the virulence gene- VdHiC-15 in the fungus (Zhang et al., 2016).

Up-regulated expression of miR168 has been shown during infection of Soybean mosaic virus G7 and elevated level of miR168 targeted the AGO1 gene (H. Chen et al., 2015; Pacheco et al., 2012; Várallyay et al., 2010). It was observed that infection with Tomato leaf curl virus in tomato resistant cultivars showed enhanced content of miR156 and did not show any signs of disease (Tousi et al., 2017). Repression of plant effective genes occurs under infection with bacterial pathogen resulting in increased disease progression. Under such unfavorable conditions, plant responds by expressing certain genes involved with miRNAs (Jin, 2008; Sunkar et al., 2012).

Two miRNAs m0001 & m0002 have been indicated to be important in imparting resistance to resistant eggplant against *V. dahliae* infection. Many conserved miRNAs including miR156, miR159, miR166, miR167, miR169, miR172, miR319 and miR396 are under-expressed after attack with fungal pathogen. These miRNAs are involved in regulating the expression of genes coding for various transcription factor families (such as ARFs, GRFs, MYBs, NACs, HD-ZIPs, SPLs, SCLs and AP2-like factors) (Yang et al., 2013).

Zhu et al. (2013) observed a down-regulation of miRNAs ghr-miR482b.2, ghr-miR482c and ghr-miR482d.2 in susceptible cultivar Sicot71 (*Gossypium hirsutum*) under infection with *V.*

dahliae (fungal pathogen), while an up-regulation in expression of several NBS-LRR targets (genes involved in pathogen sensing and host defense) of ghr-miR482c and ghr-miR482d was observed.

Gupta et al. (2012) investigated 8 miRNAs (miR159, miR164, miR167, miR171, miR444, miR408, miR1129 and miR1138), in wheat (*Triticum aestivum*) during stem rust infection (caused by *Puccinia graminis* f. sp. tritici), linked to three varied and independent defense responses at cellular level against pathogen attack.

In some studies, it is reported that miR159 (having similarity to miR156) showed a down-regulation in expression in wheat infected with *B. graminis* and pine infected with *C. quercuum* (Lu et al., 2007; Xin et al., 2010).

2.10 Universal Stress Proteins

Universal stress proteins (USP) belonging to PF00582 superfamily (COG0589) as per Pfam classification are widespread in nature and exist in a wide array of organisms such as archaea, bacteria, fungi as well as plants (Tkaczuk et al., 2013). The first USP was identified in *Escherichia coli* (and named as USP-A) and afterwards, five other members (paralogs) of the USP superfamily were discovered in *E. coli*, including USP-C, USP-D, USP-E, USP-F, and USP-G (Gustavsson et al., 2002).

USPs play a role in protein scaffolding, holding and preventing the denaturation of molten globular macromolecules, ion scavenging, response to hypoxia, modulating growth and development of cells, and cellular protein transport (Chi et al., 2019; Vollmer & Bark, 2018). In addition to this, many USPs have exhibited activities of DNA binding, repairing, and refolding

which help organisms in protection of nucleic acids (DNA/RNA) from extrinsic stresses (Drumm et al., 2009; Kvint et al., 2003).

Various kinds of stresses such as drought, high salinity, biotic stress, extreme temperatures, nutrient starvation as well as existence of uncouplers, metals and antibiotics can act as signal for induction of USP domain containing proteins (Chi et al., 2019; Tkaczuk et al., 2013). In such stressed conditions, cell growth may be arrested and USPs establish a natural biological defense mechanism (Tkaczuk et al., 2013).

A broad range of structural diversity has been observed in USPs, conforming to their functions in resistance to multiple stresses, owing to the presence of a variety of functional motifs joined to the USP domain. These catalytic motifs, in addition to rendering structural diversity to USPs, enable a range of biochemical functions already mentioned above. Even so, the molecular mechanisms, governing the function of USPs in protecting cells and resistance to stress, remain predominantly unexplained (Chi et al., 2019).

Earliest structure of a USP in Protein Data Bank is that of the hypothetical USP-A type protein (MJ0577), crystallized with a bound ATP molecule, from *Methanocaldococcus jannaschii* (PDB ID: 1MJH) (Zarembinski et al., 1998). Sousa and McKay (2001) reported that structure for a USP-A from *Haemophilus influenzae* (PDB ID: 1JMV). In the crystal structure of this protein, an asymmetric dimer was revealed having a tertiary α/β fold exhibiting similarity with the MJ0577 protein. However, the two structures differed significantly such as the triphosphate binding loop of the ATP binding motif and *H. influenzae* USP lacked the ability of binding with nucleotides (Kvint et al., 2003; Sousa & McKay, 2001). A consensus sequence G-(2X)-G-(9X)- G(S/T) for nucleotide binding was found, but ability of binding with nucleotides differs from protein to

protein. Some of the PDB structures of USPs have also been identified to have bound to AMP and dAMP (Vollmer & Bark, 2018).

Structural insights into the USP family proteins, have revealed that USP-A, USP-C, and USP-D are single-domain structures composed of 150 amino acids that represent dimerization in the crystal lattice. USP-F and G proteins structures also are made from single domains of approximately 150 residues and the domain has similarity with one of the domains in USP-E structural repertoire. USP-E structure is double-domain and composed of about 300 residues. USPs have modular structural motifs and can exist in fusion with other protein domains such as U-box domains, KdpD, kinase domains, HATPase domains and domains with unknown functions). The additional domain can also be another USP domain (Kvint et al., 2003; Tkaczuk et al., 2013).

USP classification can directly linked to *E.coli* USPs: Class I of USPs includes USP-A, USP-C and USP-D, Class II includes USP-E and USP-F while Classes III and IV contain USP-E domains 1 and 2. The difference classes of USPs have some common interface features and can undergo interaction to generate heterodimers and homodimers (Nachin et al., 2008).

USPs exist both as single-domain proteins and fusions with additional domains. The extra domain can be an additional USP domain (as in PA1789 from *Pseudomonas aeruginosa* POA1), a protein kinase domain (as in some plants) or an amino acid permease in combination with two USP domains (as in some archaea) (Kvint et al., 2003). Superfamily members of USP can be split into two major classes i.e. ATP binding USPs and non-ATP-binding USPs (Vollmer & Bark, 2018).

2.11 Universal Stress Proteins (USPs) in Biotic and Abiotic Stresses

Stress responsive genes among various gene families have been studied in the model species, *Arabidopsis thaliana*; however, only a few investigations have been conducted regarding the role of Universal Stress Proteins. Members of the USP Family have a conserved domain with 130-160 residues (Bhuria et al., 2019).

External stress response is initiated in plant by specific receptors that sense external stress and pass on the signal into the internal components of plant resulting in activation or expression of defense/ protective molecules that act as a shield to plant against the environmental stress (Chi et al., 2019). The domain of USP family (Pfam code: PF00582) folds into an α/β subdomain structure which is required for a number of signaling processes in cells that are associated with cellular defense (Kerk et al., 2003; Tkaczuk et al., 2013) and many stress-resistant metabolic pathways (Persson et al., 2007; Sen et al., 2019). All USP proteins belonging to USP family have one or more USP domains and in addition can have other catalytic motifs, which have differential expression under different parts of plant, under various stages of development or under various stress conditions (Li et al., 2010; Wang et al., 2017).

Expression of a nodule-enhanced gene, AsD243, was observed from roots of *Astragalus sinicus* (Chinese Milk Vetch) infected with leguminous bacterium, *Mesorhizobium huakuii* (having nodule inducing properties). AsD243 gene was noted to have an enhanced expression which indicated the role of AsD243 in nodule development. A significant enhancement in expression was noticed after 7 days of inoculation i.e. 2 days after the onset of leghemoglobin expression in nodules of *A. sinicus*. The gene for AsD243 encodes a protein with molecular weight of 20kD and is related to bacterial USP family and has a high similarity to ATP-binding USPs (the MJ0577-type) of bacteria (Chou et al., 2007).

Phosphorylation of USPs has been shown to render protection to plant against pathogenic attack through activation of defense signaling. Phosphorylation of AtPHOS32 and AtPHOS34 (two USPs) by mitogen-activated protein kinases (MAPKs) has been observed after Arabidopsis cells were elicited by treatment with *Phytophthora infestans* zoospores or flagellin-22 (bacterial eliciting peptide) (Lenman et al., 2008; Merkouropoulos et al., 2008).

Observations of such studies strongly suggests the role of plant USPs and their upregulation in response to environmental stresses. Functions and molecular characteristics of USPs are predominantly unknown, and future studies to gain insights into the function of USPs in defense against specific biotic as well as abiotic stresses and the accompanying molecular mechanisms. While the physiological roles and diversity of structures of USPs in microorganisms have been broadly investigated, only a few studies have in plants (though having numerous USPs) have been conducted so far (Chi et al., 2019).

Arabia et al. (2021) comprehensively characterized USPs in *Oryza sativa* L. through in-silico analysis. Profiling of gene expression under various conditions of abiotic stress revealed multiple OsUSP genes with differential expression suggesting their involvement in stress-specific modules. OsUSP12 had significant upregulation under conditions of salinity, heat, drought, dehydration and treatment with GA3.

Gou et al. (2020) spotted a novel USP (MfUSP1) from *Medicago falcata* which was induced under multiple abiotic stresses. MfUSP1 over-expression led to an enhanced tolerance to various stresses such as osmotic stress, freezing, salinity and oxidative stress induced by methyl viologen (MV). It was inferred that MfUSP1 enabled multi-stress tolerance to the plant through regulation of antioxidant defense and accumulation of proline under conditions of abiotic stress.

Bhuria et al. (2019) performed a study in which 53 proteins having USP domain were identified in the genome of *Arabidopsis thaliana*. AtUSP9 and AtUSP12 genes showed response to multiple stresses in both roots as well as shoot tissues. AtUSP9 was induced under infection with various pathogens (biotic stress) and elicitor treatments. Both proteins, AtUSP9 and AtUSP12 have the potential of being engineered into plants to improve tolerance against different stresses.

Melencion et al. (2017) reported the function of an *Arabidopsis thaliana* USP- AtUSP (At3g53990) in cold stress. Over-expression of this gene lead to an enhanced tolerance to cold stress due to its function as RNA chaperone. AtUSP localization was noticed in the nucleus and cytoplasm and had ability of binding with DNA (both single and double stranded) and luciferase mRNA and also had a strong nucleic acid melting activity. Expression of AtUSP in cold-sensitive *E. coli* mutant (BX04) conferred cold and freezing tolerance to the mutants.

Udawat et al. (2016) overexpressed a cytosolic abiotic stress responsive USP from *Salicornia brachiata* in tobacco plant and compared transgenic and wild-type tobacco plants exposed to salt and osmotic stress. Overexpression of USP gene enhanced growth, reduced accumulation of ROS (reactive oxygen species), maintained ion homeostasis and improved the physiological condition of tobacco plant under the stresses.

Elhosseiny et al. (2015) investigated the involvement of USP-A (from *Acinetobacter baumannii*) in combating the oxidative stress response induced by H₂O₂, and low pH and also involvement in the development of pneumonia and sepsis, through cloning and expression in *E. coli*. *A. baumannii*, pathogen causing infection in blood, urinary tract, and lungs, is becoming increasingly resistant to many antimicrobials. USP-A of *A. baumannii* was found as an important therapeutic target for designing novel anti-microbial therapeutics.

Attia et al. (2013) determined host response to staphylococcal infection and studied *Staphylococcus aureus* virulence factors through proteome analysis of *Staphylococcus aureus* abscess from wild-type and neutropenic mice. USP2 was found to be abundant in lesions of kidney in neutropenic mice was needed for response of *S. aureus* to stringent stress induced by antibiotic Mupirocin.

Seifart Gomes et al. (2011) studied three homologous proteins having USP-A domain in *Listeria monocytogenes* under various stress conditions and discovered that these USPs were involved in stress response and were associated with survival as tested for growth *in vitro* and *in vivo*. Three USP-A deletion mutants were exposed to sublethal concentration of hydrogen peroxide (oxidative stress) or acidic conditions. Two mutants showed a reduced tolerance to acidic stress, while all the mutants had a substantially reduced growth under acidic conditions.

2.12 *Agave sisalana* and its Applications

Agave sisalana is a monocarpic herbaceous plant with a rosette of sword-shaped leaves that can grow up to 1.5–2 meters length and a short, stout stem. The plant typically yields 200–250 leaves that are suitable for commercial use, with each leaf having about 1000 fibres on average. Only 4% of the plant's weight is made up of fibres (Kavin & Janagan, 2019). Roots can extend up to 35 cm deep. It is a short-lived perennial plant with an annual growth cycle of 150–270 days. Depending on the temperature, it can be harvested after 2-4 years of planting and can continue to survive for up to 7–12 or even 20 years (EcoPort, <http://www.EcoPort.org>). *A. sisalana* mostly reproduces via asexual reproduction through suckers that grow from the rhizomes and bulbils (Monja-Mio et al., 2019). It is one of the monocarpic plants that forms an inflorescence after 6 to 9 years before dying (Asfaw, 2011).

There are around 200 species of *Agave*, most of which are found in desert, dry, and semiarid conditions in tropical and subtropical regions of the world (Davis et al., 2011). In Central America, *Agave* exists in many varieties, and is widely used and cultivated (Gross et al., 2013). *A. sisalana* is predominantly grown on an extensive scale in Mexico (where it is native), Central America, Brazil, Asia, and Africa (Shahzad et al., 2022).

The genus *Agave* has significance in terms of social as well as economic factors. *A. sisalana*, sometimes referred to as sisal, is an important source of hard, biodegradable fiber for a number of industrial uses (i.e. providing raw materials for the cellulose, automotive, and pharmaceutical industries). The manufacturing of composite materials for autos, furniture, construction, plastics, and paper products makes use of the robust and coarse fiber of *A. sisalana* (Botura et al., 2011; Magalhães et al., 2017). It is the sixth-most significant fiber crop in the world, and *A. sisalana* is the only source of nearly 75% of the world's hard natural fiber production (Shahzad et al., 2022). Fibers from processed sisal are used to production of hats, bags, carpets, twine, and rope (Coutinho et al., 2006). Extract from *A. sisalana* has applications in the pharmaceutical industry owing to its constituents having anti-inflammatory, antimicrobial, and anthelmintic activities (Botura et al., 2011; Magalhães et al., 2017).

Agave sisalana and other *Agave* plants have garnered a lot of interest recently for their potential as a feedstock for biofuels on a worldwide scale. Leftover from *A. sisalana* fiber industry is typically dumped in open ponds and utilized as fertilizers. *Agave* may thrive in marginal soils that are unsuitable for food production due to its high productivity in dry and semi-arid climates, drought resistance features, and high water usage efficiency (Vuorinne et al., 2021).

All the plants in the genus *Agave* are obligatory CAM plants. Beyond nutrient and water use efficiency, CAM physiology has advantages in the context of biofuels. CAM plants' highly

soluble nonstructural carbohydrate reserves and low lignin content would consume less energy for production of fuels, potentially producing feedstock of higher quality (Davis et al., 2011).

2.13 Bole Rot Disease Caused by *Aspergillus* sp.

Agave sisalana has tolerance to multiple stresses including drought and is resistant to many pathogens; however, it can be attacked by fungal disease known as bole rot which can cause damaging effect on sisal and limit its production (Coutinho et al., 2006). Sisal bole rot disease is caused by black aspergilli, including *Aspergillus welwitschiae*, *Aspergillus niger*, *A. brasiliensis* and *A. tubingensis* and may lead to plant death (Coutinho et al., 2006; De Souza et al., 2021; Magalhães et al., 2017).

Bole rot disease causes the aerial parts to appear yellowish and red-colored appearance of the bole rot (Santos et al., 2014)- the internal tissues turn brown, bordered by a reddish periphery. After a period of 2-3 months of infection, the rot can penetreate into the whole area below the meristem. Leaves appear yellow, crumple and eventually disintegrate when the bole is entirely rotten. In the production areas, incidence of bole rot disease varies from 5-40% (Coutinho et al., 2006).

In the recent years, bole rot has gained significant attention as it is among the factors that cause a downturn in the production of sisal around the world (Santos et al., 2014). Cultivation of sisal is significantly important from the socio-economic perspective; thus, strategies for countering the disease are being prompted (Magalhães et al., 2017).

De Souza et al. (2021) screened a total of 497 isolates (from endophytic bacteria from various areas of normal and diseased *A. sisalana* plants) and tested them as potential biocontrol agents using *A. sisalana* discs; among those tested, nine were chosen on the basis of direct antagonistic activity. Nine isolates harboring in the genera *Bacillus*, *Brevibacterium*, *Burkholderia*

(2 isolates), *Paenibacillus*, *Pseudomonas* and *Serratia* (3 isolates) lead to a significant reduction in incidence of bole rot in two field experiments.

Damasceno et al. (2019) explored *Penicillium citrinum* as a potential biocontrol agent for sisal bole rot. *P. citrinum* was inoculated into the dual culture medium 72 hours prior to pathogen (*A. welwitschiae*) and led to inhibition of mycelium growth of pathogen by 65.8%. Moreover, upon inoculation in *A. sisalana* 48 hours prior to pathogen inoculation, it was able to diminish bole rot disease up to 90%.

Barbosa et al. (2018) tested five selected bacterial isolates for their ability of controlling bole rot in vitro and in field experiments. Bacterial isolates, individually and in combinations, lead to significant inhibition of mycelial growth and germination of *A. niger* spores. Though no synergistic effect was observed in combination, these bacteria could be potentially implemented in controlling sisal bole rot in the field.

Magalhães et al. (2017) reported biocontrol activity of Burkholderia isolates (gram-negative, obligately aerobic bacteria) from a sand dune leaf litter against bole rot of sisal. Isolates were run through various tests and various enzymes, diffusible compounds and VOCs (volatile organic compounds) and siderophores were detected in dual culture experiments; isolates had ability of solubilizing phosphate and reduce mycelial growth as well as spore germination of *A. niger* in vitro. The isolates lessend severity of bole rot by 54 to 77% as determined through two independent experiments in field.

3. MATERIALS AND METHODS

3.1 Retrieval of mature miRNA Dataset

miRNA prediction from RNA-seq data of *A. sisalana* (Section 3.1-3.6) was performed as per previously described by Jabbar et al. (2021) with some modifications. A total of 10,414 mature miRNAs belonging to Viridiplantae were downloaded from the miRNA repository miRBase (release 22), accessible from <https://www.mirbase.org/cgi-bin/browse.pl> and sequence redundancy was removed to obtain 6,880 non-redundant miRNAs sub-set, including 850 miRNA sequences of plant pathogen responsive miRNA families reported in literature (Table 3.1), which were proceeded for miRNA prediction from *A. sisalana* transcriptome.

Table 3.1: Summary of pathogen responsive miRNAs, fetched from literature, belonging to various miRNA families

Sr. No.	miRNA Family	Type of Pathogen	Reported by	Number of Non-Redundant miRNAs Fetched from miRbase
1	miR156	Fungus	(Xin et al., 2010)	108
2	miR159	Bacteria/ Fungus/Virus	(Zhang et al., 2011) (Wagaba et al., 2016; Xin et al., 2010)	71
3	miR160	Fungus	(Sun, 2012; Zhao et al., 2012)	41
4	miR164	Fungus	(Hu et al., 2020; Wang et al., 2018)	43
5	miR167	Fungus	(Salamon et al., 2021; Zhao et al., 2020)	66
6	miR171	Virus, Fungus	(Ai et al., 2011; Gupta et al., 2012; Simón-Mateo & García, 2006)	92
7	miR172	Fungus	(Li et al., 2014)	56
8	miR319	Virus, Fungus	(Jelly et al., 2012; Van Vu et al., 2013; Wu et al., 2020)	50
9	miR393	Fungus	(Pinweha et al., 2015)	26
10	miR396	Fungus	(L. Chen et al., 2015; Soto-Suárez et al., 2017)	58
11	miR398	Fungus	(Li et al., 2014)	42
12	miR408	Fungus	(Feng et al., 2013)	33
13	miR444	Fungus	(Gupta et al., 2012)	8
14	miR472	Bacteria, Fungus	(Boccaro et al., 2014; Chen & Cao, 2015)	6

15	miR482	Fungus	(Lu et al., 2008; Yin et al., 2012)	113
16	miR1138	Fungus	(Gupta et al., 2012)	1
17	miR1444	Fungus	(Chen & Cao, 2015)	3
18	miR1447	Fungus	(Chen et al., 2012)	1
19	miR1448	Fungus	(Chen et al., 2012; Yin et al., 2012)	1
20	miR1450	Fungus	(Zhao et al., 2012)	1
21	miR1917	Fungus	(Yin et al., 2012)	1
22	miR2118	Fungus	(Yin et al., 2012)	30
23	miR5300	Fungus	(Ouyang et al., 2014)	1
Total				852

3.2 RNA-Seq Data Resource

Agave sisalana transcriptome (RNA-seq) assembly was accessed from *A. sisalana* RNA-seq data reported by Sarwar et al. (2019) (raw data i.e. paired-end reads sequenced on Illumina HiSeq 2500 platform, accessible from NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>) with SRA accession IDs: SRR5137659, SRR5137661, SRR5137662, SRR5137658, SRR5137663, and SRR5137660). The transcriptome assembly included 67,328 non-redundant unigenes (Sarwar et al., 2019). A local database was generated with 67,328 unigenes for onward standalone-BLASTN analysis.

3.3 Standalone BLAST-N Analysis

Homology based identification of miRNAs was performed by BLASTN (standalone BLAST v2.8.1) by using 6,880 unique (non-redundant) miRNA sequences as query against the local database created from sequenced transcripts of *A. sisalana*, while keeping E value at 10, word size as 7, scoring (match: 2, mismatch: -3), gap cost (open: 5, extend: 2) and selecting low complexity filter. From the BLAST results, those with full (100%) query coverage were selected, having not more than 2 (i.e. less than 3) nucleotides mismatch were filtered for further processing.

3.4 Removal of Protein Coding Transcripts

From the transcripts giving positive hits, sequences 100 bp upstream and downstream of target sites were retrieved and exported in FASTA format. These sequences were used as a query in standalone BLASTx (v2.8.1) against non-redundant protein database while keeping e-value at 10, word size 6, scoring (match: 2, mismatch: -3), gap cost (open: 11, extend: 1), BLOSUM62 matrix and standard genetic code. From the output, sequences with no hits (i.e. non-coding sequences) were collected in FASTA format as potential pre-miRNA (miRNA precursor) sequences.

3.5 Prediction of pre-miRNA Stem-Loop (Secondary) Structures

Mfold accessible from <http://www.unafold.org/mfold/applications/rna-folding-form-v2.php> (RNA folding form v2.3) was implemented for the prediction of pre-miRNA secondary structures to rule out pseudo pre-miRNAs. As input parameters, linear RNA sequence was selected, folding temperature was set at 37 °C, ionic conditions of 1M NaCl without divalent ions was used, 5 percent as sub-optimality number, maximum interior/bulge loop size of 30, and energy dot plot was turned on while any other parameters were set to default.

3.6 Selection of Stable Pre-miRNA Secondary Structure

For each pre-miRNA, %G+C and A+U content, ΔG values (Minimum free energy, unit: - kcal/mol) of stem-loop structures, Adjusted Minimum Folding Energy (MFE of a 100-nucleotide length) and Minimum Free Energy Index (MFEI) were evaluated as mentioned below.

$$AMFE = MFE / LP \times 100$$

$$MFEI = AMFE / (G + C) \%$$

(Where: AMFE: Adjusted Minimum Folding Energy, MFE: Minimum Folding Energy and LP: Length of Precursor Sequence)

Each predicted structure was evaluated as per criteria specified by Jabbar et al. (2021) to find stable pre-miRNA secondary structures. miRNA sequences with stable pre-miRNA secondary structures were predicted as miRNA candidates to be further confirmed through wet-lab experimental validation analysis. Overview of miRNA prediction methods from *A. sisalana* RNA-seq data is presented in Figure 3.1.

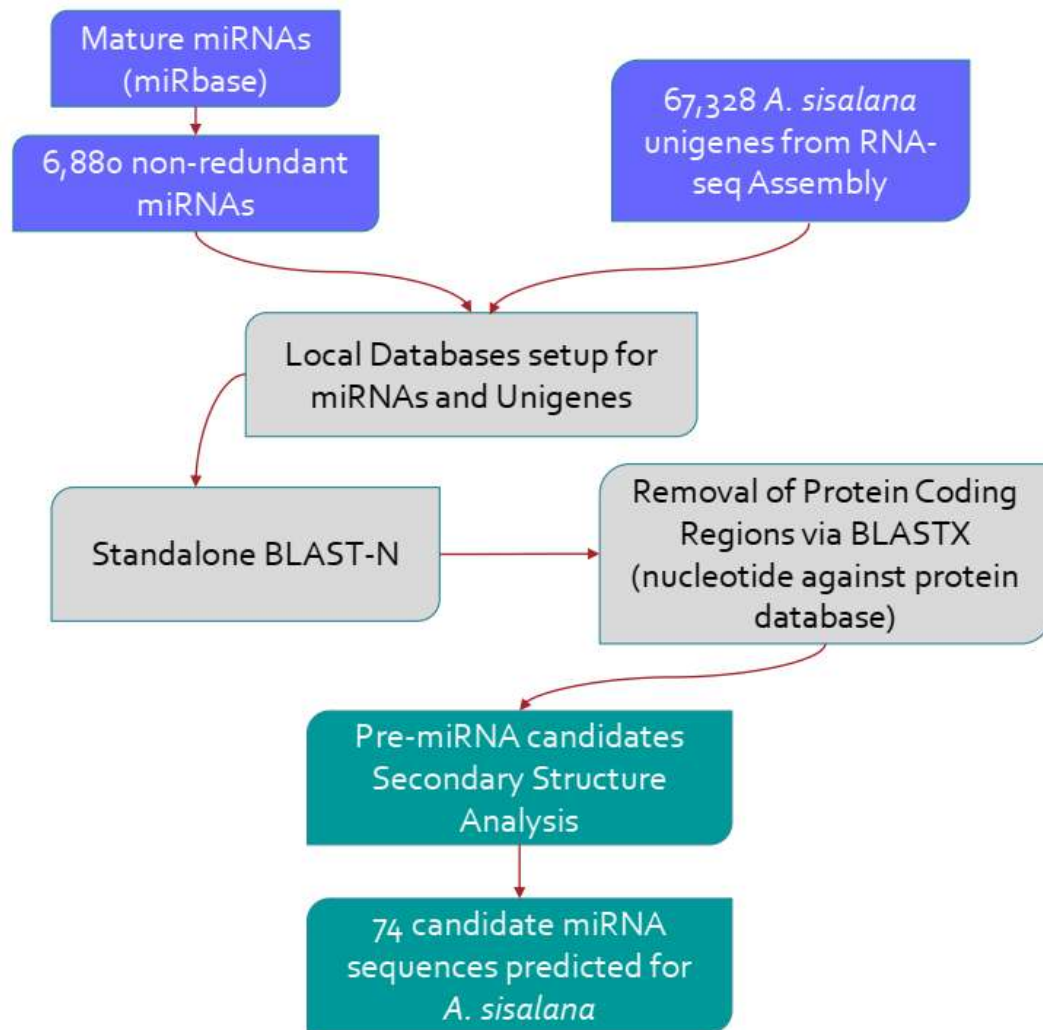


Figure 3.1: Overview of pipeline for prediction of *A. sisalana* miRNAs from RNA-seq assembly

3.7 Growth of *Agave sisalana* in Greenhouse

Agave sisalana grown in greenhouse at Centre of Excellence in Molecular Biology, Lahore, Pakistan. The plants were kept under conditions of $25\pm 2^{\circ}\text{C}$ and $65.5\pm 5\%$ relative humidity and watered regularly. Biotic stress was induced in *A. sisalana* by infecting with *A. niger* and the effect of biotic stress on various plant parameters was assessed as per methods outlined in Section 3.37-3.41.

3.8 Tissue Sampling for RNA Extraction

The leaf area targeted for excision was sprayed with 70% ethanol and wiped off the dust with clean blotting paper. The central leaf region (about 2g) was excised with sterile surgical blade and wrapped with Aluminium foil immediately and placed in liquid Nitrogen flask, for at least 10 minutes for proper sample freezing. Mortar and pestle was pre-sterilized by autoclaving (121°C , 15psi for 15 min), dried and cooled at -20°C . Sample was ground thoroughly using mortar and pestle to produce a fine powder to be used for RNA extraction.

3.9 Isolation of Total RNA from *A. sisalana*

β -mercaptoethanol (1ml/50ml of extraction buffer i.e., 2% v/v) was added to extraction buffer (Appendix-I) (pre-heated at 65°C for 5-10 min in water bath) and the resultant solution was placed in water bath at 65°C for 5-10 min. Then, buffer with β -mercaptoethanol added (10 ml buffer/ g of sample) was mixed with finely ground sample powder. After 4-5 minute incubation at room temperature, the mixture was vortexed for cell lysis (until the mixture became homogenized). To the sample-buffer mixture, an equal volume of pre-chilled chloroform: isoamylalcohol was added and vortexed for proper mixing, followed by centrifugation at 6000 rpm for 25 minutes at 4°C .

Supernatant was carefully separated and an equal volume of chloroform: isoamylalcohol was added to it and vortexed for proper mixing following by centrifugation at 6500 rpm for 30 minutes at 4°C. Again, supernatant was separated and iso-propyl alcohol equal to the volume of supernatant was added to it, inverted the tube a few times for mixing and placed at 4°C for overnight, followed by centrifugation at 12,000 rpm at 4°C for 20 min. RNA pellets obtained were processed further as mentioned in the following section.

3.10 Separation of Small and Large RNA Pool

Small RNAs (<200 bp) were isolated from Total RNA pellets with the method defined in mirVana™ miRNA Isolation Kit (Ambion, Life Technologies). Total RNA (80-100 µg) isolated from *A. sisalana* i.e. the pellets obtained in previous section were pooled and dissolved in a small volume of Lysis/Binding Buffer and made up to final volume of 750 µL with Lysis/Binding Buffer). Afterwards, 75 µL (1/10 volume) of miRNA Homogenate Additive was added to the mixture and mixed well by vortexing/ inverting the tube several times. The mixture was left on ice for 10 min. Afterwards, 1/3 volume of 100% ethanol (i.e. 1/3 of the total volume of RNA + Binding/Lysis Buffer) was added to the RNA mixture and mixed thoroughly by vortexing/ inverting the tube several times. The lysate/ethanol mixture (up to 700 µL) was pipetted onto the Filter Cartridge and centrifuged at 5000 rpm for 1 min to collect the filtrate. Centrifugation step was repeated until all the lysate/ethanol mixture has passed through the filter (at this point, filter will retain large RNA while small RNAs will pass with the filtrate through the filter; these filters were stored temporarily at 4°C till the washing step). Collected filtrates were pooled together. To the collected filtrate, 2/3 volume of 100% ethanol was added and mixed thoroughly. The filtrate/ethanol mixture (up to 700 µL) was applied onto a new Filter Cartridge and centrifuged at 5000 rpm for 1 min and discarded the flow-through. Centrifugation was repeated until all the

filtrate/ethanol mixture has passed through the filter. Filters containing large RNA and small RNA were processed further to the washing step i.e. 700 μL miRNA Wash Solution 1 was added to the Filter Cartridge and centrifuged for 1 min at 5000 rpm, and discarded the flow-through from the collection tube. Wash Solution 2/3 (500 μL) was added to the Filter Cartridge and centrifuged for 1 min at 5000 rpm and removed the flow-through. Washing step with Wash Solution 2/3 was repeated and Filter Cartridge was placed in the same collection tube and the assembly was spun 1 min at 10,000 rpm to remove residual fluid from the filter. Filter Cartridge was placed into a fresh Collection Tube and 50 μL of pre-heated (95°C) Elution Solution was applied to the center of the filter and incubated at room temperature for ~2 min, followed by centrifugation for 1 min at 10,000 rpm to recover the RNA. Collected eluates of large RNA and small RNA were stored at -80°C . RNA quantification was performed using Nanodrop Spectrophotometer, and absorbance values at 230 nm, 260 nm and 280 nm were used for assessment of RNA quality.

3.11 Polyadenylation of Small RNA Pool

A poly(A) tail was added to the small RNA fraction in poly(A) tailing reaction (50 μL) using *E. coli* Poly A Polymerase (AM2030, Invitrogen). Reaction mixture was prepared as given in Table 3.2 and incubated at 37°C for 1 hour, and held at 4°C .

Table 3.2: Reaction mixture for generation of poly(A)tailed RNA from miRNAs

Component	Volume (μL)
small RNA (~33.3 ng/ μL)	15 μL (0.5 μg)
25 mM MnCl_2	5 μL
10 mM ATP	5 μL
5X E-PAP Buffer	10 μL
2U/ μL Poly(A) Polymerase	1 μL
Ultra-Pure TM (RNase free) water	14 μL
Total	50 μL

3.12 Precipitation of Poly(A) RNA

Afterwards, the poly(A) tailed RNA was precipitated by adding double volume of ice-cold 100% ethanol and mixing several times by inversion and placed at -20°C for 30 min-1 hour. Poly(A)RNA precipitate was centrifuged at 12,000 rpm for 10 min and supernatant was discarded. Pellet was washed with 70% ethanol and centrifuged again at 12,000 rpm for 10 min and air-dried at 37°C for 5 min. Poly(A)RNA pellet was re-dissolved in 20 µL of de-ionized Ultra-Pure™ (RNase free) water.

3.13 cDNA Preparation with Reverse Adaptor

From the Poly(A) RNA transcripts, cDNA was prepared using the SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen). Approx. 1 µg of Poly(A)RNA was used in cDNA Synthesis Reaction. The following adaptor sequence was used in the cDNA synthesis reaction:

5'-CGAACATGTACAGTCCATGGATAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3'
(where V= G/A/C, and N: G,A,T,C)

For cDNA preparation, a mixture (10 µl) was prepared on ice by mixing the components as specified in Table 3.3.

Table 3.3: Reaction mixture for cDNA preparation from poly(A)tailed RNA

Component	Volume
Poly(A)RNA	7.5 µl (1 µg)
50µM Reverse Adaptor	1 µl
10 mM dNTP mix	1 µl
Ultra-Pure™ (RNase free) water	0.5 µl
Total	10 µl

The mixture was incubated at 65°C for 5 min for denaturation and placed on ice for at least 1 minute. cDNA synthesis mixture was then prepared by mixing the components on ice as mentioned in Table 3.4.

Table 3.4: Composition of the cDNA synthesis mixture

Component	Volume
10x RT Buffer	2 μ l
25mM MgCl ₂	4 μ l
0.1 M DTT	2 μ l
RNaseOUT™ (40 U/ μ l)	1 μ l
SuperScript™ III RT (200 U/ μ l)	1 μ l
Total	10 μ l

cDNA synthesis mixture was then added to the mixture containing poly(A)RNA and pipetted a few times for proper mixing, followed by cDNA synthesis reaction at 50°C for 50 min and termination at 85°C for 5 min. After a brief centrifugation, 1 μ l of RNase H was added and incubated at 37°C for 20 min for removal of remaining RNA. cDNA mixture was chilled on ice and store at -20°C. A positive control reaction was also run parallel with HeLA RNA and reverse primer of β -actin gene (5'-CAAACATGATCTGGGTCATCTTCTC-3' available in SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen)). A negative control was also run with the use of no RNA.

3.14 Primers Designing for Poly(A)RT PCR

Primers were designed for amplification of cDNA transcripts by using the miRNA sequence as forward primer (miRNA specific primer) and a reverse (adaptor) primer (common primer having complementarity to the adaptor used in cDNA synthesis from Poly(A)RNA).

Universal (reverse) adaptor sequence: 5'-CGAACATGTACAGTCCATGGATAG-3'

Table 3.5: Details of forward (miRNA specific) primers used in Poly(A)RT PCR

Sr. No.	miRNA	miRNA Specific Forward Primer Sequence	Length	Tm
1	asi-miR159b-3p	TTTGGATTGAAGGGAAGGCTCTT	23	61
2	asi-miR161	TGAAAGTGACTACATTGAGGT	21	56.7
3	asi-miR165a-5p	GACTGTTGTCTGGCTCGAGG	20	64.5
4	asi-miR166b	GGACTGTTGTCTGGCTCGAGG	21	66.5
5	asi-miR166c-5p	GGGACTGTTGTCTGGCTCGAG	21	66.5
6	asi-miR168-3p	TCTCGCCTTGCATCAAGTGAAT	22	60.8
7	asi-miR168a	TCGCTTGGTGCAGGTCGGGAA	21	66.5
8	asi-miR168a-3p	CCCGCCTTGCGCCAAGTGAA	20	66.6
9	asi-miR168b-3p	CCCGCCTTGCGCCAAGTGAAT	21	66.5
10	asi-miR168b	TCGCTTGGTGCAGGTCGGG	19	66.6
11	asi-miR168	TCGCTTGGTGCAGGTCGGGA	20	66.6
12	asi-miR168-5p	TCGCTTGGTGCAGGTCGGGAAC	22	68.3
13	asi-miR169d	TAGCCAAGGAGACTGCCTATG	21	62.6
14	asi-miR169a-3p	GGCAAGTCATCCTTGGCTAC	20	62.4
15	asi-miR169b-3p	GCAGTCTCCTTGGCTACT	18	59.9
16	asi-miR393b-3p	TCATGCGTTCCCTTTGGAATT	21	58.7
17	asi-miR395	CTGAAGTGTTTGGGGGAAGTC	21	62.6
18	asi-miR395e	TGAAGTGTTTGGGGGAAGTC	20	60.4
19	asi-miR395b-3p	AAGTGTTTGGGGGAAGTC	18	57.6
20	asi-miR395a	TGAAGTGTTTGGGGGAAGTCT	21	60.6
21	asi-miR399f	GGGAGCTCTGCTTTGGCAGG	20	66.6
22	asi-miR444b	TTGCTGCCTCAACCCTGCTGC	21	66.5
23	asi-miR528	TGGAAGGGGCATGCAGAGGAG	21	66.5
24	asi-miR535b	TGACAATGAGAGAGAGCACGC	21	62.6
25	asi-miR535-3p	GTGCTCTTTCTCGTTGTCCCC	21	64.5
26	asi-miR535d	TTGACAATGAGAGAGAGCACG	21	60.6
27	asi-miR5655	AAGAAGACACTTAAGAAGGAG	21	56.7
28	asi-miR6156	TTGAAAATTGTTCTATTTCTGT	22	51.5
29	asi-miR8005c	TTTAGAGTTTAGGGTTTAGGGTTT	24	57.7
30	asi-miR396	TCCACAGGCTTTCTTGAAC	19	58
31	asi-miR396e	TTCCACAGGCTTTCTTGAAC	20	58.4
32	asi-miR396a	TTCCACAGGCTTTCTTGAACG	21	60.6
33	asi-miR408b-5p	AGGTACGAGGCAGAGCCTG	19	64.5
34	asi-miR482e	TTACCGATTCTCCCATGCC	20	62.4
35	asi-miR482c	TTGCCCACTCCGCCATTCC	20	66.6
36	asi-miR482b-3p	TTCCCGAAACCTCCCATACCTA	22	62.7
37	asi-miR482a-3p	TTCCCGAAACCTCCCATACC	20	62.4
38	asi-miR156c-3p	TGCTTGCTCTCTATCTGTC	19	55
39	asi-miR156f-3p	TCTCTCTTCTCTGTCAT	18	51.4

40	asi-miR156a-3p	TGCTTGCGCCTCTTCTGTCAGC	22	65.8
41	asi-miR159b	TTGGATTGAAGAGAGCTCC	19	55
42	asi-miR159c	TTGGATTGAAGAGAGCTCCA	20	56.4
43	asi-miR167d-3p	GATAATGTGGTTGCTTCA	18	49.3
44	asi-miR164	GGGGAAGAGGAGCACATGAA	20	60.5
45	asi-miR164a-3p	CGTGTTCTCCTTCTCCCTC	19	59.5
46	asi-miR167c-5p	TAAGCTGCCAGCATGGATCT	20	58.4
47	asi-miR167b-3p	GTCATGCTGGAGTTGTCAT	19	55
48	asi-miR172f	AGAATCCTGATGATGCTGCA	20	56.4
49	asi-miR172b	GCAGCATTATCTAGATTCA	19	50.9
50	asi-miR172b-5p	GCAGCATCATCAGGATTCTCA	21	59.5
51	asi-miR172d-5p	AACTTCTTCAAGATTGAGA	19	48.9
52	asi-miR172c-5p	AGGATCTTCATGATTGACA	19	50.9
53	asi-miR319	TGAGTGATGGGGGAGATA	18	53.8
54	asi-miR319n	GGACCGAAGGGAGGCCCT	18	62.9
55	asi-miR319-5p	GCTGCAGATTCATTGATT	18	49.3
56	asi-miR319c-5p	GAGCTTGCTTCAGCCAC	18	58.4
57	asi-miR319d	AAACTGTTTAGTTCCACC	18	49.3
58	asi-miR398a-5p	GGGTGTTGTTTGATAACATAT	20	52.3
59	asi-miR398l-3p	ATTAGGATGCATCTCATC	18	49.3
60	asi-miR398a	TGTGTTTCTCAGGCCACCCCT	22	64.2
61	asi-miR398e	TGTGTTTCTCAGGCCACCCCT	21	63.2
62	asi-miR393-5p	TCCAAAGGGATCGCATTGAT	20	56.4
63	asi-miR393c	TCCAAAGGGATCGCATTGATCT	22	60.1
64	asi-miR398-5p	AGAGAGCTTAAGAACACAA	19	50.9
65	asi-miR444d.1	TGCTCCTTCAAGCTTGCTGC	20	60.5
66	asi-miR444b.1	TGCTGTCTCAAGCTCGCTGCC	21	65.3
67	asi-miR472b	TTTCTCAACTTCACCCATC	19	53
68	asi-miR1444d	AACGTTAAGCGAATGTGA	18	49.3
69	asi-miR1444c	TCACATTGCTTAACGTT	18	49.3
70	asi-miR2118l	CCTAAGGCTTCCCCTTCCT	19	59.5
71	asi-miR2118r	TTTCCAATGCCTCACATGC	19	55
72	asi-miR5300	CAGTCTAGGCCTTCCAAC	18	56.3
73	asi-miR2118k	CTTATGCGTCTCATTCCT	18	51.4
74	asi-miR2118b-5p	GGAGATGGGAGGTTTGGTA	19	57.5

3.15 Poly(A) RT-PCR for Validation of Predicted miRNAs

cDNA generated from HeLA RNA, was used as positive control with the following sense and anti-sense primers available in SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen) for amplification of human β -actin gene:

β -actin gene sense primer: 5'-GCTCGTCGTCGACAACGGCT C-3'

β -actin gene antisense primer: 5'-CAAACATGATCTGGGTCATCTTCTC-3

PCR reaction mixture is given in Table 3.6. A negative control reaction for PCR was also run parallel without *Agave sisalana* cDNA. RT-PCR was run with the steps as mentioned in Table 3.7.

Table 3.6: Poly(A)RT PCR reaction mixture for miRNA validation

Component	Volume
cDNA Template	3 μ l
dNTPs (2mM)	2.5 μ l
Forward Primer (10 μ M)	1 μ l
Reverse Primer (10 μ M)	1 μ l
Taq DNA Polymerase (5U/ μ l)	0.5 μ l
10x Taq Buffer (with KCl)	2.5 μ l
25 mM MgCl ₂	2 μ l
De-ionzed Ultra-Pure™ water	12.5 μ l
	Total = 25 μ l

Table 3.7: PCR profile for poly(A)RT PCR for miRNA validation

Step	Temperature	Time	Cycle
Initial Denaturation	94°C	3 min	1x
Denaturation	94°C	30 sec	40x
Annealing	set at 3-5°C lower than average T _m of primers pair	30 sec	
Extension	72°C	30 sec	
Final Extension	72°C	5 min	1x
Hold	4°C	∞	1x

3.16 Target Prediction of miRNAs Against TAIR

The targets of the identified miRNAs of *A. sisalana* were searched in the database-*Arabidopsis thaliana*, transcript, removed *MIRNA* gene, TAIR (The Arabidopsis Information Resource) v.10 under default parameters using psRNATarget webserver (<https://www.zhaolab.org/psRNATarget/analysis>). miRNAs were also checked for targeting of the USP genes of *A. thaliana* contained in the TAIR database.

3.17 Gene Ontology Analysis of Predicted Target Genes

miRNA target genes of *A. thaliana* were further analyzed for Gene Ontology using ‘GO Term Enrichment for Plants’ server available on TAIR online database and accessible from https://www.arabidopsis.org/tools/go_term_enrichment.jsp. The list of gene identifiers was input to the server in tab-delimited format and the data was sent to the PANTHER classification system (<http://pantherdb.org/>) to investigate biological processes, molecular functions, and cellular components.

3.18 Prediction of Candidate Coding Regions from *A. sisalana* RNA-seq Assembly

To find coding regions in RNA-seq (NGS) data of transcriptome of *A. sisalana* ([https://www.ncbi.nlm.nih.gov/sra/SRX2457055\[accn\]](https://www.ncbi.nlm.nih.gov/sra/SRX2457055[accn])), 67,328 unigene sequences in the transcriptome assembly were loaded in TransDecoder Module (available in Galaxy <https://usegalaxy.org/>). Parameters were selected as pre-set while minimum protein length was set at 100.

3.19 Hidden Markov Models (HMM) Search for USP Family Genes

Multiple Sequence Alignment of 62,462 protein sequences belonging to USP family (PF00582) were fetched from Pfam server (EMBL) and along with predicted complete 13,102

ORF protein sequences were loaded in Ugene (v41.0) and established Ugene workflow with HMM3 under default settings to search for proteins corresponding to USP family (Pfam code: PF00582).

3.20 cDNA Preparation from Large RNA Pool

Same method was used for cDNA preparation from large RNA pool as given in Section 3.13 with the use of large RNA fraction instead of Poly(A)RNA and oligo(dT) in place of reverse adaptor.

3.21 Primer Design for Amplification of USP Genes

Primers for amplification of full-length AsUSP genes were designed using Snap Gene software. Primer sequences for amplification of nine AsUSP genes are mentioned in Table 3.8.

Table 3.8: Primer Sequences designed for PCR amplification of AsUSP genes

Sr. No.	Primer Name	Primer Sequence	Length	T _m (°C)	Expected Product Size (bp)
1	ORF1-F	ATGCAAGACTCCGATCGC	18	56	738
2	ORF1-R	TCAAGCATCTTTATGCTCCTCATC	24	56	
3	ORF2-F	ATGCCCAGTGACGTGACAC	19	58	489
4	ORF2-R	TCACAAAGCTTCTCTCTTGCATCG	24	59	
5	ORF3-F	ATGGGAGGAGATAGGAAG	18	50	492
6	ORF3-R	TCAATGCTTGACGTTTGG	18	52	
7	ORF4-F	ATGGAGCCAGTGAACGAAGACG	22	60	438
8	ORF4-R	CTACCTTTGGAAGAGGCTCCTCC	23	59	
9	ORF5-F	ATGGCAGGCGATCGCACG	18	62	498
10	ORF5-R	CTACGCACCGTGCACTTCAGAGC	23	63	
11	ORF6-F	ATGGCAGCGGCAGTGGAG	18	61	480
12	ORF6-R	TCACCCAGAGTTCTTCACAACAAGG	25	60	
13	ORF7-F	ATGGCGGAGGCGGAGGCG	18	66	528
14	ORF7-R	TTAGTGCTTGTGCATGGGCTTGGG	24	63	
15	ORF8-F	ATGGAAATTGAAGAGATGAAGG	22	51	2421
16	ORF8-R	TCATGGATTCTTGGCTC	17	51	
17	ORF9-F	ATGGGATCGGAACCGACT	18	56	528
18	ORF9-R	TCAATCATCAACTGGATCCTGTG	23	56	
19	β-actin-F	GCCAGTGCGGTAACCAGAT	19	60	138

20	β -actin-R	CGCTCGCCTCGTTGTAGTA	19	60	
----	------------------	---------------------	----	----	--

3.22 Amplification of Universal Stress Protein Family Genes- PF00582

For amplification/ PCR reaction, reaction mixture was prepared on ice for each of the gene to be amplified as specified in Table 3.9.

Table 3.9: Reaction mixture for PCR amplification of AsUSP genes

Component	Volume
cDNA Template	3.5 μ l
PCR 2x Master Mix (Thermo Scientific)	12.5 μ l
Forward Primer (10 μ M)	1 μ l
Reverse Primer (10 μ M)	1 μ l
Nuclease free (Ultra-Pure™) water	7 μ l
Total	25 μ l

PCR reaction was carried out according to the following profile with optimized annealing temperature for each of the primer pairs and extension time set according to the base pair size of expected PCR product (i.e. 60 sec/kb). Primers for β -actin (reference gene) of *A. sisalana* were used for the positive control reaction. PCR profile for amplification of AsUSP genes is given in Table 3.10.

Table 3.10: PCR profile for amplification of AsUSP genes

Step	Temperature	Time	Cycle
Initial Denaturation	94°C	3 min	1x
Denaturation	94°C	40 sec	35x
Annealing	set at 3-5 °C lower than average T _m of primers pair	35 sec	
Extension	72°C	set according to expected length of PCR product (1min/kb)	
Final Extension	72°C	10 min	1x
Hold	4°C	∞	1x

3.23 Agarose Gel Electrophoresis

Agarose gel electrophoresis (in range of 1%-2% w/v) with 1x TBE buffer, prepared from 5x TBE buffer (Appendix-II) was used for assessment of RNA integrity as well as for checking the PCR products after amplification and for checking the products size of restriction digestion. Ethidium bromide (10 mg/ml) was used for staining the gel. After running gels at 90 V for 1 hour, results were visualized under UVP Benchtop UV Transilluminator. Thermo Scientific™ GeneRuler® 1kb DNA Ladder (Cat. No. SM0312), Thermo Scientific™ GeneRuler® 100 bp DNA Ladder (Cat. No. SM0241) and Thermo Scientific™ GeneRuler® 50 bp DNA Ladder (Cat. No. SM0371) were used as markers depending on the expected band sizes for product size estimation.

3.24 Purification of PCR Product

Bands from Agarose Gel were excised and weighed. Gel extraction was performed using protocol of Thermo Scientific GeneJET Gel extraction kit (K0692). Binding buffer (1:1 volume: weight) was added to each of the gel slice and incubated at 50-60°C for 10 min or until the gel slice was completely dissolved (tubes mixed by inversion). For DNA fragments ≤ 500 bp, 1 gel volume of 100% isopropanol was added to the solubilized gel solution. About 800 μ L Gel solubilized solution was transferred to GeneJET purification column and centrifuged at 12,000 rpm for 1 min. Flow-through discarded and column as placed back into the same collection tube. Repeated centrifugation until all of the Gel solubilized solution has passed through the column. Binding buffer (100 μ L) was then added to the GeneJET purification column and centrifuged at 12,000 rpm for 1 min. Flow-through discarded and column as placed back into the same collection tube. Wash Buffer (700 μ L) was then to the GeneJET purification column and centrifuged at 12,000 rpm for 1 min. Flow-through discarded and column as placed back into the same collection

tube. Then centrifugation was repeated for an additional 1 min to completely remove residual wash buffer. Then GeneJET purification column was transferred into a clean 1.5 ml microcentrifuge tube and 20-25 µL of Elution Buffer was added to the center of the purification column membrane and centrifuged at 12,000 rpm for 1 min to collect the eluate. Purified product was used for ligation and the rest was stored at -20°C.

3.25 Cloning with TA Cloning® Kit Dual Promoter (pCR®II)

The cleaned/eluted PCR products were run through Nanodrop Spectrophotometer to determine concentration of each of the PCR product. Concentration of the insert to be used for cloning was calculated as mentioned below:

$$\text{Concentration of insert} = \frac{\text{size of insert (bp)} \times \text{concentration of vector (ng/}\mu\text{L)} \times 3}{\text{size of vector (bp)}}$$

Ligation of insert into the vector as performed using TA Cloning™ Kit, Dual Promoter, with pCR®II Vector (map given in Figure 3.1). Reaction mixture (10 µL) for ligation of PCR products with pCR®II vector was prepared on ice with components as outlined in Table 3.11. Plasmid map of pCR®II vector including restriction sites, origin of replication and other details is presented in Figure 3.2.

Table 3.11: Reaction mixture for ligation of PCR products with pCR®II vector

Component	Volume
Fresh PCR product	x µL
5X ExpressLink™ T4 DNA Ligase Buffer	2 µL
pCR®II vector (25 ng/µL)	2 µL
Water to a total volume of 9 µL	y µL
ExpressLink™ T4 DNA Ligase (5 units)	1 µL
Total	10 µl

The mixture was incubated at room temperature for overnight and kept at -20°C to be used later for transformation into competent cells.

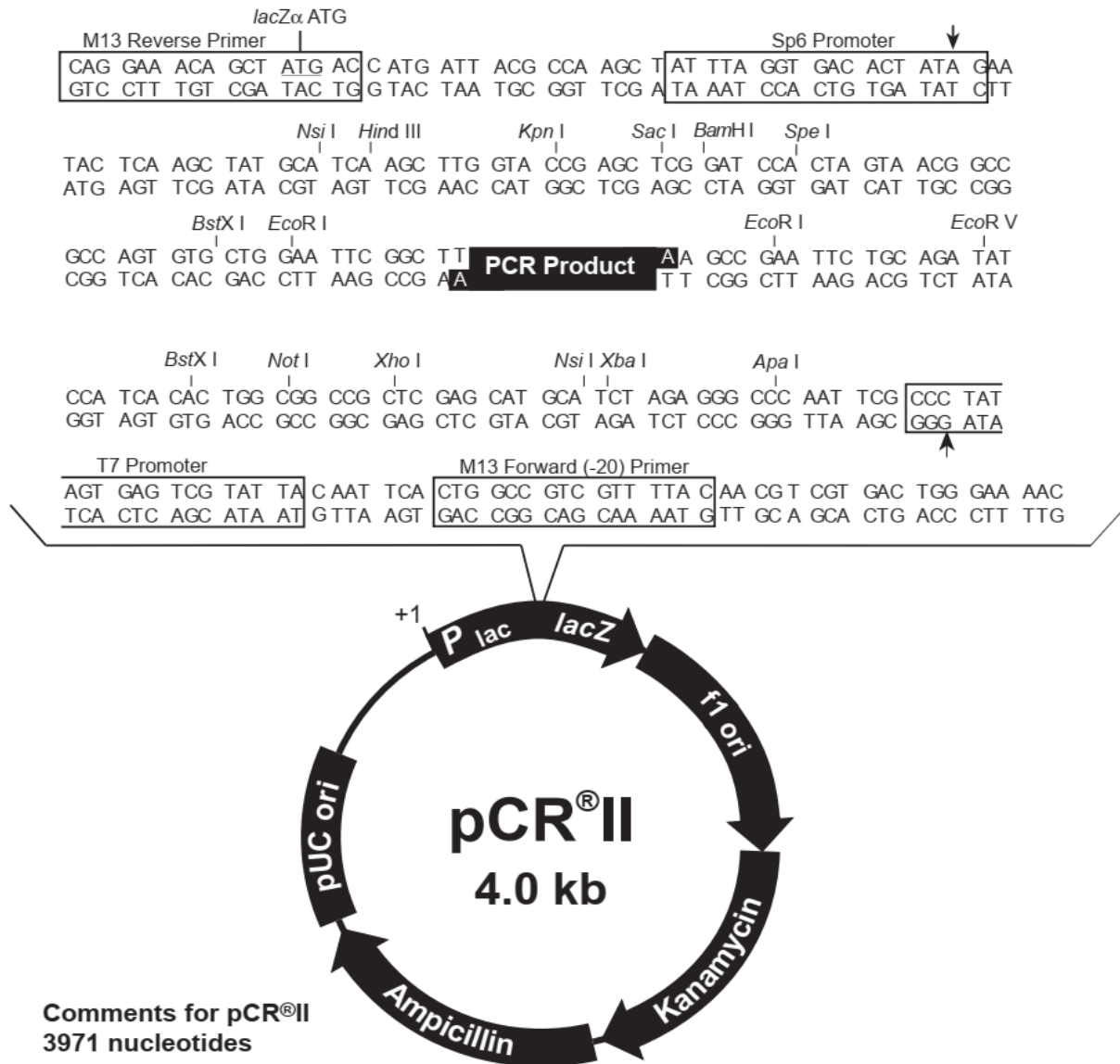


Figure 3.2: Plasmid map of pCR®II vector including restriction sites, origin of replication and other details

3.26 Preparation of *Escherichia coli* TOP 10 Competent Cells

Escherichia coli Top 10 strain was streaked on LB agar (Appendix-III) plate and incubated at 37°C for overnight. A single colony was picked from Top 10 plate and suspended in 5ml LB broth (Appendix-IV) having tetracycline (with final concentration of 12.5 µg/ml from 12.5 mg/ml sterile stock solution) (Appendix-V). The culture was incubated in shaking incubator at 37°C for overnight. Afterwards, the overnight culture was sub cultured (1:100) for 3-4 hours under the same conditions as used for the previous culture, followed by centrifugation at 3000 rpm for 10 min at 4°C. Supernatant was discarded and pellet was resuspended in 5 ml transformation buffer (Appendix-VI). This suspension was kept on ice for 30 min and centrifugation was again performed at 3000 rpm for 10 min at 4°C. Supernatant was discarded and pellet was again resuspended in 5 ml transformation buffer and kept on ice for 30 minutes. Afore-mentioned centrifugation was repeated, supernatant was discarded and glycerol + transformation buffer mixture (i.e. 300µl of 80% glycerol + 700µl transformation buffer) was added to dissolve the pellet. Aliquots of 100µl competent cells were formed and stored at -80°C for long term use.

3.27 Transformation of Recombinant Vector in Competent *E. coli* TOP10

Competent cell aliquot was taken, and cells were thawed on ice. In a 1.5 ml microcentrifuge tube, 2µl of ligation mixture was added and 100µl of competent cells were added the mixture followed by gentle mixing by pipetting. The mixture was placed on ice for 45 min and then subjected to heat shock at 42°C for 90 sec in water bath. LB broth (900µl) was poured into 15 ml falcon tube and placed on shaking incubator at 37°C at 250rpm for 60 min. 100 µl of mixture was spread on agar plate and incubated for overnight at 37°C.

3.28 Blue-White Screening

X-gal (50 μ l) was added into 100 ml LB media and poured into petri plates with selected anti-biotics i.e., ampicillin (to a final concentration of 100 μ g/ml from 100 mg/ml sterile stock solution) (Appendix-VII) and tetracycline (to a final concentration of 12.5 μ g/ml). The colonies formed by recombinant cells appeared as white while colonies produced by non-recombinant cells appeared blue; thus, the desired recombinant colony was picked and cultured.

3.29 Plasmid Extraction with GeneJET Plasmid Miniprep

GeneJET plasmid miniprep kit (catalog no. K0502, Thermo Fischer) was used for extraction of recombinant plasmids from *E. coli* TOP10. After blue-white screening, a white colony was picked from the freshly streaked transformed cells plate and inoculated in 5ml sterile LB broth with tetracycline (with final concentration of 12.5 μ g/ml from 12.5 mg/ml sterile stock solution) and ampicillin (to a final concentration of 100 μ g/ml from 100 mg/ml sterile stock solution). The culture was incubated in shaking incubator for 16-20 hours at 37°C, and 3 ml culture was transferred to microcentrifuge tube and centrifuged at 13,000 rpm to pellet down the cells.

The pellet was resuspended in 250 μ l of resuspension buffer and dissolved properly until no clumps remained. Lysis solution (250 μ l) was then added and mixed thoroughly by inverting the tube 5-6 times until the solution became clear and viscous. Neutralization solution (350 μ l) was added to the tube and mixed immediately 5-6 times and centrifuged at 13,000 rpm for 5 min to pellet down chromosomal DNA and cell debris. Supernatant was transferred to spin column carefully, not disturbing the white precipitates and centrifuged at 12,000 rpm for 1 min. Flow-through was discarded and column was placed back into the collection tube. Wash Solution was added into the spin column and centrifuged for 1 min at 12,000 rpm. The washing step was repeated and then centrifugation was repeated for 2 minutes after discarding the flow-through to

remove residual ethanol. The column was shifted to a fresh microcentrifuge tube and added 30 μ l of elution buffer and centrifuged at 12,000 rpm for 2 min to collect the eluate containing purified plasmid and stored at -20°C.

3.30 Restriction Digestion with *EcoRI*

To confirm the proper ligation of genes into the pCR®II vector, restriction analysis was performed with enzyme *EcoRI*. Restriction digestion mixture (Table 3.12) was prepared and incubated at 37°C for 15 min.

Table 3.12: Restriction digestion mixture for *EcoRI* digestion of the recombinant plasmids

Component	Volume
Plasmid	7 μ l
<i>EcoRI</i>	1 μ l
<i>EcoRI</i> buffer	1 μ l
DEPC water	1 μ l
Total	10 μ l

3.31 Sanger Sequencing Analysis

The recombinant plasmids, harboring the genes of interest, were submitted to the sequencing lab for Sanger Sequencing using ABI PRISM 3100 Sequencer Genetic Analyzer. Sequencing output/results were analyzed using Ugene software (v41.0).

3.32 Target Prediction of miRNAs Against AsUSPs

To predict miRNA targets among AsUSP gene sequences, a consensus approach was implemented whereby the three tools RNAhybrid, psRNATarget and psRobot were run and the common hits were recorded.

3.32.1 RNAhybrid

For prediction of *A. sisalana* miRNA targets in AsUSP gene sequences, standalone version of the Tapir software (<http://bioinformatics.psb.ugent.be/webtools/tapir/tapir-1.2.tar.gz>) was procured and run in Linux OS with score value as 10 and free energy ratio as 0.4 using the precise mode of Tapir i.e. RNAhybrid, while sequences of miRNAs and genes were input in FASTA format.

3.32.2 psRNATarget

Both asi-miRNAs and AsUSP gene sequences (in FASTA format) were submitted to psRNATarget web-server (<https://www.zhaolab.org/psRNATarget/analysis>) and were run through the algorithm with the following settings: expectation = 8, penalty for opening gap = 2, penalty for extending gap = 0.5, penalty for G:U pair = 0.5, seed region = 2–13 nt, extra weight in the seed region = 1.5, mismatches allowed in the seed region = 2 and HPS size = 19.

3.32.3 psRobot

psRobot standalone (v1.2) was downloaded from psRobot server (http://omicslab.genetics.ac.cn/psRobot/program/WebServer/psRobot_v1.2.tar.gz) and the required script was run on Linux OS with the following parameters: target penalty score threshold=4, three prime boundary of essential sequence=8, position after which with gap/bulge permit=1, number of gaps/bulges permitted=2, while other settings were kept to default.

3.33 miRNA 3D Structure Modelling (MC-Fold)

MC-Fold (<https://major.iric.ca/MC-Fold/>) was used for building the 3D structure of each of the miRNA, identified through consensus of target prediction results. The whole sequence of miRNA in FASTA format was submitted to the server, under default parameters, and the top-most MC-Sym formatted structure in MC-Fold was proceeded for 3D structure prediction.

3.34 3D Structure Modelling of AsUSP Genes mRNA (RNA Composer)

Target AsUSP sequence in FASTA format was trimmed 100 bp up-stream and down-stream of the predicted miRNA target start position, and the resultant sequence was fed to RNAComposer (<https://rnacomposer.cs.put.poznan.pl/>) for 3D structure modelling of mRNA with RNAfold selection (for 2D structure to be used for 3D structure building) in RNAComposer.

3.35 miRNA-mRNA Molecular Docking Analysis

Molecular Docking for each of miRNA-mRNA target pair, predicted from consensus of the three target prediction algorithms, was carried out using Haddock server (<https://wenmr.science.uu.nl/haddock2.4>) under default parameters. 3D structures of miRNA and mRNA were input to the Haddock server with information of binding site in the receptor i.e., mRNA structure.

3.36 Multiple Sequence Alignment of pre-miRNAs and Phylogenetic Analysis

Sequences of miRNA precursors of asi-miRNAs targeting the *A. sisalana* USP genes were further studied for their evolutionary relationship with homologous miRNAs (in the same miRNA family) of other species. Sequences pre-miRNAs in FASTA format were first aligned through Clustal omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and the alignment file was imported in MEGA-X (Kumar et al., 2018). Phylogenetic trees were built subsequently by using Maximum likelihood as statistical method, bootstrap method as a test of phylogeny, bootstrap replication number as 1000, substitution model: kimura 2-parameter, uniform rates among sites and tree inference using nearest-neighbor-interchange (NNI).

3.37 Biotic Stress Induction

3.37.1 *Aspergillus niger* Culture

Aspergillus niger grown on agar plates was obtained from the culture bank of Institute of Agricultural Sciences (IAGS), University of the Punjab, Lahore. Potato dextrose agar (PDA) medium was prepared by mixing required quantities of agar, dextrose and potato extract as mentioned in Appendix-VIII and made up the final volume up to 100 ml. Media was autoclaved (121°C, 15psi for 15 min) and slants were prepared in sterile test tubes. A sterile inoculating loop was used to transfer a small amount of fungal growth from plates to PDA slants aseptically. The inoculated media were incubated in static incubator at 30°C for 7-10 days until fungal growth was observed.

3.37.2 Preparation of *Aspergillus niger* Suspension

To an *A. niger* slant, 20 ml of sterile water was added and gently mixed a few times. 10 ml of this suspension was added to 10 ml of 5% Tween-20 in a 50 ml falcon tube. The suspension was adjusted to 10^5 conidia/ml and the volume was raised up to 50 ml with sterile distilled water.

3.37.3 *Aspergillus niger* Inoculation in *Agave sisalana*

Bole region of *A. sisalana* plants (grown to 7 months of age) was cleaned with 70% ethanol using a blotting paper and 10 ml *A. niger* suspension was injected using syringe. After inoculation, plants were watered regularly and placed in growth chamber at $25\pm 2^\circ\text{C}$ and a relative humidity of $95\pm 5\%$. Inoculation was repeated after 7 days and plants were kept under same conditions of $25\pm 2^\circ\text{C}$ and a relative humidity of $95\pm 5\%$.

3.38 Estimation of Chlorophyll Content

About 100 mg of fresh leaves (from stressed/infected and control plants) were ground with 10 ml of 80% acetone and kept overnight in the dark. Afterwards, supernatant was extracted by

centrifugation at 5000 rpm, and absorbance values were noted at 663 nm and 645 nm for Chlorophyll a and b content estimation, respectively. Experiments for test and control were run in triplicates. The following equations were used for calculation of the Chlorophyll a, b and total Chlorophyll content as mg g⁻¹ FW:

$$Chla = \frac{0.0127(OD663)-0.00269(OD645)}{FW(g)} \times V(ml)$$

$$Chlb = \frac{0.229(OD645)-0.00488(OD663)}{FW(g)} \times V(ml)$$

$$Total\ Chl = Chla + Chlb$$

(Where OD 633: Absorbance at 663 nm

OD 645: Absorbance at 645 nm

V: volume of acetone (ml)

FW: Fresh weight of the tissue sample (g))

3.39 Estimation of Proline Content

Proline content (µg g⁻¹ FW of leaf tissue) was determined as described by Bates et al. (1973). Fresh leaf samples (1g) from control and stressed plants were ground with 10 ml of 3% Sulfosalicylic acid and centrifuged (or filtered through Whatman #1 filter paper) to remove residue. Filtrate (2 ml) was mixed with 2 ml of acid ninhydrin solution (Appendix-IX) and tubes were placed in boiling water bath for 1 hour. After incubation, tubes were cooled down and the contents were vigorously mixed with 4 ml of absolute toluene for 30 seconds and left at room temperature for phase separation. Absorbance at 520 nm was taken using Spectrophotometer to measure the red color intensity of upper layer, using pure toluene as blank. Experiments for test and control were performed in triplicates. A standard curve was also prepared using different concentrations of proline.

3.40 Microscopic Examination of Leaf Tissues

Leaf tissue was cleaned with 70% ethanol and excised with help of sterile blades and preserved in FAA (Appendix-X) at 4°C. Tissue was dehydrated, treated with xylene, embedded in paraffin wax and transverse sections were cut using microtomy followed by staining with H&E (hematoxylin and eosin) stain. Slides of the control and infected leaves were observed under optical microscope to observe morphological differences.

3.41 Statistical Data Analyses

Biochemical parameters were evaluated by statistical analysis for determination of statistically significant differences between plants under control and biotic stress conditions. Data was analyzed in Graphpad Prism software (v.9.0) by implementation of one-way ANOVA (Analysis of Variance) test. Variations between means were compared by using least significance difference (LSD) test; p-value 0.05, 0.01 and 0.001 were indicated with asterisks *, ** and ***, respectively while non-significant values were indicated by 'ns'.

4. RESULTS

A comparative genomics-based approach was applied on the RNA-seq data of *A. sisalana* transcriptome (accessible from NCBI, SRA: <https://www.ncbi.nlm.nih.gov/sra>). In this study 6,880 non-redundant miRNA query sequences were obtained from miRbase including 852 pathogen responsive miRNAs (retrieved by review of literature). 67,328 non-redundant unigenes assembled from the raw data of RNA-seq analysis performed on Illumina HiSeq 2500 (Sarwar et al., 2019) platform was used as transcriptome data resource for miRNA prediction.

4.1 Standalone BLAST-N Result

From the results of standalone BLAST analysis, homology-based pair-wise sequence alignments of selective miRNA query sequences (one from each miRNA family) with the respective hit transcripts (unigene sequences) are indicated in Figure 4.1. miRNA sequences were either completely aligned with perfect match with the hits or with a few mismatches. However, BLAST results indicating no more than not more than 2 (i.e. less than 3) nucleotides mismatches were only retained in the results for further analysis. miRNAs predicted were labelled arbitrarily based on the name of miRNA homologue.

Consensus 1 10 20 23
 ⚡ ath-miR159b-3p
 TRINITY_DN8805_c0_g2_i1

Consensus 1 10 20
 ⚡ aly-miR165a-5p
 TRINITY_DN42410_c0_g1_i1

Consensus 1 10 21
 ⚡ pab-miR168a
 TRINITY_DN33542_c0_g1_i1

Consensus 1 10 21
 ⚡ csi-miR393b-3p
 TRINITY_DN44123_c0_g1_i1

Consensus 1 10 20
 ⚡ pab-miR399f
 TRINITY_DN9291_c1_g1_i1

Consensus 1 10 21
 ⚡ sbi-miR528
 TRINITY_DN36065_c0_g1_i1

Consensus 1 10 21
 ⚡ ath-miR5655
 TRINITY_DN42765_c0_g1_i1

Consensus 1 10 20 24
 ⚡ stu-miR8005c
 TRINITY_DN33453_c0_g1_i1

Consensus 1 10 19
 ⚡ zma-miR408b-5p
 TRINITY_DN9858_c0_g1_i1

Consensus 1 10 19
 ⚡ mtr-miR156c-3p
 TRINITY_DN12866_c0_g2_i1

Consensus 1 10 20
 ⚡ rgl-miR164
 TRINITY_DN17932_c0_g1_i1

Consensus 1 10 21
 ⚡ ath-miR161.1
 TRINITY_DN25914_c0_g1_i1

Consensus 1 10 21
 ⚡ cas-miR166b
 TRINITY_DN42410_c0_g1_i1

Consensus 1 10 21
 ⚡ zma-miR169d
 TRINITY_DN29275_c0_g1_i1

Consensus 1 10 21
 ⚡ atr-miR395
 TRINITY_DN46165_c0_g1_i1

Consensus 1 10 21
 ⚡ tae-miR444b
 TRINITY_DN15906_c0_g2_i1

Consensus 1 10 21
 ⚡ ppt-miR535b
 TRINITY_DN18597_c0_g1_i4

Consensus 1 10 20 22
 ⚡ nta-miR6156
 TRINITY_DN30748_c0_g1_i1

Consensus 1 10 20
 ⚡ fve-miR396e
 TRINITY_DN18615_c1_g2_i1

Consensus 1 10 20
 ⚡ ppe-miR482e
 TRINITY_DN13721_c0_g1_i1

Consensus 1 10 18
 ⚡ stu-miR167d-3p
 TRINITY_DN17251_c1_g1_i1

Consensus 1 10 20
 ⚡ sbi-miR172f
 TRINITY_DN12549_c0_g1_i1

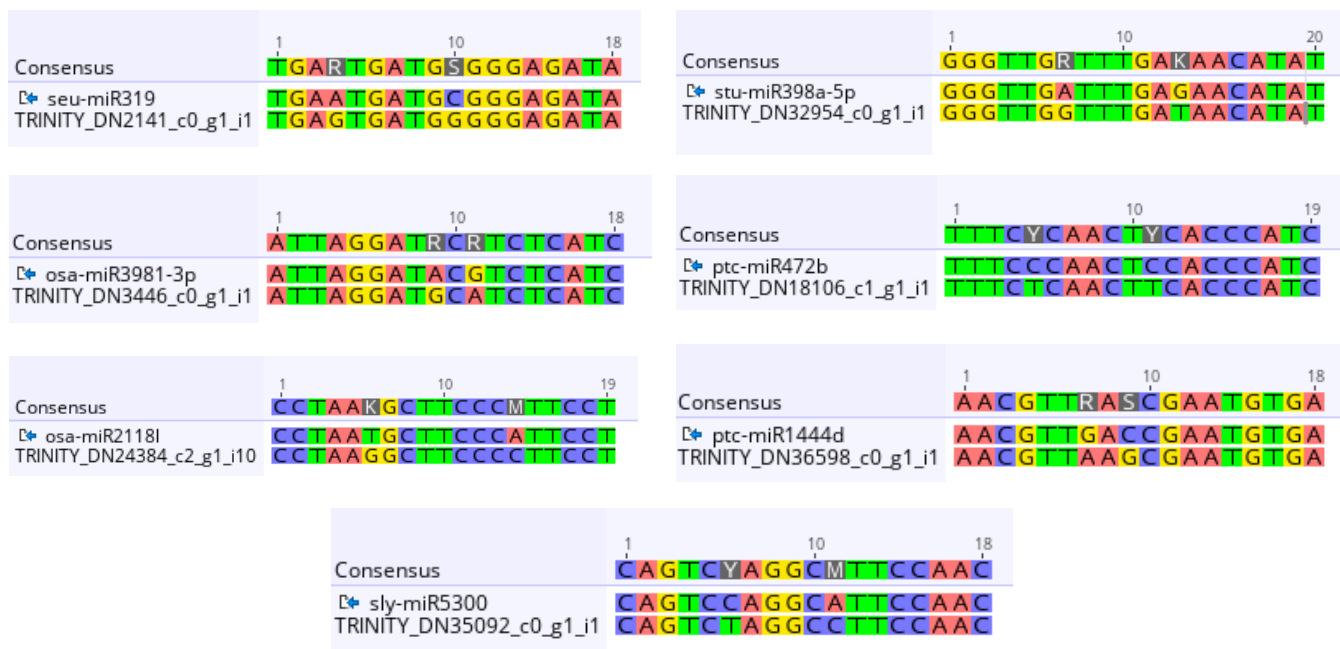


Figure 4.1: Representation of pair-wise nucleotide sequence alignments of selective query miRNAs (one from each miRNA family), with *A. sisalana* transcriptome assembly unigenes, obtained from the results of standalone BLAST-N

4.2 Pre-miRNA Secondary Structures

Pre-miRNAs were analyzed on a pre-defined criteria in order to uncover potential pre-miRNAs with stable secondary structures. Some of the most stable representative secondary structures of pre-miRNAs are shown in Figure 4.2.

Table 4.1 represents the list of 74 potential miRNAs predicted in the study including miRNA sequences, homologues, names of transcripts in RNA-seq assembly giving positive hits and various characteristics of miRNAs and pre-miRNAs.

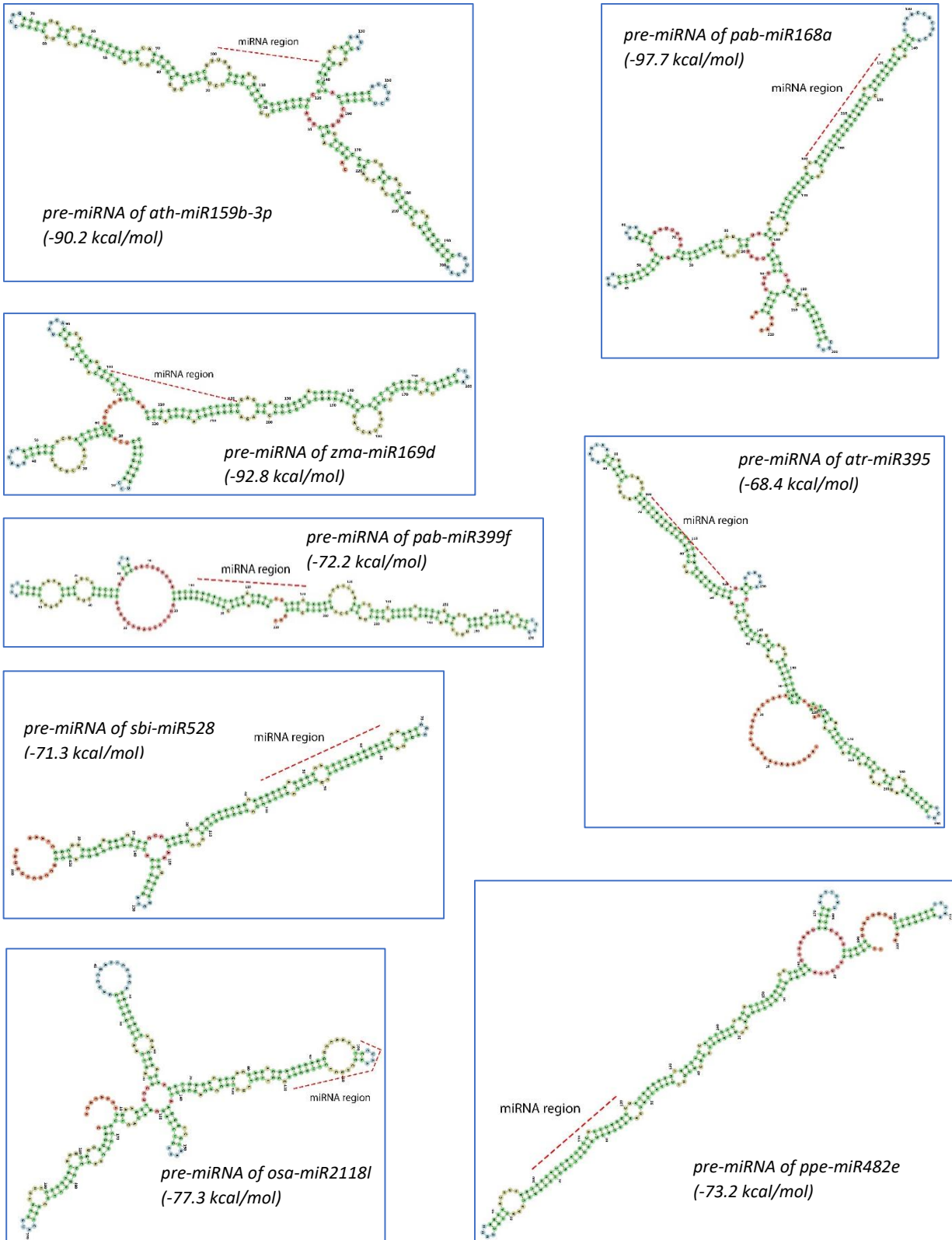


Figure 4.2: Secondary structures of selective pre-miRNAs with high stability of structures as indicated by the minimum free energy (ΔG) values

Table 4.1: List of 74 potential miRNAs predicted in the study including miRNA sequences, homologues, names of transcripts in RNA-seq assembly of *A. sisalana* giving positive hits and various characteristics of miRNAs and pre-miRNA

Sr. No.	miRNA	Sequence (5' to 3')	Homologue	Transcript	% GC Content	% AU Content	miRNA Length	pre-miRNA Length	MFE	AMFE	MFEI	Hit start	Hit end	Strand
1	asi-miR159b-3p	UUUGGAUUGAAGGGAAGGCUCUU	ath-miR159b-3p	TRINITY_DN8805_c0_g2_i1	58	42	23	221	-90.15	-40.79	0.70	227	249	+
2	asi-miR161	UGAAAGUGACUACAUGAGGU	ath-miR161.1	TRINITY_DN25914_c0_g1_i1	45	55	21	128	-40.20	-31.41	0.70	8	28	+
3	asi-miR165a-5p	GACUGUUGUCUGGCUCGAGG	aly-miR165a-5p	TRINITY_DN42410_c0_g1_i1	43	57	20	138	-43.05	-31.20	0.73	19	38	+
4	asi-miR166b	GGACUGUUGUCUGGCUCGAGG	cas-miR166b	TRINITY_DN42410_c0_g1_i1	43	57	21	138	-43.05	-31.20	0.73	18	38	+
5	asi-miR166c-5p	GGGACUGUUGUCUGGCUCGAG	csi-miR166c-5p	TRINITY_DN42410_c0_g1_i1	42	58	21	137	-43.05	-31.42	0.75	17	37	+
6	asi-miR168-3p	UCUCGCCUUGCAUCAAGUGAAU	lja-miR168-3p	TRINITY_DN25267_c0_g1_i1	51	49	22	222	-77.50	-34.91	0.68	153	132	-
7	asi-miR168a	UCGCUUGGUGCAGGUCGGGA	pab-miR168a	TRINITY_DN33542_c0_g1_i1	53	47	21	221	-97.68	-44.20	0.83	115	135	+
8	asi-miR168a-3p	CCC GCCUUGCGCCAAGUGAA	zma-miR168a-3p	TRINITY_DN33542_c0_g1_i1	55	45	20	220	-95.40	-43.36	0.79	160	179	+
9	asi-miR168b-3p	CCC GCCUUGCGCCAAGUGAAU	vca-miR168b-3p	TRINITY_DN33542_c0_g1_i1	55	45	21	221	-95.40	-43.17	0.78	160	180	+
10	asi-miR168b	UCGCUUGGUGCAGGUCGGG	gma-miR168b	TRINITY_DN33542_c0_g1_i1	53	47	19	219	-97.68	-44.60	0.84	115	133	+
11	asi-miR168	UCGCUUGGUGCAGGUCGGGA	cme-miR168	TRINITY_DN33542_c0_g1_i1	53	47	20	220	-97.68	-44.40	0.84	115	134	+
12	asi-miR168-5p	UCGCUUGGUGCAGGUCGGGAAC	lja-miR168-5p	TRINITY_DN33542_c0_g1_i1	54	46	22	222	-97.68	-44.00	0.81	115	136	+
13	asi-miR169d	UAGCCAAGGAGACUGCCUAUG	zma-miR169d	TRINITY_DN29275_c0_g1_i1	57	43	21	221	-92.80	-41.99	0.74	187	207	+
14	asi-miR169a-3p	GGCAAGUCAUCCUUGGCUAC	ath-miR169a-3p	TRINITY_DN29275_c0_g1_i1	58	42	20	195	-72.00	-36.92	0.64	296	315	+
15	asi-miR169b-3p	GCAGUCUCCUUGGCUACU	stu-miR169b-3p	TRINITY_DN9327_c0_g1_i1	47	53	18	201	-58.30	-29.00	0.62	84	101	+
16	asi-miR393b-3p	UCAUGCGUCCCCUUGGAAU	csi-miR393b-3p	TRINITY_DN44123_c0_g1_i1	38	62	21	221	-62.34	-28.21	0.74	142	162	+
17	asi-miR395	CUGAAGUGUUUGGGGAACUC	atr-miR395	TRINITY_DN46165_c0_g1_i1	40	60	21	221	-68.40	-30.95	0.77	141	121	-
18	asi-miR395e	UGAAGUGUUUGGGGAACUC	cpa-miR395e	TRINITY_DN46165_c0_g1_i1	40	60	20	220	-68.40	-31.09	0.78	140	121	-
19	asi-miR395b-3p	AAGUGUUUGGGGAACUC	ata-miR395b-3p	TRINITY_DN46165_c0_g1_i1	40	60	18	218	-68.40	-31.38	0.78	138	121	-
20	asi-miR395a	UGAAGUGUUUGGGGAACUCU	aof-miR395a	TRINITY_DN46165_c0_g1_i1	40	60	21	221	-69.20	-31.31	0.78	140	120	-
21	asi-miR399f	GGGAGCUCUGCUUUGGCAGG	pab-miR399f	TRINITY_DN9291_c1_g1_i1	60	40	20	220	-72.19	-32.81	0.55	227	246	+
22	asi-miR444b	UUGCUGCCUCAACCCUGCUGC	tae-miR444b	TRINITY_DN15906_c0_g2_i1	40	60	21	221	-55.95	-25.32	0.63	269	289	+
23	asi-miR528	UGGAAGGGGCAUGCAGAGGAG	sbi-miR528	TRINITY_DN36065_c0_g1_i1	60	40	21	163	-71.30	-43.74	0.73	43	63	+

24	asi-miR535b	UGACAAUGAGAGAGAGCACGC	ppt-miR535b	TRINITY_DN18597_c0_g1_i4	45	55	21	221	-85.20	-38.55	0.86	1105	1085	-
25	asi-miR535-3p	GUGCUCUUUCUCGUUGUCCCC	vca-miR535-3p	TRINITY_DN18597_c0_g1_i4	47	53	21	221	-83.00	-37.56	0.80	1039	1019	-
26	asi-miR535d	UUGACAAUGAGAGAGAGCACG	mes-miR535d	TRINITY_DN18597_c0_g1_i4	46	54	21	221	-85.00	-38.46	0.84	1106	1086	-
27	asi-miR5655	AAGAAGACACUUAAGAAGGAG	ath-miR5655	TRINITY_DN42765_c0_g1_i1	42	58	21	221	-53.20	-24.07	0.57	489	509	+
28	asi-miR6156	UUGAAAAUUGUUCUAUUUCUGU	nta-miR6156	TRINITY_DN30748_c0_g1_i1	32	68	22	222	-43.80	-19.73	0.62	106	127	+
29	asi-miR8005c	UUUAGAGUUUAGGGUUUAGGGUUU	stu-miR8005c	TRINITY_DN33453_c0_g1_i1	38	62	24	224	-43.29	-19.33	0.51	313	290	-
30	asi-miR396	UCCACAGGCUUUCUUGAAC	far-miR396	TRINITY_DN18615_c1_g2_i1	53	47	19	122	-41.65	-34.14	0.64	22	4	-
31	asi-miR396e	UUCCACAGGCUUUCUUGAAC	fve-miR396e	TRINITY_DN18615_c1_g2_i1	53	47	20	123	-41.65	-33.86	0.64	23	4	-
32	asi-miR396a	UUCCACAGGCUUUCUUGAACG	ppe-miR396a	TRINITY_DN18615_c1_g2_i1	53	47	21	123	-41.65	-33.86	0.64	23	3	-
33	asi-miR408b-5p	AGGUACGAGGCAGAGCCUG	zma-miR408b-5p	TRINITY_DN9858_c0_g1_i1	39	61	19	201	-49.73	-24.74	0.63	153	171	+
34	asi-miR482e	UUACCGAUUCCUCCCAUGCC	ppe-miR482e	TRINITY_DN13721_c0_g1_i1	40	60	20	220	-73.15	-33.25	0.83	116	135	+
35	asi-miR482c	UUGCCCACUCCGCCCAUUCC	aof-miR482c	TRINITY_DN17345_c1_g1_i2	43	57	20	220	-89.60	-40.73	0.95	115	134	+
36	asi-miR482b-3p	UUCCCGAAACCUCUCAUACCUA	ppe-miR482b-3p	TRINITY_DN17345_c1_g1_i2	39	61	22	222	-72.88	-32.83	0.84	321	342	+
37	asi-miR482a-3p	UUCCCGAAACCUCUCAUACC	ppe-miR482a-3p	TRINITY_DN17345_c1_g1_i2	40	60	20	220	-72.68	-33.04	0.83	321	340	+
38	asi-miR156c-3p	UGCUUGCUCUCUAUCUGUC	mtr-miR156c-3p	TRINITY_DN12866_c0_g2_i1	48	52	19	219	-40.41	-18.45	0.38	756	738	-
39	asi-miR156f-3p	UCUCUCUUCUCCUGUCAU	csi-miR156f-3p	TRINITY_DN19170_c0_g1_i2	39	61	18	218	-37.63	-17.26	0.44	732	715	-
40	asi-miR156a-3p	UGCUUGCGCCUCUUCUGUCAGC	csi-miR156a-3p	TRINITY_DN23526_c1_g2_i4	49	51	22	222	-59.99	-27.02	0.55	493	514	+
41	asi-miR159b	UUGGAUUGAAGAGAGCUCC	pta-miR159b	TRINITY_DN32656_c0_g1_i1	35	65	19	197	-27.64	-14.03	0.40	79	97	+
42	asi-miR159c	UUGGAUUGAAGAGAGCUCCA	osa-miR159c	TRINITY_DN32656_c0_g1_i1	35	65	20	198	-27.64	-13.96	0.40	79	98	+
43	asi-miR167d-3p	GAUAAUGUGGUUGCUUCA	stu-miR167d-3p	TRINITY_DN17251_c1_g1_i1	46	54	18	218	-49.67	-22.78	0.50	608	625	+
44	asi-miR164	GGGGAAGAGGAGCACAUCAA	rgl-miR164	TRINITY_DN17932_c0_g1_i1	60	40	20	220	-75.10	-34.14	0.57	1397	1416	+
45	asi-miR164a-3p	CGUGUUCUCCUUCUCCUC	zma-miR164a-3p	TRINITY_DN21370_c0_g1_i2	51	49	19	219	-59.50	-27.17	0.53	937	955	+
46	asi-miR167c-5p	UAAGCUGCCAGCAUGGAUCU	aly-miR167c-5p	TRINITY_DN24701_c0_g1_i1	45	55	20	220	-55.44	-25.20	0.56	272	291	+
47	asi-miR167b-3p	GUCAUGCUGGAGUUGUCAU	ata-miR167b-3p	TRINITY_DN24758_c3_g2_i2	45	55	19	219	-62.63	-28.60	0.64	370	352	-
48	asi-miR172f	AGAAUCCUGAUGAUGCUGCA	sbi-miR172f	TRINITY_DN12549_c0_g1_i1	56	44	20	144	-46.90	-32.57	0.58	25	44	+
49	asi-miR172b	GCAGCAUUAUCUAGAUAU	gra-miR172b	TRINITY_DN21831_c1_g1_i2	41	59	19	219	-48.95	-22.35	0.55	228	246	+
50	asi-miR172b-5p	GCAGCAUCAUCAGGAUUCUCA	csi-miR172b-5p	TRINITY_DN25099_c0_g1_i1	56	44	21	221	-61.03	-27.62	0.49	117	137	+
51	asi-miR172d-5p	AACUUCUUAAGAUAUCAGA	ath-miR172d-5p	TRINITY_DN35440_c0_g1_i1	41	59	19	170	-37.60	-22.12	0.54	70	52	-

52	asi-miR172c-5p	AGGAUCUUCAUGAUUCACA	stu-miR172c-5p	TRINITY_DN43552_c0_g1_i1	42	58	19	219	-49.30	-22.51	0.54	109	127	+
53	asi-miR319	UGAGUGAUGGGGGAGAUUA	seu-miR319	TRINITY_DN2141_c0_g1_i1	37	63	18	214	-32.74	-15.30	0.41	203	220	+
54	asi-miR319n	GGACCGAAGGGAGGCCCU	gma-miR319n	TRINITY_DN22078_c0_g2_i3	66	34	18	121	-37.84	-31.27	0.47	4	21	+
55	asi-miR319-5p	GCUGCAGAUUCAUUCAUU	ata-miR319-5p	TRINITY_DN26876_c0_g1_i1	43	57	18	158	-35.29	-22.34	0.52	41	58	+
56	asi-miR319c-5p	GAGCUUGCUUCAGCCCAC	sly-miR319c-5p	TRINITY_DN28525_c0_g1_i1	47	53	18	131	-25.13	-19.18	0.41	31	14	-
57	asi-miR319d	AAACUGUUUAGUCCACC	sly-miR319d	TRINITY_DN48048_c0_g1_i1	34	66	18	200	-31.84	-15.92	0.47	224	207	-
58	asi-miR398a-5p	GGGUUGGUUUGAUACAUAU	stu-miR398a-5p	TRINITY_DN32954_c0_g1_i1	34	66	20	130	-22.30	-17.15	0.50	30	11	-
59	asi-miR3981-3p	AUUAGGAUGCAUCUCAUC	osa-miR3981-3p	TRINITY_DN3446_c0_g1_i1	37	63	18	196	-41.33	-21.09	0.57	122	139	+
60	asi-miR398a	UGUGUUUCUCAGGCCACCCCUU	gma-miR398a	TRINITY_DN36330_c0_g1_i1	50	50	22	222	-68.30	-30.77	0.62	139	118	-
61	asi-miR398e	UGUGUUUCUCAGGCCACCCCU	lus-miR398e	TRINITY_DN36330_c0_g1_i1	50	50	21	221	-67.40	-30.50	0.61	139	119	-
62	asi-miR393-5p	UCCAAAGGGAUCGCAUUGAU	ata-miR393-5p	TRINITY_DN44123_c0_g1_i1	42	58	20	189	-66.95	-35.42	0.84	70	89	+
63	asi-miR393c	UCCAAAGGGAUCGCAUUGAUCU	mes-miR393c	TRINITY_DN44123_c0_g1_i1	42	58	22	191	-67.29	-35.23	0.84	70	91	+
64	asi-miR398-5p	AGAGAGCUUAAGAACACAA	lja-miR398-5p	TRINITY_DN8652_c0_g1_i1	36	64	19	219	-25.40	-11.60	0.32	233	251	+
65	asi-miR444d.1	UGCUCUUAAGCUUGCUGC	osa-miR444d.1	TRINITY_DN15558_c0_g1_i1	48	52	20	220	-54.50	-24.77	0.52	173	192	+
66	asi-miR444b.1	UGCUGUCUCAAGCUCGCGCC	osa-miR444b.1	TRINITY_DN1809_c0_g1_i1	36	64	21	211	-31.99	-15.16	0.42	91	111	+
67	asi-miR472b	UUUCUCAACUACCCAUC	ptc-miR472b	TRINITY_DN18106_c1_g1_i1	37	63	19	219	-49.85	-22.76	0.62	177	195	+
68	asi-miR1444d	AACGUUAAGCGAAUGUGA	ptc-miR1444d	TRINITY_DN36598_c0_g1_i1	35	65	18	171	-34.00	-19.88	0.57	54	71	+
69	asi-miR1444c	UCACAUUCGCUUAACGUU	ptc-miR1444c	TRINITY_DN36598_c0_g1_i1	35	65	18	171	-26.14	-15.29	0.44	71	54	-
70	asi-miR2118l	CCUAAGGCUUCCCUUCCU	osa-miR2118l	TRINITY_DN24384_c2_g1_i10	60	40	19	219	-77.30	-35.30	0.59	266	284	+
71	asi-miR2118r	UUUCCAUGCCUCACAUGC	osa-miR2118r	TRINITY_DN34829_c0_g1_i1	46	54	19	161	-46.60	-28.94	0.63	43	61	+
72	asi-miR5300	CAGUCUAGGCCUCCAAC	sly-miR5300	TRINITY_DN35092_c0_g1_i1	38	62	18	196	-33.26	-16.97	0.45	313	330	+
73	asi-miR2118k	CUUAUGCGUCUAUCCU	osa-miR2118k	TRINITY_DN50328_c0_g1_i1	44	56	18	163	-39.00	-23.93	0.54	216	233	+
74	asi-miR2118b-5p	GGAGAUGGGAGGUUGGUA	gma-miR2118b-5p	TRINITY_DN5080_c0_g1_i1	50	50	19	187	-53.08	-28.39	0.57	201	183	-

4.3 Characteristics of the predicted miRNA and pre-miRNA Candidates

Figure 4.3 shows box and whisker plot of miRNA length and pre-miRNA parameters (including % AU content, pre-miRNA length, MFE, AFME and MFEI).

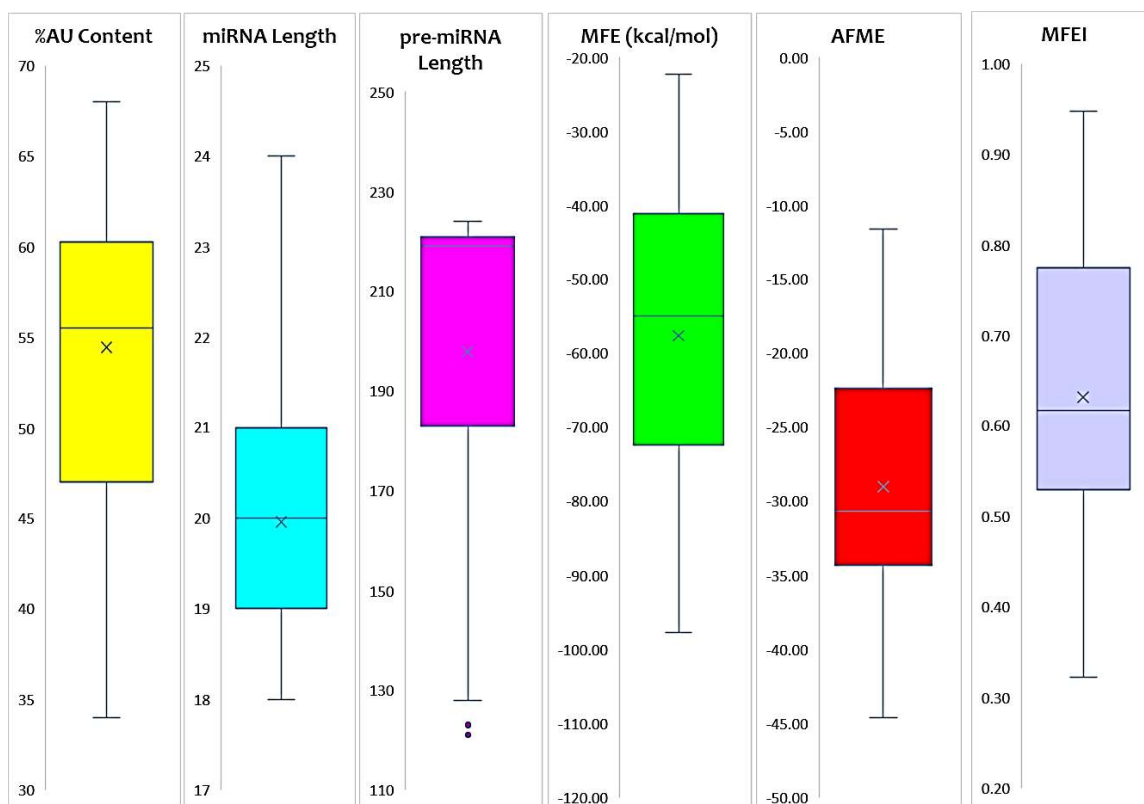


Figure 4.3: Box and whisker plots showing distribution of values for miRNA length and various parameters analyzed for pre-miRNAs

Length of miRNAs varied from 18-24 nucleotides while the median length was found to be 20 with lower quartile (LQ) as 19 and upper quartile (UQ) as 21. % AU content of pre-miRNAs ranged from 34-68% with median value of 56% (LQ: 47%, UQ: 60.3%). Pre-miRNA lengths (ranging from 121-224 nucleotides) indicated a largely positively skewed distribution of values as a majority of pre-miRNAs were >200 nucleotides in length (median pre-miRNA length was 219 while LQ as 183 and UQ as 221). Values of pre-miRNAs MFE (kcal/mol) varied from -97.7 to -

22.3 (with a median of -55.0, LQ as -72.3 and UQ as -41.1). Lower values of MFE indicated more stable pre-miRNA structures which were more probable pre-miRNA candidates. Values of AFME indicated a negatively skewed distribution with lowest value as -44.6 and the highest as -11.6. Median AFME was -30.6 while LQ as -34.3 and UQ as -22.4. MFEI values ranged from 0.32 to 0.95 with median value as 0.62, LQ as 0.53 and UQ as 0.77. All of the pre-miRNA evaluation parameters of the selected miRNAs were verified to be within the recommended range of values, suitable for further evaluation.

Figure 4.4 represents the bar chart with number of miRNAs predicted in each miRNA family chosen in the current study. Highest number of miRNAs (i.e. 7) were predicted in the miR-168 family while only one miRNA each were predicted in the families miR-161, miR-165, miR-399, miR-528, miR-5655, miR-6156, miR-8005, miR-408, miR-3981, miR-472 and miR-5300. From the overall analysis of miRNA prediction, a total of 74 miRNA candidates were identified that were further analyzed for target prediction and experimental validation studies.

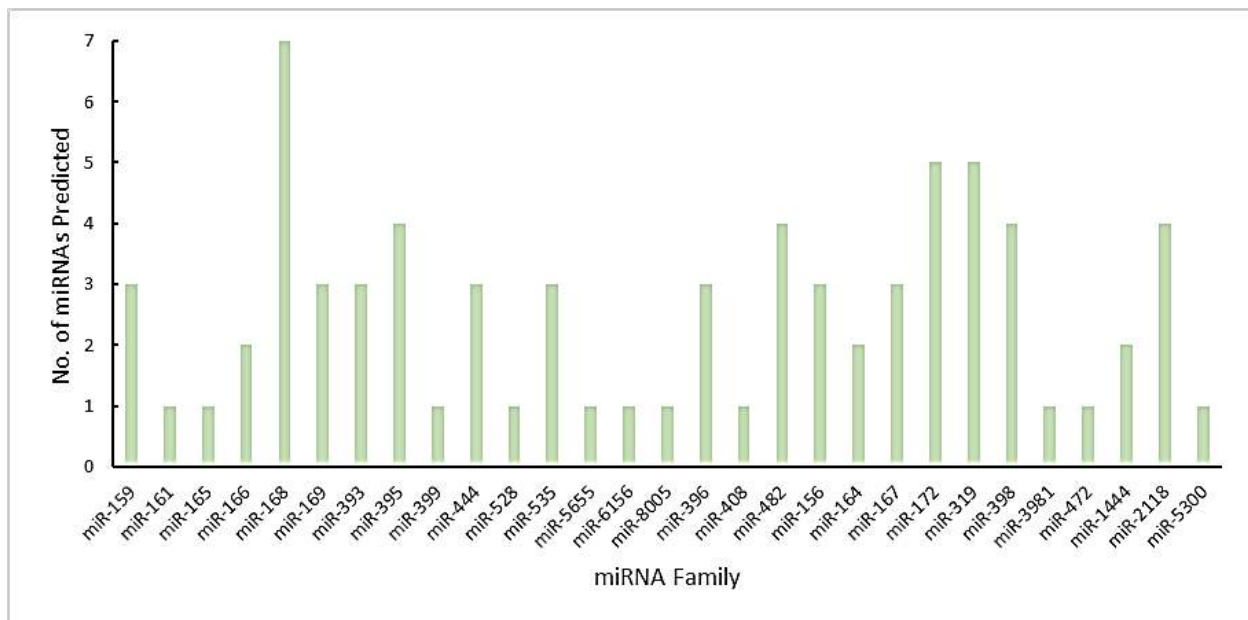


Figure 4.4: Bar chart representing the number of miRNAs predicted in each miRNA family

4.4 Electrophoretic Analysis of Extracted RNA

Upon Agarose Gel Electrophoresis (2%) in 1x TBE buffer (Appendix-II) of the extracted RNA from *A. sisalana* under biotic stress, clearly distinguishable bands of 28s and 18s ribosomal RNAs were observed on the gel (Figure 4.5) indicating the integrity of the extracted RNA and suitability for downstream analysis.

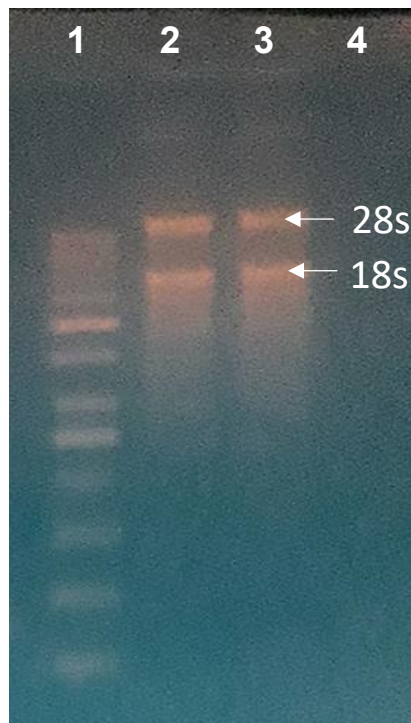


Figure 4.5: Agarose Gel Electrophoresis (2%) of the large RNA fraction. Lane 1: Gene ruler (100 bp), Lane 2-3: RNA sample loaded

4.5 Quantification and Quality Assessment of RNA via NanoDrop

For quantification and further quality assessment, 1 μ L of each RNA sample was loaded in spectrophotometer with the corresponding elution buffer as blank, and the following results were obtained as depicted in Table 4.2.

Table 4.2: Quality assessment and quantification of small and large RNA fraction

Sr. No.	Sample	Ratio of absorbance (260/280)	Ratio of absorbance (260/230)	Concentration (ng/μl)
1	Large RNA fraction	2.05	2.26	1478.8
2	Small RNA fraction	2.08	2.01	57.1

4.6 Poly(A)RT PCR for Experimental Validation of miRNAs

After Poly(A)tail addition and cDNA preparation as per methods outlined in Section 3.11-3.13, Poly(A)RT PCR was performed, and the results are indicated in Figures 4.6 A-F. From the results of miRNA validation, a total of 48 miRNAs were experimentally validated (as giving an approximately 100 bp band in Agarose Gel Electrophoresis) as result of Poly(A)RT PCR analysis out of 74 predicted sequences in this study: asi-miR165a-5p, asi-miR166b, asi-miR166c-5p, asi-miR168-3p, asi-miR168a-3p, asi-miR168b-3p, asi-miR168, asi-miR168-5p, asi-miR169d, asi-miR169a-3p, asi-miR169b-3p, asi-miR395b-3p, asi-miR399f, asi-miR444b, asi-miR528, asi-miR535b, asi-miR396, asi-miR396e, asi-miR396a, asi-miR408b-5p, asi-miR482e, asi-miR482c, asi-miR482b-3p, asi-miR482a-3p, asi-miR156c-3p, asi-miR156a-3p, asi-miR159c, asi-miR167d-3p, asi-miR164, asi-miR164a-3p, asi-miR167c-5p, asi-miR172b-5p, asi-miR319n, asi-miR319c-5p, asi-miR319d, asi-miR398a-5p, asi-miR3981-3p, asi-miR398a, asi-miR398e, asi-miR393-5p, asi-miR393c, asi-miR398-5p, asi-miR444d.1, asi-miR444b.1, asi-miR2118l, asi-miR2118r, asi-miR5300, asi-miR2118b-5p (Figure 4.6 A-F).

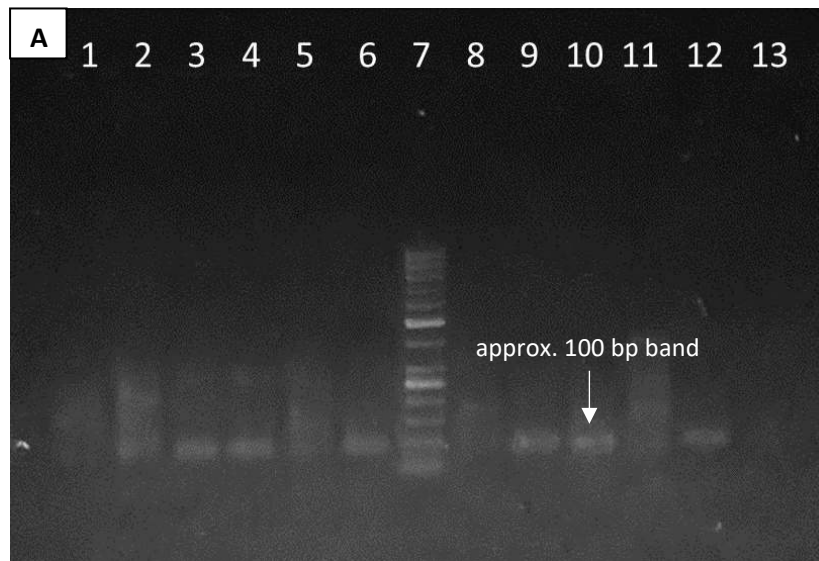


Figure 4.6 (A-F): Poly(A)RT PCR analysis for experimental validation of miRNAs.

Figure 4.6 A: Lane 3: asi-miR165a-5p, Lane 4: asi-miR166b, Lane 6: asi-miR168-3p, Lane 7: 50 bp ladder, Lane 9: asi-miR168a-3p, Lane 10: asi-miR168b-3p, Lane 12: Positive Control (β -actin gene from HeLa cDNA), Lane 13: Negative Control

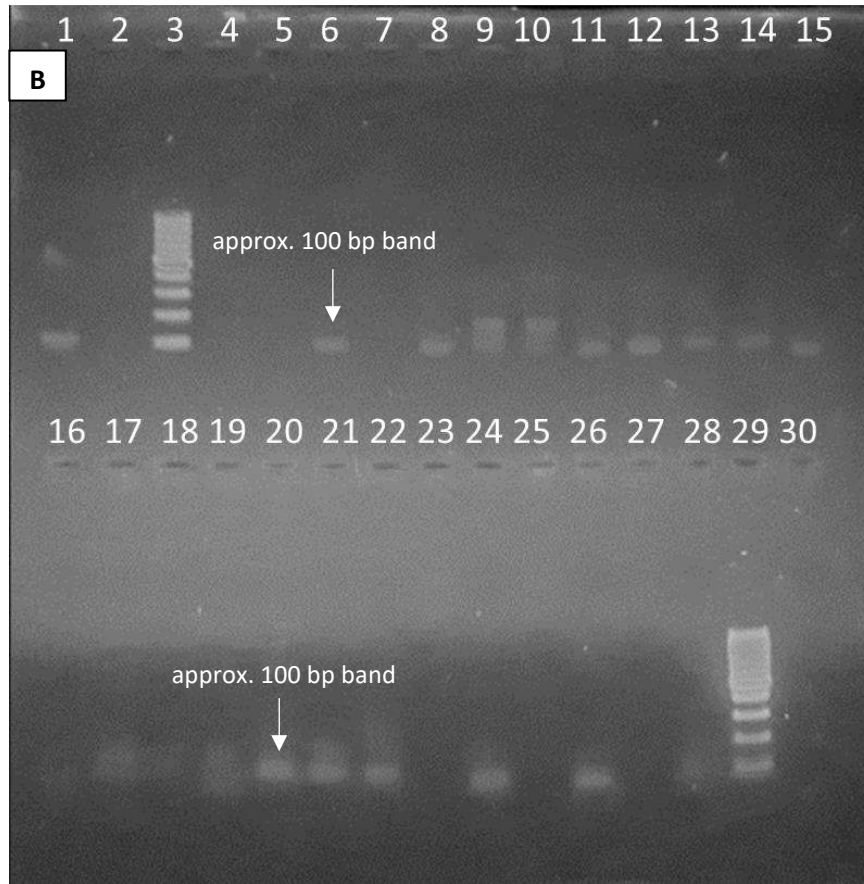


Figure 4.6 B: Poly(A)RT PCR analysis for experimental validation of miRNAs. Lane 1: Positive Control (β-actin gene from HeLa cDNA), Lane 2: Negative Control, Lane 3: 100 bp ladder, Lane 6 and Lane 8: asi-miR166c-5p, Lane 11: asi-miR168, Lane 12: asi-miR168-5p, Lane 13: asi-miR169d, Lane 14: asi-miR169a-3p, Lane 15: asi-miR169b-3p, Lane 20: asi-miR399f, Lane 21: asi-miR444b, Lane 22: asi-miR528, Lane 24: asi-miR535b, Lane 26: asi-miR5300, Lane 29: 100 bp ladder

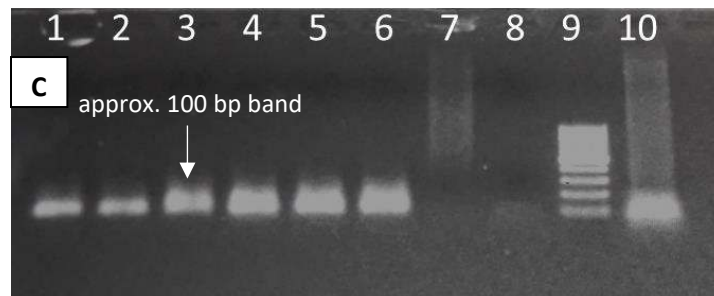


Figure 4.6 C: Poly(A)RT PCR analysis for experimental validation of miRNAs. Lane 1: asi-miR482e, Lane 2: asi-miR482c, Lane 3: asi-miR396, Lane 4: asi-miR396e, Lane 5: asi-miR396a, Lane 8: Negative Control, Lane 9: 100 bp ladder, Lane 10: Positive Control (β-actin gene from HeLa cDNA)

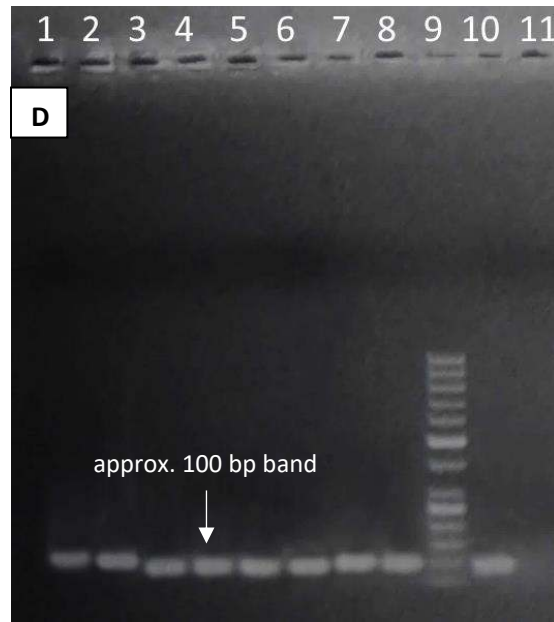


Figure 4.6 D: Poly(A)RT PCR analysis for experimental validation of miRNAs. Lane 1: asi-miR156a-3p, Lane 2: asi-miR164, Lane 3: asi-miR159c, Lane 4: asi-miR482b-3p, Lane 5: asi-miR408b-5p, Lane 6: asi-miR482a-3p, Lane 7: asi-miR164a-3p, Lane 8: asi-miR2118b-5p, Lane 9: 50 bp ladder, Lane 10: Positive Control (β -actin gene from HeLa cDNA), Lane 11: Negative Control

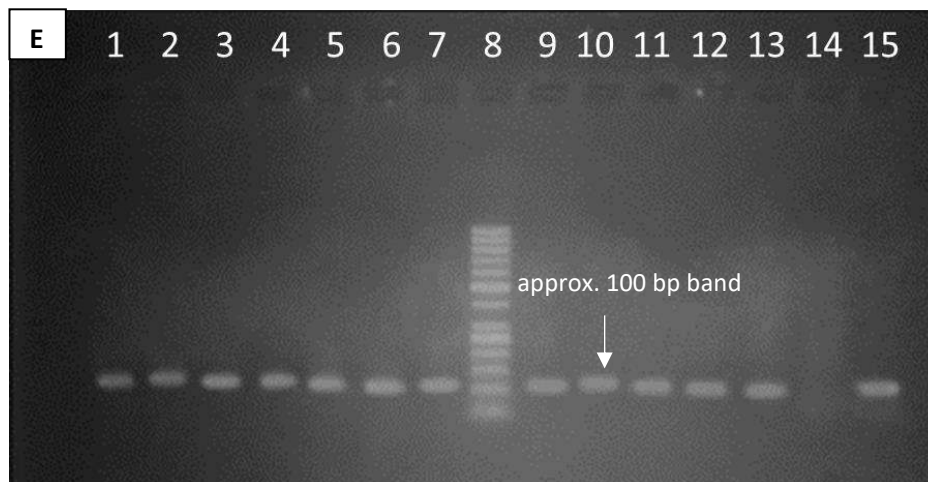


Figure 4.6 E: Poly(A)RT PCR analysis for experimental validation of miRNAs. Lane 1: asi-miR167c-5p, Lane 2: asi-miR172b-5p, Lane 3: asi-miR319n, Lane 4: asi-miR319c-5p, Lane 5: asi-miR398a, Lane 6: asi-miR398e, Lane 7: asi-miR393c, Lane 8: 50 bp ladder, Lane 9: asi-miR444d.1, Lane 10: asi-miR444b.1, Lane 11: asi-miR2118r, Lane 12: asi-miR2118l, Lane 13: asi-miR395b-3p, Lane 14: Negative Control, Lane 15: Positive Control (β -actin gene from HeLa cDNA)

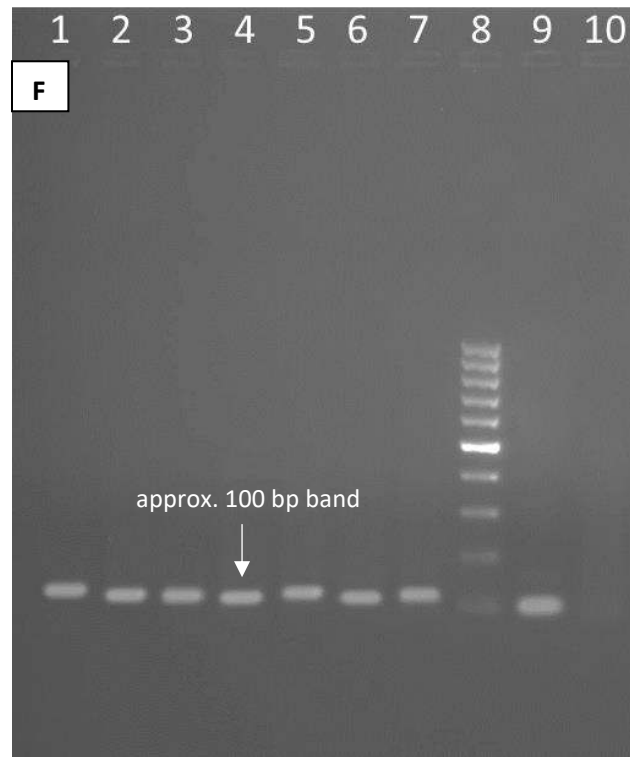


Figure 4.6 F: Poly(A)RT PCR analysis for experimental validation of miRNAs. Lane 1: asi-miR167d-3p, Lane 2: asi-miR319d, Lane 3: asi-miR398a-5p, Lane 4: asi-miR3981-3p, Lane 5: asi-miR393-5p, Lane 6: asi-miR156c-3p, Lane 7: asi-miR398-5p, Lane 8: 100 bp ladder, Lane 9: Positive Control (β -actin gene from HeLa cDNA), Lane 10: Negative Control

4.7 Predicted Targets of *A. sisalana* miRNAs against TAIR v.10

The 48 miRNA sequences of *A. sisalana* which gave positive results with Poly(A)RT PCR were run through psRNATarget server (<https://www.zhaolab.org/psRNATarget/analysis>) under the default parameters i.e. expectation = 5, penalty for opening gap = 2, penalty for extending gap = 0.5, penalty for G:U pair = 0.5, seed region = 2-13 nt, extra weight in the seed region = 1.5, mismatches allowed in the seed region = 2 and HPS size = 19. From the output of psRNATarget algorithm, 5,680 total targets of *A. sisalana* miRNAs (with 4334 unique/non-repeating targets) were identified against TAIR database v.10.

4.8 Gene Ontology Analysis of Predicted Targets in TAIR v.10

4.8.1 GO Biological Processes

Results of GO biological processes (Figure 4.7) showed involvement of asi-miRNAs targeted TAIR genes in various biological processes such as protein phosphorylation, metabolic processes, and developmental processes, and 651 genes were involved in response to stress (GO:0006950; $-\log_{10}$ P-value: 1.83) and 1079 genes involved in response to stimulus (GO:0050896; $-\log_{10}$ P-value: 1.32). The highest number of genes i.e. 1786 were found to be involved in cellular process (GO:0009987; $-\log_{10}$ P-value: 15.04).

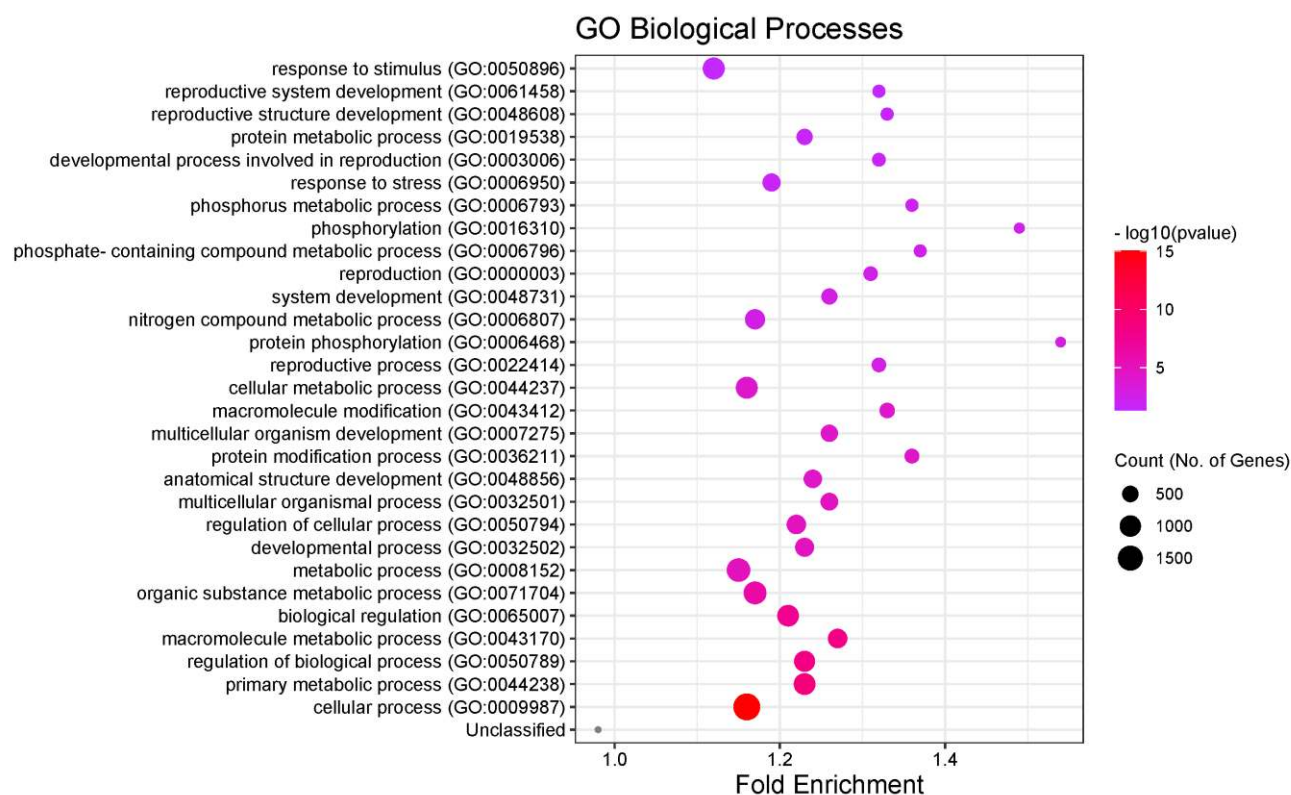


Figure 4.7: Gene Ontology classification into biological processes of the identified miRNA targets in TAIR v.10

4.8.2 GO Molecular Function

Results of GO molecular functions (Figure 4.8) showed involvement of asi-miRNAs targeted TAIR genes in various molecular functions such as nucleotide binding, protein binding, catalytic activity, and transferase activity. The highest number of genes i.e. 1332 were found to be belonging to function of binding (GO: 0005488; $-\log_{10}$ P-value: 17.7).

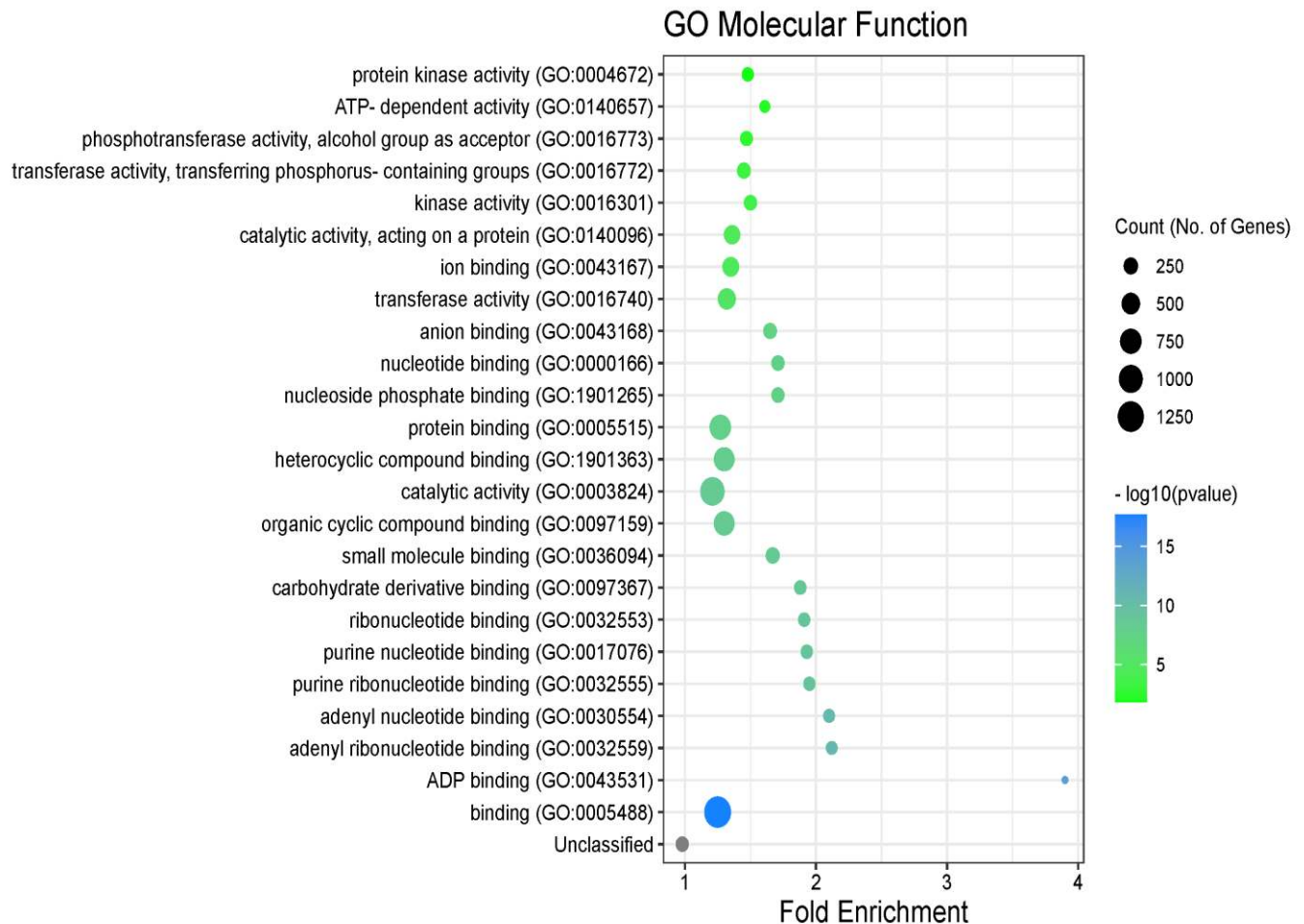


Figure 4.8: Gene Ontology classification into molecular functions of the identified miRNA targets in TAIR v.10

4.8.3 GO Cellular Component

The identified GO cellular components of asi-miRNAs targeted TAIR genes are shown in Figure 4.9, which showed that the targeted genes were localized in various cellular components such as mitochondrion, nucleus, endomembrane system and cullin-RING ubiquitin ligase complex. The greatest number of hits i.e., 1236 were recorded for the cellular component- nucleus (GO:0005634) with $-\log_{10}$ P-value of 3.48.

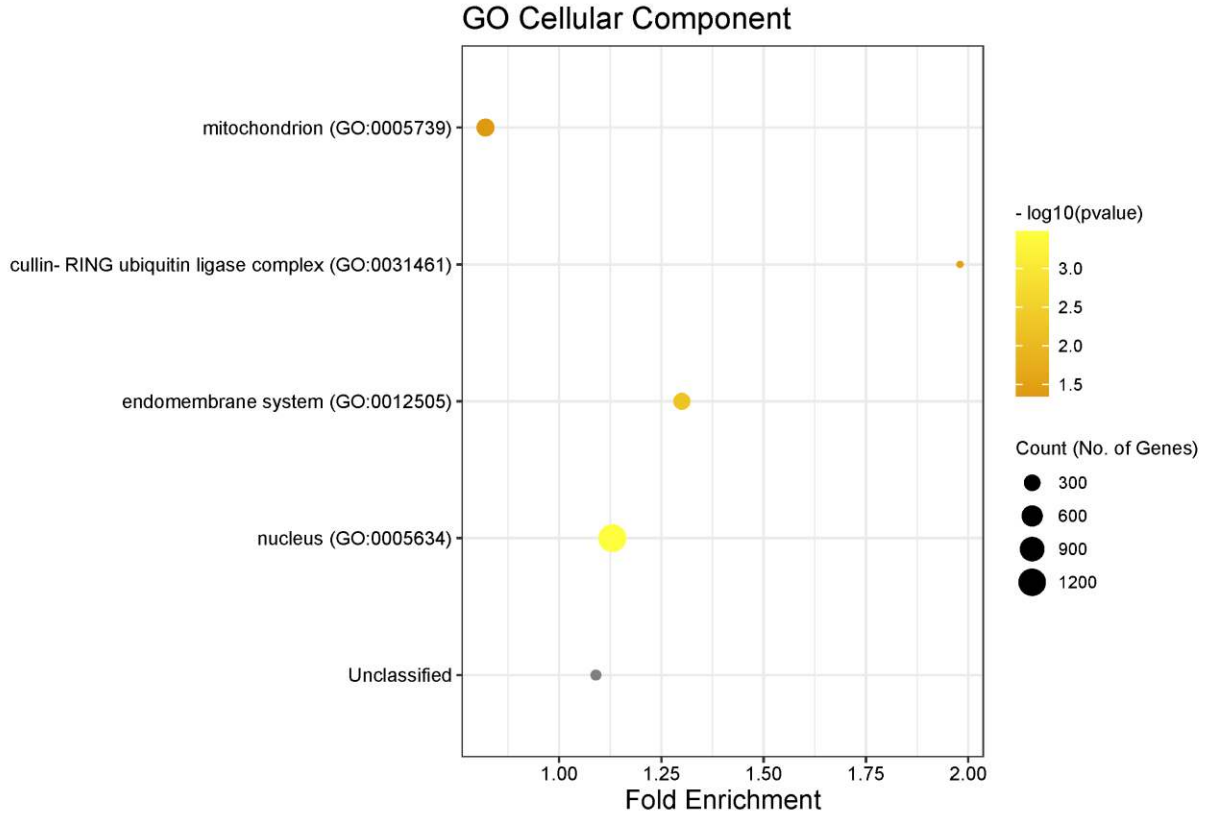


Figure 4.9: Gene Ontology classification into cellular components of the identified miRNA targets in TAIR v.10

4.9 Targets of *A. sisalana* miRNAs among USP Genes in TAIR

To confirm the targeting of USP family genes by asi-miRNAs, results were scanned to search for USP family genes present in the targeted gene sequences in TAIR v.10. From this analysis it was found that among the *Arabidopsis thaliana* (TAIR v.10) targeted genes (i.e. 4334), 23 Genes belong to the family PF00582 i.e. USP gene family. Table 4.3 shows the summary of output from target prediction of *A. sisalana* miRNAs against USP genes in TAIR v.10.

Table 4.3: Summary of output from target prediction of *A. sisalana* miRNAs against USP genes in TAIR v.10

miRNA_Acc.	Expectation	Target_Acc.	Target_start	Target_end	miRNA_aligned_fragment	Alignment	Target_aligned_fragment	Inhibition
asi-miR398a	2	AT2G19410.1.p1	2266	2287	UGUGUUUCUCAGGCCACCCCUU	::	UUAGCGUGGCUUGAGAAACAUA	Cleavage
asi-miR398e	2	AT2G19410.1.p1	2267	2287	UGUGUUUCUCAGGCCACCCCU	::	UAGCGUGGCUUGAGAAACAUA	Cleavage
asi-miR2118b-5p	2.5	AT3G21210.1.p1	674	692	GGAGAUGGGAGGUUUGGUA	:	CACCAAUUCUUUAUCUCC	Cleavage
asi-miR172b-5p	3.5	AT3G53990.1.p1	324	344	GCAGCAUCAUCAGGAUUCUCA	 :	AGAGAAGCUUGUUGAUGCUGU	Translation
asi-miR164a-3p	4	AT2G19410.1.p1	1342	1360	CGUGUUCUCCUUCUCCCUC	:	AUUGGAGAAGGAGGAUACG	Cleavage
asi-miR165a-5p	4	AT5G57035.1.p1	1458	1477	GACUGUUGUCUGGCUCGAGG	... : :	CCUCAAAACCAGGCAACAUUC	Cleavage
asi-miR172b-5p	4	AT3G03270.1.p1	324	344	GCAGCAUCAUCAGGAUUCUCA	 :	GGAGAAACUUUGUGAUGCUGU	Translation
asi-miR172b-5p	4	AT3G03270.2.p1	324	344	GCAGCAUCAUCAGGAUUCUCA	 :	GGAGAAACUUUGUGAUGCUGU	Translation
asi-miR398a	4	AT4G25160.1.p1	2389	2410	UGUGUUUCUCAGGCCACCCCUU	:	GAGGAGUGGCUUAAGGAGCACA	Translation
asi-miR398e	4	AT4G25160.1.p1	2390	2410	UGUGUUUCUCAGGCCACCCCU	:	AGGAGUGGCUUAAGGAGCACA	Translation
asi-miR164	4.5	AT3G21210.1.p1	1677	1696	GGGGAAGAGGAGCACAUGAA	 :	AACAUGGCACAUCUCCUC	Translation
asi-miR164a-3p	4.5	AT1G69080.1.p1	304	322	CGUGUUCUCCUUCUCCCUC	: : :	AAAGAAGAAGGAGAAGACG	Cleavage
asi-miR164a-3p	4.5	AT1G69080.2.p1	304	322	CGUGUUCUCCUUCUCCCUC	: : :	AAAGAAGAAGGAGAAGACG	Cleavage
asi-miR166b	4.5	AT5G57035.1.p1	1458	1478	GGACUGUUGUCUGGCUCGAGG	... : :	CCUCAAAACCAGGCAACAUUCU	Cleavage
asi-miR166c-5p	4.5	AT5G57035.1.p1	1459	1479	GGGACUGUUGUCUGGCUCGAG	... : :	CUCAAAACCAGGCAACAUUCUC	Cleavage
asi-miR396a	4.5	AT5G61550.2.p1	1201	1221	UUCCACAGGCUUUCUUGAACG	: ::	GAAUCCGAGAAGCUUGUGGAA	Cleavage
asi-miR396a	4.5	AT5G61550.1.p1	1201	1221	UUCCACAGGCUUUCUUGAACG	: ::	GAAUCCGAGAAGCUUGUGGAA	Cleavage
asi-miR396e	4.5	AT5G61550.2.p1	1202	1221	UUCCACAGGCUUUCUUGAAC	: ::	AAUCCGAGAAGCUUGUGGAA	Cleavage
asi-miR396e	4.5	AT5G61550.1.p1	1202	1221	UUCCACAGGCUUUCUUGAAC	: ::	AAUCCGAGAAGCUUGUGGAA	Cleavage
asi-miR482a-3p	4.5	AT5G51270.1.p1	1399	1418	UUCCCGAAACCUCCCAUACC	: : :	GGAAUGGGAGCUUACGGGGA	Translation
asi-miR482b-3p	4.5	AT5G51270.1.p1	1397	1418	UUCCCGAAACCUCCCAUACCUA	: : :	UCGGAAUGGGAGCUUACGGGGA	Translation
asi-miR156a-3p	5	AT1G01660.1.p1	1020	1041	UGCUUGCGCCUCUUCUGUCAGC	: ::	GGAGAUCAAGAAGUGCGAGCC	Translation
asi-miR164	5	AT5G47740.2.p1	207	226	GGGGAAGAGGAGCACAUGAA	... : :	UUCUUCUACUCUUCUUUUC	Cleavage
asi-miR164	5	AT5G47740.1.p1	207	226	GGGGAAGAGGAGCACAUGAA	... : :	UUCUUCUACUCUUCUUUUC	Cleavage
asi-miR164a-3p	5	AT1G16760.1.p1	1366	1384	CGUGUUCUCCUUCUCCCUC	:	GUUGGAGAAGGAGGGUAUG	Cleavage
asi-miR164a-3p	5	AT5G63940.1.p1	16	34	CGUGUUCUCCUUCUCCCUC	: : :	GAAGGAGGAGGGGAAGUCG	Cleavage

asi-miR164a-3p	5	AT1G01660.1.p1	1140	1158	CGUGUUCUCCUUCUCCCUC	:: ::::: : ::	GAAGGAGAAGGAAGAAGCG	Cleavage
asi-miR166b	5	AT2G24370.1.p1	1686	1706	GGACUGUUGUCUGGCUCGAGG	::: : :::: ::::	UUUCAGACUAGGAAACAGUCC	Cleavage
asi-miR166c-5p	5	AT4G25160.1.p1	711	731	GGGACUGUUGUCUGGCUCGAG	: : :::: ::::	UGCAGGACAGGCAUCAGUUC	Cleavage
asi-miR172b-5p	5	AT3G13690.1.p1	825	845	GCAGCAUCAUCAGGAUUCUCA	:: ::::: ::::	UGAUGGUCUUGAUGAUUCUGG	Cleavage
asi-miR172b-5p	5	AT4G27320.1.p1	618	638	GCAGCAUCAUCAGGAUUCUCA	::: ::::: ::::	UGAUGAUCGUGAUGGUCCUGC	Cleavage
asi-miR2118b-5p	5	AT5G61550.2.p1	2450	2468	GGAGAUGGGAGGUUUGGUA	: : ::::: ::::	UGACGAAUCUCCAUUGCC	Cleavage
asi-miR2118b-5p	5	AT5G61550.1.p1	2450	2468	GGAGAUGGGAGGUUUGGUA	: : ::::: ::::	UGACGAAUCUCCAUUGCC	Cleavage
asi-miR2118r	5	AT4G23700.1.p1	407	425	UUUCCAAUGCCUCACAUGC	::: : ::::: ::::	UCGUUUAGGCAUUGGAAC	Cleavage
asi-miR2118r	5	AT1G01660.1.p1	1071	1089	UUUCCAAUGCCUCACAUGC	::: ::: ::::	GCAGCUGAGACAUAGGAAA	Translation
asi-miR393-5p	5	AT5G57670.2.p1	37	56	UCCAAAGGGAUCGCAUUGAU	: : ::::: ::::	CUUGUGGCGAUCUCGUUGGA	Cleavage
asi-miR396	5	AT5G63940.1.p1	235	253	UCCACAGGCUUUCUUGAAC	::::: ::::: :::	GUUUUAUGAAAGCUUCUGCA	Cleavage
asi-miR396a	5	AT5G63940.1.p1	234	254	UCCACAGGCUUUCUUGAACG	::::: ::::: ::::	UGUUUAUGAAAGCUUCUGCAA	Cleavage
asi-miR396a	5	AT2G19410.1.p1	2110	2130	UCCACAGGCUUUCUUGAACG	::::: : ::::: ::::	GUUUUGAAACGGCUUGUGGAG	Cleavage
asi-miR396e	5	AT5G63940.1.p1	235	254	UCCACAGGCUUUCUUGAAC	::::: ::::: ::::	GUUUUAUGAAAGCUUCUGCAA	Cleavage
asi-miR396e	5	AT2G19410.1.p1	2111	2130	UCCACAGGCUUUCUUGAAC	::::: : ::::: ::::	UUUUGAAACGGCUUGUGGAG	Cleavage
asi-miR398-5p	5	AT2G47710.1.p1	222	240	AGAGAGCUUAAGAACACAA	::: : : ::::: ::::	UUAUCCUUUUAAGCUCUUC	Cleavage
asi-miR399f	5	AT1G16760.1.p1	628	647	GGGAGCUCUGCUUUGGCAGG	: : ::::: ::::: ::::	GUUACCAGAUUAGGGCUUCC	Translation
asi-miR528	5	AT2G24370.1.p1	191	211	UGGAAGGGGCAUGCAGAGGAG	::::: ::::: ::::	CUUCUCUUCAUGCUUCCUCUG	Cleavage
asi-miR528	5	AT1G16760.1.p1	140	160	UGGAAGGGGCAUGCAGAGGAG	::::: : ::::: ::::	CUCUUUUUCAUGUCCUCCCA	Cleavage
asi-miR528	5	AT5G57670.2.p1	1360	1380	UGGAAGGGGCAUGCAGAGGAG	::: ::::: ::::: ::::	GCAUUCGGCAUUCUCCUUCUA	Translation

4.10 *A. sisalana* USP Family Genes Predicted by TransDecoder and HMM3 Search

From TransDecoder results, a total of 38,802 sequences including 13,102 complete ORFs (peptides and corresponding CDS) were predicted. On HMM3 search in Ugene (v41.0), with MSA of 62,462 protein sequences of USP family (PF00582) and 13,102 complete ORFs (predicted through TransDecoder), nine (9) full length proteins of USP family were found in *A. sisalana*.

4.11 PCR Amplification of Predicted *A. sisalana* USP Family Genes

PCR amplification of AsUSP Genes using primers for full-length gene amplification with PCR profile as specified in Table 3.9. After a few reactions for optimization of annealing temperatures, products appeared on analysis with 1.2% Agarose Gel Electrophoresis. Figures 4.10-4.12 show bands on electrophoretic analysis of PCR products representing amplification of AsUSP Genes 1-9.

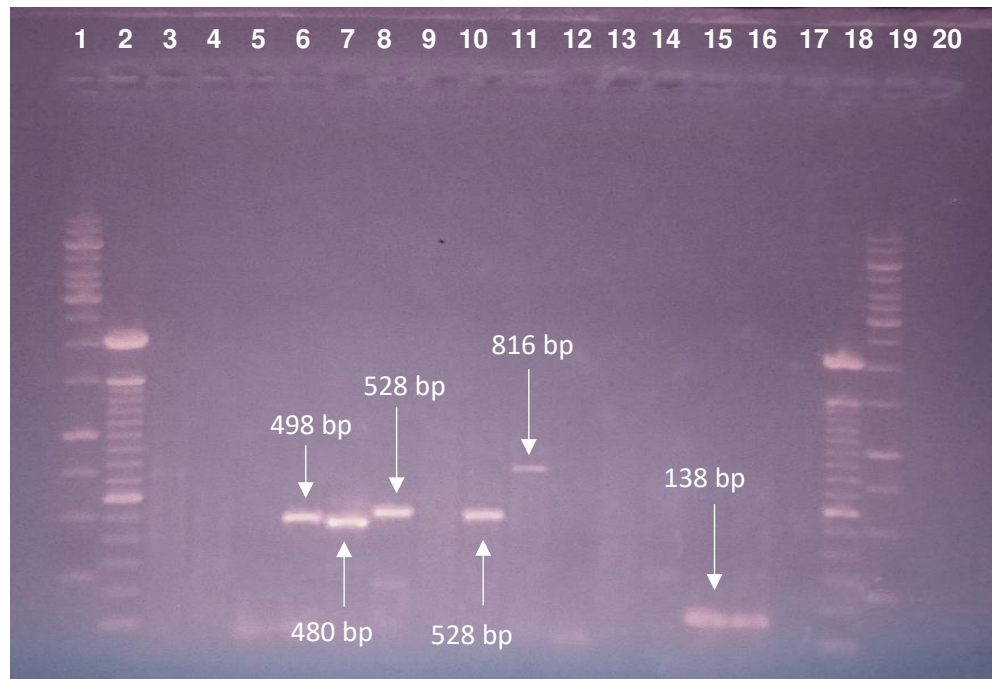


Figure 4.10: 1.2% Agarose Gel Electrophoresis for analysis of PCR products from amplification of AsUSP Gene 5, AsUSP Gene 6, AsUSP Gene 7, AsUSP Gene 9, AsUSP Gene 10

Lane 1: 1 kb ladder, Lane 2: 100 bp ladder, Lane 6: AsUSP Gene 5, Lane 7: AsUSP Gene 6, Lane 8: AsUSP Gene 7, Lane 10: AsUSP Gene 9, Lane 11: AsUSP Gene 10, Lane 15-16: Positive Control (β -actin), Lane 17: Negative Control, Lane 18: 100 bp ladder, Lane 19: 1kb ladder

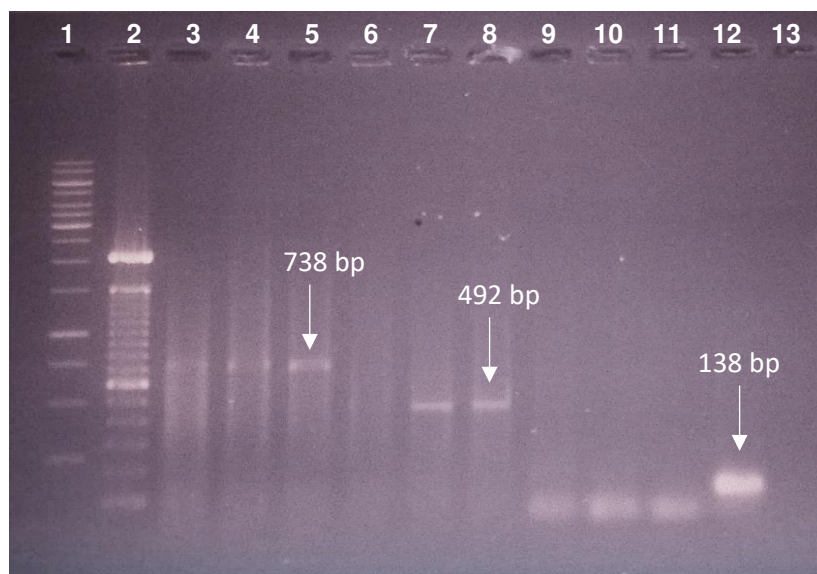


Figure 4.11: 1.2% Agarose Gel Electrophoresis for analysis of PCR products from amplification of AsUSP Gene 1 and AsUSP Gene 3

Lane 1: 1 kb ladder, Lane 2: 100 bp ladder, Lanes 3-5: AsUSP Gene 1, Lanes 7-8: AsUSP Gene 3, Lane 12: Positive Control (β -actin), Lane 13: Negative Control

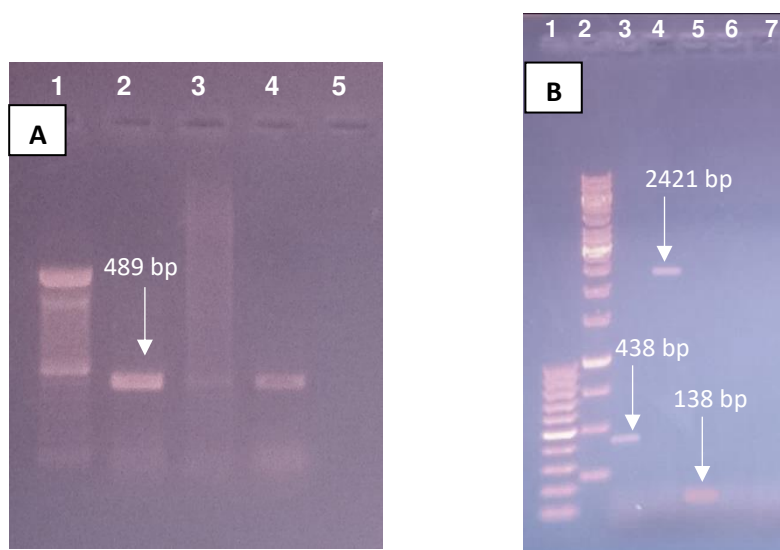


Figure 4.12: 1.2% Agarose Gel Electrophoresis for analysis of PCR products from amplification of AsUSP Gene 2, AsUSP Gene 4 and AsUSP Gene 8

A): Lane 1: 100 bp ladder, Lane 2 and 4: AsUSP Gene 2, Lane 5: Negative Control

B): Lane 1: 100 bp ladder, Lane 2: 1kb ladder, Lane 3: AsUSP Gene 4, Lane 4: AsUSP Gene 8, Lane 5: Positive Control (β -actin), Lanes 6 and 7: Negative Control

4.12 Ligation with pCR®II vector, Transformation in Competent *E. coli* TOP 10 and Blue White Screening

For each of the 9 USP genes of *A. sisalana*, gel extraction was performed for purification of fragment and afterwards ligated with pCR®II vector and successfully transformed in *E. coli* TOP10 as colonies were obtained after blue-white screening. Figure 4.13 represents results obtained after blue-white screening of clones with AsUSP Gene 3 and AsUSP Gene 9.

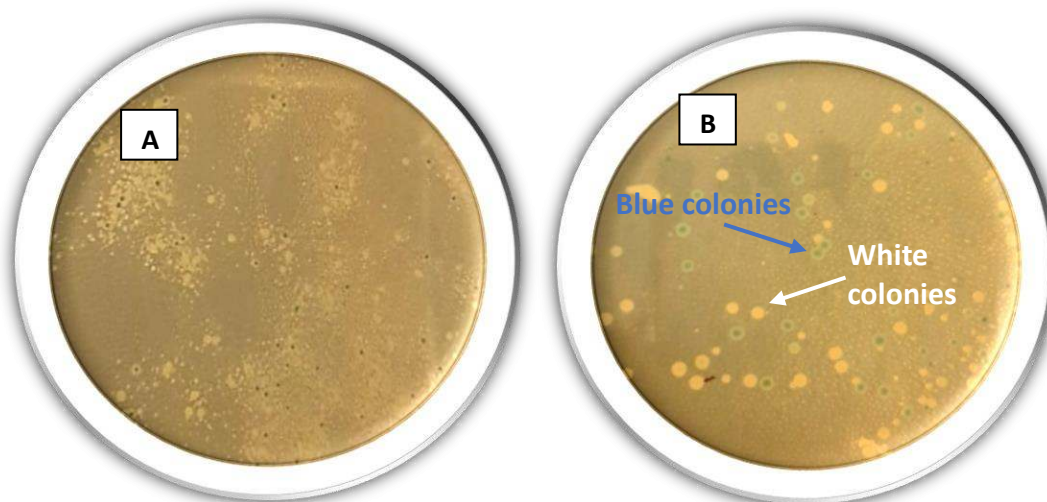


Figure 4.13: Transformation results of pCR®II vector ligated with AsUSP Gene 3 (A) and pCR®II vector ligated with AsUSP Gene 9 (B). Blue and white colonies can be observed in the respective figures

4.13 Confirmation of Recombinant Clones by Restriction with *EcoRI*

After transformants of all amplified genes were obtained, white colonies obtained from blue-white screening were picked (a single colony) from the plates and sub-cultured, followed by recombinant plasmids extraction and restriction analysis with *EcoRI* and analysis by 1 % Agarose Gel Electrophoresis. Results of some clones (i.e. those obtained for AsUSP Gene 3, AsUSP Gene 7 and AsUSP Gene 9) are presented in Figures 4.14 A-B.

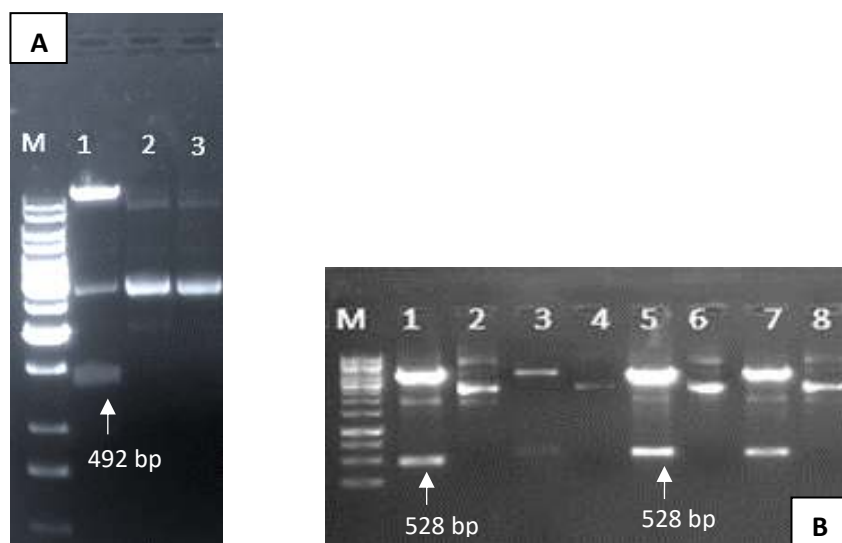


Figure 4.14: 1% Agarose Gel Electrophoresis for analysis of restriction products with *EcoRI* digestion

Figure 4.14 A: Lane M: 1 kb ladder, Lane 1: recombinant pCR®II with AsUSP Gene 3 digested with *EcoRI*, Lanes 2 and 3: recombinant pCR®II with AsUSP Gene 3 undigested

Figure 4.14 B: Lane M: 1 kb ladder, Lanes 1 and 3: recombinant pCR®II with AsUSP Gene 7 digested with *EcoRI*, Lanes 2 and 4: recombinant pCR®II with AsUSP Gene 7 undigested, Lanes 5 and 7: recombinant pCR®II with AsUSP Gene 9 digested with *EcoRI*, Lanes 6 and 8: recombinant pCR®II with AsUSP Gene 9 undigested

4.14 Sequencing Analysis and Submission to NCBI

After transformation of recombinant clones, those giving the expected band sizes of respective genes after restriction digestion with *EcoRI*, were further submitted for sequencing analysis. All AsUSP genes (i.e. AsUSP Genes 1-9) were successfully sequenced and submitted to NCBI, Genbank. Accession Numbers for each of the genes under study are as follows and are expected to be publicly released in Feb-Mar, 2023: AsUSP Gene 1: OP243714, AsUSP Gene 2: OP243715, AsUSP Gene 3: OP243718, AsUSP Gene 4: OP243720, AsUSP Gene 5: OP243719, AsUSP Gene 6: OP169013, AsUSP Gene 7: OP243716, AsUSP Gene 8: OP492076 and AsUSP Gene 9: OP243717. Figures 4.15 A-I represent nucleotide sequences of AsUSP Gene 1-9 along with relevant information submitted to NCBI Genbank.

LOCUS	AsUSP(1)	738 bp	mRNA	linear	PLN 16-AUG-2022		
DEFINITION	Agave sisalana Universal Stress Protein Gene PHOS32-like.						
ACCESSION	AsUSP(1)						
VERSION							
KEYWORDS	.						
SOURCE	Agave sisalana						
ORGANISM	Agave sisalana						
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
	Spermatophyta; Magnoliopsida; Liliopsida; Asparagales;						
	Asparagaceae; Agavoideae; Agave.						
REFERENCE	1 (bases 1 to 738)						
AUTHORS	Jabbar,B., Hassan,S. and Rashid,B.						
TITLE	Direct Submission						
JOURNAL	Submitted (16-AUG-2022) Centre of Excellence in Molecular Biology						
	(CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz						
	Baig Sector 1, Lahore, Punjab 53700, Pakistan						
COMMENT	Bankit Comment: TOTAL # OF SEQS:1.						
FEATURES	Location/Qualifiers						
source	1..738						
	/organism="Agave sisalana"						
	/mol_type="mRNA"						
	/db_xref="taxon:442491"						
	/identified_by="Basit Jabbar"						
gene	1..738						
	/gene="universal stress protein PHOS32-like gene"						
CDS	1..738						
	/gene="universal stress protein PHOS32-like gene"						
	/codon_start=1						
	/product="universal stress protein PHOS32-like"						
	/translation="MQSDSRHQSPKSAAAAAAQTIQSSPRFFFSSAAANPSAASH						
	RRIAIAVDLSDESAYAVKWAVQNYLRPGDAVILLHVRPTSVLYGADWGSIDLSPSET						
	PTEESQQKLEDDFDAFTTTKAQDLAQPLIDSQIPFKIHIVKDHDMKERLCLEVERLGL						
	SAVIMGSRGFGASRRGNKGRLGVSVDYCVHHCVCVVVVRFPDEGTEAAGGVAGPVTE						
	EKEEAELHPVLEEELEYHDADEEHKDA"						
BASE COUNT	141 a	247 c	222 g	128 t			
ORIGIN							
	1	atgcaagact	ccgatcgcca	tcagtcccca	aaatccgccg	ccgccgccgc	cgccgcgcag
	61	accatccagc	cctcctcccc	ccgcttcttc	ttctcctcgg	ccgccgccgc	caaccctctg
	121	gccgcctccc	accgccgcat	cgccatcgcc	gtcgacctct	ccgacgagtc	cgccctacgcc
	181	gtcaagtggg	ccgtgcagaa	ctacctccgc	cctggcgacg	ccgtcatcct	cctgcacgtc
	241	cgccccacct	ccgtcctcta	tggcgccgac	tggggatcca	tcgacctctc	ccctccctcc
	301	gaaaccccca	ccgaggagtc	acagcagaag	ctcgaggacg	atttcgacgc	cttcaccacc
	361	accaaggccc	aagatctcgc	tcagccgctg	atcgactcgc	agatcccctt	caagatccac
	421	atcgtcaagg	atcatgatat	gaaagagagg	ctgtgcctcg	aggtggagag	gctggggctg
	481	agtgcggtga	tcatgggtag	tcgagggttt	ggagcttcga	gaaggggcaa	caaggggagg
	541	cttggggagt	tcagcgatta	ctgcgtgcat	cattgcgtgt	gcccgttgtt	ggtggtgaga
	601	tttctgatg	aggggactga	agccgccgga	ggggtggcgg	gaccctgac	cgaggagaaa
	661	gaggaggcgg	agctgcaccc	ggtgctggag	gaggagctgg	agtatcatga	tgctgatgag
	721	gagcataaag	atgcttga				
	//						

Figure 4.15 A-I: Nucleotide sequences of AsUSP Gene 1-9 along with relevant information submitted to NCBI Genbank

Figure 4.15 A: Nucleotide sequence of AsUSP Gene 1 along with relevant information submitted to NCBI Genbank

LOCUS AsUSP(2) 489 bp mRNA linear PLN 16-AUG-2022
 DEFINITION Agave sisalana Universal Stress Protein Gene PHOS32-like.
 ACCESSION AsUSP(2)
 VERSION B
 KEYWORDS .
 SOURCE Agave sisalana
 ORGANISM Agave sisalana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; Liliopsida; Asparagales;
 Asparagaceae; Agavoideae; Agave.
 REFERENCE 1 (bases 1 to 489)
 AUTHORS Jabbar,B., Hassan,S. and Rashid,B.
 TITLE Direct Submission
 JOURNAL Submitted (16-AUG-2022) Centre of Excellence in Molecular Biology
 (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz
 Baig Sector 1, Lahore, Punjab 53700, Pakistan
 COMMENT Bankit Comment: TOTAL # OF SEQS:1
 ##Assembly-Data-START##
 Sequencing Technology :: Sanger dideoxy sequencing
 ##Assembly-Data-END##
 FEATURES
 source Location/Qualifiers
 1..489
 /organism="Agave sisalana"
 /mol_type="mRNA"
 /db_xref="taxon:442491"
 /identified_by="Basit Jabbar"
 gene 1..489
 /gene="universal stress protein PHOS32-like gene"
 CDS 1..489
 /gene="universal stress protein PHOS32-like gene"
 /codon_start=1
 /product="universal stress protein PHOS32-like"
 /translation="MPDDVTLKSNEERINSPSLNFTSDPSSKTPVRIHSHSQLILVS
 FLKWPEMGGNRTVGIGMDYSPMSKTA VRWVIGNLVAEGDRIVLIHVQPSKSESPQKTL
 WGD TGTPLIPLDEFKEMNLPKQYGINPDPEVLQLLEDASKTKKGEGGVQGLLGRCKRE
 AL"
 BASE COUNT 135 a 121 c 133 g 100 t
 ORIGIN
 1 atgcccgatg acgtgacact caaaagcaac gaggagagaa taaattcccc ttcgctactc
 61 aatttcactt cagaccccag tagcaaaacg cctgtacgca tacattcaca tcccagctc
 121 attctcgtga gtttcctcaa gtggccggag atgggaggaa atcggacggt cgggatcggg
 181 atggactact cgcccatgag caaaacggcg gtccggtggg tgatcggcaa ccttgtcgcg
 241 gaaggagatc gcattgttct gatccatgtg cagcccagca agtcggagag ccctcagaaa
 301 actctgtggg gtgacacagg aacacctctg atccctcttg acgagttcaa ggaaatgaac
 361 ttgccgaagc agtacggaat taatccagat cccgaggtgc tccaactcct tgaggatgca
 421 tccaagacaa aaaaaggtga aggtggtgtc caaggtctat tggggcgatg caagagagaa
 481 gctttgtga
 //

Figure 4.15 B: Nucleotide sequence of AsUSP Gene 2 along with relevant information submitted to NCBI Genbank

LOCUS AsUSP(3) 492 bp mRNA linear PLN 17-AUG-2022
 DEFINITION Agave sisalana universal stress protein PHOS32, mRNA.
 ACCESSION AsUSP(3)
 VERSION C
 KEYWORDS .
 SOURCE Agave sisalana
 ORGANISM Agave sisalana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; Liliopsida; Asparagales;
 Asparagaceae; Agavoideae; Agave.
 REFERENCE 1 (bases 1 to 492)
 AUTHORS Jabbar,B., Hassan,S., Maqsood,I. and Rashid,B.
 TITLE Direct Submission
 JOURNAL Submitted (17-AUG-2022) Centre of Excellence in Molecular Biology
 (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz
 Baig Sector 1, Lahore, Punjab 53700, Pakistan
 COMMENT Bankit Comment: TOTAL # OF SEQS:1
 ##Assembly-Data-START##
 Sequencing Technology :: Sanger dideoxy sequencing
 ##Assembly-Data-END##
 FEATURES
 source Location/Qualifiers
 1..492
 /organism="Agave sisalana"
 /mol_type="mRNA"
 /db_xref="taxon:442491"
 /identified_by="Basit Jabbar"
 gene 1..492
 /gene="universal stress protein PHOS32 gene"
 CDS 1..492
 /gene="universal stress protein PHOS32 gene"
 /codon_start=1
 /product="universal stress protein PHOS32"
 /translation="MGGDRKIGVALDFSKSSKMALQWAVDNLDDKGDTLVVIHVTPAK
 PDESKHALWSESGSPLIPMSELRDPAALKHYDLECDMEVLDLLDTSRQKEVTIVMKL
 YWGDAREKLCEAVHDLKLDLSVMGSRGLSQIRRIILLGSVTNYVLSNASCPVTIIKDPN
 VKH"
 BASE COUNT 114 a 127 c 161 g 90 t
 ORIGIN
 1 atgggaggag ataggaagat cggggtggcg ctggacttct cgaagagcag caagatggcg
 61 ctgcagtggg cgggtggacaa cctcctcgac aagggcgaca ccctcgctcg gatccacgtc
 121 acgccggcga agcccgacga gtccaagcat gccctctggt ccgaatccgg atctcctctg
 181 attccgatgt cggagctccg cgatccggcg gcgctgaagc actacgatct ggagtgcgac
 241 atggaggtgc tcgacttgct agacaccact tcgagacaga aggaggtgac gattgtgatg
 301 aagctgtact ggggagacgc gagggagaag ctgtgcgagg cgggtgcacga tctgaagctc
 361 gactcgctgg tcatggggag cagaggcctc agccagatcc ggaggattct gctcggcagt
 421 gtaacaaact atgtgctgtc aaatgcgtcg tgccctgtga caatcatcaa agatccaaac
 481 gtcaagcatt ga
 //

Figure 4.15 C: Nucleotide sequence of AsUSP Gene 3 along with relevant information submitted to NCBI Genbank

LOCUS	AsUSP(4)	438 bp	mRNA	linear	PLN 17-AUG-2022
DEFINITION	Agave sisalana universal stress protein PHOS34.				
ACCESSION	AsUSP(4)				
VERSION					
KEYWORDS	.				
SOURCE	Agave sisalana				
ORGANISM	Agave sisalana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Asparagales; Asparagaceae; Agavoideae; Agave.				
REFERENCE	1 (bases 1 to 438)				
AUTHORS	Jabbar,B., Hassan,S., Maqsood,I. and Rashid,B.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-AUG-2022) Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz Baig Sector 1, Lahore, Punjab 53700, Pakistan				
COMMENT	Bankit Comment: TOTAL # OF SEQS:1				
	##Assembly-Data-START## Sequencing Technology :: Sanger dideoxy sequencing ##Assembly-Data-END##				
FEATURES	Location/Qualifiers				
source	1..438 /organism="Agave sisalana" /mol_type="mRNA" /db_xref="taxon:442491" /identified_by="Basit Jabbar"				
gene	1..438 /gene="universal stress protein PHOS34 gene"				
CDS	1..438 /gene="universal stress protein PHOS34 gene" /codon_start=1 /product="universal stress protein PHOS34" /translation="MEPVNEDVEEYSWREVILPSLIPVVPPEPELERETGERRRGRDLL VAVDHGPNSQHAFRWALHLCRLADTVHLIHAVSSVQNEVVYEKSRELMETLAVEAFQ VSLVKTVARIVEGDAGKAICREAERVKPAAVIMGTRGRSLFQR"				
BASE COUNT	87 a	128 c	154 g	69 t	
ORIGIN	1 atggagccag tgaacgaaga cgtggaggag tacagctggc gggaggtgat cttaccgtcg 61 ctgatcccg tggccccga gcccgagctc gagcgggaga ccggcgagcg gcggcgcggc 121 cgcgacctcc tcgtcgccgt cgaccacggg cccaacagcc agcacgcctt cagatgggcc 181 ctggcccacc tctgccgcct cgccgacacc gtccacctca tccacgccgt ctccagtgtg 241 cagaatgagg tgggtgatga gaagagcagg gaactgatgg agactctggc cgttgaggcc 301 ttccaggtct cactggtgaa gacagttgct cggattgttg aaggagacgc ggggaaggcc 361 atatgcaggg aggccgagag ggtgaagcca gcggccgtta tcatgggaac ccgtgggagg 421 agcctcttcc aaaggtag				
	//				

Figure 4.15 D: Nucleotide sequence of AsUSP Gene 4 along with relevant information submitted to NCBI Genbank

LOCUS AsUSP(5) 498 bp mRNA linear PLN 17-AUG-2022

DEFINITION Agave sisalana universal stress protein PHOS32.

ACCESSION AsUSP(5)

VERSION

KEYWORDS .

SOURCE Agave sisalana

ORGANISM Agave sisalana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Asparagales; Asparagaceae; Agavoideae; Agave.

REFERENCE 1 (bases 1 to 498)

AUTHORS Jabbar,B., Hassan,S., Maqsood,I. and Rashid,B.

TITLE Direct Submission

JOURNAL Submitted (17-AUG-2022) Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz Baig Sector 1, Lahore, Punjab 53700, Pakistan

COMMENT Bankit Comment: TOTAL # OF SEQS:1

##Assembly-Data-START##

Sequencing Technology :: Sanger dideoxy sequencing

##Assembly-Data-END##

FEATURES

Location/Qualifiers

source 1..498

/organism="Agave sisalana"

/mol_type="mRNA"

/db_xref="taxon:442491"

/identified_by="Basit Jabbar"

gene 1..498

/gene="universal stress protein PHOS32-like gene"

CDS 1..498

/gene="universal stress protein PHOS32-like gene"

/codon_start=1

/product="universal stress protein PHOS32-like"

/translation="MAGDRTVGVAVDSECSKAALRWAVEHLVRDGDHLVLVIVQKQV
FYEGGEMQLWEATGSPLIPLTELSDPGIEKKYGVKPD AETLDIVNTIARQKEIVVGMK
IYWGDPREKICEAIEKIPLSCLVIGNRGLGTLKRIIMGSVSNYVVNNGACPVTVVKSS
EVHGA"

BASE COUNT 119 a 109 c 158 g 112 t

ORIGIN

1 atggcaggcg atgcacggt gggcgtggcg gtggacttct cggagtgcag caaggcggcg

61 cttcgggtggg cggtgagca cctggtccgc gacggcgacc acctcgctct cgtcatcgtc

121 caaaagcagg tcttctatga ggggggcgag atgcagcttt gggaaagccac aggatctccc

181 ctaatccctc tgactgagtt gtcggaccct ggcattgaaa agaaatatgg agtgaagcct

241 gatgctgaga ccctagatat agttaatacc attgcgaggc aaaaagagat tgttggtgga

301 atgaagatct actggggaga cccccgtgaa aagatatgtg aagccattga aaagattcct

361 ttgagctgtc ttgttattgg taatcggggg cttggcacac tcaagaggat tattatgggc

421 agtgtcagca actacgtggt gaacaatggc gcgtgcccag tcacagttgt caagagctct

481 gaagtgcacg gtgcgtag

//

Figure 4.15 E: Nucleotide sequence of AsUSP Gene 5 along with relevant information submitted to NCBI Genbank

LOCUS AsUSP_Gene_6 480 bp mRNA linear PLN 06-AUG-2022

DEFINITION Agave sisalana isolate Universal Stress Protein A-like Protein Gene (6).

ACCESSION AsUSP_Gene_6 F

VERSION

KEYWORDS .

SOURCE Agave sisalana

ORGANISM Agave sisalana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Asparagales;
Asparagaceae; Agavoideae; Agave.

REFERENCE 1 (bases 1 to 480)

AUTHORS Jabbar,B., Maqsood,I., Hassan,S. and Rashid,B.

TITLE Expression Analysis of Agave sisalana Universal Stress Protein (USP) Genes under Biotic Stress

JOURNAL unpublished

REFERENCE 2 (bases 1 to 480)

AUTHORS Jabbar,B., Maqsood,I., Hassan,S. and Rashid,B.

TITLE Direct Submission

JOURNAL Submitted (06-AUG-2022) Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz Baig Sector 1, Lahore, Punjab 53700, Pakistan

COMMENT Bankit Comment: TOTAL # OF SEQS:1

##Assembly-Data-START##
Sequencing Technology :: Sanger dideoxy sequencing
##Assembly-Data-END##

FEATURES

source Location/Qualifiers
1..480
/organism="Agave sisalana"
/mol_type="mRNA"
/isolate="Universal Stress Protein A-like Protein Gene (6)"
/db_xref="taxon:442491"
/identified_by="Basit Jabbar"

CDS
1..480
/codon_start=1
/product="universal stress protein A-like protein"
/translation="MAAAVESNEKKFMVAIDSECSHHALNWTLTNLRNSLTSPVLVY
TVQPLSSLAFLSAASMGAAPELIQSVAEHQKKVSMALLEKAKEICAQHGVVAETVTE
VGDPKEAICDAVEKFKINLLILGNHGRGAFQRAFLGSVSNYCAHNAKCPVLVVKNSG"

BASE COUNT 115 a 121 c 130 g 114 t

ORIGIN

1 atggcagcgg cagtggagag caatgagaag aagtttatgg tggccatcga cgagagcgag
61 tgcagccacc acgctctcaa ctggaccctc accaatctcc gcaactctct cacctcgcc
121 ctgctccttt atacgggtcca gcccctctcc agtctcgctt ttctctccgc tgcttccatg
181 ggcgccgctc ctgctggagt gattcagtcg gtggcagagc accagaagaa ggtgtcaatg
241 gctcttttgg agaaagccaa ggagatctgt gctcaacatg gggtagtcgc tgaaacagtt
301 acagaggtag gggatcctaa agaagcaata tgtgacgcag ttgaaaagtt caaatcaat
361 ctgctaatat tgggtaacca tggctgagga gccttccaaa gggctttctt ggggagcgtc
421 agcaattact gtgcgcataa tgccaagtgt cctgtccttg ttgtgaagaa ctctgggtga

//

Figure 4.15 F: Nucleotide sequence of AsUSP Gene 6 along with relevant information submitted to NCBI Genbank

LOCUS	AsUSP(7)	528 bp	mRNA	linear	PLN 16-AUG-2022
DEFINITION	Agave sisalana Universal Stress Protein Gene PHOS34-like.				
ACCESSION	AsUSP(7)	G			
VERSION					
KEYWORDS	.				
SOURCE	Agave sisalana				
ORGANISM	Agave sisalana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Asparagales; Asparagaceae; Agavoideae; Agave.				
REFERENCE	1 (bases 1 to 528)				
AUTHORS	Jabbar,B., Hassan,S. and Rashid,B.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-AUG-2022) Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz Baig Sector 1, Lahore, Punjab 53700, Pakistan				
COMMENT	Bankit Comment: TOTAL # OF SEQS:1				
	##Assembly-Data-START## Sequencing Technology :: Sanger dideoxy sequencing ##Assembly-Data-END##				
FEATURES	Location/Qualifiers				
source	1..528 /organism="Agave sisalana" /mol_type="mRNA" /db_xref="taxon:442491" /identified_by="Basit Jabbar"				
gene	1..528 /gene="universal stress protein PHOS34-like gene"				
CDS	1..528 /gene="universal stress protein PHOS34-like gene" /codon_start=1 /product="universal stress protein PHOS34-like" /translation="MAEAEAAAVVETKTEGVTVMVVGIDDSEHSYYALQWTLHLHFFSG RASAASPFKLIVVHAKPSPSSVSVSLAGPGAADVLPYVESDLRKIAARVLEKAKELCAA NSVGEVLLEVVEGDARNVLCEAVEKHHADMLVVGSHGYAIKRVVLGSVSDYCAHHAH CTVMIVKKPKPMHKH"				
BASE COUNT	105 a	137 c	176 g	110 t	
ORIGIN	1 atggcggagg cggaggcggc agcggtaggta gagacgaaga cggaggggggt gacggtgatg 61 gtggtgggga tcgacgacag cgagcacagc tactacgccc tccagtggac cctgctccac 121 ttcttctccg gcagggcctc cgccgcctcc cctttcaagc tcatcgctcg tcaagccaag 181 ccctcgccct cctccgctcg cagcctcgcc ggccctgggg ctgctgatgt gctgccgtac 241 gtggagtccg atctgaggaa gatcgccgcc aggttcttgg agaaggccaa ggagctctgc 301 gccgcaaatt ccgttggtga agtgctgctt gaagtgggtg aaggagatgc aaggaatgtg 361 ctttgtgagg ctgtggagaa gcacatgctt gatattgttg ttgtgggcag ccatggctat 421 ggagctatca aaagggttgt ttgggaagc gtcagcgatt actgtgctca tcatgcgcac 481 tgcactgtga tgatagtga gaagccaag cccatgcaca agcactaa				
	//				

Figure 4.15 G: Nucleotide sequence of AsUSP Gene 7 along with relevant information submitted to NCBI Genbank

LOCUS AsUSP(8) 2421 bp mRNA linear PLN 05-SEP-2022
 DEFINITION Agave sisalana U-box domain-containing protein 35-like.
 ACCESSION AsUSP(8)
 VERSION
 KEYWORDS
 SOURCE Agave sisalana
 ORGANISM Agave sisalana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; Liliopsida; Asparagales;
 Asparagaceae; Agavoideae; Agave.
 REFERENCE 1 (bases 1 to 2421)
 AUTHORS Jabbar,B., Hassan,S. and Rashid,B.
 TITLE Direct Submission
 JOURNAL Submitted (05-SEP-2022) Centre of Excellence in Molecular Biology
 (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz
 Baig Sector 1, Lahore, Punjab 53700, Pakistan
 COMMENT Bankit Comment: TOTAL # OF SEQS:1

H

##Assembly-Data-START##
 Sequencing Technology :: Sanger dideoxy sequencing
 ##Assembly-Data-END##

FEATURES Location/Qualifiers
 source 1..2421
 /organism="Agave sisalana"
 /mol_type="mRNA"
 /db_xref="taxon:442491"
 /identified_by="Basit Jabbar"
 gene 1..2421
 /gene="U-box domain-containing protein 35-like gene"
 CDS 1..2421
 /gene="U-box domain-containing protein 35-like gene"
 /codon_start=1
 /product="U-box domain-containing protein 35-like"
 /translation="MEIEEMKERGEPLTPSLLTVAVAVHGTKNSKYALKWALEKFIPE
 GRIFFRLLYVRPKITMVPTPMGNYLPISKVRDDVASAYMKDVEMQTNMFLPYKKMCT
 QKKVEAEAVVIEADDVAEALSAEVNKFVSKLVIGVSSRNIFTRKFKGSKISSLISEC
 IPTFCTLYVVSCKGLSSVHASTSGAVESTGHESISNVSCGSLTHSSSERSEWINSGT
 SNFSLSRSSSLNQRDQTLKVTNQGFVDMIPCPGKHLTRALSFSYEDFMGSCSNYS
 EIQCYNSEFSSYISCHTDLPWNDSQASADTVKYSSLSQNEVNIDFELERLRIELKH
 LQGMYLEAENESSEASRRLNELSARREKEAVRLNEIRIREYKAREMAKKERERREAAE
 KEAKYVRVCAEKEVVQRKDAEDNAAREATEKQRLERALSAGDGQYKKYTWEDIESATS
 SFSDALKIGSGANGTVYKASFRHTIAAVKILHSNEGRGTEQYKQELEILGRIRHPHLL
 LLLGACPENGCLVYEMENGLQDRLQKDDTPPLPWQYRYRIAWELIASLVFLHNSK
 PDPIIHRDLKPANILLDSNFVSKIGDVGLSTLLPAMNFTMSTMYKDTAPVGTFFYIDP
 EYQRTGQVSPKSDTYALGMVILQLLTGRPPMGLAHIVEKAVEGGFLVDILDPEAGQWP
 IVESQELAFGLSCVELRRDRPELKDMILPALEKLDKFADKARDLVCDVPSATPGHF
 ICPILQEVMEPCVASDGYTYERRAIERWFGMNDNSPLTNMKLPHKTLITNQSLLSAI
 KDWRARNP"

BASE COUNT 725 a 489 c 573 g 634 t
 ORIGIN

```

1 atggaattg aagagatgaa ggaaaggggt gaacccttaa ctccctcact gttaactggt
61 gcagtggctg tacacggaac caaaaacagc aagtatgcac ttaaattgggc attggagaag
121 ttcattcctg aggggagaat tttttttatg cttttatagc tgcggccgaa gatcacaatg
181 gtccctacac caatgggtaa ctatcttcct atttcaaaag tgcgggatga tgtcgctct
241 gcatacatga aagatgtaga aatgcacaac aattccatgt ttctacccta caagaaaatg
301 tgtactcaga aaaaggtaga agcgggaagct gtggtaatg aagcagatga tgtagcagag
361 gcattgtcag cagaggtcaa taaatttgca gtcagcaagc ttgtaatagg tgtttcttcg
421 agaaatatct ttactaggaa atttaagggc agcaaaatat cctcaagaat ctcagaatgc
481 attccaacct tttgcacact gtatgttggt tccaaaggaa aattgtcatc tgtacatgca
541 tcaacgtctg gggcagttga gagtaccggg catgaaagta tttcaagcaa tgtttctctg

```

```

601 ggcagtttaa cccattcttc cagtgaacgg tcagaatgga taaactcagg gacgtcaaat
661 ttctctcttt ctctgtcttc ttctctttca aatcagcgag accaaactct taaagttaca
721 aaccagggct ttgttgatat gataccatgc ccaggcgaga agctccatac tagagcccta
781 tctttttcta cttatgaaga ttttatgggt tcctgttcta attcatatga gattcaatgc
841 gtcaatagtg aattttctag ctacataagc tgtcatacag atttgccctc atggaattca
901 gatcaggcct ccgcatcaga tactgtgaag tactcttcac tctctcagaa cgagggtaat
961 attgattttg agcttgaaag actgagaatt gaacttaagc atcttcaggg gatgtatgaa
1021 cttgctgaga atgaatcttc tgaggcatcc cgtcggttga atgaattaag tgcacgtcgt
1081 gaaaaggaag cagtgaact caatgagata aggattagag aatataaggc cagagaaatg
1141 gccaaaaaag agagggaag acggaagct gcagaaaaag aggcaaaata tgttagggta
1201 tgtgctgaaa aggaggtggt acaaaggaag gacgctgagg acaatgctgc tcgtgaagcc
1261 actgagaaac agaggctgga gagggctctt gcatctggtg atgggcaata caagaagtat
1321 acatgggaag acattgaatc tgctacctca tcattctctg atgctcttaa aattggatca
1381 ggggcaaatg gaacagtta taaagcaagc ttccgccaca ctattgcagc tgtgaaaatt
1441 ctccactcca atgagggcag gggaactgag caatacaagc aagagcttga aatcctgggt
1501 agaattcggc accctcatct actgctcctc ctagggtgct gtcccagagaa tggttgcctg
1561 gtgtatgagt acatggaaaa tggcagcctc caagatagggt tacagcgcaa ggatgataca
1621 cctcctcttc catggcagta ccgctacaga atagcttggg aaattgcctc cgcactcgta
1681 tttctacaca actcaaagcc agatcccata atccaccgtg atctcaagcc cgcgaacatt
1741 ttgcttgata gcaatttcgt gagcaagata ggggatgttg gcctctctac gttgttacc
1801 gctatgaatt tcacaatgtc tacaatgtac aaggacacgg caccagttgg cacattcttc
1861 tacatagacc ctgagtacca gagaacgggt caagtctcac caaagtcaga tacctatgcc
1921 ttagggatgg tcatcttgca attacttact ggaaggccac caatgggact tgcccacatt
1981 gtagaaaagg cagtgaagg tggtttcttg gtagatattt tggatcctga agctggacaa
2041 tggcctatag tagagtctca ggaattggct tttcttgggc taagttgtgt ggaactgaga
2101 cgaagggacc ggctgagtt gaaggacatg attcttccgg cacttgaaaa attgaaagat
2161 tttgctgaca aagcacgaga tttggtatgc gatgtgccct ctgcaacacc aggccatttc
2221 atctgtccaa tacttcagga ggtgatggaa gatccttgtg ttgcttctga tggatataca
2281 tacgagcgca gagcaatcga gagatggttt gggatgaatg acaattctcc actgacaaac
2341 atgaagttgc ctcaaaaac tctcataact aatcaatcac ttctttctgc gattaaggat
2401 tggagagcca ggaatccatg a

```

//

Figure 4.15 H: Nucleotide sequence of AsUSP Gene 8 along with relevant information submitted to NCBI Genbank

```

LOCUS      AsUSP(9)                      528 bp    mRNA    linear    PLN 17-AUG-2022
DEFINITION Agave sisalana universal stress protein A-like protein gene, mRNA.
ACCESSION  AsUSP(9)
VERSION
KEYWORDS   .
SOURCE     Agave sisalana
  ORGANISM Agave sisalana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; Liliopsida; Asparagales;
            Asparagaceae; Agavoideae; Agave.
REFERENCE  1 (bases 1 to 528)
  AUTHORS  Jabbar,B., Maqsood,I., Hassan,S. and Rashid,B.
  TITLE    Direct Submission
  JOURNAL   Submitted (17-AUG-2022) Centre of Excellence in Molecular Biology
            (CEMB), University of the Punjab, 87 West Canal Rd, Thoker Niaz
            Baig Sector 1, Lahore, Punjab 53700, Pakistan
COMMENT    Bankit Comment: TOTAL # OF SEQS:1

            ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source           1..528
                     /organism="Agave sisalana"
                     /mol_type="mRNA"
                     /db_xref="taxon:442491"
                     /identified_by="Basit Jabbar"
     gene             1..528
                     /gene="universal stress protein A-like protein gene"
     CDS              1..528
                     /gene="universal stress protein A-like protein gene"
                     /codon_start=1
                     /product="universal stress protein A-like protein"
                     /translation="MGSEPTRILMGVNESSIKGYPHASISSRSSFEWTLEKIVRANVG
                     GFKLLFVHVQIPDEDDGFDDMSIYASPEDFKRAKRRDKVRGLHLLEYFVGRCHEYGIA
                     CESWIKKGDPEVICSEVKRVRPDLLEVGSRLGHFQRVFVGTVSEFCVKHADCPVVA
                     IKRKAETPQDPVDD"
BASE COUNT      131 a      108 c      166 g      123 t
ORIGIN
    1 atgggacgcg aaccgactcg gatcttgatg ggggtgaacg agtcgtcgat caagggctac
   61 cctcacgcgt cgatcagcag caggctcgcg ttcgagtgga ccttgaaaaa gatcgtagca
  121 gcaaatgtcg gcgggttcaa gctccttttc gtccacgtcc agattcctga cgaagacggt
  181 ttgatgaca tggatagcat ctatgcctca cctgaagatt ttaagagagc gaagaggagg
  241 gacaaagtca gagggcttca cttgctggag tactttgttg ggcgctgtca cgaatatggg
  301 attgcttgcg aatcatggat taagaaaggt gatccaaaag aagtaatctg cagcgagggtg
  361 aagagagttc gtcctgacct tctggtcggt gggagccgtg ggcttggaac cttccagagg
  421 gtgtttgttg ggacagtgag cgagttctgc gtgaagcatg cagattgccc ggtggtcgcc
  481 atcaagcgaa aggcagaaga aactccacag gatccagttg atgattga
//

```

Figure 4.15 I: Nucleotide sequence of AsUSP Gene 9 along with relevant information submitted to NCBI Genbank

4.15 miRNA Target Prediction Against AsUSP Genes

4.15.1 Predicted Target Sites with RNAhybrid

With the parameters input to TAPIR, RNAhybrid was run, and the output returned 266 predicted target sites of *A. sisalana* miRNAs in AsUSP genes. Score, minimum free energy (mfe), minimum free energy ratio (mfe ratio), gaps, mismatches, sequence alignments etc. were present in the output file. The results were tabulated, and Figure 4.16 shows the predicted target sites with the use of RNAhybrid; target site positions in each gene targeted are indicated against the minimum free energies.

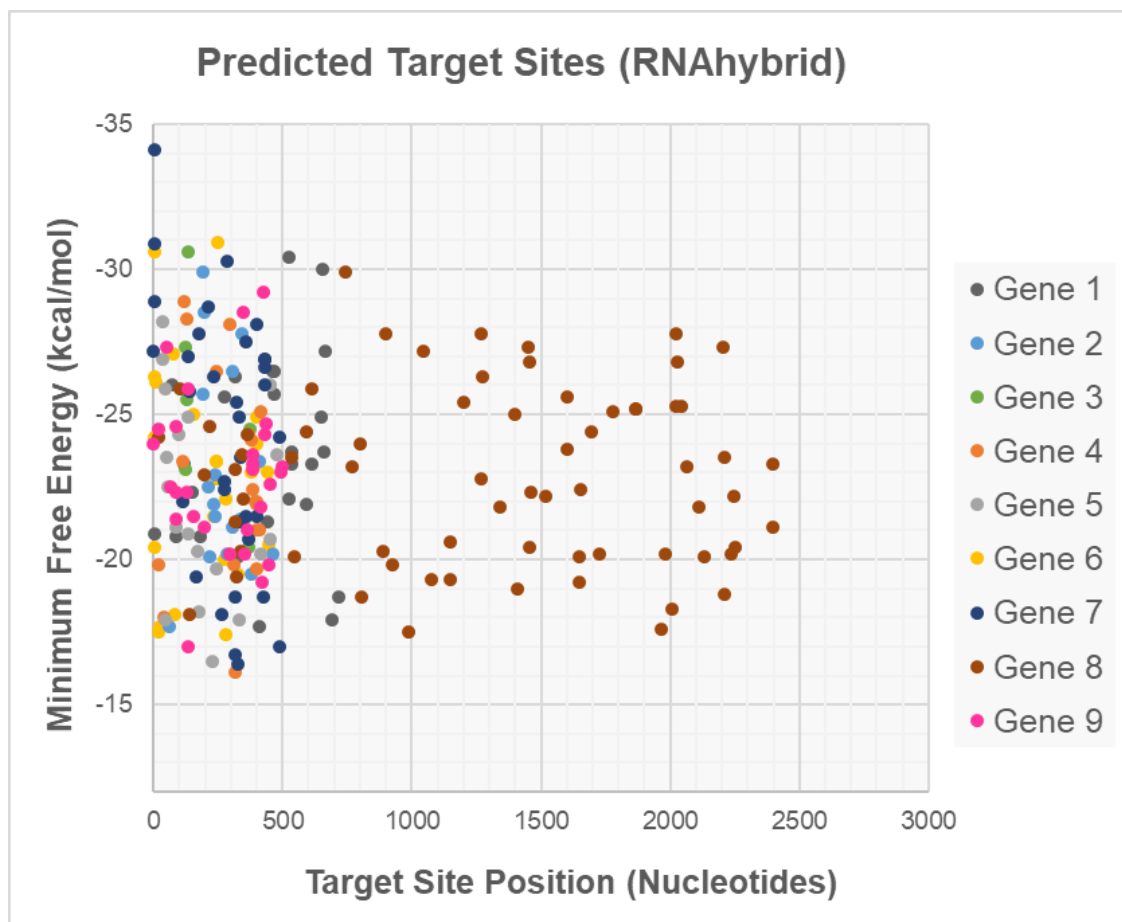


Figure 4.16: Plot of predicted miRNA target sites in AsUSP genes using RNAhybrid

4.15.2 Predicted Target Sites with psRNATarget

Output of psRNATarget under the user-defined parameters resulted in a total of 138 target sites of *A. sisalana* miRNAs in AsUSP genes along with the details such as expectation, miRNA start and end positions, target gene start and end positions, sequence of the aligned fragments, sequence alignments and the type of mRNA translational repression i.e. translational inhibition or target mRNA cleavage. Figure 4.17 shows plot of predicted target sites with psRNATarget algorithm; target site positions in each gene targeted are indicated against the expectation values.

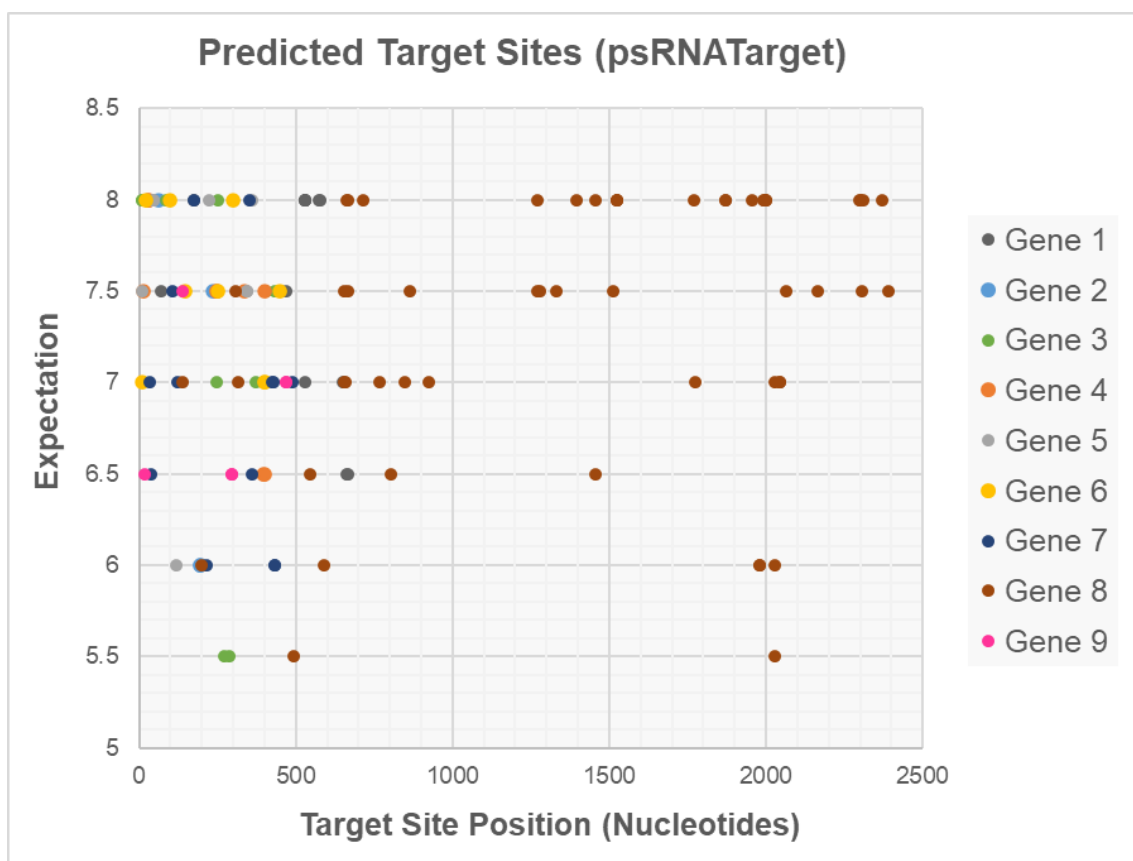


Figure 4.17: Plot of predicted miRNA target sites in AsUSP genes using psRNATarget

4.15.3 Predicted Target Sites with psRobot

From psRobot algorithm, the user defined script was run in Linux OS, and it returned a total of 139 miRNA target predictions. The output file indicated miRNAs and targeted genes, scores and query-subject alignments; results were tabulated. Figure 4.18 depicts the predicted target sites with the use of psRobot; target site positions in each gene targeted are indicated against the respective scores.

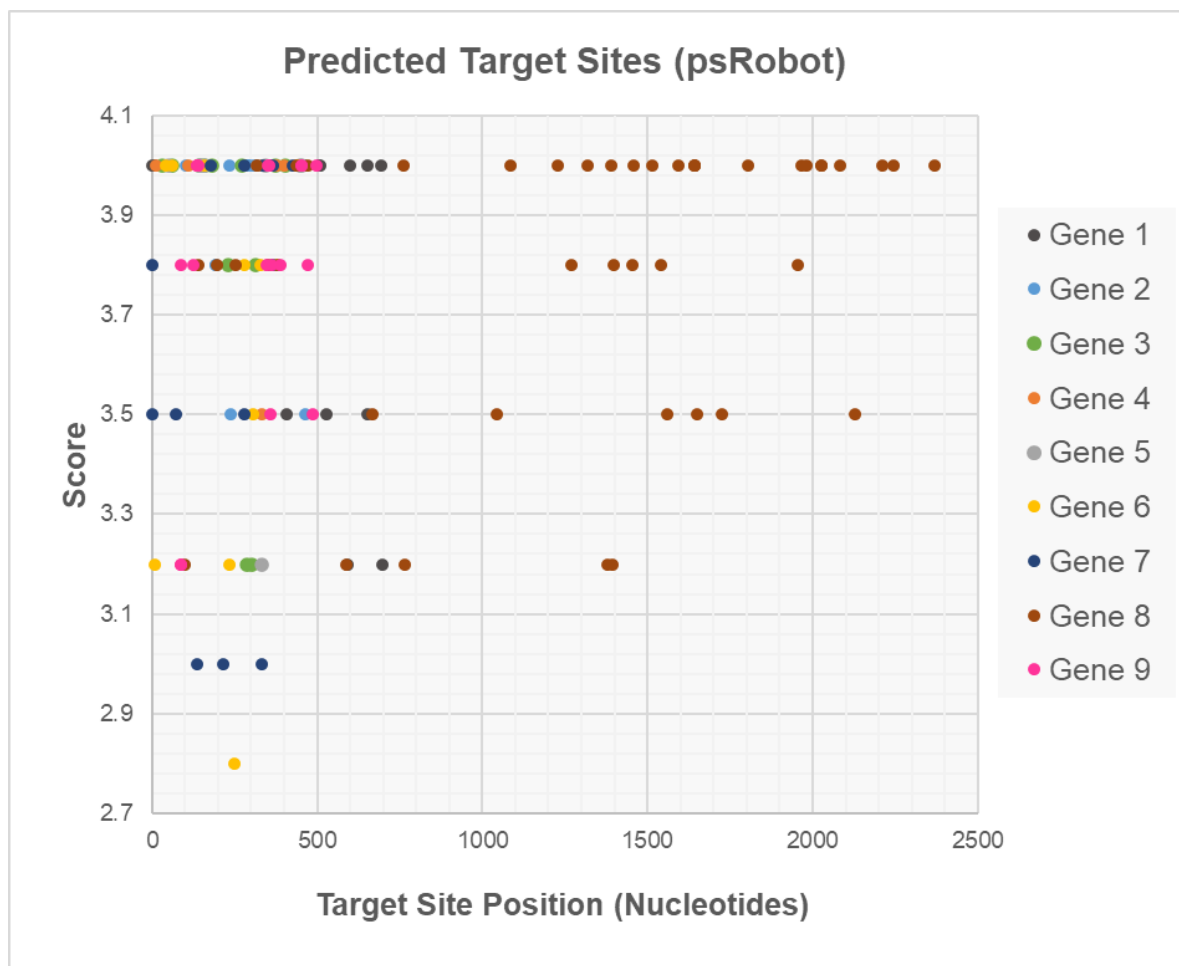


Figure 4.18: Plot of predicted miRNA target sites in AsUSP genes using psRobot

4.15.4 Consensus of miRNA Target Prediction Algorithms

From the consensus of the three miRNA target prediction algorithms i.e., RNAhybrid, psRNATarget and psRobot, a total of 406 unique target site were predicted in AsUSP Genes 1-9. Figure 4.19 shows consensus miRNA target sites identified with the use of three algorithms; target site positions are plotted against minimum free energies (MFE). As can be observed from Figure 4.19, 18 target sites were commonly identified from the three algorithms. Figure 4.20 shows a Venn Diagram indicating union and intersection of the output from the three tools; while the details of intersection are outlined in Table 4.4. It can be observed from Table 4.6 that only AsUSP Gene1, Gene 2, Gene 3, Gene 6, Gene 7 and Gene 8, among all (i.e. nine) AsUSP genes, are targeted considering the consensus of the three miRNA target prediction algorithms, while the miRNA involved in these interactions are asi-miR156c-3p, asi-miR172b-5p, asi-miR444b, asi-miR482a-3p, asi-miR165a-5p, asi-miR166b, asi-miR167c-5p, asi-miR2118b-5p, asi-miR398e, asi-miR399f, asi-miR164a-3p and asi-miR2118l.

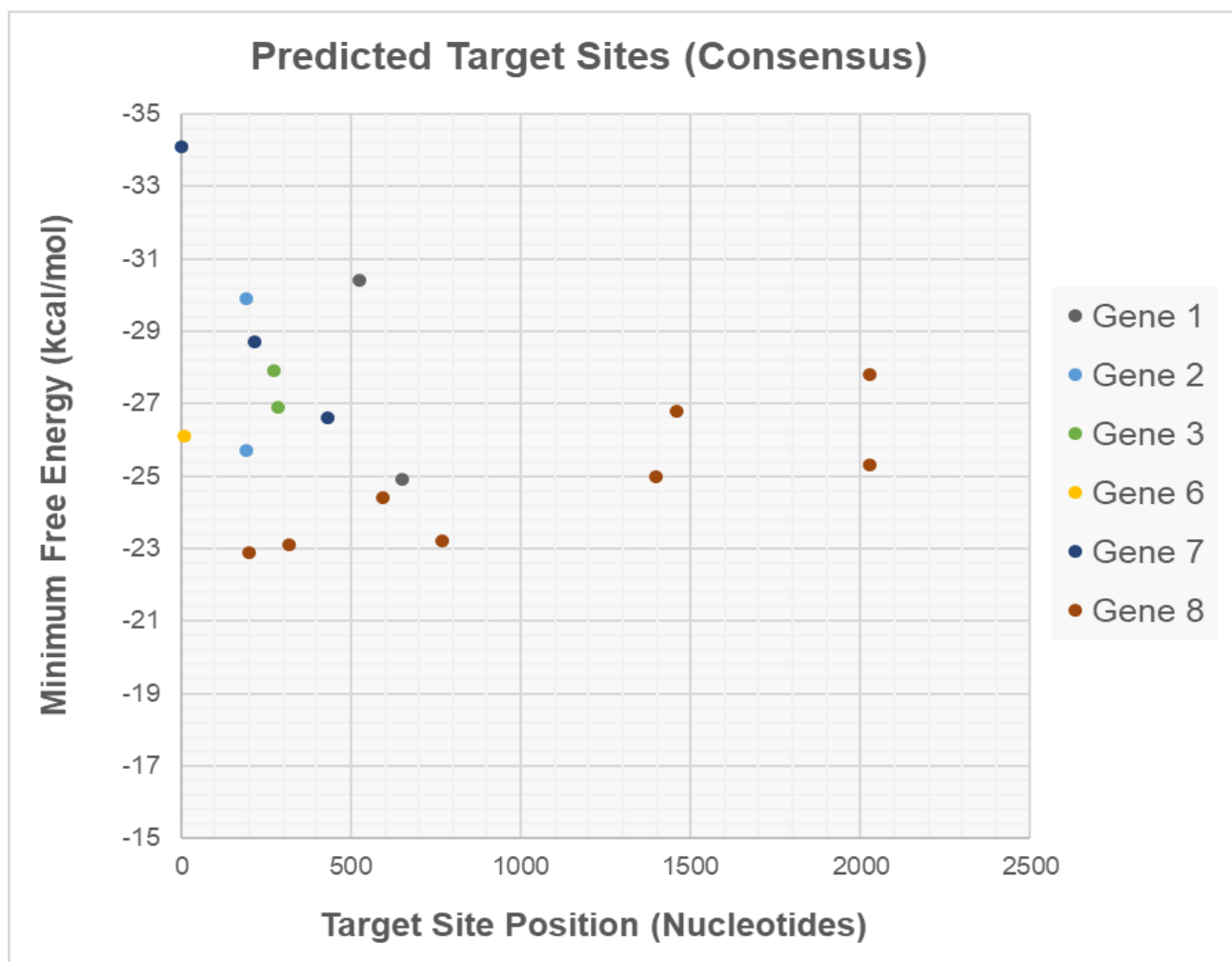


Figure 4.19: Plot of predicted miRNA target sites in AsUSP genes using consensus of three miRNA target prediction algorithms

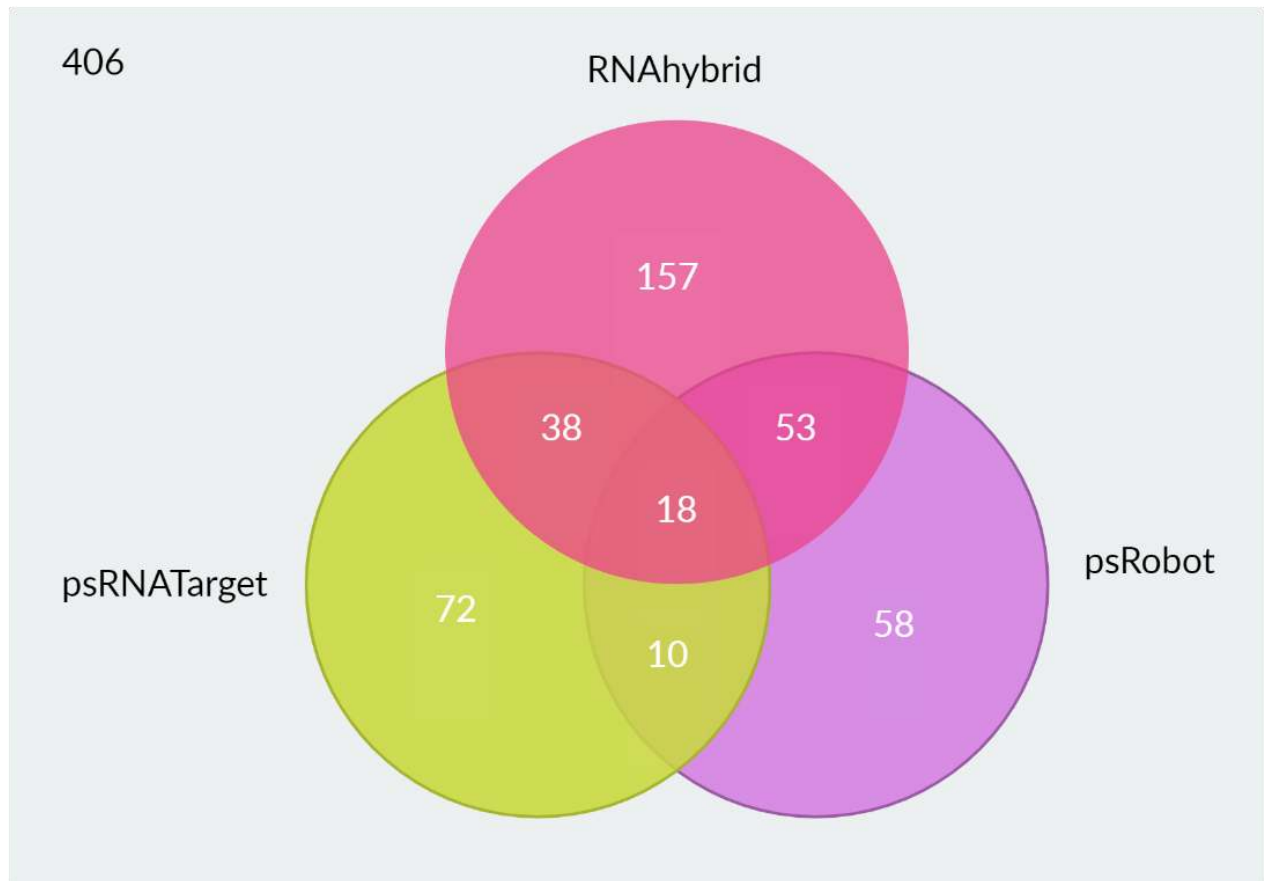


Figure 4.20: Venn Diagram representing union and intersection of the miRNA target sites predicted in AsUSP genes

Table 4.4: Intersection of results of the three miRNA target prediction algorithms (RNAhybrid, psRNATarget and psRobot)

Sr. No.	miRNA	Target	Position (RNAhybrid)	Position (psRNATarget)	Position (psRobot)	Score (RNAhybrid)	mfe (RNAhybrid)	mfe_ratio (RNAhybrid)	Expectation (psRNATarget)	Target_end (psRNATarget)	Inhibition (psRNATarget)	Score (psRobot)
1	asi-miR164a-3p	TRINITY_DN23127_c4_g1_i2.p1(Gene1)	651	651	651	7	-24.9	0.60	7	669	Translation	3.5
2	asi-miR2118l	TRINITY_DN23127_c4_g1_i2.p1(Gene1)	523	529	528	10	-30.4	0.72	7	547	Cleavage	3.5
3	asi-miR165a-5p	TRINITY_DN17561_c0_g1_i1.p1(Gene2)	193	194	193	8.5	-25.7	0.58	6	213	Translation	3.8
4	asi-miR166b	TRINITY_DN17561_c0_g1_i1.p1(Gene2)	193	194	193	7.5	-29.9	0.63	6	214	Translation	3.8
5	asi-miR164a-3p	TRINITY_DN18247_c1_g1_i1.p1(Gene3)	272	273	272	7	-27.9	0.67	5.5	291	Cleavage	4
6	asi-miR172b-5p	TRINITY_DN18247_c1_g1_i1.p1(Gene3)	286	287	286	5.5	-26.9	0.65	5.5	307	Cleavage	3.2
7	asi-miR156c-3p	TRINITY_DN17522_c0_g2_i2.p1(Gene6)	8	9	8	7.5	-26.1	0.70	7	26	Cleavage	3.2
8	asi-miR172b-5p	TRINITY_DN18335_c0_g3_i1.p2(Gene7)	215	216	215	5	-28.7	0.69	6	235	Cleavage	3
9	asi-miR444b	TRINITY_DN18335_c0_g3_i1.p2(Gene7)	2	1	3	9	-34.1	0.74	7	22	Translation	3.5
10	asi-miR482a-3p	TRINITY_DN18335_c0_g3_i1.p2(Gene7)	433	429	428	10	-26.6	0.61	6.5	448	Translation	4
11	asi-miR165a-5p	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	2025	2026	2025	8	-25.3	0.57	5.5	2045	Cleavage	4
12	asi-miR399f	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	1396	1396	1396	7	-25	0.56	8	1414	Cleavage	3.8
13	asi-miR398e	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	1458	1457	1458	7	-26.8	0.57	6.5	1477	Cleavage	4
14	asi-miR2118l	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	318	317	317	8	-23.1	0.55	7	335	Cleavage	4
15	asi-miR2118b-5p	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	199	198	197	7.5	-22.9	0.57	6	216	Cleavage	3.8
16	asi-miR2118b-5p	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	767	768	767	7.5	-23.2	0.57	7	785	Translation	3.2
17	asi-miR167c-5p	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	594	591	589	7	-24.4	0.60	6	609	Translation	3.2
18	asi-miR166b	TRINITY_DN21559_c0_g1_i2.p1(Gene8)	2025	2026	2025	7.5	-27.8	0.59	6	2046	Cleavage	4

4.15.5 Estimations of the miRNA-Gene Hybridization Potential

MFE ratios, as a measure of miRNA-mRNA hybridization potential, were obtained from RNAhybrid algorithm. Figure 4.21 represents bar chart with miRNAs against respective mfe ratio of each miRNA-target interaction. All the 18 miRNA-mRNA pairs had mfe ratio greater than 0.5. miRNA-targets with greater value of mfe ratio indicated a better hybridization potential compared to those with lower values.

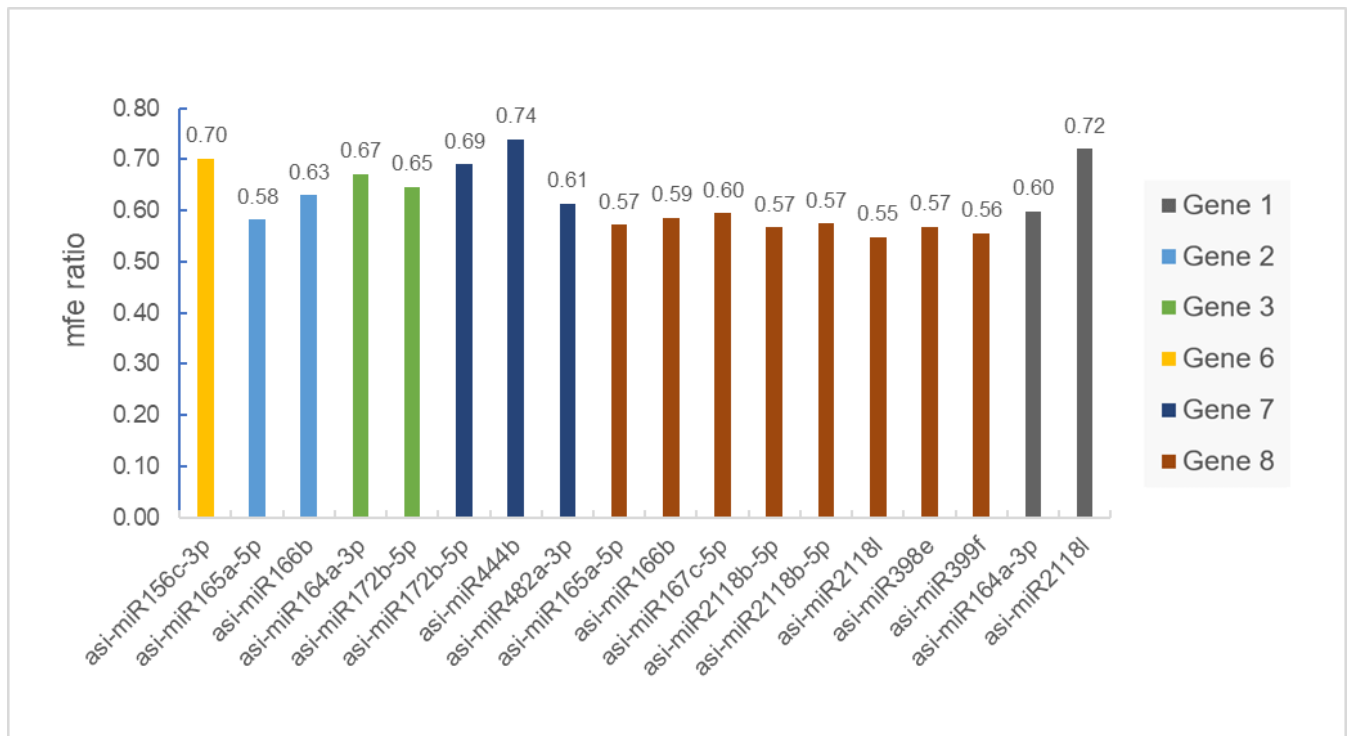


Figure 4.21: Bar chart representing miRNAs and their respective mfe ratio as a measure of miRNA-gene hybridization potential

4.16 Evaluation of asi-miRNA-AsUSP Gene Complexes through Molecular Docking Analysis

From results of molecular docking analysis using Haddock for exploring miRNA-gene interactions in 3D platform, cluster selection for each result was based on lowest Haddock scores, RMSD values and Z-scores. Table 4.5 summarizes the molecular docking results for all 18 miRNA-Gene pairs found through consensus of the three target prediction algorithms. All miRNA-mRNA pairs had negative Haddock scores indicating potential for miRNA-mRNA interactions except for asi-miR172b-5p-AsUSP Gene 7 and asi-miR444b-AsUSP Gene 7 which had Haddock scores of 7.6 +/- 3.5 and 4.8 +/- 8.0, respectively.

Considering overall docking scores and RMSD values of complexes, the following were among the complexes with the most negative docking scores as well as lowest RMSD values:

4.16.1 Docking complex of asi-miR2118l and AsUSP Gene 1

The docking result of asi-miR2118l with AsUSP Gene 1 indicated a Haddock score of -24.5 +/- 10.2 and RMSD value of 14 +/- 0.8 Å (Table 4.7) which indicated strong binding interactions between the two RNA molecules. PDB file of the complex was analyzed using Discovery studio (Accelrys) and molsoft (<https://www.molsoft.com/pdbv.html#>). Figure 4.22 A shows the complex generated from molecular docking of asi-miR2118l and AsUSP Gene 1. Figure 4.22 B indicates hydrogen bonding of Guanine 506 and Adenine 532 of AsUSP Gene 1 with Cytosine 11 and Uracil 10, respectively of asi-miR2118l. All the intermolecular hydrogen bonding interactions in this complex are outlined in Table 4.6.

4.16.2 Docking complex of asi-miR165a-5p and AsUSP Gene 2

Molecular docking results of asi-miR165a-5p with AsUSP Gene 2 represents a Haddock score of -22.8 +/- 8.9 and RMSD value of 1.3 +/- 1.0 Å (Table 4.7). PDB file was analyzed in the

same way as outlined in Section 4.16.1 and Figure 4.23 A depicts the binding of the two nucleic acid molecules. Selective intermolecular hydrogen bonds are indicated in Figure 4.23 B, which shows Guanine 204 and 205 of AsUSP Gene 2 interacting with Cytosine 14 and 16 respectively of asi-miR165a-5p, while all the intermolecular hydrogen bonds are indicated in Table 4.7.

4.16.3 Docking complex of asi-miR166b and AsUSP Gene 2

Output from the docking of asi-miR166b with AsUSP Gene 2 indicated that the two RNAs bind with each other with Haddock score of -37.2 +/- 11.6 and RMSD value of 1.1 +/- 0.9 Å (Table 4.7). This complex represented the strongest binding interaction, with the most negative Haddock score, among all the complexes, as well as the lowest RMSD value among all the complexes analyzed. Figure 4.24 A represents the complex generated from molecular docking of asi-miR166b with AsUSP Gene 2, while the intermolecular hydrogen bonds between Guanine 204 and Cytosine 132 of AsUSP Gene 2 with Cytosine 17 and Cytosine 15, respectively of asi-miR166b are shown in Figure 4.24 B. All the intermolecular hydrogen bonding interactions with distances are indicated in Table 4.8.

4.16.4 Docking complex of asi-miR482a-3p and AsUSP Gene 7

The result of molecular docking of asi-miR482a-3p with AsUSP Gene 7 indicated the interaction between the two RNA molecules with Haddock score of -31.8 +/- 7.4 and RMSD value of 1.2 +/- 0.7 Å (Table 4.7). This was the second strongest complex among all the complexes based on docking score, as well as having a low RMSD value. Figure 4.25 A depicts the binding of asi-miR482a-3p with AsUSP Gene 7. Figure 4.25 B represents hydrogen bonding interactions between Guanine 436 of AsUSP Gene 7 with Cytosine 11 of asi-miR482a-3p, and interactions of Guanine 434 of AsUSP Gene 7 with Uracil 1 and Cytosine 10 of asi-miR482a-3p. All the intermolecular hydrogen bonds along with bond distances are represented in Table 4.9.

Table 4.5: Molecular docking results of miRNAs and genes from Haddock. miRNA-Gene pairs along with the docking scores, RMSD, van der Waals energies, electrostatic energies, z-scores and other output parameters have been indicated

Sr. #	miRNA-Gene pair	HADDOCK Score	RMSD (Å)	Van der Waals Energy (kcal/mol)	Electrostatic Energy (kcal/mol)	Z-Score	Desolvation Energy	Cluster size	Buried Surface Area
1	asi-miR164a-3p-Gene1	-13.6 +/- 6.9	1.1 +/- 0.7	-48.9 +/- 10.5	5.3 +/- 29.1	-1.8	21.6 +/- 4.2	6	1377.5 +/- 140.7
2	asi-miR2118l-Gene1	-24.5 +/- 10.2	1.4 +/- 0.8	-65.9 +/- 7.9	-7.2 +/- 18.6	-2.3	26.1 +/- 2.1	7	1638.9 +/- 69.0
3	asi-miR165a-5p-Gene2	-22.8 +/- 8.9	1.3 +/- 1.0	-62.2 +/- 9.6	-55.1 +/- 19.7	-1.6	27.7 +/- 4.7	4	1801.4 +/- 206.7
4	asi-miR166b-Gene2	-37.2 +/- 11.6	1.1 +/- 0.9	-78.8 +/- 10.2	-13.8 +/- 21.5	-2.2	24.5 +/- 1.5	12	1782.8 +/- 56.4
5	asi-miR164a-3p-Gene3	-8.9 +/- 1.3	7.2 +/- 0.1	-46.0 +/- 9.2	-8.9 +/- 17.2	-1.8	19.1 +/- 2.2	23	1240.6 +/- 137.1
6	asi-miR172b-5p-Gene3	-5.7 +/- 5.9	8.4 +/- 0.6	-47.0 +/- 2.1	-42.2 +/- 20.2	-1.7	16.9 +/- 0.7	19	1322.9 +/- 88.8
7	asi-miR156c-3p-Gene6	-16.2 +/- 1.9	8.6 +/- 0.1	-61.4 +/- 4.8	-14.4 +/- 5.2	-1.3	31.8 +/- 3.4	14	1833.0 +/- 102.2
8	asi-miR172b-5p-Gene7	7.6 +/- 3.5	7.7 +/- 0.3	-31.7 +/- 8.4	-5.2 +/- 19.1	-1.0	17.8 +/- 2.6	9	1128.3 +/- 174.1
9	asi-miR444b-Gene7	4.8 +/- 8.0	6.8 +/- 0.7	-40.4 +/- 6.5	28.2 +/- 17.7	-0.9	19.1 +/- 2.3	12	1214.0 +/- 157.8
10	asi-miR482a-3p-Gene7	-31.8 +/- 7.4	1.2 +/- 0.7	-76.1 +/- 8.4	-35.9 +/- 31.3	-2.3	32.0 +/- 3.5	13	2082.7 +/- 138.0
11	asi-miR165a-5p-Gene8	-22.9 +/- 3.5	6.9 +/- 0.3	-64.2 +/- 8.1	-30.0 +/- 24.0	-1.3	28.9 +/- 1.2	13	1572.4 +/- 94.8
12	asi-miR399f-Gene8	-14.5 +/- 5.3	8.0 +/- 0.1	-53.1 +/- 4.3	-29.3 +/- 9.4	-1.2	22.9 +/- 2.5	6	1488.5 +/- 85.8
13	asi-miR398e-Gene8	-14.4 +/- 3.0	1.8 +/- 1.1	-59.4 +/- 1.4	-2.3 +/- 8.8	-1.9	29.8 +/- 1.1	10	1787.1 +/- 52.0
14	asi-miR2118l-Gene8	-20.5 +/- 3.2	4.7 +/- 0.2	-48.2 +/- 3.9	-33.1 +/- 8.8	-0.9	17.9 +/- 3.8	12	1331.8 +/- 157.2
15	asi-miR2118b-5p-Gene8	-22.2 +/- 1.0	6.9 +/- 0.3	-54.8 +/- 9.0	-52.8 +/- 33.7	-2.2	29.5 +/- 5.5	26	1540.6 +/- 212.2
16	asi-miR2118b-5p-Gene8	-21.1 +/- 4.9	6.7 +/- 0.6	-56.8 +/- 12.0	-23.5 +/- 38.2	-2.2	21.9 +/- 2.7	20	1374.4 +/- 153.9
17	asi-miR167c-5p-Gene8	-14.2 +/- 4.3	6.6 +/- 0.2	-62.5 +/- 9.7	4.4 +/- 27.6	-1.4	24.1 +/- 3.2	21	1662.3 +/- 143.8
18	asi-miR166b-Gene8	-18.9 +/- 9.5	1.7 +/- 1.2	-53.3 +/- 12.6	-16.3 +/- 11.6	-1.2	25.3 +/- 2.4	5	1455.8 +/- 161.2

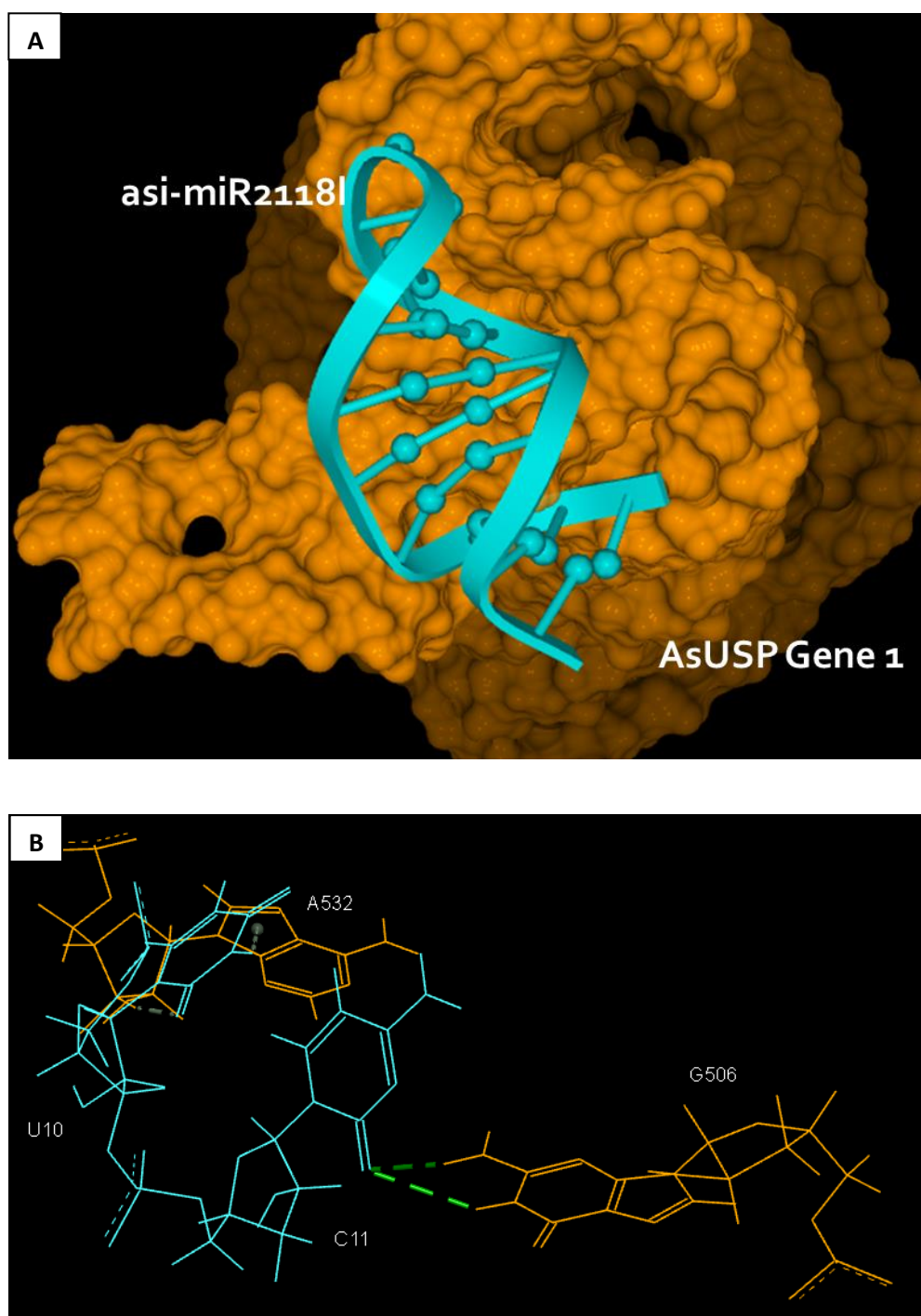


Figure 4.22: A) Complex generated from molecular docking of asi-miR2118l and AsUSP Gene 1; B) View of selective hydrogen bonding interactions between nucleotides of the complex

Table 4.6: Intermolecular hydrogen bonding interactions between asi-miR2118l and AsUSP Gene 1 along with bond distances (Å)

Hydrogen bonds; AsUSP Gene 1 (A), asi-miR2118l (B)	Distance (Å)
A:G506:H22 - B:C11:O2	2.40
A:G506:H1 - B:C11:O2	2.87
A:G540:H21 - B:A4:O1P	2.36
A:G540:HO2' - B:U3:O3'	2.54
A:U549:HO2' - B:A5:O1P	2.55
A:U551:HO2' - B:C2:O1P	1.77
A:C552:HO2' - B:C2:O2P	1.75
A:A553:H61 - B:C13:O2P	2.68
A:G554:H1 - B:C13:O2P	2.34
B:U10:H3 - A:A533:N7	2.82
B:U10:HO2' - A:A533:O2P	1.76
B:C11:HO2' - A:G534:O2P	2.13
B:C14:H41 - A:U551:O1P	2.55
A:G506:H2' - B:C13:O2'	2.41
A:A532:H3' - B:U10:O2	2.12
A:A533:H8 - B:U10:O2	2.02
A:G534:H3' - B:C12:O2P	2.06
A:G539:H8 - B:U3:O1P	1.68
A:C541:H1' - B:A4:O1P	1.94
A:U551:H1' - B:U3:O2P	2.29
A:C552:H4' - B:C2:O1P	2.45
A:C552:H1' - B:C2:O2P	2.89
B:C2:H5'' - A:U551:O2	2.61
B:U3:H5' - A:G540:N3	2.35
B:U3:H4' - A:G540:O2'	1.93
B:U10:H1' - A:A533:O2P	2.51
B:C11:H6 - A:A533:N7	2.41
B:C11:H2' - A:G534:N7	2.40
B:C12:H4' - A:G506:O6	2.00
B:C13:H4' - A:G506:O2P	2.13
B:U10:H3 - A:A532	2.79

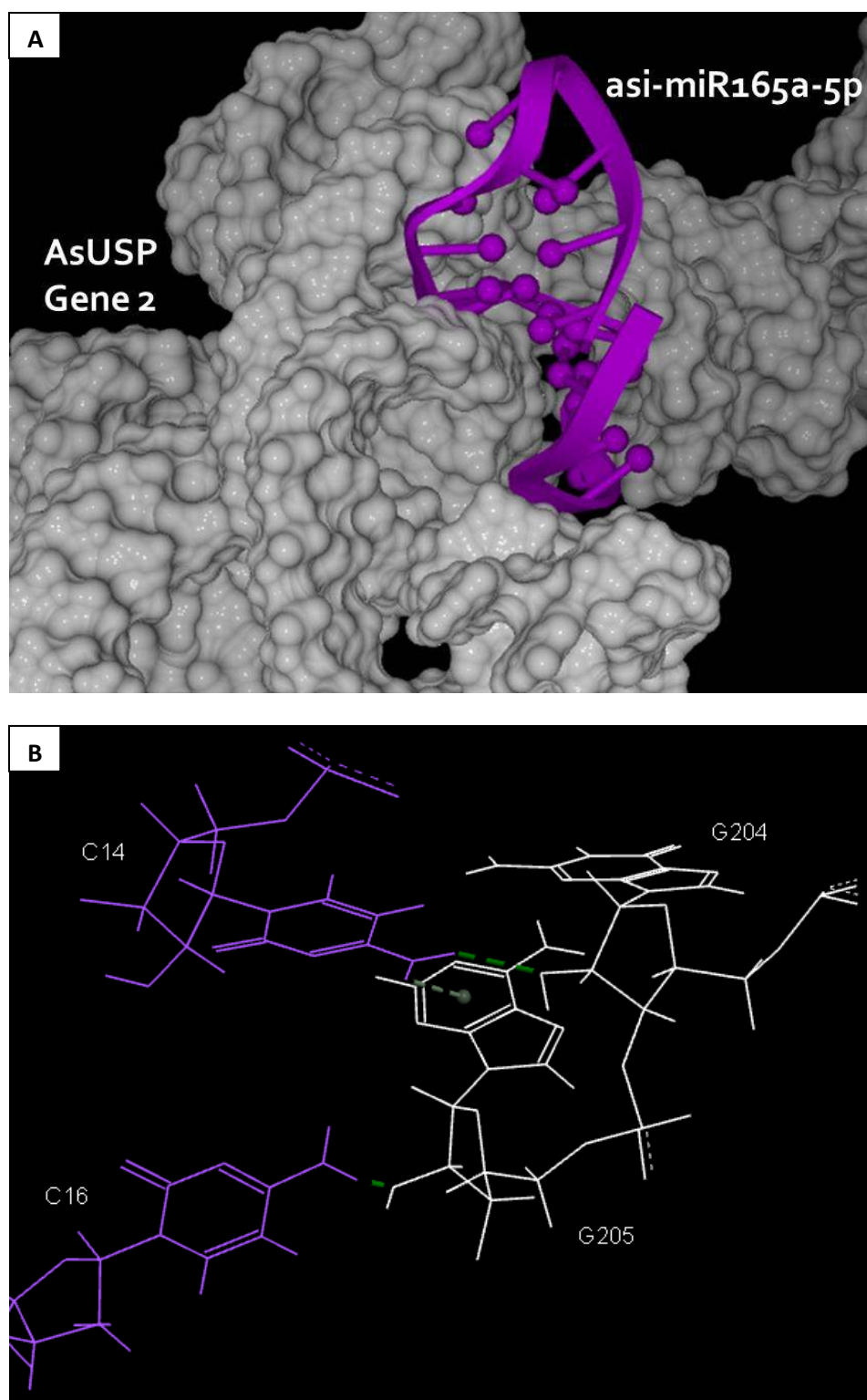


Figure 4.23: A) Complex generated from molecular docking of asi-miR165a-5p and AsUSP Gene 2; B) View of selective hydrogen bonding interactions between nucleotides of the complex

Table 4.7: Intermolecular hydrogen bonding interactions between asi-miR165a-5p and AsUSP Gene 2 along with bond distances (Å)

Hydrogen bonds; AsUSP Gene 2 (A), asi-miR165a-5p (B)	Distance (Å)
A:A128:HO2' - B:U4:O2P	1.99
A:A131:HO2' - B:G20:O6	1.81
A:G204:H21 - B:G13:O1P	2.53
A:G204:H22 - B:G12:O1P	2.01
A:G204:H1 - B:G12:O1P	1.99
A:A205:H62 - B:G12:O1P	2.18
A:G207:H22 - B:C10:O1P	2.72
A:G242:HO2' - B:U9:O4'	2.63
B:C3:HO2' - A:C129:O4'	2.59
B:U6:HO2' - A:G244:O1P	2.09
B:U7:H3 - A:C119:O2P	2.89
B:G8:HO2' - A:C243:O1P	2.75
B:C14:H41 - A:G204:O2'	2.93
B:C16:H41 - A:A205:O2'	1.84
B:G20:H22 - A:C132:O1P	1.84
B:G20:H1 - A:A131:O3'	1.91
A:G118:H5'' - B:U7:O4	2.35
A:C129:H5' - B:C3:O3'	2.85
A:U130:H4' - B:A2:O2'	2.76
A:A131:H4' - B:G20:O6	1.91
A:A131:H1' - B:G12:O4'	2.39
A:C132:H1' - B:G12:O2'	2.29
A:G204:H1' - B:G13:O2P	2.23
A:U206:H1' - B:G17:O6	2.68
A:A208:H2 - B:C10:O1P	1.86
A:C243:H5' - B:G8:O2'	2.80
B:A2:H1' - A:A131:O1P	2.75
B:U9:H4' - A:G242:O3'	2.42
B:U11:H5' - A:U130:O2	2.72
B:U11:H5'' - A:U130:O2	2.50
B:U11:H4' - A:U130:O2'	2.67
B:G12:H1' - A:A131:O2'	2.09
B:C14:H42 - A:A205	3.27

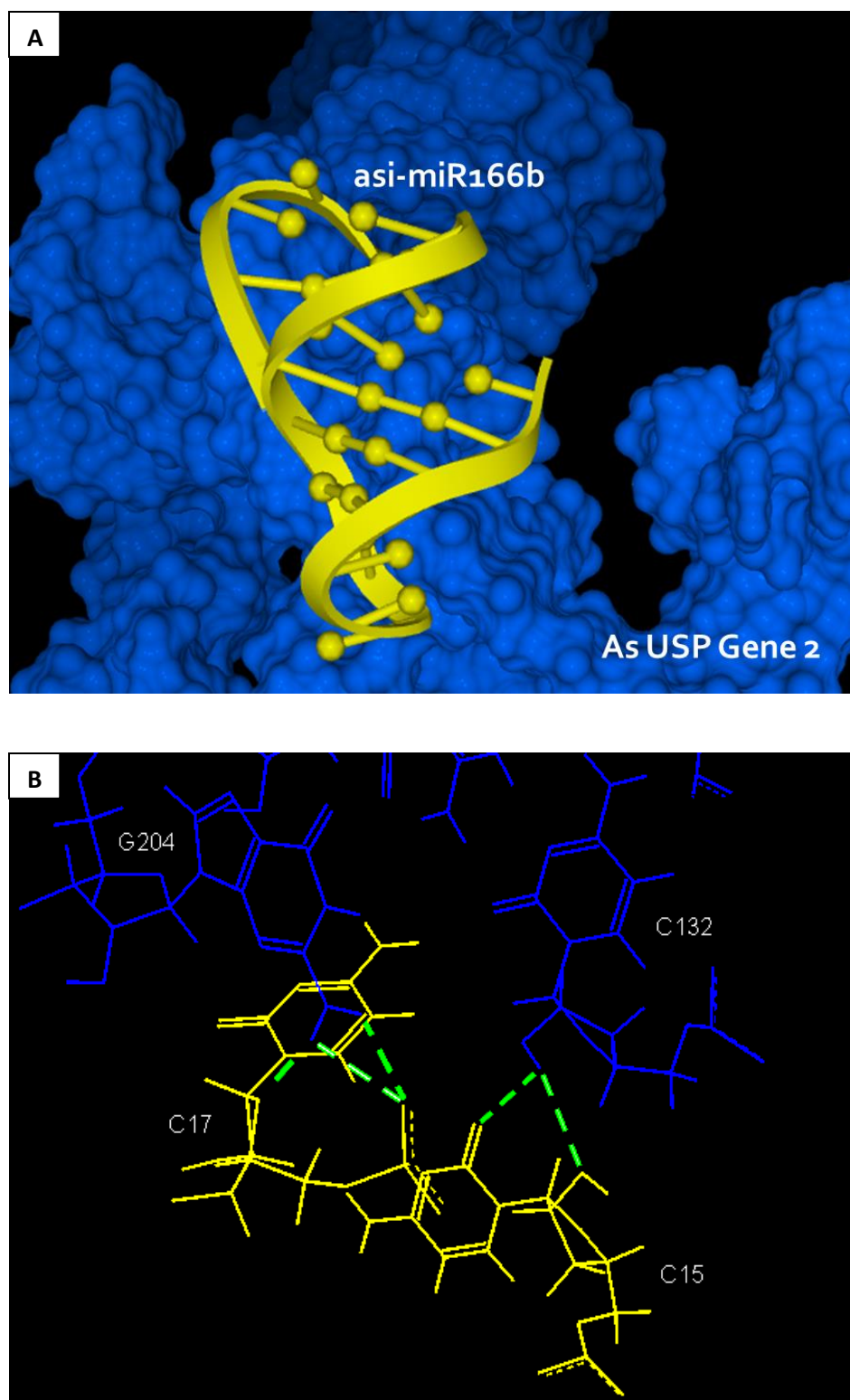


Figure 4.24: A) Complex generated from molecular docking of asi-miR166b and AsUSP Gene 2;
B) View of selective hydrogen bonding interactions between nucleotides of the complex

Table 4.8: Intermolecular hydrogen bonding interactions between asi-miR166b and AsUSP Gene 2 along with bond distances (Å)

Hydrogen bonds; AsUSP Gene 2 (A), asi-miR166b (B)	Distance (Å)
A:C132:HO2' - B:C15:O2	2.77
A:C132:HO2' - B:C15:O2'	2.95
A:G133:H22 - B:G9:O2P	2.37
A:A134:HO2' - B:G9:N7	2.84
A:A134:HO2' - B:U10:O4	2.24
A:U135:HO2' - B:G6:O3'	2.40
A:U201:HO2' - B:G9:O2P	1.66
A:G202:HO2' - B:G21:O2P	2.81
A:U203:HO2' - B:G20:O3'	3.07
A:G204:H21 - B:C17:O2P	2.72
A:G204:H21 - B:C17:O4'	1.94
A:G204:H22 - B:C17:O2P	2.49
B:A3:H61 - A:U135:O1P	2.11
B:U8:H3 - A:U135:O4'	1.81
B:C11:H41 - A:A134:O1P	2.62
B:U12:H3 - A:A134:O1P	2.26
B:G13:H22 - A:A134:O2P	1.81
B:G13:H1 - A:A134:O2P	1.93
B:C17:H41 - A:C132:O2	2.10
B:G21:HO2' - A:G133:O2'	2.86
A:A134:H5' - B:G2:O6	2.84
A:A134:H2 - B:G9:O2P	2.65
A:U135:H5' - B:G9:O6	2.00
A:U135:H4' - B:U8:O4	2.62
A:C136:H5'' - B:G6:O2'	2.79
A:C136:H1' - B:U7:O1P	2.98
A:U201:H1' - B:G9:O2P	2.68
A:U203:H5' - B:G21:O2P	3.01
A:U203:H4' - B:G21:O2P	2.17
A:G204:H1' - B:C17:O2	2.37
A:A205:H1' - B:G18:O2P	2.07
B:U8:H5' - A:U200:O2'	2.47
B:U8:H5'' - A:U200:O2	2.87
B:G9:H5'' - A:G202:O4'	2.81
B:G18:H5' - A:A205:O2'	2.38
B:G18:H4' - A:A205:O4'	2.18
B:G21:H2' - A:G133:O2'	3.05

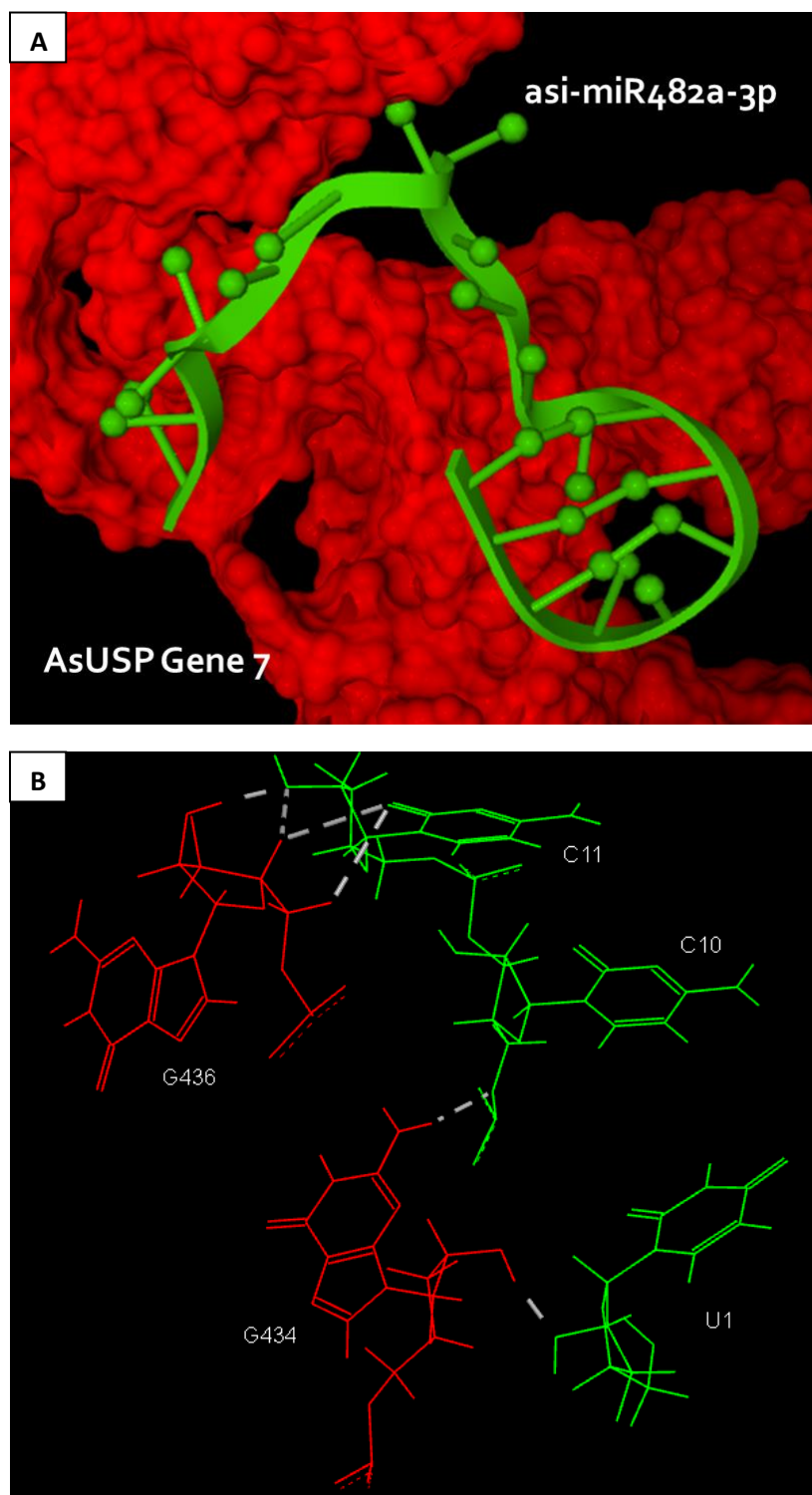


Figure 4.25: A) Complex generated from molecular docking of asi-miR482a-3p and AsUSP Gene 7; B) View of selective hydrogen bonding interactions between nucleotides of the complex

Table 4.9: Intermolecular hydrogen bonding interactions between asi-miR482a-3p and AsUSP Gene 7 along with bond distances (Å)

Hydrogen bonds; AsUSP Gene 7 (A), asi-miR482a-3p (B)	Distance (Å)
A:G434:H21 - B:C10:O5'	2.16
A:G434:HO2' - B:U1:O2'	1.81
A:G436:HO2' - B:C11:O2'	1.83
A:G439:H21 - B:A18:O1P	2.48
A:G439:H1 - B:C19:O2P	2.53
A:G439:HO2' - B:U17:O3'	2.81
A:G439:HO2' - B:A18:N7	2.04
A:G466:HO2' - B:C14:O2	2.37
A:U494:HO2' - B:C15:O2P	2.67
A:C507:H41 - B:C20:O2	3.02
A:A508:H61 - B:C20:O2	2.61
A:C513:HO2' - B:C10:O2P	2.85
B:U1:HO2' - A:A433:O2'	1.98
B:C4:H41 - A:A431:O1P	1.86
B:C14:H42 - A:C467:O4'	2.39
B:U17:H3 - A:U440:O4	2.03
B:U17:HO2' - A:U440:O1P	2.67
B:C20:H42 - A:U441:O2	2.79
B:C20:HO2' - A:A508:O2P	1.90
B:C20:H3T - A:A508:O2P	1.97
A:A430:H5' - B:C4:O1P	3.05
A:A432:H2 - B:U2:O2'	1.80
A:G436:H5' - B:C11:O2	2.86
A:G436:H4' - B:C11:O2	2.91
A:G436:H4' - B:C11:O2'	2.53
A:G439:H2' - B:A18:O2P	2.44
A:G466:H4' - B:C14:O2	2.59
A:U494:H4' - B:C15:O1P	2.36
A:A508:H8 - B:C20:O2'	2.40
A:A514:H5' - B:A9:O2P	2.56
A:A514:H4' - B:A9:O2P	2.50
B:U2:H4' - A:A433:O4'	2.24
B:A9:H5' - A:A514:O3'	2.73
B:C15:H5' - A:A495:O1P	2.33
B:C15:H6 - A:A495:O1P	2.81
B:C15:H3' - A:A495:O1P	2.44
B:U17:H2' - A:U440:O1P	2.53
B:C20:H4' - A:A508:O2P	2.70
B:C20:H2' - A:A508:N7	2.48
A:A432:H62 - B:A9	2.64
B:C19:H41 - A:G439	2.50

Hybridization potential of miRNAs evaluated in Section 4.15.5 were considered in combination with overall docking scores (and RMSD values) and based on these interactions, it was construed that the miRNA-gene pairs asi-miR2118l-AsUSP Gene 1 (mfe ratio 0.72, Haddock score -24.5 +/- 10.2), asi-miR166b-AsUSP Gene 2 (mfe ratio 0.63, Haddock score -37.2 +/- 11.6) and asi-miR482a-3p-AsUSP Gene 7 (mfe ratio 0.61, Haddock score -31.8 +/- 7.4) could hybridize with a greater likelihood and strength of interactions (also lower RMSD values for these interactions i.e. 1.4 +/- 0.8 Å, 1.1 +/- 0.9 Å and 1.2 +/- 0.7 Å respectively showed a better reliability of docking results for these complexes).

4.17 MSA and Phylogenetic Analysis of asi-miRNAs

Results of multiple sequence alignments and phylogenetic analyses are presented in Figures 4.26-4.28. From the phylogenetic analysis of asi-miR2118l (Figure 4.26), it can be observed that it is distantly related to all homologues i.e. pre-miRNAs of 2118 family from *Oryza sativa*, *Zea mays* and *Arabidopsis thaliana*; however, it is more closely related to osa-MIR2118m and osa-MIR2118j. Phylogenetic analysis of miR-166 family (Figure 4.27) indicated that asi-pre-miRNA166b lies in the same clade with gma-MIR166l (from *Glycine max*) and more distantly related to members pre-miRNAs of *Camelina sativa*, *Arabidopsis lyrata*, *Oryza sativa*, *Zea mays* and *Arabidopsis thaliana*. From the phylogenetic relationship of miRNA family 482 (Figure 4.28), pre-miRNA of asi-miR482a-3p is more closely linked to pab-MIR482k (from *Picea abies*) and exists in the same clade; other miR-482 members which are distantly related included pre-miRNAs from *Glycine max*, *Glycine soja*, *Manihot esculenta*, *Picea abies*, *Aquilegia caerulea*, *Prunus persica* and *Nicotiana tabacum*.

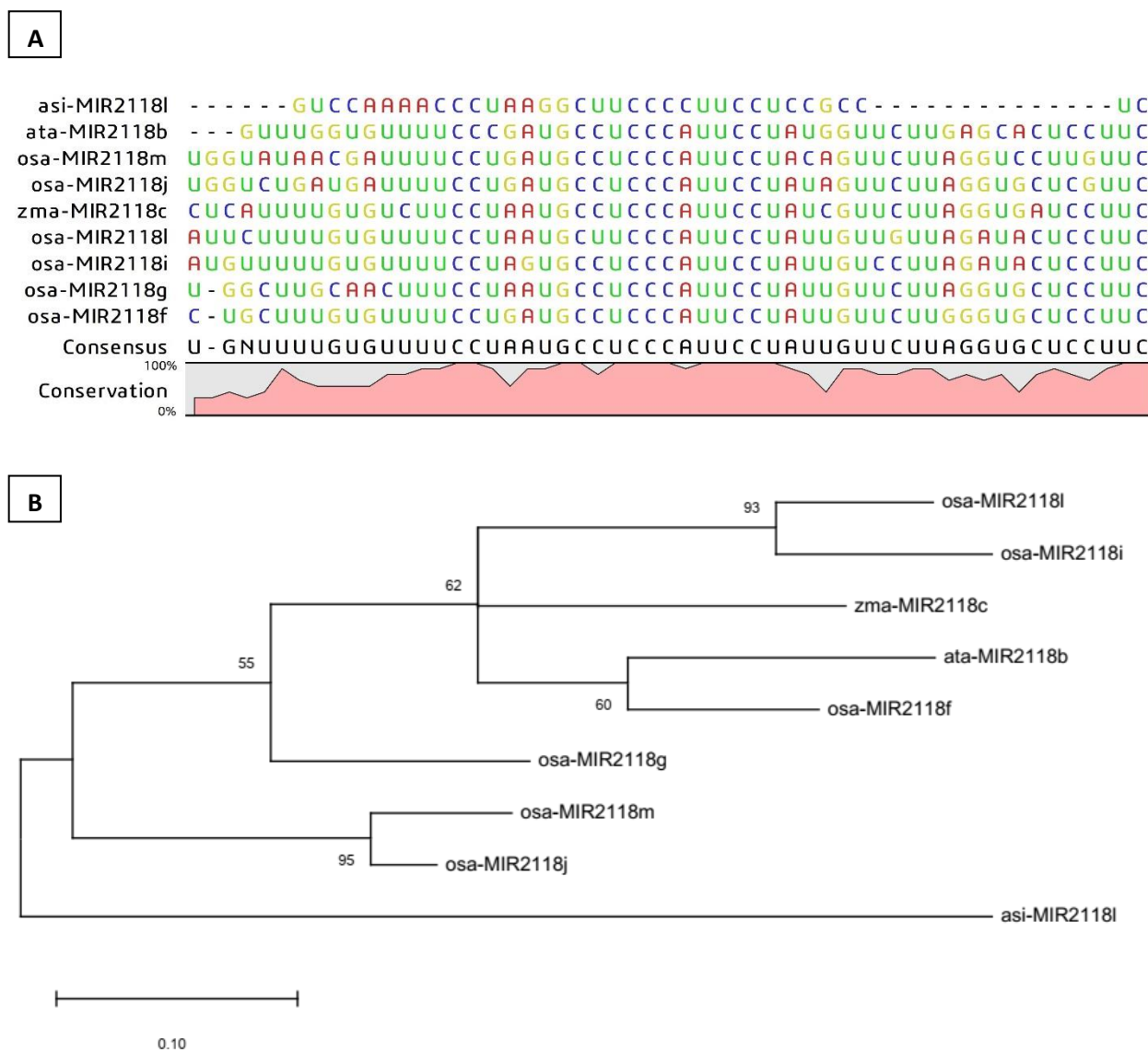


Figure 4.26: Multiple sequence alignment (A) and phylogenetic relationship (B) between asi-miR2118l and homologues in miR-2118 family

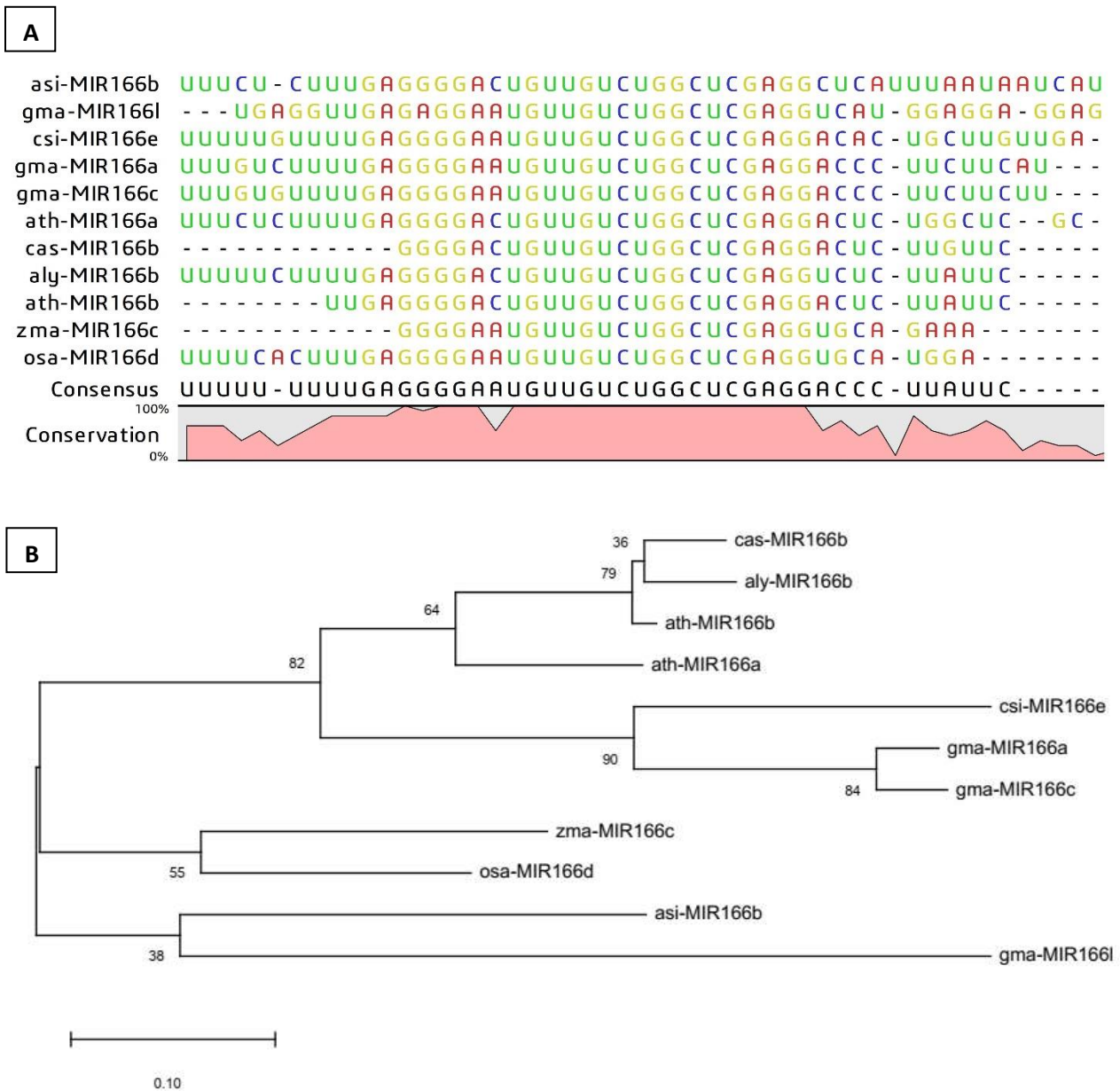


Figure 4.27: Multiple sequence alignment (A) and phylogenetic relationship (B) between asi-miR166b and homologues in miR-166 family

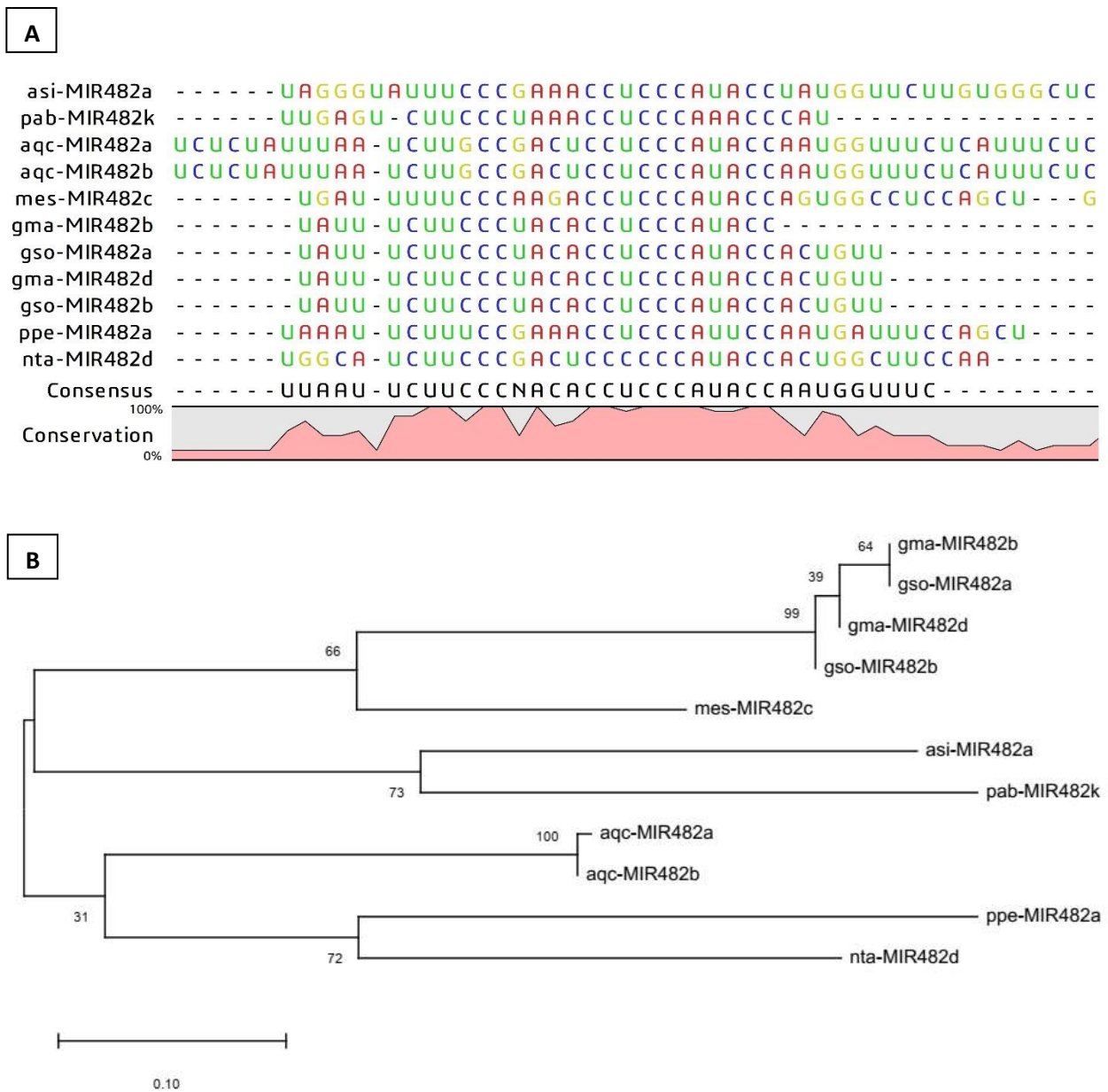


Figure 4.28: Multiple sequence alignment (A) and phylogenetic relationship (B) between asi-miR482a-3p and homologues in miR-482 family

4.18 Microscopic Analysis of *Aspergillus niger* Culture

From the culture of *A. niger* grown on PDA slants, as explained in Section 3.37.1, a small amount of fungal growth was picked using loop and mixed into a drop of sterile water placed on glass slide, followed by microscopic examination under optical microscope (100x). Results of microscopic analysis revealed various morphological features of *A. niger* as presented in the Figure 4.29. Conidiophores were observed as shown in Figure 4.29 A and various conidia could be seen as in Figure 4.29 B. Shape of conidia was globose (spherical) to subglobose (imperfectly spherical) while diameter was found within a range of 3.5-5.0 μm .

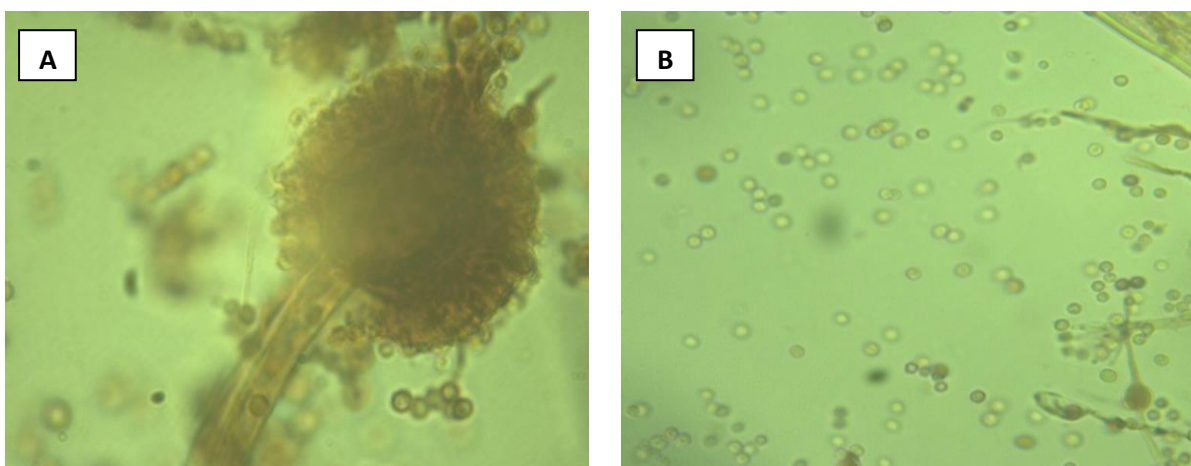


Figure 4.29: Microscopic examination of *Aspergillus niger* growth as observed under optical microscope (with magnification of 100x). Conidiophore (A) and various conidia (B) could be observed under microscope

4.19 *Agave sisalana* Growth under Control and Biotic Stress

A. sisalana plants grown to seven months of age were inoculated with the suspension of *A. niger* culture as described in Section 3.37.3. Control plants were kept parallel (without *A. niger* inoculation). After 14 days, physical signs of bole rot disease were noted in infected plants as

shown in Figure 4.30; rotting in the bole region of plants was clearly observed indicating the damaging effect of *A. niger* inoculation.

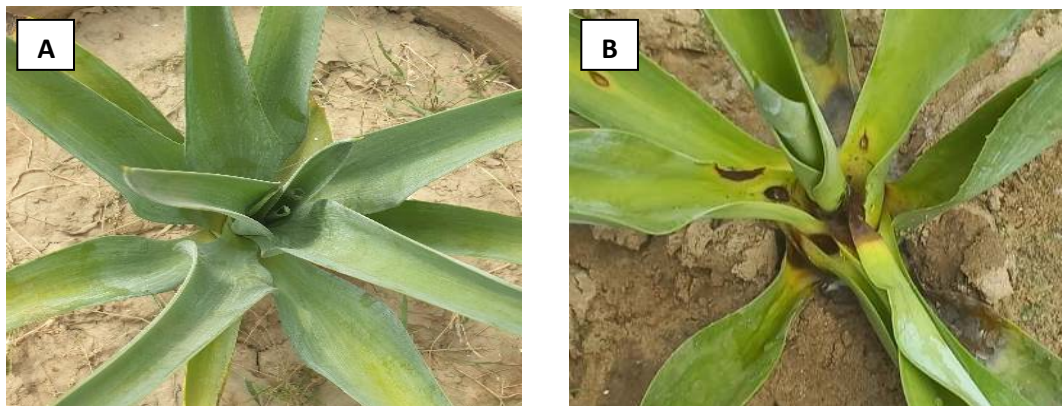


Figure 4.30: (A) *A. sisalana* Control and (B) infected with *A. niger* suspension after 14 days of infection

4.20 Estimation of Chlorophyll Content

Mean Chlorophyll-a content in control was 0.065 mg/g FW while in infected (biotic stress condition) was observed to be 0.022 mg/g FW. However, the difference (i.e. the reduction in Chlorophyll-a content between control and stress) was not found to be statistically significant ($P > 0.05$) (Figure 4.31 A). Mean Chlorophyll-b content in control was calculated to be 1.69 mg/g FW while in infected (biotic stress) was found to be 0.74 mg/g FW and the difference was found to be statistically significant ($P < 0.01$) (Figure 4.31 B). Total Chlorophyll content in control was 1.75 mg/g FW while in infected plant was 0.77 mg/g FW and the difference was statistically significant ($P < 0.01$) (Figure 4.31 C) (Table 4.10).

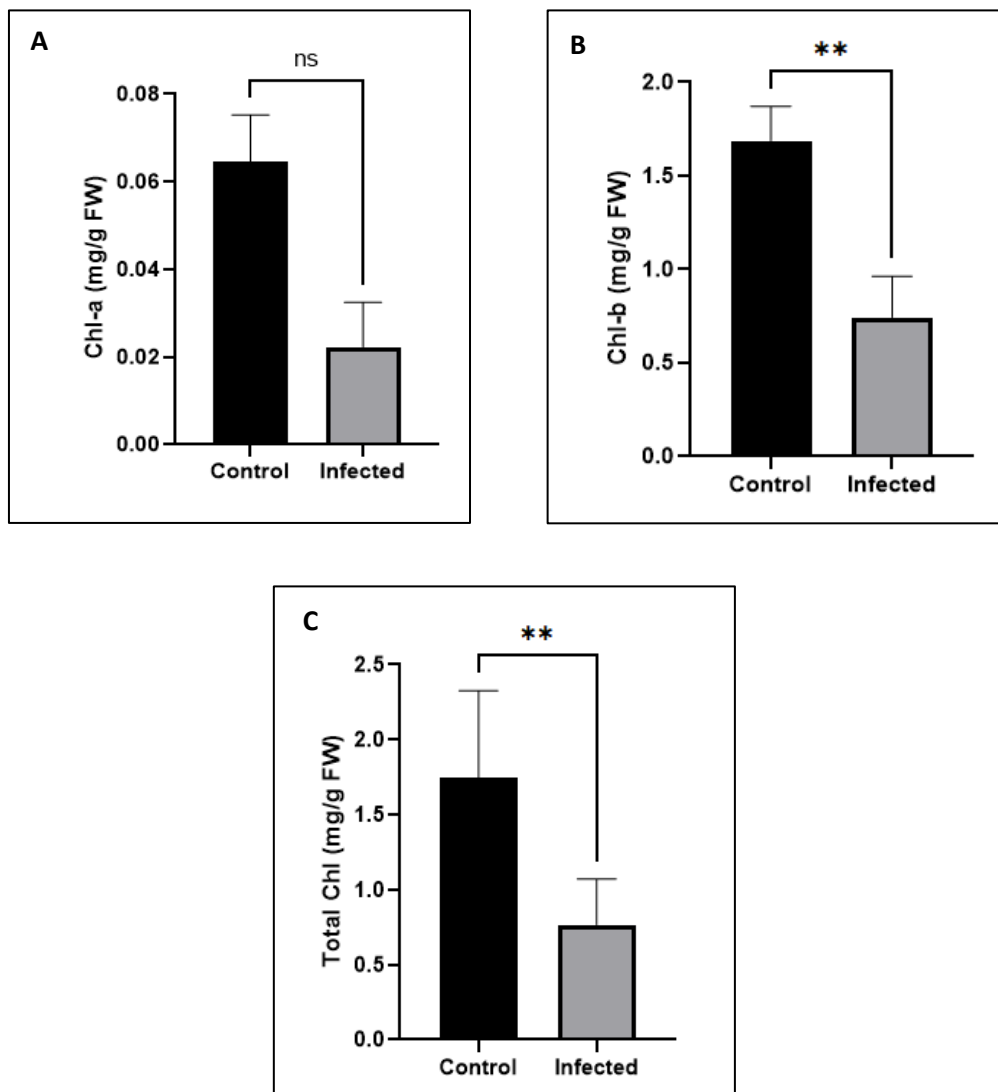


Figure 4.31: (A) Statistical comparison of Chlorophyll-a, (B) Chlorophyll-b and (C) total Chlorophyll contents in *A. sisalana* control and infected

Table 4.10: Analysis of variance (ANOVA) and mean performance analysis for Chlorophyll-a, Chlorophyll-b and total Chlorophyll contents in control and biotic stressed (infected) *A. sisalana*

Alpha	0.05			
Bonferroni's multiple comparisons test	Mean Difference	95.00% CI of diff.	Significant?	Adj. P-value
Control vs Infected				
Chlorophyll-a	0.04233	-0.5014 to 0.5861	ns	>0.9999
Chlorophyll-b	0.9433	0.3996 to 1.487	***	0.0006
Total Chlorophyll	0.9827	0.4389 to 1.526	***	0.0004

4.21 Estimation of Proline Content

Mean Proline Content in control was found to be 1.80 $\mu\text{g/g}$ FW while in infected (biotic stress) was 1.58 $\mu\text{g/g}$ FW. Although a difference was noted in the proline contents of control and stressed plants, the difference was not found to be statistically significant ($P>0.05$) (Figure 4.32) (Table 4.11).

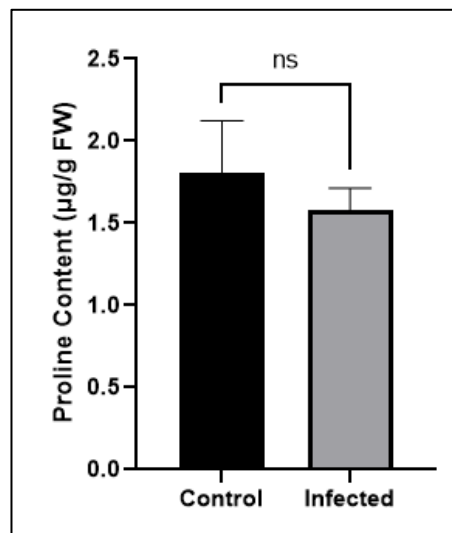


Figure 4.32: Statistical comparison of proline contents in *A. sisalana* control and infected

Table 4.11: Analysis of variance (ANOVA) and mean performance analysis for total proline content in control and biotic stressed (infected) *A. sisalana*

Alpha	0.05			
Bonferroni's multiple comparisons test	Mean Difference	95.00% CI of diff.	Significant?	Adj. P-value
Control vs Infected				
Proline Content	0.22	-0.2141 to 0.6541	ns	0.3031

4.22 Microscopic Examination of Transverse Leaf Sections

Leaf tissues of *A. sisalana* were excised from control and infected plants after two weeks of inoculation; transverse sections were prepared as described in Section 3.40 stained with H&E stain and observed under optical microscope (Olympus BX61). Images are shown in Figure 4.33 and 4.34 at 10x and 20x magnification, respectively. Section of infected leaf (Figure 4.33 B and 4.34B) indicated various infected areas (with fungal particles/aggregates) among cells compared to control including vascular bundles (Figure 4.33 A and 4.34A).

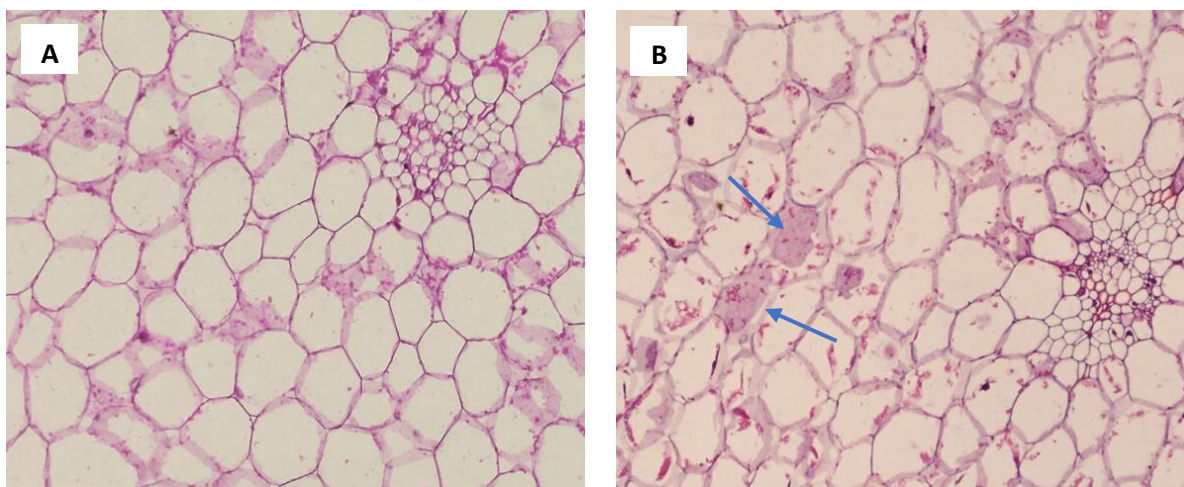


Figure 4.33: Microscopic view of cross section produced from microtomy of *A. sisalana* control (A) and infected (B) leaf samples at 10x magnification (infected regions marked with blue arrows)

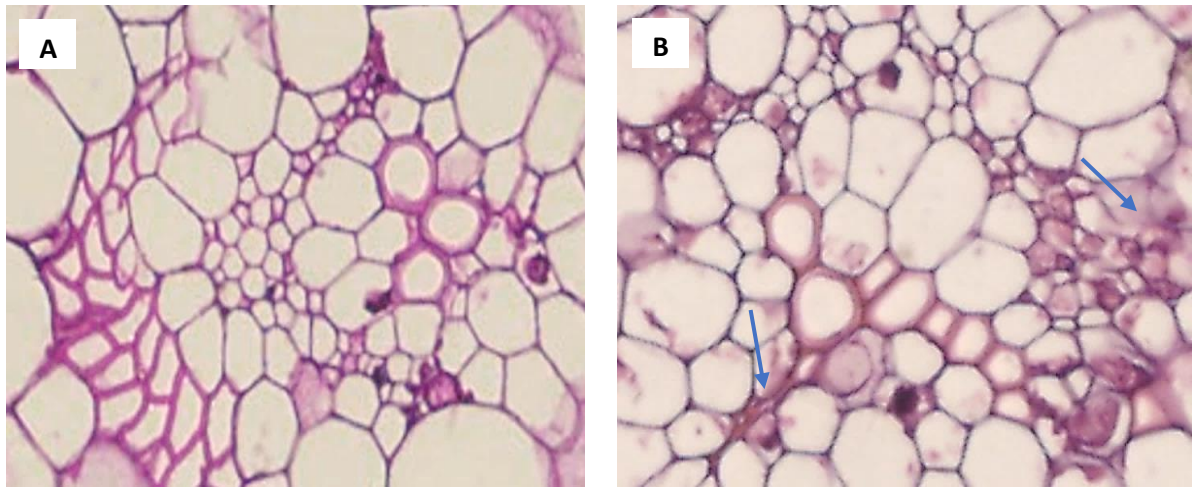


Figure 4.34: Microscopic view of leaf cross section produced from microtomy of *A. sisalana* control (A) and infected (B) leaf samples at 20x magnification (infected regions marked with blue arrows)

5. DISCUSSION

The current study was ambitious to explore the array of miRNAs that could be identified with the use of computational tools and validating the predicted miRNAs, that can potentially be involved in biotic stress response in *A. sisalana* (i.e. under infection by *A. niger*). Moreover, members of the target gene family of Universal Stress Protein (USP) were also predicted and studied through various tools for their potential regulation by the identified miRNAs.

Agave species exhibit exceptional tolerance to environmental stresses due to CAM photosynthesis, wax deposition, and protective leaf morphology. Such characteristics makes this species an excellent candidate for the exploration of genomic information governing mechanisms underlying such extra-ordinary features. Research in the area of miRNAome and gene silencing mechanisms in *A. sisalana* is needed for getting insights into the stress tolerance mechanisms. Excellent features of multi-stress tolerance in *Agave* sp. makes it an ideal candidate to study the mechanisms underlying stress tolerance. Bole rot disease has been reported in *A. sisalana* and is caused by *Aspergillus* sp. which act as a biotic stress factor limiting growth and production of *A. sisalana*. Unavailability of complete genomic information of *A. sisalana* remains one of the reasons for lack of knowledge regarding the stress tolerance mechanisms.

This study attempted to harness the available data on the transcriptome (RNA-seq data) of *A. sisalana* (Sarwar et al., 2019) to fish out the possible miRNA candidates hidden within the non-coding repertoire of its transcriptome. A total of 74 candidate miRNAs were predicted on the basis of homology and secondary structure analysis (Table 4.1). Similar pipelines for prediction of miRNAs from RNA-seq data have previously been implemented by Kalariya et al. (2019) on *Gymnema sylvestre*, Gadhavi et al. (2020) on *Bacopa monnieri*, Gupta et al. (2015) on *Helianthus annuus* and Jabbar et al. (2021) on *Agave deserti*. To filter out pseudo pre-miRNAs from all

potential pre-miRNAs, their properties were evaluated which included %G+C content and %A+U content, secondary structures, and minimum free energies. MFEI values of the selected pre-miRNAs of *A. sisalana* ranged from 0.32-0.95 while data previously reported on *A. deserti* has shown pre-miRNAs MFEI ranging from 0.5 to 1.0 (Jabbar et al., 2021). While the prediction of miRNA from the sequenced transcriptome such as ESTs, conitgs and unigenes using homology based and comparative can reveal conserved miRNA sequences from various plant species (Najafabadi & Naghavi, 2018; Singh & Sharma, 2017), the process of small RNA sequencing (sRNA-seq) can discover a vast majority of miRNAs, including both conserved and novel sequences, through detection of more than a billion RNA molecules in a single run (Kang & Friedländer, 2015). miRNA prediction pipeline implemented in this study identified 74 miRNAs out of which 48 were confirmed through Poly(A)RT PCR; small RNA sequencing effort for *A. sisalana* could reveal much more information and novel miRNA sequences in addition to the conserved sequences. The data analysis of small-RNA sequencing need a reference genome, as required by most of the tools that predict miRNA from sRNA-Seq data, to allow for excision of miRNA hairpins; however, high quality reference genomes are available for only a few of the model species. For the species without reference genomes, new methods are required for discovery of novel miRNAs for getting insights into their function. One solution to overcome this limitation is to implement genome of a closely related species also referred to as proxy reference sequence (Kang & Friedländer, 2015). However, some tools are now available that allow miRNA prediction from sRNA-Seq with or without the reference genome (Vitsios et al., 2017).

Agarose gel electrophoresis run for large RNA fraction (Figure 4.5) showed intact 28s and 18s rRNA bands indicating good quality of extracted RNA without degradation. A similar banding pattern for plant RNA has been reported by Chang et al. (2016) and Ghawana et al. (2011). The

ratios of A260/A280 and A260/A230 were >2 for small and large RNA fractions (Table 4.2), indicating that there was no polysaccharide, polyphenol, EDTA or protein contamination in the purified RNA samples. For validation of targets, poly(T) adaptor sequence as reported by Xie et al. (2007) was designed for cDNA preparation after the poly(A) tailing reaction (Mohanpuria et al., 2018). Figure 4.6 A-F showed Poly(A) RT PCR results; these results indicated that PCR with miRNA specific forward primers for 74 predicted *A. sisalana* miRNAs showed a band of approx. 100 bp for 48 *A. sisalana* miRNAs. Similar types of bands upon miRNA amplification using PCR have been shown by D. Chen et al. (2019) for *Apis cerena* miRNAs and F. Gao et al. (2021) for *Galdieria sulphuraria* miRNAs.

USP genes from *A. sisalana* transcriptome (RNA-seq) assembly were first identified using bioinformatic approaches which resulted in 9 full length open reading frames belonging to USP family-PF00582. Shokry et al. (2014) used de novo assembled genome contigs of the high-throughput sequencing dataset of *Calotropis procera* and used protein-protein BLAST and multiple sequence alignment tools for prediction of USP gene using bioinformatic approaches. In this study, HMM based approach that identified USPs using protein profile hidden Markov models (profile-HMMs) which are sensitive tools for remote protein homology detection (Johnson et al., 2010). Identified USP genes (i.e. AsUSP genes 1-9) were amplified after designing primers for full length gene amplification; amplified fragments (Figures 4.10- 4.12) were further cloned, sequenced and submitted to NCBI, Genbank (Figure 4.15 A-I).

A consensus based approach was utilized for prediction of miRNA targets among AsUSP genes for prediction of reliable miRNA-target gene interactions (Iqbal et al., 2017; Jabbar et al., 2019). Results of mRNA targeting using RNAhybrid showed a total of 266 targets of *A. sisalana* miRNAs among AsUSP genes (Figure 4.16). Precise mode of Tapirhybrid (Bonnet et al., 2010)

uses RNAhybrid algorithm for miRNA target prediction and takes into consideration various parameters such as seed pairing, target site accessibility, multiple target sites, and it predicts target sites with a higher accuracy than many other tools (Srivastava et al., 2014). By using psRNATarget, (Dai et al., 2011), 138 target sites were predicted (Figure 4.17); through integration of two important analysis functions (i.e. complementarity scoring and analysis of secondary structure), this tool searches for miRNA targets in the provided nucleotide sequence of mRNA. It also finds target-site accessibility by computation of unpaired energy (UPE) needed to remove folds from secondary structure around miRNA target site on mRNA (Brodersen et al., 2008). A total of 139 miRNA targets were predicted in the sequences of AsUSP genes using psRobot target prediction algorithm (Figure 4.18). psRobot implements a modified Smith–Waterman algorithm with the defined system for scoring for calculating alignment scores between small RNAs and target sequences. It estimates the penalty score of each alignment by subtraction of actual alignment score from the ideal perfect global pairing score; the alignments corresponding to penalty score cut-off are then output as a result. It intensively searches each mRNA sequence for detection of multiple target sites (Wu et al., 2012). Merging the results of the target prediction algorithms indicated that a total of 406 non-redundant target positions were prediction in nine AsUSP genes; and 18 common target sites were found from consensus of the three algorithms (Figure 4.19-4.20; Table 4.4). Plot of miRNAs against mfe ratio is represented in Figure 4.21; the mfe ratio is proportionate with the strength of miRNA-target hybridization, and higher values (i.e. closer to 1.0) may represent a greater probability of miRNA leading to the predicted effect i.e. translation inhibition or cleavage (Thody et al., 2020) (Table 4.4). While through earlier studies, it was construed that plant miRNAs function by cleavage of the target genes; however, studies on regulation of AP2 protein expression by miR172 of Arabidopsis without fluctuations in mRNA

expression led to the understanding that plant miRNAs can also bring about silencing of the target genes through repression of translation and is an extensive mechanism of plant miRNA based gene silencing (Ariizumi et al., 2011; Aukerman & Sakai, 2003).

Based on earlier studies, it was believed that plant miRNAs function by cleaving target genes. However, based on a study revealing regulation of AP2 protein expression by *Arabidopsis* miR172 without changes in mRNA expression, it was later proposed that plant miRNAs can also silence their target genes through translational repression (Aukerman and Sakai, 2003).

To explore binding interactions of miRNAs with respective target gene (mRNA) in 3D, RNA structures were modelled and docking analysis was performed as described in Section 3.33-3.35. From analysis of Haddock output (Table 4.5), the complexes asi-miR2118l-AsUSP Gene 1 (Haddock score -24.5 +/- 10.2, RMSD 1.4 +/- 0.8 Å), asi-miR165a-5p-AsUSP Gene 2 (Haddock score -22.8 +/- 8.9, RMSD 1.3 +/- 1.0 Å), asi-miR166b-AsUSP Gene 2 (Haddock score -37.2 +/- 11.6, RMSD 1.1 +/- 0.9 Å), and asi-miR482a-3p-AsUSP Gene 7 (Haddock score -31.8 +/- 7.4, RMSD 1.2 +/- 0.7 Å) were selected on the basis of lower (more negative) Haddock scores and low RMSD values. After considering both the results of mfe ratio and molecular docking analysis (Table 4.5), the complexes asi-miR2118l-AsUSP Gene 1 (mfe ratio 0.72, Haddock score -24.5 +/- 10.2) (Figure 4.22), asi-miR166b-AsUSP Gene 2 (mfe ratio 0.63, Haddock score -37.2 +/- 11.6) (Figure 4.24) and asi-miR482a-3p-AsUSP Gene 7 (mfe ratio 0.61, Haddock score -31.8 +/- 7.4) (Figure 4.25) represent more stable binding interactions and are more likely to cause the predicted effect of bringing about target mRNA cleavage or translational inhibition. Phylogenetic analysis revealed evolutionary relationship between these miRNAs and their respective homologues (Figures 4.26-4.28); similar types of phylogenetic relationships/trees have been built up for miRNAs of *Agave deserti* and their homologues by Jabbar et al. (2021).

Microscopic analysis of culture picked from *A. niger* PDA slants (Figure 4.29) indicated various morphological features which were concordant with the observations as reported by (Andge et al., 2017). After 14 days of *A. niger* inoculation, a reddish-dark brown rotting in the bole region was observed in *A. sisalana* similar to that reported by Santos et al. (2014). Yellowing of leaves near the bole region of *A. sisalana* plants infected with *A. niger* was already somewhat evident as can be seen in Figure 4.30 which hinted towards a decrease in Chlorophyll content. A statistically significant decrease was observed in Chlorophyll-b content (0.74 mg/g FW in infected plants) and total Chlorophyll content (0.77 mg/g FW in infected plants) (Figure 4.31). A decrease in Chlorophyll content upon infection (biotic stress) in plants have also been observed in some other studies such as reported by Guo et al. (2005) in which Chlorophyll content was significantly reduced in infected *Brassica juncea* plants after 2 weeks of turnip mosaic virus infection. While accumulation of proline in abiotic stresses in plant is well known and discussed by many authors (Kaur & Asthir, 2015; Mansour & Salama, 2020; Yaish, 2015), effect on this indicator under biotic stresses is not well established. In these results (Figure 4.32), proline content reduced in infected plants (1.58 µg/g FW) compared to control (1.80 µg/g FW); however, the decrease was not found to be statistically significant. Microscopic examination of transverse leaf sections from infected and control *A. sisalana* revealed visible differences in cells and vascular bundles in which various morphological differences could be observed due to *A. niger* infection (Figure 4.33 and Figure 4.34); however, a more detailed analysis can be done to further investigate histopathological changes in *A. sisalana* leaves upon *A. niger* infection. Fungi are capable to colonize plant tissues leading to cell damage and resulting in disease symptoms in their host plants. A significant cell proliferation and wall thickening were observed in corn plants infected by endophytic *F. sacchari* (Terna et al., 2022); such detailed study on *A. sisalana* infected with *A. niger* using light

microscopy using various staining techniques and transmission electron microscopy could reveal a more detailed analysis of histopathological effects.

The present research focuses on harnessing the available transcriptome resource of *A. sisalana* to identify miRNAs encoded by the non-coding repertoire of transcriptome assembly. Role of miRNAs in regulating gene expression in the conditions of pathogen stress that will help in understanding the molecular mechanism of gene regulation in adverse conditions. The results of the current study will pave the way in application of miRNAome research and gaining insights in the regulation of defense mechanism in xeric plants. Through this study, 48 miRNAs of *A. sisalana* have been experimentally verified which is initiatory towards gaining insights into the stress response and gene regulation mechanisms under biotic and abiotic stresses as miRNAs act as key regulators of various biological processes. In this study, the possible interactions between As-miRNAs identified in this study and the USP genes of *A. sisalana* have been predicted, further confirmation of these miRNA-USP gene interactions and functional investigations of these miRNAs can be performed through in-vitro, wet-lab experimentation to further uncover their roles and function in biotic stress in *A. sisalana*. Moreover, small RNA-seq analysis of *A. sisalana* under control and biotic stress can be performed which can reveal novel miRNA sequences in *A. sisalana* and their differential expression under control and biotic stress can unveil miRNAs which are differentially expressed and play role in gene regulation of *A. sisalana* under biotic stress. Differential expression profiling of As-USP genes identified in this study, under control and biotic stress, can be conducted to further investigate their role in defense to infection by *A. niger*.

Till date the control strategies that have been presented by various researchers to counter the bole rot disease of *A. sisalana*, caused by *A. niger*, have been focused on application of bacterial isolates that have the potential to limit growth of *A. niger*. While the physiological roles and

diversity of structures of USPs in microorganisms have been broadly investigated, only a few studies have in plants (though having numerous USPs) have been conducted so far (Chi et al., 2019). The role of USPs in various biotic and abiotic stresses have been reported by some researchers in plant species such as *Arabidopsis thaliana* (Bhuria et al., 2019; Melencion et al., 2017), *Medicago falcata* (Gou et al., 2020), *Oryza sativa* L (Arabia et al., 2021); however, no such studies on USPs in *A. sisalana* have been performed hitherto. Exploring molecular mechanism of stress tolerance in plants is needed to overcome the deterioration of yield and quality of crop plants to meet the food security challenges of the growing population. Crops are frequently exposed to various biotic and abiotic stresses (such as extreme temperatures, salinity, drought and pathogens) that compromise the yield (Pandey et al., 2017). Results of this study will also pave the way towards development of molecular based strategies to improve tolerance to *A. sisalana* to the bole rot disease caused by *A. niger* through over-expression of USP genes after confirmation of their role in pathogen defense.

6. REFERENCES

- Ai, T., Zhang, L., Gao, Z., Zhu, C., & Guo, X. (2011). Highly efficient virus resistance mediated by artificial microRNAs that target the suppressor of PVX and PVY in plants. *Plant Biology*, 13(2), 304-316.
- Ameres, S. L., & Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology*, 14(8), 475-488.
- Andge, R., Parate, R., Sawai, H., & Kalaskar, R. (2017). In vitro studies on management of collar rot caused by *Aspergillus niger* in groundnut. *Journal of Soils and Crops*, 27(1), 80-83.
- Arabia, S., Sami, A. A., Akhter, S., Sarker, R. H., & Islam, T. (2021). Comprehensive in-silico characterization of Universal Stress Proteins in rice (*Oryza sativa* L.) with insight into their stress-specific transcriptional modulation. *Frontiers in Plant Science*, 12, 712607.
- Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., & Baulcombe, D. C. (2005). Cloning and characterization of micro-RNAs from moss. *The Plant Journal*, 43(6), 837-848.
- Ariizumi, T., Lawrence, P. K., & Steber, C. M. (2011). The role of two F-box proteins, SLEEPY1 and SNEEZY, in Arabidopsis gibberellin signaling. *Plant Physiology*, 155(2), 765-775.
- Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339(6123), 1074-1077.
- Asfaw, K. G. (2011). Investigation of the reasons for the unique growth and development of agave species (*Agave sisalana* and *Agave americana*) crop plants at the southern, central,

- north western and eastern parts of Tigray, Ethiopia. *Current Research Journal of Biological Sciences*, 3(4), 273-281.
- Attia, A. S., Cassat, J. E., Aranmolate, S. O., Zimmerman, L. J., Boyd, K. L., & Skaar, E. P. (2013). Analysis of the *Staphylococcus aureus* abscess proteome identifies antimicrobial host proteins and bacterial stress responses at the host–pathogen interface. *Pathogens and Disease*, 69(1), 36-48.
- Aukerman, M. J., & Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *The Plant Cell*, 15(11), 2730-2741.
- Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biology*, 12(4), 1-13.
- Bano, N., Fakhrah, S., Nayak, S. P., Bag, S. K., & Mohanty, C. S. (2022). Identification of miRNA and their target genes in *Cestrum nocturnum* L. and *Cestrum diurnum* L. in stress responses. *Physiology and Molecular Biology of Plants*, 28(1), 31-49.
- Barbosa, L. O., Lima, J. S., Magalhães, V. C., Gava, C. A. T., Soares, A. C. F., Marbach, P. A. S., & de Souza, J. T. (2018). Compatibility and combination of selected bacterial antagonists in the biocontrol of sisal bole rot disease. *BioControl*, 63(4), 595-605.
- Batcho, A., Jabbar, B., Sarwar, M., Rashid, B., Hassan, S., & Husnain, T. (2022). Transient Expression Analysis of *Agave sisalana* Heat Shock Protein Gene (AsHSP70) in Model Species (*Nicotiana benthamiana*) under Heat Stress. *Biology Bulletin*, 49(3), 160-168.
- Bates, L. S., Waldren, R. P., & Teare, I. (1973). Rapid determination of free proline for water-stress studies. *Plant and Soil*, 39(1), 205-207.
- Baulcombe, D. (2004). RNA silencing in plants. *Nature*, 431(7006), 356-363.

- Bazzini, A. A., Almasia, N. I., Manacorda, C. A., Mongelli, V. C., Conti, G., Maroniche, G. A., Rodriguez, M. C., Distéfano, A. J., Hopp, H. E., & Del Vas, M. (2009). Virus infection elevates transcriptional activity of miR164a promoter in plants. *BMC Plant Biology*, 9(1), 1-12.
- Bebber, D. P., & Gurr, S. J. (2015). Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genetics and Biology*, 74, 62-64.
- Bhuria, M., Goel, P., Kumar, S., & Singh, A. K. (2019). Genome-wide identification and expression profiling of genes encoding universal stress proteins (USP) identify multi-stress responsive USP genes in *Arabidopsis thaliana*. *Plant Physiology Reports*, 24(3), 434-445.
- Boccaro, M., Sarazin, A., Thiebeauld, O., Jay, F., Voinnet, O., Navarro, L., & Colot, V. (2014). The Arabidopsis miR472-RDR6 silencing pathway modulates PAMP-and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS Pathogens*, 10(1), e1003883.
- Bonnet, E., He, Y., Billiau, K., & Van de Peer, Y. (2010). TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, 26(12), 1566-1568.
- Botura, M. B., Silva, G., Lima, H., Oliveira, J., Souza, T., Santos, J. D. G. d., Branco, A., Moreira, E., Almeida, M. A. O. d., & Batatinha, M. J. M. (2011). In vivo anthelmintic activity of an aqueous extract from sisal waste (*Agave sisalana* Perr.) against gastrointestinal nematodes in goats. *Veterinary Parasitology*, 177(1-2), 104-110.

- Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., & Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, 320(5880), 1185-1190.
- Calviño, M., Bruggmann, R., & Messing, J. (2011). Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. *BMC Genomics*, 12(1), 1-13.
- Chang, E., Zhao, Y., Wei, Q., Shi, S., & Jiang, Z. (2016). Isolation of high-quality RNA from *Platycladus orientalis* and other Cupressaceae plants. *Electronic Journal of Biotechnology*, 23, 21-27.
- Chellappan, P., Mason, M. V., Vanitharani, R., Taylor, N. J., & Fauquet, C. M. (2004). Broad spectrum resistance to ssDNA viruses associated with transgene-induced gene silencing in cassava. *Plant Molecular Biology*, 56(4), 601-611.
- Chen, D., Du, Y., Chen, H., Fan, Y., Fan, X., Zhu, Z., Wang, J., Xiong, C., Zheng, Y., & Hou, C. (2019). Comparative identification of microRNAs in *Apis cerana cerana* workers' midguts in response to *Nosema ceranae* invasion. *Insects*, 10(9), 258.
- Chen, H., Zhang, L., Yu, K., & Wang, A. (2015). Pathogenesis of Soybean mosaic virus in soybean carrying Rsv1 gene is associated with miRNA and siRNA pathways, and breakdown of AGO1 homeostasis. *Virology*, 476, 395-404.
- Chen, L., Heikkinen, L., Wang, C., Yang, Y., Sun, H., & Wong, G. (2019). Trends in the development of miRNA bioinformatics tools. *Briefings in Bioinformatics*, 20(5), 1836-1852.

- Chen, L., Luan, Y., & Zhai, J. (2015). Sp-miR396a-5p acts as a stress-responsive genes regulator by conferring tolerance to abiotic stresses and susceptibility to *Phytophthora nicotianae* infection in transgenic tobacco. *Plant Cell Reports*, 34(12), 2013-2025.
- Chen, L., Ren, Y., Zhang, Y., Xu, J., Zhang, Z., & Wang, Y. (2012). Genome-wide profiling of novel and conserved *Populus* microRNAs involved in pathogen stress response by deep sequencing. *Planta*, 235(5), 873-883.
- Chen, M., & Cao, Z. (2015). Genome-wide expression profiling of microRNAs in poplar upon infection with the foliar rust fungus *Melampsora larici-populina*. *BMC Genomics*, 16(1), 1-13.
- Chen, X. (2005). MicroRNA biogenesis and function in plants. *FEBS Letters*, 579(26), 5923-5931.
- Chi, Y. H., Koo, S. S., Oh, H. T., Lee, E. S., Park, J. H., Phan, K. A. T., Wi, S. D., Bae, S. B., Paeng, S. K., & Chae, H. B. (2019). The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. *Frontiers in Plant Science*, 10, 750.
- Chou, M.-X., Wei, X.-Y., Chen, D.-S., & Zhou, J.-C. (2007). A novel nodule-enhanced gene encoding a putative universal stress protein from *Astragalus sinicus*. *Journal of Plant Physiology*, 164(6), 764-772.
- Conde, L., Bracci, P. M., Richardson, R., Montgomery, S. B., & Skibola, C. F. (2013). Integrating GWAS and expression data for functional characterization of disease-associated SNPs: an application to follicular lymphoma. *The American Journal of Human Genetics*, 92(1), 126-130.

- Coutinho, W. M., Suassuna, N. D., Luz, C. M., Suinaga, F. A., & Silva, O. R. (2006). Bole rot of sisal caused by *Aspergillus niger* in Brazil. *Fitopatologia Brasileira*, *31*, 605-605.
- Creighton, C. J., Reid, J. G., & Gunaratne, P. H. (2009). Expression profiling of microRNAs by deep sequencing. *Briefings in Bioinformatics*, *10*(5), 490-497.
- Cui, X., & Cao, X. (2014). Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Current Opinion in Plant Biology*, *21*, 83-88.
- Dai, X., Zhuang, Z., & Zhao, P. X. (2011). Computational analysis of miRNA targets in plants: current status and challenges. *Briefings in Bioinformatics*, *12*(2), 115-121.
- Damasceno, C. L., De Sa, J. O., Da Silva, R. M., do Carmo, C. O., Haddad, L., Soares, A. C. F., & das Almas, C. (2019). *Penicillium citrinum* as a potential biocontrol agent for sisal bole rot disease. *Journal of Agricultural Science*, *11*, 206-216.
- Davis, S. C., Dohleman, F. G., & Long, S. P. (2011). The global potential for Agave as a biofuel feedstock. *GCB Bioenergy*, *3*(1), 68-78.
- De Souza, J. T., Silva, A. C. M., de Jesus Santos, A. F., Santos, P. O., Alves, P. S., Cruz-Magalhaes, V., Marbach, P. A. S., & Loguercio, L. L. (2021). Endophytic bacteria isolated from both healthy and diseased *Agave sisalana* plants are able to control the bole rot disease. *Biological Control*, *157*, 104575.
- Dezulian, T., Remmert, M., Palatnik, J. F., Weigel, D., & Huson, D. H. (2006). Identification of plant microRNA homologs. *Bioinformatics*, *22*(3), 359-360.
- Domingues, L. F., Botura, M. B., Cruz, A. C. F. G. d., Yuki, C. C., Silva, G. D. d., Costa, M. S., Murphy, G., Moreira, E. L. T., Meneses, Í. D. S. d., & Almeida, M. d. G. Á. R. d. (2010). Evaluation of anthelmintic activity of liquid waste of *Agave sisalana* (sisal) in goats. *Revista Brasileira de Parasitologia Veterinária*, *19*, 270-272.

- Drumm, J. E., Mi, K., Bilder, P., Sun, M., Lim, J., Bielefeldt-Ohmann, H., Basaraba, R., So, M., Zhu, G., & Tufariello, J. M. (2009). *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. *PLoS Pathogens*, 5(5), e1000460.
- El Mokni, R., & Verloove, F. (2021). Additional species of Agave (Agavoideae/Agavaceae, Asparagaceae *sensu lat.*) introduced and naturalising in Tunisia and North Africa. *Bradleya*, 2021(39), 221-235.
- Eldem, V., Okay, S., & Ünver, T. (2013). Plant microRNAs: new players in functional genomics. *Turkish Journal of Agriculture and Forestry*, 37(1), 1-21.
- Elhosseiny, N. M., Amin, M. A., Yassin, A. S., & Attia, A. S. (2015). *Acinetobacter baumannii* universal stress protein A plays a pivotal role in stress response and is essential for pneumonia and sepsis pathogenesis. *International Journal of Medical Microbiology*, 305(1), 114-123.
- Ergin, S., Kherad, N., & Alagoz, M. (2022). RNA sequencing and its applications in cancer and rare diseases. *Molecular Biology Reports*, 1-9.
- Escamilla-Treviño, L. L. (2012). Potential of plants from the genus Agave as bioenergy crops. *Bioenergy Research*, 5(1), 1-9.
- Fahlgren, N., Howell, M. D., Kasschau, K. D., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., Givan, S. A., Law, T. F., Grant, S. R., & Dangl, J. L. (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PloS One*, 2(2), e219.
- Faller, M., & Guo, F. (2008). MicroRNA biogenesis: there's more than one way to skin a cat. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1779(11), 663-667.

- Feng, H., Zhang, Q., Wang, Q., Wang, X., Liu, J., Li, M., Huang, L., & Kang, Z. (2013). Target of tae-miR408, a chemocyanin-like protein gene (TaCLP1), plays positive roles in wheat response to high-salinity, heavy cupric stress and stripe rust. *Plant Molecular Biology*, 83(4), 433-443.
- Flood, J. (2010). The importance of plant health to food security. *Food Security*, 2(3), 215-231.
- Frazier, T. P., Xie, F., Freistaedter, A., Burklew, C. E., & Zhang, B. (2010). Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). *Planta*, 232(6), 1289-1308.
- Gadhavi, H., Patel, M., Mangukia, N., Shah, K., Bhadresha, K., Patel, S. K., Rawal, R. M., & Pandya, H. A. (2020). Transcriptome-wide miRNA identification of *Bacopa monnieri*: a cross-kingdom approach. *Plant Signaling & Behavior*, 15(1), 1699265.
- Gao, F., Nan, F., Feng, J., Lv, J., Liu, Q., & Xie, S. (2021). Discovery of Conserved and Novel MicroRNAs in *Galdieria sulphuraria*. *Iranian Journal of Biotechnology*, 19(2), e2671.
- Gao, Z., Nie, J., & Wang, H. (2021). MicroRNA biogenesis in plant. *Plant Growth Regulation*, 93(1), 1-12.
- Gautam, A. K., Sharma, S., Avasthi, S., & Bhadauria, R. (2011). Diversity, pathogenicity and toxicology of *A. niger*: an important spoilage fungi. *Research Journal of Microbiology*, 6(3), 270-280.
- Ghawana, S., Paul, A., Kumar, H., Kumar, A., Singh, H., Bhardwaj, P. K., Rani, A., Singh, R. S., Raizada, J., & Singh, K. (2011). An RNA isolation system for plant tissues rich in secondary metabolites. *BMC Research Notes*, 4(1), 1-5.
- Gomes, C. P., Cho, J.-H., Hood, L., Franco, O. L., Pereira, R. W., & Wang, K. (2013). A review of computational tools in microRNA discovery. *Frontiers in Genetics*, 4, 81.

- Gou, L., Zhuo, C., Lu, S., & Guo, Z. (2020). A Universal Stress Protein from *Medicago falcata* (MfUSP1) confers multiple stress tolerance by regulating antioxidant defense and proline accumulation. *Environmental and Experimental Botany*, 178, 104168.
- Gross, S. M., Martin, J. A., Simpson, J., Abraham-Juarez, M. J., Wang, Z., & Visel, A. (2013). De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics*, 14(1), 1-14.
- Guo, D.-P., Guo, Y.-P., Zhao, J.-P., Liu, H., Peng, Y., Wang, Q.-M., Chen, J.-S., & Rao, G.-Z. (2005). Photosynthetic rate and chlorophyll fluorescence in leaves of stem mustard (*Brassica juncea* var. *tsatsai*) after turnip mosaic virus infection. *Plant Science*, 168(1), 57-63.
- Gupta, H., Tiwari, T., Patel, M., Mehta, A., & Ghosh, A. (2015). An approach to identify the novel miRNA encoded from H. Annuus EST sequences. *Genomics Data*, 6, 139-144.
- Gupta, O. P., Permar, V., Koundal, V., Singh, U. D., & Praveen, S. (2012). MicroRNA regulated defense responses in *Triticum aestivum* L. during *Puccinia graminis* f. sp. *tritici* infection. *Molecular Biology Reports*, 39(2), 817-824.
- Gustavsson, N., Diez, A., & Nyström, T. (2002). The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Molecular Microbiology*, 43(1), 107-117.
- Han, J., Kong, M., Xie, H., Sun, Q., Nan, Z., Zhang, Q., & Pan, J. (2013). Identification of miRNAs and their targets in wheat (*Triticum aestivum* L.) by EST analysis. *Genetics and Molecular Research*, 12(3793), 805.
- Han, Y., Gao, S., Muegge, K., Zhang, W., & Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinformatics and Biology Insights*, 9, BBI. S28991.

- Hu, G., Lei, Y., Liu, J., Hao, M., Zhang, Z., Tang, Y., Chen, A., & Wu, J. (2020). The ghr-miR164 and GhNAC100 modulate cotton plant resistance against *Verticillium dahlia*. *Plant Science*, 293, 110438.
- Huang, J., Yang, M., & Zhang, X. (2016). The function of small RNAs in plant biotic stress response. *Journal of Integrative Plant Biology*, 58(4), 312-327.
- Iqbal, M. S., Jabbar, B., Sharif, M. N., Ali, Q., Husnain, T., & Nasir, I. A. (2017). In silico MCMV silencing concludes potential host-derived miRNAs in maize. *Frontiers in Plant Science*, 8, 372.
- Jabbar, B., Anicet, B. A., Sarwar, M. B., Rashid, B., Hassan, S., & Husnain, T. (2021). RNA-Seq Data Analysis Unveils Potential Conserved Micro-RNAs in *Agave deserti*. *Current Proteomics*, 18(2), 248-263.
- Jabbar, B., Iqbal, M. S., Batcho, A. A., Nasir, I. A., Rashid, B., Husnain, T., & Henry, R. J. (2019). Target prediction of candidate miRNAs from *Oryza sativa* for silencing the RYMV genome. *Computational Biology and Chemistry*, 83, 107127.
- Jelly, N. S., Schellenbaum, P., Walter, B., & Maillot, P. (2012). Transient expression of artificial microRNAs targeting Grapevine fanleaf virus and evidence for RNA silencing in grapevine somatic embryos. *Transgenic Research*, 21(6), 1319-1327.
- Jin, H. (2008). Endogenous small RNAs and antibacterial immunity in plants. *FEBS Letters*, 582(18), 2679-2684.
- Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, 11(1), 1-8.
- Jones-Rhoades, M. W., & Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell*, 14(6), 787-799.

- Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323-329.
- Kalariya, K. A., Meena, R. P., Saran, P. L., & Manivel, P. (2019). Identification of microRNAs from transcriptome data in gurmar (*Gymnema sylvestre*). *Horticulture, Environment, and Biotechnology*, 60(3), 383-397.
- Kang, W., & Friedländer, M. R. (2015). Computational prediction of miRNA genes from small RNA sequencing data. *Frontiers in Bioengineering and Biotechnology*, 3, 7.
- Kar, M. M., & Raichaudhuri, A. (2021). Role of microRNAs in mediating biotic and abiotic stress in plants. *Plant Gene*, 26, 100277.
- Kaur, G., & Asthir, B. (2015). Proline: a key player in plant abiotic stress tolerance. *Biologia Plantarum*, 59(4), 609-619.
- Kavin, T., & Janagan, S. (2019). Studies on Natural Fibres as Fixed Aerated Beds for Domestic Wastewater Treatment. *International Research Journal of Engineering and Technology*, 6(4), 3449-3452.
- Kawaguchi, R., Girke, T., Bray, E. A., & Bailey-Serres, J. (2004). Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *The Plant Journal*, 38(5), 823-839.
- Kerk, D., Bulgrien, J., Smith, D. W., & Gribskov, M. (2003). Arabidopsis proteins containing similarity to the universal stress protein domain of bacteria. *Plant Physiology*, 131(3), 1209-1219.
- Khraiweh, B., Zhu, J.-K., & Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819(2), 137-148.

- Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 2015(11), pdb. top084970.
- Kulcheski, F. R., de Oliveira, L. F., Molina, L. G., Almerão, M. P., Rodrigues, F. A., Marcolino, J., Barbosa, J. F., Stolf-Moreira, R., Nepomuceno, A. L., & Marcelino-Guimarães, F. C. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics*, 12(1), 1-17.
- Kulshrestha, C., Pathak, H., Kumar, D., Dave, S., & Sudan, J. (2020). Elucidating micro RNAs role in different plant–pathogen interactions. *Molecular Biology Reports*, 47, 8219-8227.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547.
- Kvint, K., Nachin, L., Diez, A., & Nyström, T. (2003). The bacterial universal stress protein: function and regulation. *Current Opinion in Microbiology*, 6(2), 140-145.
- Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294(5543), 858-862.
- Lenman, M., Sörensson, C., & Andreasson, E. (2008). Enrichment of phosphoproteins and phosphopeptide derivatization identify universal stress proteins in elicitor-treated *Arabidopsis*. *Molecular Plant-Microbe Interactions*, 21(10), 1275-1284.
- Li, W.-T., Wei, Y.-M., Wang, J.-R., Liu, C.-J., Lan, X.-J., Jiang, Q.-T., Pu, Z.-E., & Zheng, Y.-L. (2010). Identification, localization, and characterization of putative USP genes in barley. *Theoretical and Applied Genetics*, 121(5), 907-917.

- Li, Y., Lu, Y.-G., Shi, Y., Wu, L., Xu, Y.-J., Huang, F., Guo, X.-Y., Zhang, Y., Fan, J., & Zhao, J.-Q. (2014). Multiple rice microRNAs are involved in immunity against the blast fungus *Magnaporthe oryzae*. *Plant Physiology*, *164*(2), 1077-1092.
- Liu, J. (2008). Control of protein synthesis and mRNA degradation by microRNAs. *Current Opinion in Cell Biology*, *20*(2), 214-221.
- Llave, C., Kasschau, K. D., Rector, M. A., & Carrington, J. C. (2002). Endogenous and silencing-associated small RNAs in plants. *The Plant Cell*, *14*(7), 1605-1619.
- Lu, S., Sun, Y. H., Amerson, H., & Chiang, V. L. (2007). MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. *The Plant Journal*, *51*(6), 1077-1098.
- Lu, S., Sun, Y. H., & Chiang, V. L. (2008). Stress-responsive microRNAs in Populus. *The Plant Journal*, *55*(1), 131-151.
- Magalhães, V. C., de Oliveira Barbosa, L., Andrade, J. P., Soares, A. C. F., de Souza, J. T., & Marbach, P. A. S. (2017). Burkholderia isolates from a sand dune leaf litter display biocontrol activity against the bole rot disease of *Agave sisalana*. *Biological Control*, *112*, 41-48.
- Mansour, M. M. F., & Salama, K. H. A. (2020). Proline and abiotic stresses: Responses and adaptation. In *Plant Ecophysiology and Adaptation under Climate Change: Mechanisms and Perspectives II* (pp. 357-397). Springer.
- Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, *9*(1), 387-402.
- Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, *470*(7333), 198-203.

- Matts, J., Jagadeeswaran, G., Roe, B. A., & Sunkar, R. (2010). Identification of microRNAs and their targets in switchgrass, a model biofuel plant species. *Journal of Plant Physiology*, 167(11), 896-904.
- McCombie, W. R., McPherson, J. D., & Mardis, E. R. (2019). Next-generation sequencing technologies. *Cold Spring Harbor Perspectives in Medicine*, 9(11), a036798.
- Meera Krishna, B., Khan, M. A., & Khan, S. T. (2019). Next-generation sequencing (NGS) platforms: an exciting era of genome sequence analysis. In *Microbial Genomics in Sustainable Agroecosystems* (pp. 89-109). Springer.
- Melencion, S. M. B., Chi, Y. H., Pham, T. T., Paeng, S. K., Wi, S. D., Lee, C., Ryu, S. W., Koo, S. S., & Lee, S. Y. (2017). RNA chaperone function of a universal stress protein in *Arabidopsis* confers enhanced cold stress tolerance in plants. *International Journal of Molecular Sciences*, 18(12), 2546.
- Meng, F., Liu, H., Wang, K., Liu, L., Wang, S., Zhao, Y., Yin, J., & Li, Y. (2013). Development-associated microRNAs in grains of wheat (*Triticum aestivum* L.). *BMC Plant Biology*, 13(1), 1-14.
- Merkouropoulos, G., Andreasson, E., Hess, D., Boller, T., & Peck, S. C. (2008). An *Arabidopsis* protein phosphorylated in response to microbial elicitation, AtPHOS32, is a substrate of MAP kinases 3 and 6. *Journal of Biological Chemistry*, 283(16), 10493-10499.
- Mette, M. F., van der Winden, J., Matzke, M., & Matzke, A. J. (2002). Short RNAs can identify new candidate transposable element families in *Arabidopsis*. *Plant Physiology*, 130(1), 6-9.
- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31-46.

- Millar, A. A., & Waterhouse, P. M. (2005). Plant and animal microRNAs: similarities and differences. *Functional & Integrative Genomics*, 5(3), 129-135.
- Mohanpuria, P., Duhan, N., Sarao, N. K., Kaur, M., & Kaur, M. (2018). In silico identification and validation of potential microRNAs in Kinnow Mandarin (*Citrus reticulata* Blanco). *Interdisciplinary Sciences: Computational Life Sciences*, 10(4), 762-770.
- Monja-Mio, K. M., Herrera-Alamillo, M. A., Sánchez-Teyer, L. F., & Robert, M. L. (2019). Breeding strategies to improve production of Agave (Agave spp.). *Advances in Plant Breeding Strategies: Industrial and Food Crops*, 319-362.
- Mousumi, D., Mukeshwar, P., Rohit, S., Gulab, S. T., & Pushpa, L. (2010). Biotechnological intervention of *Agave sisalana*: a unique fiber yielding plant with medicinal property. *Journal of Medicinal Plants Research*, 4(3), 177-187.
- Murtaza, S., Tabassum, B., Tariq, M., Riaz, S., Yousaf, I., Jabbar, B., Khan, A., Samuel, A. O., Zameer, M., & Nasir, I. A. (2022). Silencing a *Myzus persicae* Macrophage Inhibitory Factor by Plant-Mediated RNAi Induces Enhanced Aphid Mortality Coupled with Boosted RNAi Efficacy in Transgenic Potato Lines. *Molecular Biotechnology*, 64(10), 1152-1163. <https://doi.org/10.1007/s12033-022-00498-w>
- Nachin, L., Brive, L., Persson, K.-C., Svensson, P., & Nystrom, T. (2008). Heterodimer formation within universal stress protein classes revealed by an in silico and experimental approach. *Journal of Molecular Biology*, 380(2), 340-350.
- Naganeeswaran, S., Fayas, T., Rachana, K., & Rajesh, M. (2015). Computational prediction and characterization of miRNA from coconut leaf transcriptome. *Journal of Applied Horticulture*, 17(1), 12-17.

- Najafabadi, A. S., & Naghavi, M. R. (2018). Mining *Ferula gummosa* transcriptome to identify miRNAs involved in the regulation and biosynthesis of terpenes. *Gene*, 645, 41-47.
- Nobuta, K., McCormick, K., Nakano, M., & Meyers, B. C. (2010). Bioinformatics analysis of small RNAs in plants using next generation sequencing technologies. *Plant MicroRNAs: Methods and Protocols*, 89-106.
- Olliff, J. C., Mei, J. A., Shirley, K. M., & Hanson, S. J. (2022). A Perspective on RNA. In *Fundamentals of RNA Structure and Function* (pp. 91-122). Springer.
- Ouyang, S., Park, G., Atamian, H. S., Han, C. S., Stajich, J. E., Kaloshian, I., & Borkovich, K. A. (2014). MicroRNAs suppress NB domain genes in tomato that confer resistance to *Fusarium oxysporum*. *PLoS Pathogens*, 10(10), e1004464.
- Pacheco, R., García-Marcos, A., Barajas, D., Martiáñez, J., & Tenllado, F. (2012). PVX–potyvirus synergistic infections differentially alter microRNA accumulation in *Nicotiana benthamiana*. *Virus Research*, 165(2), 231-235.
- Pan, X., Zhang, B., Francisco, M. S., & Cobb, G. P. (2007). Characterizing viral microRNAs and its application on identifying new microRNAs in viruses. *Journal of Cellular Physiology*, 211(1), 10-18.
- Pandey, P., Irulappan, V., Bagavathiannan, M. V., & Senthil-Kumar, M. (2017). Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Frontiers in Plant Science*, 8, 537.
- Pentimone, I., Lebrón, R., Hackenberg, M., Rosso, L. C., Colagiero, M., Nigro, F., & Ciano, A. (2018). Identification of tomato miRNAs responsive to root colonization by endophytic *Pochonia chlamydosporia*. *Applied Microbiology and Biotechnology*, 102(2), 907-919.

- Persson, Ö., Valadi, Å., Nyström, T., & Farewell, A. (2007). Metabolic control of the *Escherichia coli* universal stress protein response through fructose-6-phosphate. *Molecular Microbiology*, 65(4), 968-978.
- Pöggeler, S. (2002). Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. *Current Genetics*, 42(3), 153-160.
- Pradhan, B., Naqvi, A. R., Saraf, S., Mukherjee, S. K., & Dey, N. (2015). Prediction and characterization of Tomato leaf curl New Delhi virus (ToLCNDV) responsive novel microRNAs in *Solanum lycopersicum*. *Virus Research*, 195, 183-195.
- Prakash, P., Rajakani, R., & Gupta, V. (2016). Transcriptome-wide identification of *Rauvolfia serpentina* microRNAs and prediction of their potential targets. *Computational Biology and Chemistry*, 61, 62-74.
- Pritchard, C. C., Cheng, H. H., & Tewari, M. (2012). MicroRNA profiling: approaches and considerations. *Nature Reviews Genetics*, 13(5), 358-369.
- Ramachandran, V., & Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science*, 321(5895), 1490-1492.
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., & Bartel, D. P. (2002). MicroRNAs in plants. *Genes & Development*, 16(13), 1616-1626.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H. M., & Qian, J. Q. (2010). De novo assembly and analysis of RNA-seq data. *Nature Methods*, 7(11), 909-912.
- Salamon, S., Żok, J., Gromadzka, K., & Błaszczyk, L. (2021). Expression Patterns of miR398, miR167, and miR159 in the Interaction between Bread Wheat (*Triticum aestivum* L.) and

- Pathogenic *Fusarium culmorum* and Beneficial Trichoderma Fungi. *Pathogens*, 10(11), 1461.
- Santos, P. O. d., Silva, A. C. M. d., Corrêa, É. B., Magalhães, V. C., & Souza, J. T. d. (2014). Additional species of *Aspergillus* causing bole rot disease in *Agave sisalana*. *Tropical Plant Pathology*, 39, 331-334.
- Sarwar, M. B., Ahmad, Z., Rashid, B., Hassan, S., Gregersen, P. L., Leyva, M. D. I. O., Nagy, I., Asp, T., & Husnain, T. (2019). De novo assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Scientific Reports*, 9(1), 1-14.
- Schuster, E., Dunn-Coleman, N., Frisvad, J., & Van Dijck, P. (2002). On the safety of *Aspergillus niger*—a review. *Applied Microbiology and Biotechnology*, 59(4), 426-435.
- Šečić, E., Kogel, K.-H., & Ladera-Carmona, M. J. (2021). Biotic stress-associated microRNA families in plants. *Journal of Plant Physiology*, 263, 153451.
- Seifart Gomes, C., Izar, B., Pazan, F., Mohamed, W., Mraheil, M. A., Mukherjee, K., Billion, A., Aharonowitz, Y., Chakraborty, T., & Hain, T. (2011). Universal stress proteins are important for oxidative and acid stress resistance and growth of *Listeria monocytogenes* EGD-e in vitro and in vivo. *PLoS One*, 6(9), e24965.
- Sen, S., Rai, R., Chatterjee, A., Rai, S., Yadav, S., Agrawal, C., & Rai, L. (2019). Molecular characterization of two novel proteins All1122 and Alr0750 of *Anabaena* PCC 7120 conferring tolerance to multiple abiotic stresses in *Escherichia coli*. *Gene*, 685, 230-241.
- Shahzad, S., Hussain, M., Munir, H., & Arfan, M. (2022). Proximate composition and spatio-temporal heterogeneity of phytochemicals in *Agave sisalana* Perrine (sisal) adapted in

- different agro-ecological zones of Punjab, Pakistan. *Environmental Science and Pollution Research*, 29(32), 48869-48879. <https://doi.org/10.1007/s11356-022-19260-5>
- Sharma, R. (2012). Pathogenecity of *Aspergillus niger* in plants. *CIBTech Journal of Microbiology*, 1(1), 47-51.
- Shokry, A. M., Al-Karim, S., Ramadan, A., Gadallah, N., Al Attas, S. G., Sabir, J. S., Hassan, S. M., Madkour, M. A., Bressan, R., & Mahfouz, M. (2014). Detection of a Usp-like gene in *Calotropis procera* plant from the de novo assembled genome contigs of the high-throughput sequencing dataset. *Comptes Rendus Biologies*, 337(2), 86-94.
- Simón-Mateo, C., & García, J. A. (2006). MicroRNA-guided processing impairs Plum pox virus replication, but the virus readily evolves to escape this silencing mechanism. *Journal of Virology*, 80(5), 2429-2436.
- Singh, N., & Sharma, A. (2017). Turmeric (*Curcuma longa*): miRNAs and their regulating targets are involved in development and secondary metabolite pathways. *Comptes Rendus Biologies*, 340(11-12), 481-491.
- Soto-Suárez, M., Baldrich, P., Weigel, D., Rubio-Somoza, I., & San Segundo, B. (2017). The Arabidopsis miR396 mediates pathogen-associated molecular pattern-triggered immune responses against fungal pathogens. *Scientific Reports*, 7(1), 1-14.
- Sousa, M. C., & McKay, D. B. (2001). Structure of the universal stress protein of *Haemophilus influenzae*. *Structure*, 9(12), 1135-1141.
- Sripathi, V. R., Anche, V. C., Gossett, Z. B., & Walker, L. T. (2021). Recent applications of RNA-sequencing in food and agriculture. *Applications of RNA-Seq in Biology and Medicine*, 97.

- Srivastava, P. K., Moturu, T. R., Pandey, P., Baldwin, I. T., & Pandey, S. P. (2014). A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics*, *15*(1), 1-15.
- Sun, B., Liu, C., Li, H., Zhang, L., Luo, G., Liang, S., & Lü, M. (2020). Research progress on the interactions between long non-coding RNAs and microRNAs in human cancer. *Oncology Letters*, *19*(1), 595-605.
- Sun, G. (2012). MicroRNAs and their diverse functions in plants. *Plant Molecular Biology*, *80*(1), 17-36.
- Sunkar, R., Li, Y.-F., & Jagadeeswaran, G. (2012). Functions of microRNAs in plant stress responses. *Trends in Plant Science*, *17*(4), 196-203.
- Terna, T. P., Mohamed Nor, N. M. I., & Zakaria, L. (2022). Histopathology of Corn Plants Infected by Endophytic Fungi. *Biology*, *11*(5), 641.
- Thody, J., Moulton, V., & Mohorianu, I. (2020). PAREameters: a tool for computational inference of plant miRNA–mRNA targeting rules using small RNA and degradome sequencing data. *Nucleic Acids Research*, *48*(5), 2258-2270.
- Tkaczuk, K. L., A. Shumilin, I., Chruszcz, M., Evdokimova, E., Savchenko, A., & Minor, W. (2013). Structural and functional insight into the universal stress protein family. *Evolutionary Applications*, *6*(3), 434-449.
- Tousi, N., Eini, O., Ahmadvand, R., Carra, A., Miozzi, L., Noris, E., & Accotto, G. (2017). In silico prediction of miRNAs targeting ToLCV and their regulation in susceptible and resistant tomato plants. *Australasian Plant Pathology*, *46*(4), 379-386.

- Udawat, P., Jha, R. K., Sinha, D., Mishra, A., & Jha, B. (2016). Overexpression of a cytosolic abiotic stress responsive universal stress protein (Sb USP) mitigates salt and osmotic stress in transgenic tobacco plants. *Frontiers in Plant Science*, 7, 518.
- Üner, C. (2020). *Next Generation Sequencing For Disease Diagnosis* Fen Bilimleri Enstitüsü].
- Van Vu, T., Choudhury, N. R., & Mukherjee, S. K. (2013). Transgenic tomato plants expressing artificial microRNAs for silencing the pre-coat and coat proteins of a begomovirus, Tomato leaf curl New Delhi virus, show tolerance to virus infection. *Virus Research*, 172(1-2), 35-45.
- Várallyay, É., Válóczy, A., Ágyi, Á., Burgyán, J., & Havelda, Z. (2010). Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *The EMBO Journal*, 29(20), 3507-3519.
- Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Development*, 20(7), 759-771.
- Vitsios, D. M., Kentepozidou, E., Quintais, L., Benito-Gutiérrez, E., Van Dongen, S., Davis, M. P., & Enright, A. J. (2017). Mirnovo: genome-free prediction of microRNAs from small RNA sequencing data and single-cells using decision forests. *Nucleic Acids Research*, 45(21), e177-e177.
- Vollmer, A. C., & Bark, S. J. (2018). Twenty-five years of investigating the universal stress protein: function, structure, and applications. In *Advances in Applied Microbiology* (Vol. 102, pp. 1-36). Elsevier.
- Vuorinne, I., Heiskanen, J., & Pellikka, P. K. (2021). Assessing leaf biomass of *Agave sisalana* using sentinel-2 vegetation indices. *Remote Sensing*, 13(2), 233.

- Wagaba, H., Patil, B. L., Mukasa, S., Alicai, T., Fauquet, C. M., & Taylor, N. J. (2016). Artificial microRNA-derived resistance to *Cassava brown streak disease*. *Journal of Virological Methods*, 231, 38-43.
- Wang, J., Mei, J., & Ren, G. (2019). Plant microRNAs: biogenesis, homeostasis, and degradation. *Frontiers in Plant Science*, 10, 360.
- Wang, X.-F., Su, J., Yang, N., Zhang, H., Cao, X.-Y., & Kang, J.-F. (2017). Functional characterization of selected universal stress protein from *Salvia miltiorrhiza* (SmUSP) in *Escherichia coli*. *Genes*, 8(9), 224.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63.
- Wang, Z., Xia, Y., Lin, S., Wang, Y., Guo, B., Song, X., Ding, S., Zheng, L., Feng, R., & Chen, S. (2018). Osa-miR164a targets Os NAC 60 and negatively regulates rice immunity against the blast fungus *Magnaporthe oryzae*. *The Plant Journal*, 95(4), 584-597.
- Wilhelm, B. T., Briau, M., Austin, P., Faubert, A., Boucher, G., Chagnon, P., Hope, K., Girard, S., Mayotte, N., & Landry, J.-R. (2011). RNA-seq analysis of 2 closely related leukemia clones that differ in their self-renewal capacity. *Blood, The Journal of the American Society of Hematology*, 117(2), e27-e38.
- Wilson, N. K., Kent, D. G., Buettner, F., Shehata, M., Macaulay, I. C., Calero-Nieto, F. J., Castillo, M. S., Oedekoven, C. A., Diamanti, E., & Schulte, R. (2015). Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell*, 16(6), 712-724.
- Wu, F., Qi, J., Meng, X., & Jin, W. (2020). miR319c acts as a positive regulator of tomato against *Botrytis cinerea* infection by targeting TCP29. *Plant Science*, 300, 110610.

- Wu, F., Shu, J., & Jin, W. (2014). Identification and validation of miRNAs associated with the resistance of maize (*Zea mays* L.) to *Exserohilum turcicum*. *PloS One*, 9(1), e87251.
- Wu, H.-J., Ma, Y.-K., Chen, T., Wang, M., & Wang, X.-J. (2012). PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Research*, 40(W1), W22-W28.
- Wu, J., Yang, Z., Wang, Y., Zheng, L., Ye, R., Ji, Y., Zhao, S., Ji, S., Liu, R., & Xu, L. (2015). Viral-inducible Argonaute18 confers broad-spectrum virus resistance in rice by sequestering a host microRNA. *eLife*, 4, e05733.
- Xie, F. L., Huang, S. Q., Guo, K., Xiang, A. L., Zhu, Y. Y., Nie, L., & Yang, Z. M. (2007). Computational identification of novel microRNAs and targets in *Brassica napus*. *FEBS Letters*, 581(7), 1464-1474.
- Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S. A., & Carrington, J. C. (2005). Expression of Arabidopsis MIRNA genes. *Plant Physiology*, 138(4), 2145-2154.
- Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z., & Sun, Q. (2010). Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). *BMC Plant Biology*, 10(1), 1-11.
- Xu, J., Xian, Q., Zhang, N., Wang, K., Zhou, X., Li, Y., Dong, J., & Chen, X. (2021). Identification of miRNA-Target Gene Pairs Responsive to Fusarium Wilt of Cucumber via an Integrated Analysis of miRNA and Transcriptome Profiles. *Biomolecules*, 11(11), 1620.
- Yaish, M. (2015). Proline accumulation is a general response to abiotic stress in the date palm tree (*Phoenix dactylifera* L.). *Genetics and Molecular Research*, 14(3), 9943-9950.

- Yang, L., Jue, D., Li, W., Zhang, R., Chen, M., & Yang, Q. (2013). Identification of MiRNA from eggplant (*Solanum melongena* L.) by small RNA deep sequencing and their response to *Verticillium dahliae* infection. *PloS One*, 8(8), e72840.
- Yin, Z., Li, Y., Han, X., & Shen, F. (2012). Genome-wide profiling of miRNAs and other small non-coding RNAs in the *Verticillium dahliae*-inoculated cotton roots. *PloS One*, 7(4), e35765.
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Steward, R., & Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science*, 307(5711), 932-935.
- Yu, X., Hou, Y., Chen, W., Wang, S., Wang, P., & Qu, S. (2017). *Malus hupehensis* miR168 targets to ARGONAUTE1 and contributes to the resistance against *Botryosphaeria dothidea* infection by altering defense responses. *Plant and Cell Physiology*, 58(9), 1541-1557.
- Zarembinski, T. I., Hung, L.-W., Mueller-Dieckmann, H.-J., Kim, K.-K., Yokota, H., Kim, R., & Kim, S.-H. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proceedings of the National Academy of Sciences*, 95(26), 15189-15193.
- Zhang, B., Pan, X., Cannon, C. H., Cobb, G. P., & Anderson, T. A. (2006). Conservation and divergence of plant microRNA genes. *The Plant Journal*, 46(2), 243-259.
- Zhang, B., Wang, Q., & Pan, X. (2007). MicroRNAs and their regulatory roles in animals and plants. *Journal of Cellular Physiology*, 210(2), 279-289.

- Zhang, B. H., Pan, X. P., Wang, Q. L., Cobb, G. P., & Anderson, T. A. (2005). Identification and characterization of new plant microRNAs using EST analysis. *Cell Research*, 15(5), 336-360.
- Zhang, L., Xiang, Y., Chen, S., Shi, M., Jiang, X., He, Z., & Gao, S. (2022). Mechanisms of MicroRNA Biogenesis and Stability Control in Plants. *Frontiers in Plant Science*, 13, 844149.
- Zhang, T., Zhao, Y.-L., Zhao, J.-H., Wang, S., Jin, Y., Chen, Z.-Q., Fang, Y.-Y., Hua, C.-L., Ding, S.-W., & Guo, H.-S. (2016). Cotton plants export microRNAs to inhibit virulence gene expression in a fungal pathogen. *Nature Plants*, 2(10), 1-6.
- Zhang, W., Gao, S., Zhou, X., Chellappan, P., Chen, Z., Zhou, X., Zhang, X., Fromuth, N., Coutino, G., & Coffey, M. (2011). Bacteria-responsive microRNAs regulate plant innate immunity by modulating plant hormone networks. *Plant Molecular Biology*, 75(1), 93-105.
- Zhao, J. P., Jiang, X. L., Zhang, B. Y., & Su, X. H. (2012). Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. *PLoS One*, 7(9), e44968.
- Zhao, Z. X., Feng, Q., Cao, X. L., Zhu, Y., Wang, H., Chandran, V., Fan, J., Zhao, J. Q., Pu, M., & Li, Y. (2020). Osa-miR167d facilitates infection of *Magnaporthe oryzae* in rice. *Journal of Integrative Plant Biology*, 62(5), 702-715.
- Zhu, Q.-H., Fan, L., Liu, Y., Xu, H., Llewellyn, D., & Wilson, I. (2013). miR482 regulation of NBS-LRR defense genes during fungal pathogen infection in cotton. *PLoS One*, 8(12), e84390.

Zuo, J., Wang, Y., Liu, H., Ma, Y., Ju, Z., Zhai, B., Fu, D., Zhu, Y., Luo, Y., & Zhu, B. (2011).
MicroRNAs in tomato plants. *Science China Life Sciences*, 54(7), 599-605.

APPENDICES

Appendix-I

RNA Extraction Buffer

	Stock Solution	Working Solution
CTAB	20%	2%
PVPP	20%	2%
EDTA	0.5M	0.025M
NaCl	5M	2M
β -mercaptoethanol	-	2% (v/v)

Appendix-II

5x TBE Buffer (1000 ml)

Tris base	54 g
Boric acid	27.5 g
0.5 M EDTA (pH 8.0)	20 ml

Appendix-III

Luria Bertani (LB) Agar (1000 ml) (pH 7.5)

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g

Appendix-IV

Luria Bertani (LB) Broth (1000 ml) (pH 7.5)

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Appendix-V

Tetracycline (12.5 mg/ml)

Tetracycline powder	12.5 mg
70% Ethanol	1 ml

Appendix-VI

Transformation Buffer

MgCl ₂	80 mM
CaCl ₂	100 mM

Appendix-VII

Ampicillin (100 mg/ml)

Ampicillin powder	1 g
Sterile water	10 ml

Appendix-VIII

PDA medium (100 ml)

Agar	1.5 g
Dextrose	2.0 g
Potato Extract	0.4 g

Appendix-IX

Acid ninhydrin solution

Ninhydrin	1.25 g
Glacial acetic acid	30 ml
6 M H ₃ PO ₄	20 ml

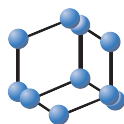
(Kept at 4°C, the reagent remains stable for 24 hours)

Appendix-X

FAA

Formaldehyde	10%
Ethanol	50%
Glacial acetic acid	5%
Water	35%

RESEARCH ARTICLE

BENTHAM
SCIENCERNA-Seq Data Analysis Unveils Potential Conserved Micro-RNAs in *Agave deserti*

Basit Jabbar¹, Batcho Agossa Anicet¹, Muhammad Bilal Sarwar¹, Bushra Rashid^{1,*}, Sameera Hassan¹ and Tayyab Husnain¹

¹Centre of Excellence in Molecular Biology, Faculty of Life Sciences, University of the Punjab, Lahore-53700, Pakistan

Abstract: Aims: Exploring molecular mechanism of abiotic stress tolerance in plants is needed to overcome the deterioration of yield and quality of crop plants to meet the food security challenges of the growing population.

Background: MicroRNAs (miRNAs) are small non-coding RNA molecules that regulate target gene expression for modulating plant growth, development, and response to different stresses. *Agave* belonging to CAM plants' has remarkable tolerance to extreme conditions of drought and heat; however, molecular mechanisms underlying this excellence are yet to explore.

Objective: This study applies comparative genomics approach on available Transcriptome (RNA-Seq) data of *Agave deserti* to identify potential miRNAs, and miRNA targets.

Methods: Transcriptome datasets consisting of 128,869 *Agave* contigs was processed to create local database, for nucleotide homology analysis with 6,028 non-redundant plant miRNAs as query sequences. Protein coding sequences were removed, and potential pre-miRNA sequences were tested for stability analysis based on a variety of factors, including but not limited to %G+C content and minimum free energy ($-\Delta G$), as a filter to remove pseudo pre-miRNAs.

Results: This study identified 30 unique miRNAs of *Agave deserti* harboring 14 different categories of precursors. Phylogenetic analysis revealed evolutionary relationship between newly identified pre-miRNAs with corresponding pre-miRNA homologues. Target genes of miRNAs were predicted subsequently, and possible functions were defined by functional annotation analysis.

Conclusion: The results of this study will pave the way for further research, exploring the molecular mechanisms in *Agave deserti* and the role of miRNAs in gene regulation under abiotic stresses.

Keywords: Micro-RNA, *Agave deserti*, gene homology, minimum free energy, precursors, target gene.

1. INTRODUCTION

Micro-RNAs are single-stranded (ss) non-coding RNAs (ncRNAs) 18-25 nucleotides, also involved in post-transcriptional regulation of gene expression in plants by a phenomenon called RNA interference (RNAi) [1]. MiRNAs are intensively complementary to hybridization sites harbored in coding regions and bring about cleavage or translational repression of specific genes [2]. MiRNAs are generated from pri-miRNA (primary miRNA), which ranges in size from hundreds to thousand nucleotides and can be modified by post-transcriptional processes such as capping and polyadenylation. A long stem-loop structure (pre-miRNA) located within pri-miRNA is excised in the nucleus before it is

transported to the cytoplasm and undergoes processing by RNase III [3]. As a result, a mature miRNA, a short double-stranded RNA (dsRNA) and a miRNA duplex (miRNA:miRNA*) [4]. Consequently, the miRNA duplex is processed by RNA-Induced Silencing Complex (RISC) to generate one strand of mature miRNA while the complementary miRNA* strand undergoes degradation. The mature miRNA can down-regulate gene expression by hybridization with 3'-UTR of the transcripts, thereby leading to cleavage of target mRNA (in case of perfect base-pairing) or repression of translation (usually imperfect base-pairing) [5].

Various transcription factor genes are particularly targeted by plant miRNAs and include Myeloblastosis-related proteins (MYB), Squamosa Promoter binding Protein-box (SBP-box), Growth-Regulating Factor (GRF), Apetala2 (AP2), and MADS-box gene families [6]. The key role of miRNAs in plants includes differentiation of leaf and flower, auxin response and flowering time as well as in biotic and abiotic stress responses and nutrient deficiency [7-9].

*Address correspondence to this author at the Centre of Excellence in Molecular Biology, Faculty of Life Sciences, University of the Punjab, P.O. Box: 53700, Lahore, Pakistan; Tel: +92-42-3529-3141; Fax: +92-42-3529-3148; E-mail: bushra.cemb@pu.edu.pk

Different approaches have emerged for the identification of miRNAs. Initially, genetic screening strategy based on forward genetics was used, such as that by Lee *et al.* [10] for the identification of *lin-4*, a small non-coding RNA in *Caenorhabditis elegans*, and Reinhart *et al.* [11] describing *let-7*, another miRNA encoded in *C. elegans*. It is an inefficient method due to time and expenses. Direct cloning procedure uses RT-PCR and sequencing and is more efficient than genetic screening; however, it cannot work for miRNAs that are difficult to clone owing to the sequence composition, or post-transcriptional features *e.g.* methylation [12]. In the contemporary era, sequence information such as GSS (Genome Survey Sequences), ESTs (Expressed Sequenced Tags) and transcriptome data (unigenes) as the target database enabled genomic and transcriptomic data analysis leading to successful prediction of miRNAs and miRNA targets. The process uses Computational prediction based on the detection of sequence features of miRNAs, having high evolutionary conservation between closely related, and in some cases distant species, by homology analysis. Another major feature is the existence of a stable pre-miRNA hairpin-like secondary structure with high-negative minimum folding energy [8, 13].

Agave, residing in monocot family of Agavaceae, is among the succulent plants having a cluster of enormous-sized leaves- thick and fleshy, and having sharp, pointed ending, denizens of arid and semi-arid regions in the North and South America [14]. It exhibits exceptional tolerance to drought and heat stress due to CAM photosynthesis, wax deposition, and protective leaf morphology [15]. *Agave deserti* is indigenous to the Sonoran desert regions of the Southwestern United States and Northwestern Mexico and can survive for a year without rainfall, and over a temperature range of -16.1°C to 61.4°C. These characteristics make this species an excellent candidate for the exploration of genomic information governing mechanisms of heat tolerance and stress responses [16]. This research involves the identification of miRNAs in *A. deserti* by implementing homology-based complementarity from previously available transcriptome data yielded from next generation (Illumina) sequencing. The approach entails retrieving the repertoire of non-redundant miRNAs available in miRBase, and screening through a stringent criterion involving homology search within certain parameters, the stability of pre-miRNA secondary structure and minimum folding energy. These miRNAs were processed further to predict the potential miRNA targets, and functional annotation of the predicted loci. Phylogenetic analysis elucidated the evolutionary relationship between pre-miRNAs of *A. deserti* and related pre-miRNA sequences.

2. MATERIALS AND METHODS

2.1. MiRNA Retrieval from MiRNA Database

A total of 10,414 mature miRNAs belonging to *Viridiplantae* were downloaded from the miRNA repository miRBase (release 22), accessible from <http://www.mirbase.org/cgi-bin/browse.pl> and removed miRNA redundancy to ob-

tain 6,028 non-redundant miRNAs sub-set and proceeded for miRNA prediction of *Agave deserti*. The length of these miRNAs ranged from 19 to 23 nucleotides.

2.1.1. Agave Deserti Transcriptome Data Source

Transcriptome assembly of *Agave deserti*, as reported by Gross *et al.* [16] sequenced by using Illumina HiSeq 2000 and de novo transcriptome assembly using Rnnotator [17], accessed from Dryad Digital Repository was used. The sequenced transcriptome contained 88,718 Agave loci, 128,869 Agave contigs and 35,086 protein coding loci.

2.2. Homology Based Prediction of Potential MiRNAs

Homology based identification of miRNAs was performed by BLASTN (standalone BLAST v2.8.1) by using unique miRNA sequences as query against the local database created from sequenced transcripts of *A. deserti*, while keeping E value at $1e^{-1}$, word size as 7, scoring (match: 2, mismatch: -3), gap cost (open: 5, extend: 2) and low complexity filter. From the BLAST results, those with full (100%) query coverage were selected, having not more than 2 (*i.e.* less than 3) nucleotides mismatch were filtered for further processing.

2.2.1. Elimination of Protein Coding Transcripts

From the transcripts giving positive hits, sequences 100 bp upstream and downstream of target sites were retrieved and exported in FASTA format. These sequences were used as a query in standalone BLASTx (v2.8.1) against non-redundant protein data-base while keeping e-value at 10, word size 6, scoring (match: 2, mismatch: -3), gap cost (open: 11, extend: 1), BLOSUM62 matrix and standard genetic code. From the output, sequences with no hits (*i.e.* non-coding sequences) were collected in FASTA as pre-miRNA (miRNA precursor) sequences.

2.2.2. Prediction of Pre-miRNA Secondary Structure

Mfold server (mfold.rit.albany.edu), RNA folding form was implemented for the prediction of pre-miRNA secondary structure to rule out pseudo pre-miRNAs. As input parameters, linear RNA sequence was selected, folding temperature was set at 37°C, ionic conditions of 1M NaCl without divalent ions was used, 5 percent as sub-optimality number, maximum interior/bulge loop size of 30, and energy dot plot was turned on while any other parameters were set at default.

2.2.3. Selection of Stable Pre-miRNA Structures

For each pre-miRNA, %G+C and A+U content, ΔG values (Minimum free energy, unit: -kcal/mol) of stem-loop structures, Adjusted Minimum Folding Energy (MFE of a 100-nucleotide length) and Minimum Free Energy Index (MFEI) were evaluated as mentioned below in equation 1.

$$AMFE = \frac{MFE}{\text{Length of precursor sequence (LP)}} \times 100 \quad (1)$$

The generated secondary structures were screened against the following criteria:

- 30–70% A+U content
- minimum folding free energy ≤ -40 kcal/mol
- no more than six nucleotide mismatches between the predicted mature miRNA and the opposite miRNA sequence (miRNA*)
- no more than 7 bases bulge size in the mature miRNA region.

2.3. Phylogenetic Analysis of *Agave Deserti* (Ade-) MiRNAs

Sequence homologues of identified ade-miRNAs were recognized by performing standalone BLASTN by using 10,414 mature miRNAs (from mirBase) as a target database. Criteria were selected for BLASTN: E-value 0.001, BLOSUM62 Matrix, scoring (match: 1, mismatch: -3), and filtered sequences for low-complexity regions.

After the identification of homologous miRNA sequences, corresponding pre-miRNA sequences were obtained from miRBase to be used for phylogenetic analysis in MEGA-X. Multiple sequence alignment was executed by using Clustal-w contained in MEGA-X with the following parameters: Gap Open Penalty 15, Gap Extension Penalty 6.66, DNA Weight Matrix IUB, kept off the negative matrix. Following the sequence alignment, phylogenetic trees were built subsequently by using Maximum likelihood as statistical method, bootstrap method as a test of phylogeny, bootstrap replication number as 1000, substitution model: kimura 2-parameter, uniform rates among sites and tree inference using Nearest-Neighbor-Interchange (NNI) [18, 19].

2.4. Prediction of Potential MiRNA Targets

For miRNA target prediction, psRNATarget [20] was used with predicted *A. deserti* miRNAs as query sequences, and target databases of *A. thaliana* DFCI Gene Index (AGI), and The Arabidopsis Information Resource (TAIR) transcript removed miRNA gene. Expectation value was set to 2.5, penalty for G:U pair 0.5, penalty for other mismatches 1, extra weight in seed region 1.5, seed region length 2-13 nucleotides, mismatches allowed in the seed region 2, length for complementary scoring (HSP score) 18, dis-allowed bulge (gap), target site accessibility with 25 Maximum UPE, 17 bp upstream and 13 bp downstream flanking length around target site and 10-11 nt as the range of central mismatch that will lead to translational inhibition.

2.5. Functional Annotation of Target Genes

For functional annotation of genes targeted by ade-miRNAs, Gene ontology prediction from QuickGO (version 2019-02-07, and Annotation set from 2019-01-28) [21] as well as Uniprot [22] and Interpro [23] were used as resources to identify putative functions of target genes. Biological process, molecular function and cellular component of

each gene targeted by ade-miRNAs were defined. Workflow for the whole process has been depicted in Fig. (1).

3. RESULTS AND DISCUSSION

3.1. Identification of MiRNA with *Agave Deserti* Transcriptome

Agave deserti transcriptome NGS data (Illumina HiSeq 2000) available from Gross *et al.* [16] was used to develop target database for nucleotide BLAST analysis and processed to generate targets that were further filtered by BLASTx analysis to curtail the hits from protein coding transcripts. The non-coding transcripts were further proceeded for pre-miRNA secondary structure analysis. Fig. (2) indicates pair-wise sequence alignment (obtained from BLASTN results) of selected miRNAs from various species in Viridiplantae category with respective sequences (final miRNA sequences found in the study) of *Agave deserti* transcriptome.

3.2. Analysis of Potential Pre-miRNAs Secondary Structure

To screen out the pseudo pre-miRNAs from the potential pre-miRNAs, %G+C content and %A+U content was estimated for each pre-miRNA and MFOLD algorithm [24] was used to predict minimum free energy ($-\Delta G$: kcal/mol) and secondary structures, and pre-miRNAs were identified based on a selected criteria. MFEI values of the selected miRNAs ranged from 0.5 to 1.0 (data shown in Table 1). MFEI for miRNAs is higher than other coding and non-coding RNA molecules (*i.e.* tRNAs, rRNAs and mRNAs) are lower than miRNAs [25]. Among the *A. deserti* miRNAs identified, all miRNAs have MFEI greater than 0.60 while miRNAs ade-miR167 and ade-miR11451 were 0.55 and 0.56, respectively. *A. deserti* miRNA with the MFEI reaching 0.9 and above include ade-miR393c-5p (0.92), ade-miR396-3p (0.9), ade-miR396b-3p (0.91), ade-miR396g-3p (0.91), ade-miR482b (0.94) and ade-miR482e-3p (1.01).

After pre-miRNA testing against the stringent criteria, 30 potential miRNAs were identified in *A. deserti* with loci, corresponding homologues, and other associated data (Table 1). The resultant 30 conserved miRNAs were identified and named as: ade-miR156, ade-miR156b-3p, ade-miR156c, ade-miR156g, ade-miR156x, ade-miR160f, ade-miR167, ade-miR167a-3p, ade-miR167b-3p, ade-miR168, ade-miR168a, ade-miR168b-3p, ade-miR169b-5p, ade-miR169c, ade-miR169l, ade-miR171h, ade-miR172a-5p, ade-miR172c-3p, ade-miR393c-5p, ade-miR396-3p, ade-miR396a-5p, ade-miR396b-3p, ade-miR396e-5p, ade-miR396g-3p, ade-miR398a-3p, ade-miR482b, ade-miR482e-3p, ade-miR528, ade-miR828 and ade-miR11451. The identified mature miRNAs belonged to 14 different categories of precursors *i.e.* MIR156, MIR160, MIR167, MIR168, MIR169, MIR171, MIR172, MIR393, MIR396, MIR398, MIR482, MIR528, MIR828 and MIR11451. Fig. (3) shows secondary structures of pre-miRNAs with miRNA region highlighted.

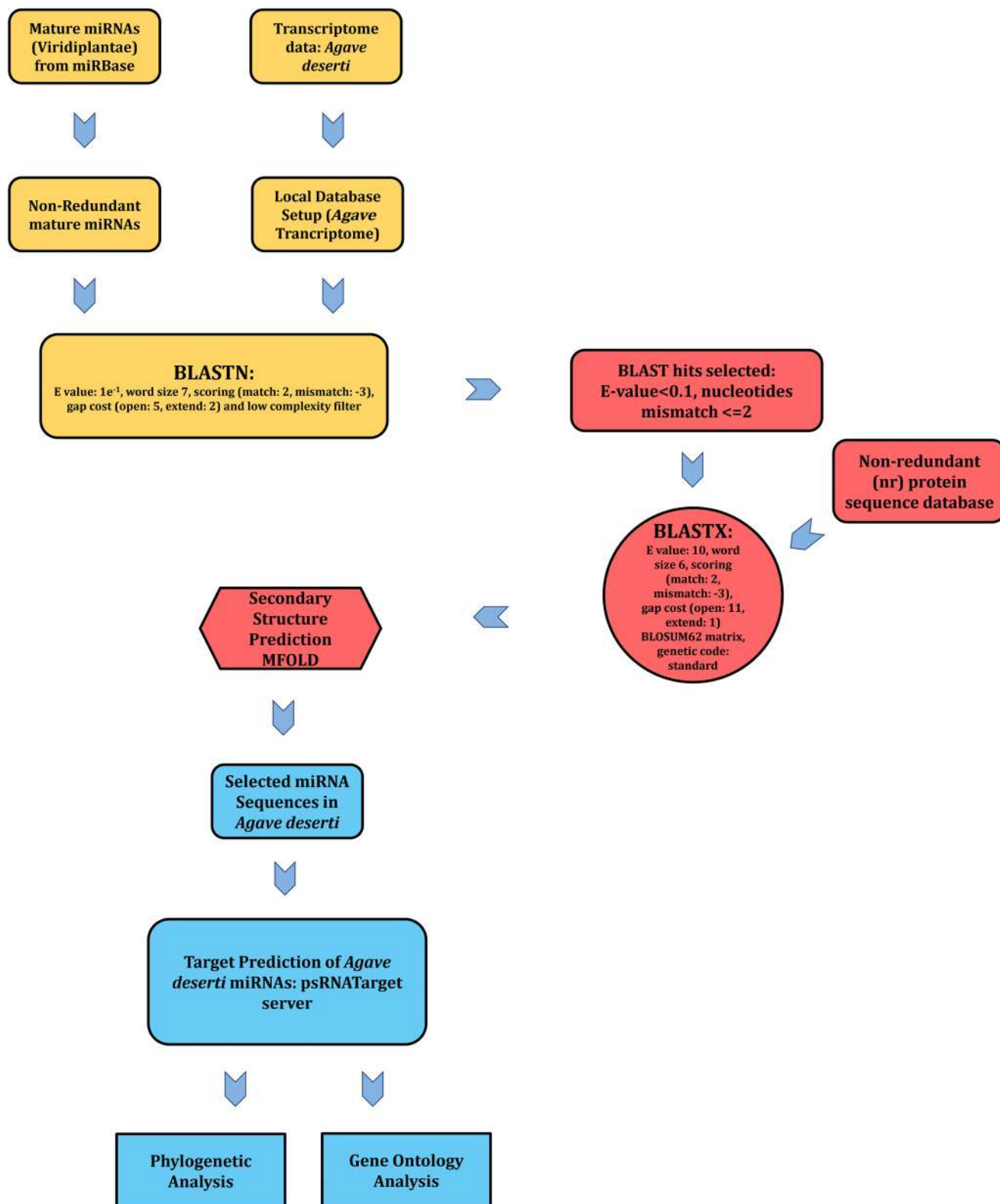


Fig. (1). Schematic workflow indicating the steps involved in miRNA identification from *Agave deserti* transcriptome data.

hbr-miR156	1	UUGACAGAAGAUAGAGAGC	19	fve-miR171h	1	UUGAGCCGCGUCAUAUCUCC	21
Locus38596v1rpk2.21	147	UUGACAGAAGAUAGAGAGC	165	Locus60820v1rpk0.61	285	UUGAGCCGCGUCAUAUCUCC	305
vca-miR156b-3p	1	GCUCUCUAUGCUCUGUCAUC	21	ata-miR172a-5p	1	GCGGCACCAUCAAGAUUCACA	21
Locus38596v1rpk2.21	214	GCUCUCUAUGCUCUGUCAUC	234	Locus51432v1rpk0.89	357	GCGGCACCAUCAAGAUUCACA	377
atr-miR156c	1	UUGACAGAAGAUAGAGAGCAC	21	stu-miR172c-3p	1	AGAAUCUUGAUGAUGCUGC	19
Locus38596v1rpk2.21	147	UUGACAGAAGAUAGAGAGCAC	167	Locus51432v1rpk0.89	413	AGAAUCUUGAUGAUGCUGC	431
gma-miR156g	1	ACAGAAGAUAGAGAGCACAG	20	zma-miR393c-5p	1	UCCAAAGGGAUCGCAUUGAUCU	22
Locus38596v1rpk2.21	150	ACAGAAGAUAGAGAGCACAG	169	Locus61622v1rpk0.59	164	UCCAAAGGGAUCGCAUUGAUCU	185
pab-miR156x	1	UGACAGGAGAUAGAGAGCAC	20	vca-miR396-3p	1	GUUCAUAAAGCUGUGGGAAA	21
Locus38596v1rpk2.21	148	UGACAGGAGAUAGAGAGCAC	167	Locus72271v1rpk0.44	136	GUUCAUAAAGCUGUGGGAAA	156
ppt-miR160f	1	UGCCUGGCUCUCCUGUAUGCCA	21	eun-miR396a-5p	1	UUCCACGGCUUUCUUGAACUG	21
Locus44839v1rpk1.34	448	UGCCUGGCUCUCCUGUAUGCCA	468	Locus72271v1rpk0.44	81	UUCCACGGCUUUCUUGAACUG	101
atr-miR167	1	UGAAGCUGCCAGCAUGAUCUG	21	osa-miR396b-3p	1	GUUCAUAAAGCUGUGGGAA	20
Locus38849v1rpk2.16	199	UGAAGCUGCCAGCAUGAUCUG	219	Locus72271v1rpk0.44	136	GUUCAUAAAGCUGUGGGAA	155
vca-miR167a-3p	1	CAGAUCAUGUGGAGUUUCAU	21	csi-miR396e-5p	1	UUCCACGGCUUUCUUGAACU	20
Locus38849v1rpk2.16	252	CAGAUCAUGUGGAGUUUCAU	272	Locus72271v1rpk0.44	81	UUCCACGGCUUUCUUGAACU	100
vca-miR167b-3p	1	AGAUCAUUGUAGAGUUUCAUC	21	zma-miR396g-3p	1	GUUCAAGAAAGCUGUGGAAGA	21
Locus38849v1rpk2.16	253	AGAUCAUUGUAGAGUUUCAUC	273	Locus72271v1rpk0.44	99	GUUCAAGAAAGCUGUGGAAGA	79
cme-miR168	1	UCGCUUGGUGCAGGUCGGGA	20	ath-miR398a-3p	1	UGUGUUCUCAGGUCACCCCUU	21
Locus49688v1rpk0.98	153	UCGCUUGGUGCAGGUCGGGA	172	Locus1256v1rpk204.51	80	UGUGUUCUCAGGUCACCCCUU	100
pab-miR168a	1	UCGCUUGGUGCAGGUCGGGAA	21	aof-miR482b	1	UUUCCAAAACCUCCAUUCCGA	22
Locus49688v1rpk0.98	153	UCGCUUGGUGCAGGUCGGGAA	173	Locus48842v1rpk1.03	118	UUUCCAAUACCUCCAUUCCGA	139
vca-miR168b-3p	1	CCCGCCUUGCACCAACUGAAU	21	csi-miR482e-3p	1	UUGCCAACUCCUCCAUUGCCGA	22
Locus49688v1rpk0.98	208	CCCGCCUUGCACCAACUGAAU	228	Locus48842v1rpk1.03	286	UUCCCAACCCUCCAUUGCCGA	307
ata-miR169b-5p	1	UAGCCAAGGAUGAUUUGCCUGUG	23	sbi-miR528	1	UGGAAGGGGCAUGCAGAGGAG	21
Locus71563v1rpk0.45	756	UAGCCAAGGAUGACUUGCCUGUG	734	Locus20212v1rpk10.84	43	UGGAAGGGGCAUGCAGAGGAG	63
atr-miR169c	1	UAGCCAAGGAUGACUUGCCU	20	ath-miR828	1	UCUUGCUUAAAUGAGUAUCCA	22
Locus71563v1rpk0.45	756	UAGCCAAGGAUGACUUGCCU	737	Locus47984v1rpk1.08	28	UCUUGCUUAAAUGGGUGUCCA	49
ath-miR169l	1	UAGCCAAGGAUGACUUGCCUG	21	pab-miR11451	1	CAAUUACUAUGUGGAUUGCU	20
Locus71563v1rpk0.45	756	UAGCCAAGGAUGACUUGCCUG	736	Locus59898v1rpk0.63	490	CAAUUACUAUUGGAUUGCU	471

Fig. (2). Pair-wise sequence alignment of selected miRNAs from various species in Viridiplantae category with respective sequences of *Agave deserti* transcriptome. (A higher resolution/colour version of this figure is available in the electronic copy of the article).

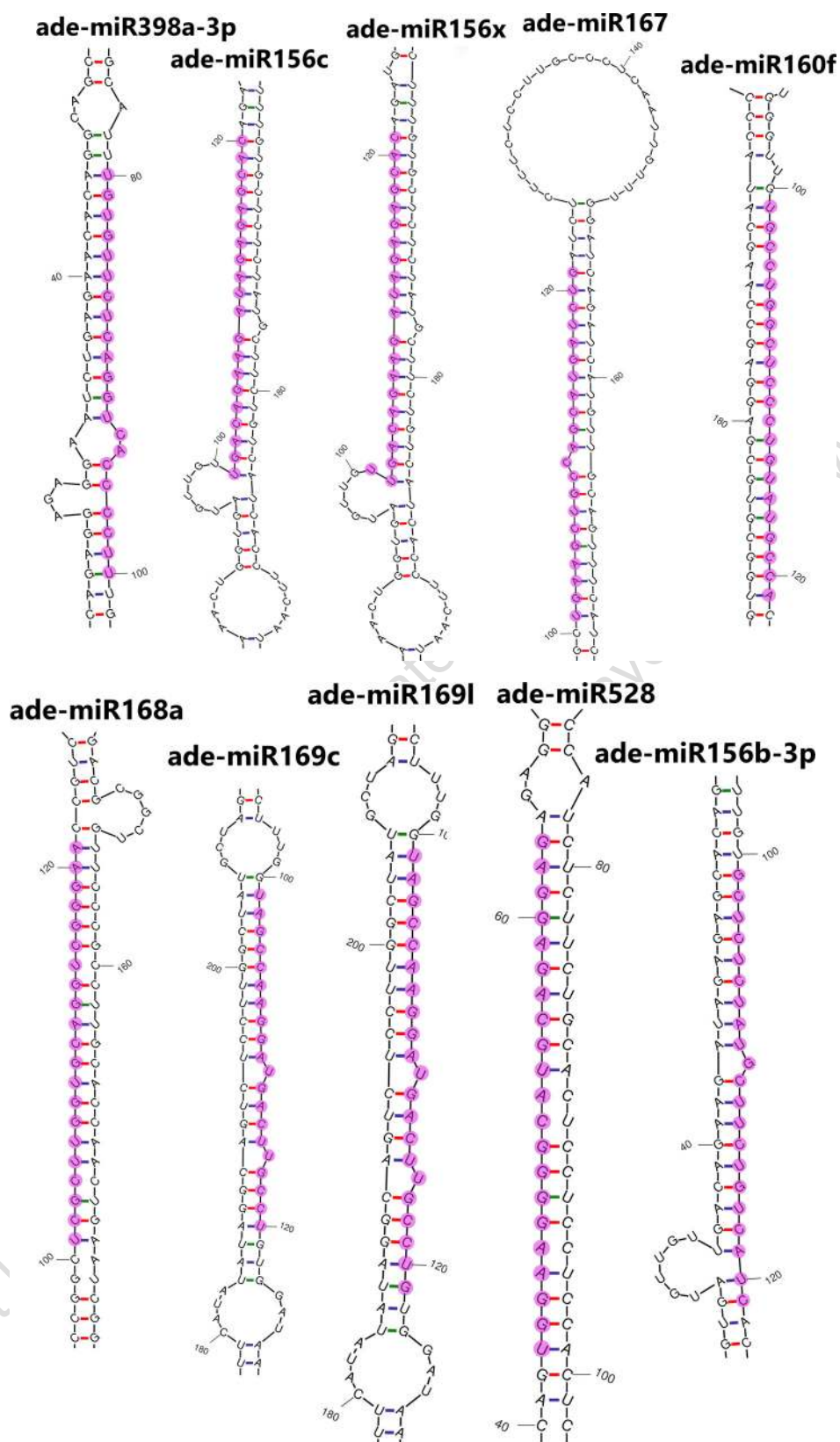
Table 1. miRNAs identified in *Agave deserti* are shown; also mature miRNA sequences, homologue miRNAs, loci, % GC content and % AU content of pre-miRNAs and other relevant details are indicated.

miRNA	miRNA Sequence	Homologue	Locus	% GC (pre-miRNA)	% AU (pre-miRNA)	miRNA Length	Length (pre-miRNA)	MFE	AMFE	MFEI	Hit start	Hit end	Strand
ade-miR156	UUGACAGAAGA-UAGAGAGC	hbr-miR156	Locus38596v1rpkm2.21	41	59	19	219	-78.1	35.7	0.87	147	165	+
ade-miR156b-3p	GCUCUCUAUGCU-UCUGUCAUC	vca-miR156b-3p	Locus38596v1rpkm2.21	44	56	21	221	-78.6	35.6	0.81	214	234	+
ade-miR156c	UUGACAGAAGAU-AGAGAGC	atr-miR156c	Locus38596v1rpkm2.21	41	59	21	221	-78.2	35.4	0.86	147	167	+
ade-miR156g	ACAGAAGAUAGA-GAGCACAG	gma-miR156g	Locus38596v1rpkm2.21	40	60	20	220	-78.1	35.5	0.89	150	169	+
ade-miR156x	UGACAGAAGAUAGAGAGC	pab-miR156x	Locus38596v1rpkm2.21	40	60	20	220	-77.8	35.4	0.88	148	167	+
ade-miR160f	UGCCUGGCUCC-CUGAUGCCA	ppt-miR160f	Locus44839v1rpkm1.34	50	50	21	221	-78.4	35.5	0.71	448	468	+
ade-miR167	UGAAGCUGCCAGCAUGAUCUG	atr-miR167	Locus38849v1rpkm2.16	41	59	21	221	-50.1	22.7	0.55	199	219	+
ade-miR167a-3p	CAGAUCAUGUUGCAGUUU-CAU	vca-miR167a-3p	Locus38849v1rpkm2.16	43	57	21	221	-64.9	29.4	0.68	252	272	+
ade-miR167b-3p	AGAUAUGUUGCAGUUU-CAUC	vca-miR167b-3p	Locus38849v1rpkm2.16	43	57	21	221	-66.9	30.3	0.70	253	273	+
ade-miR168	UCGCUUGGUGCA-GGUCGGGA	cme-miR168	Locus49688v1rpkm0.98	55	45	20	220	-89.9	40.9	0.74	153	172	+
ade-miR168a	UCGCUUGGUGCAGGUCGG-GAA	pab-miR168a	Locus49688v1rpkm0.98	55	45	21	221	-89.9	40.7	0.74	153	173	+
ade-miR168b-3p	CCCGCCUUGCACCACU-GAAU	vca-miR168b-3p	Locus49688v1rpkm0.98	57	43	21	221	-92.1	41.7	0.73	208	228	+
ade-miR169b-5p	UAGCCAAGGAU-GACUUGCCUGUG	ata-miR169b-5p	Locus71563v1rpkm0.45	41	59	23	223	-80.2	36.0	0.88	756	734	-
ade-miR169c	UAGCCAAGGAU-GACUUGCCU	atr-miR169c	Locus71563v1rpkm0.45	41	59	20	220	-76.6	34.8	0.85	756	737	-
ade-miR169l	UAGCCAAGGAUACUUGC-CUG	ath-miR169l	Locus71563v1rpkm0.45	41	59	21	221	-76.6	34.7	0.85	756	736	-
ade-miR171h	UUGAGCCGGGCUAAUUCUCC	fve-miR171h	Locus60820v1rpkm0.61	46	54	21	221	-71.3	32.3	0.70	285	305	+
ade-miR172a-5p	GCGGCACCAUAAGAUAU-CACA	ata-miR172a-5p	Locus51432v1rpkm0.89	41	59	21	221	-59.0	26.7	0.65	357	377	+
ade-miR172c-3p	AGAAUCUUGA-UGAUGCUG	stu-miR172c-3p	Locus51432v1rpkm0.89	45	55	19	219	-66.6	30.4	0.68	413	431	+
ade-miR393c-5p	UCCAAAGGGAUCGCAUUGAUCU	zma-miR393c-5p	Locus61622v1rpkm0.59	43	57	22	222	-87.8	39.5	0.92	164	185	+
ade-miR396-3p	GUUCAAAAGCUGUGG-GAAA	vca-miR396-3p	Locus72271v1rpkm0.44	38	62	21	221	-76.0	34.4	0.90	136	156	+
ade-miR396a-5p	UUCCACAGCUU-UCUUGAACUG	eun-miR396a-5p	Locus72271v1rpkm0.44	40	60	21	201	-66.3	33.0	0.82	81	101	+
ade-miR396b-3p	GUUCAAAAGCUGUGGGA	osa-miR396b-3p	Locus72271v1rpkm0.44	38	62	20	220	-75.9	34.5	0.91	136	155	+
ade-miR396e-5p	UUCCACAGCUU-UCUUGAACU	csi-miR396e-5p	Locus72271v1rpkm0.44	40	60	20	200	-66.3	33.2	0.83	81	100	+
ade-miR396g-3p	GUUCAAGAAAGCUGUG-GAAGA	zma-miR396g-3p	Locus72271v1rpkm0.44	40	60	21	199	-72.3	36.3	0.91	99	79	-
ade-miR398a-3p	UGUGUUCUCAGGUCACCC-CUU	ath-miR398a-3p	Locus1256v1rpkm204.51	36	64	21	200	-50.1	25.1	0.70	80	100	+
ade-miR482b	UUUCCAAUACCUCCAUUC-CGA	aof-miR482b	Locus48842v1rpkm1.03	39	61	22	222	-81.6	36.7	0.94	118	139	+
ade-miR482e-3p	UUCCCAACCCUCCCAUGC-CGA	csi-miR482e-3p	Locus48842v1rpkm1.03	41	59	22	222	-91.8	41.4	1.01	286	307	+
ade-miR528	UGGAAGGGGCAUGCAGAG-GAG	sbi-miR528	Locus20212v1rpkm10.84	60	40	21	163	-63.3	38.8	0.65	43	63	+
ade-miR828	UCUUGCUUAAAUGGGUGU-UCCA	ath-miR828	Locus47984v1rpkm1.08	42	58	22	149	-47.9	32.1	0.77	28	49	+
ade-miR11451	CAAUUACUAUA-UGGAUUGCU	pab-miR11451	Locus59898v1rpkm0.63	34	66	20	220	-41.9	19.1	0.56	490	471	-

3.3. Phylogenetic Analysis of Predicted Sequences

Corresponding pre-miRNA sequences from the homologous miRNA sequences were retrieved from miRBase and used to build multiple sequence alignment and phylogenetic trees in MEGA-X [19]. The multiple sequence alignment (Fig. 4a-n) indicated that the mature miRNAs can be highly conserved among different species; this observations point

towards the fact that they could be regulating same/similar genes in different species and phylogenetic analysis revealed the possible evolutionary pathways that could generate orthologous pre-miRNAs over the course of evolution [26]; also, certain miRNAs emerging at different rates, could emanate from a common evolutionary ancestor [27, 28].

**Fig. (3).** contd....

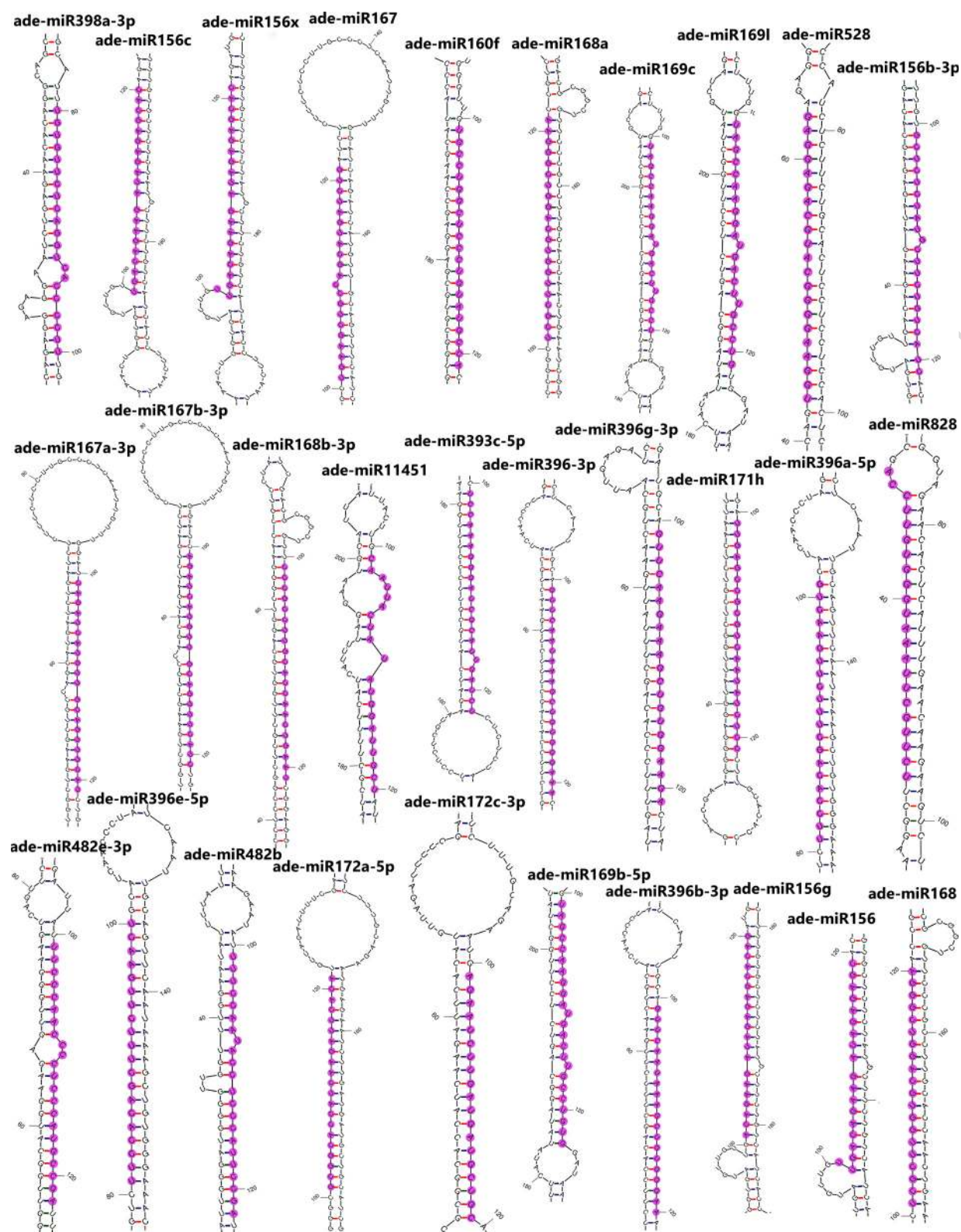


Fig. (3). Secondary structures of potential pre-miRNAs of *Agave deserti*. Mature miRNA sequence within the pre-miRNA backbone are highlighted in magenta (labelled mature miRNAs). (A higher resolution/colour version of this figure is available in the electronic copy of the article).

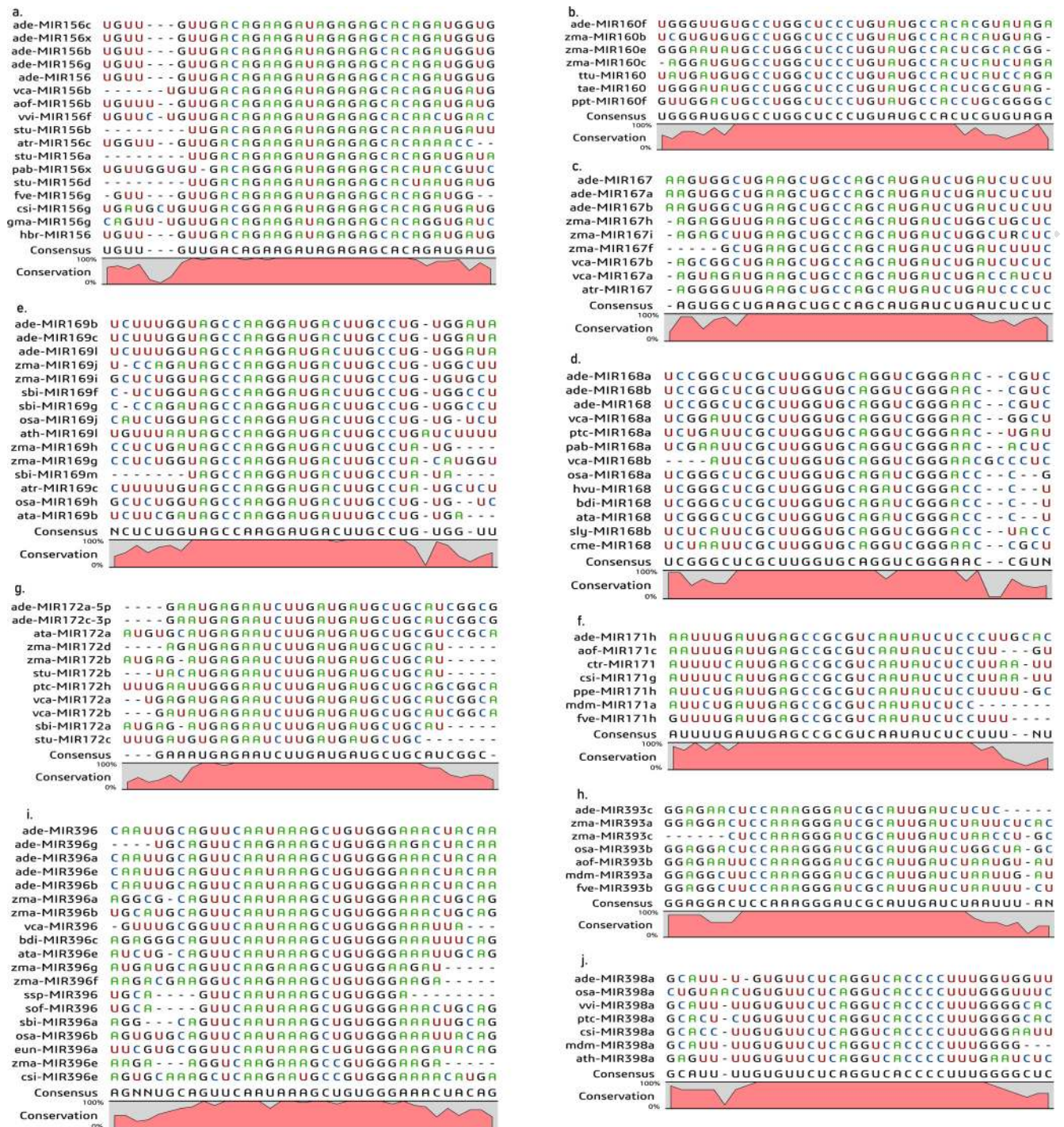


Fig. (4). contd....

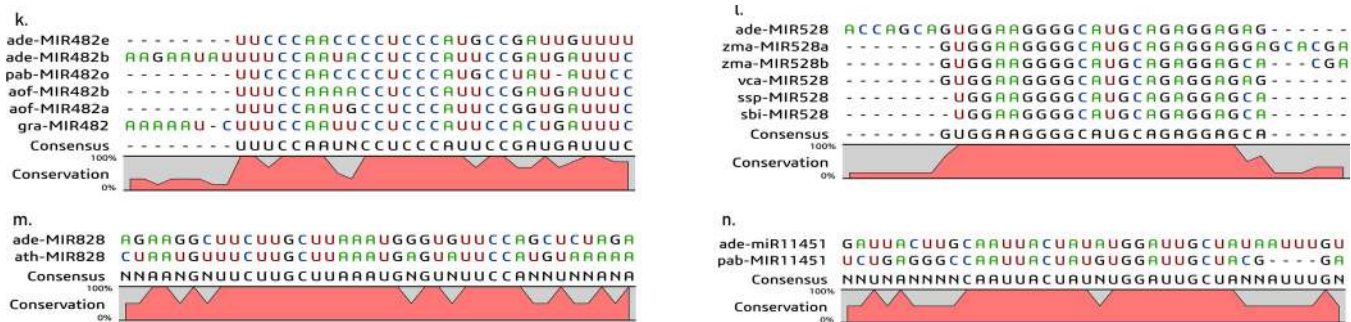


Fig. (4). Clustal-w Multiple Sequence Alignments of pre-miRNA sequences from *Agave deserti* and pre-miRNA sequences of corresponding homologue miRNAs identified. **a)** ade-MIR156, ade-MIR156b, ade-MIR156c, ade-MIR156g, ade-MIR156x. **b)** ade-MIR160f. **c)** ade-MIR167, ade-MIR167a, ade-MIR167b. **d)** ade-MIR168, ade-MIR168a, ade-MIR168b. **e)** ade-MIR169b, ade-MIR169c, ade-MIR169l. **f)** ade-MIR171h. **g)** ade-MIR-172a, ade-MIR172c. **h)** ade-MIR393c. **i)** ade-MIR396, ade-MIR396a, ade-MIR396b, ade-MIR396e, ade-MIR396g. **j)** ade-MIR398a. **k)** ade-MIR482b, ade-MIR482e. **l)** ade-MIR528. **m)** ade-MIR828. **n)** ade-MIR11451. Conserved sites (indicating region of existence of mature miRNA) are shown as asterisk. (A higher resolution/colour version of this figure is available in the electronic copy of the article).

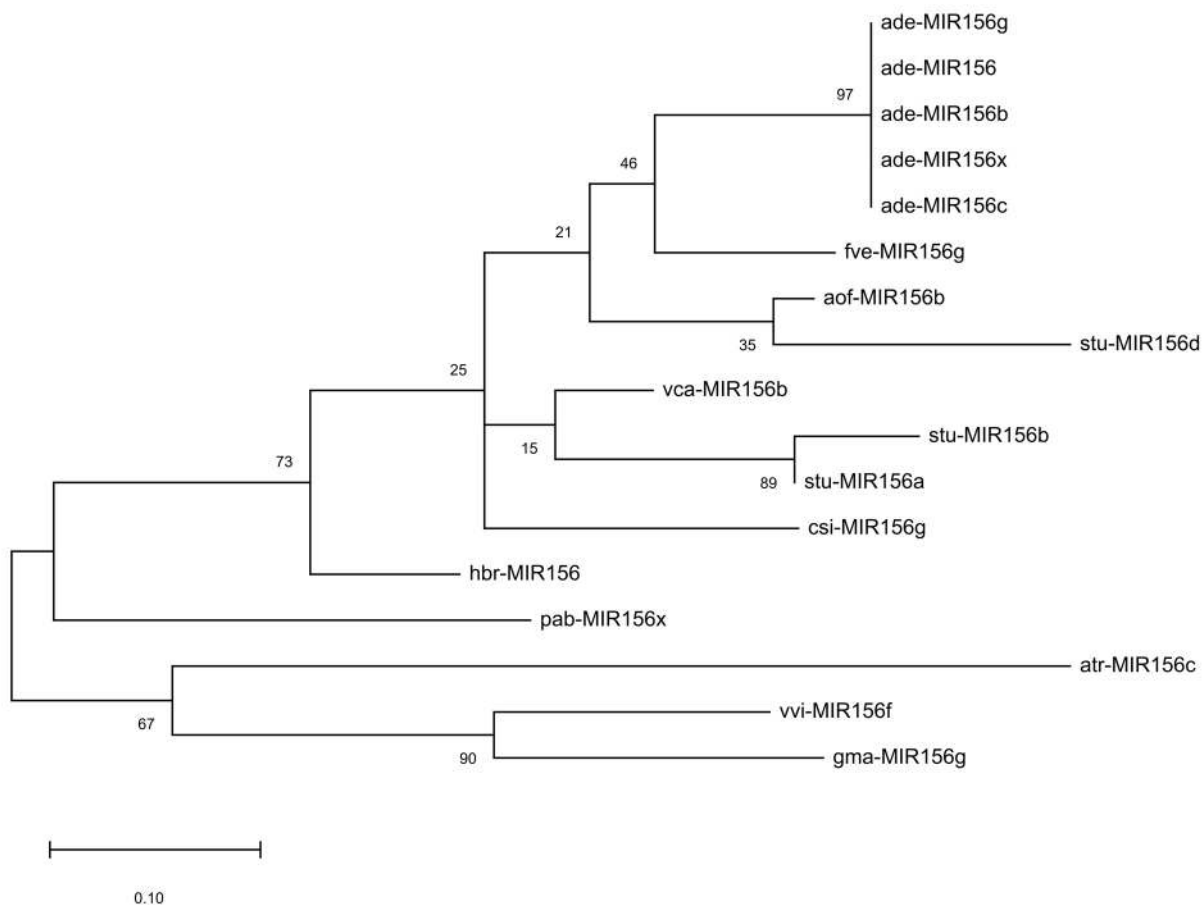


Fig. (5). Phylogenetic Tree of pre-miRNA sequences ade-MIR156, ade-MIR156b, ade-MIR156c, ade-MIR156g and ade-MIR156x from *Agave deserti* and pre-miRNA sequences of identified homologue miRNAs.

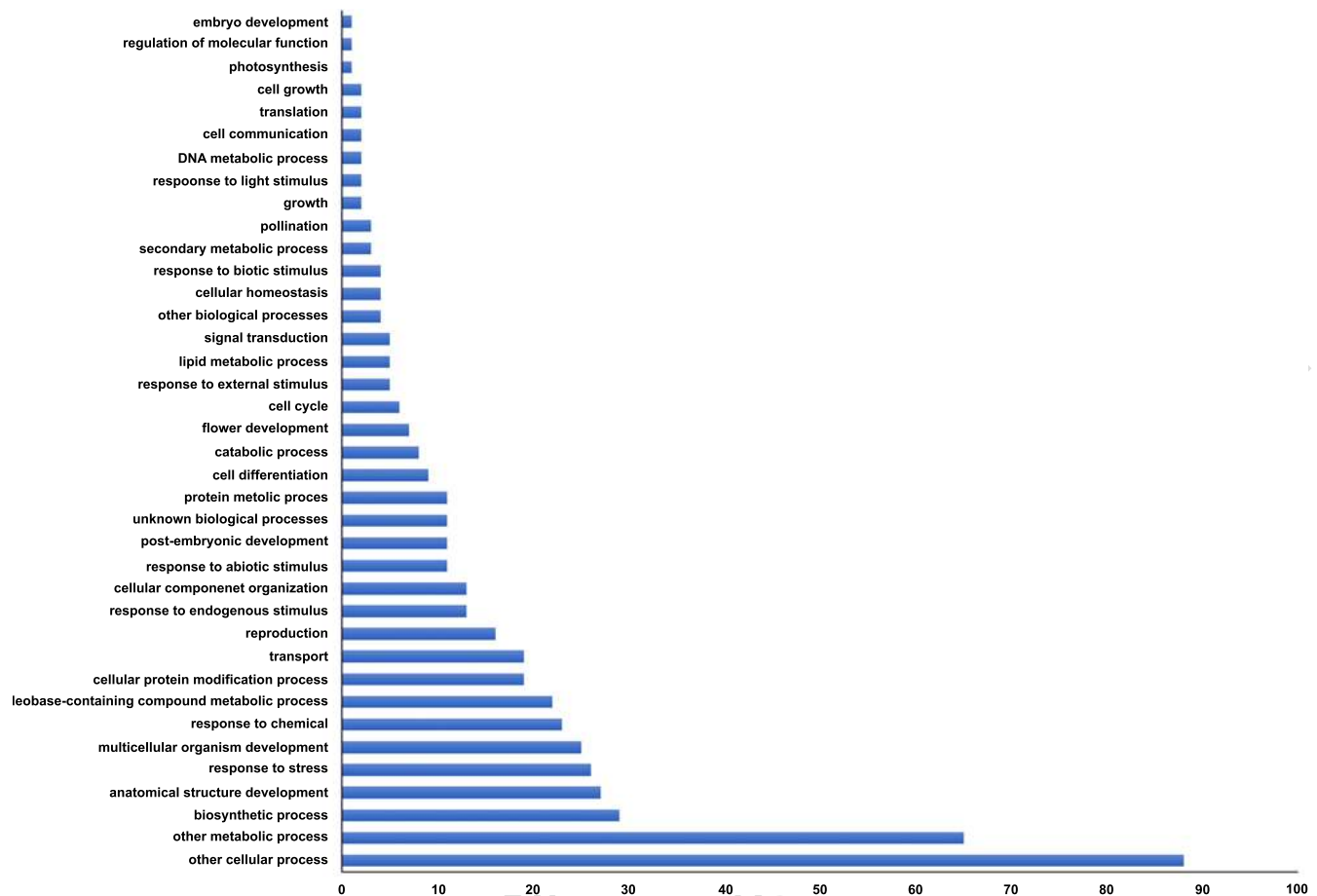


Fig. (6). Clustered bar chart showing gene ontology classification of biological processes of the target loci. (A higher resolution/colour version of this figure is available in the electronic copy of the article).

Phylogenetic analysis revealed an evolutionary relationship between ade-pre-miRNAs and the pre-miRNA sequences from other plant species having miRNA homologous to ade-miRNAs. It was noticed that *Agave deserti* pre-miRNAs have additional members in certain categories of pre-miRNAs, e.g. pre-miRNAs MIR156, MIR167, MIR168, MIR169 and MIR396 all of which had at least three members of *Agave deserti* pre-miRNAs. In the phylogenetic tree of MIR156 (with five *A. deserti* pre-miRNAs), ade-MIR156, ade-MIR156b, ade-MIR156c, ade-MIR156g and ade-MIR156x were found to be more closely related to fve-MIR156g (*Fragaria vesca*), aof-MIR156b (*Asparagus officinalis*) and stu-MIR156d (*Solanum tuberosum*) (Fig. 5). Pre-miRNAs ade-MIR167a, ade-MIR167b and ade-MIR167 were found closer to the pre-miRNAs vca-MIR167a and vca-MIR167f (*Vriesea carinata*) (Supplementary Fig. S2). It was observed that in the tree of MIR168, ade-MIR168, ade-MIR168a and ade-MIR168b were closely linked to vca-MIR168b and vca-MIR168a (Supplementary Fig. S3). In the tree of MIR169: MIR169b, ade-MIR169b, ade-MIR169c and ade-MIR169l were more closely related to sbi-MIR169m (*Sorghum bicolor*), zma-MIR169i and ata-MIR169b (*Aegilops tauschii*) (Supplementary Fig. S4). The

output for MIR396 family showed that ade-MIR396g was in the closest neighborhood with zma-MIR396g while ade-MIRs 396, 396a, 396e and 396b were in close neighborhood with vca-MIR396 and osa-MIR396b (*Oryza sativa*) (Supplementary Fig. S8). Phylogenetic trees of other pre-miRNAs are also attached as Supplementary Fig. S1, S5-7 and S9-11).

3.4. miRNA Targets and Functional Annotation

Complementarity of miRNA with the target gene enables prediction of miRNA targets in genome. *A. thaliana* (model) genes (from AGI and TAIR) were used for miRNA target identification with user-defined parameters in psRNA-Target server [20]. Target genes identified together with their functional annotation (Gene Ontology) are given in Figs. (6-8) and Table S1. Genes regulated by miRNAs are involved in diverse biological processes, molecular functions and different sub-cellular localization. Conservation of miRNAs across plant species is an implication towards conservation of their functions e.g. in development and differentiation processes, hormone signaling and stress responses [7, 29].

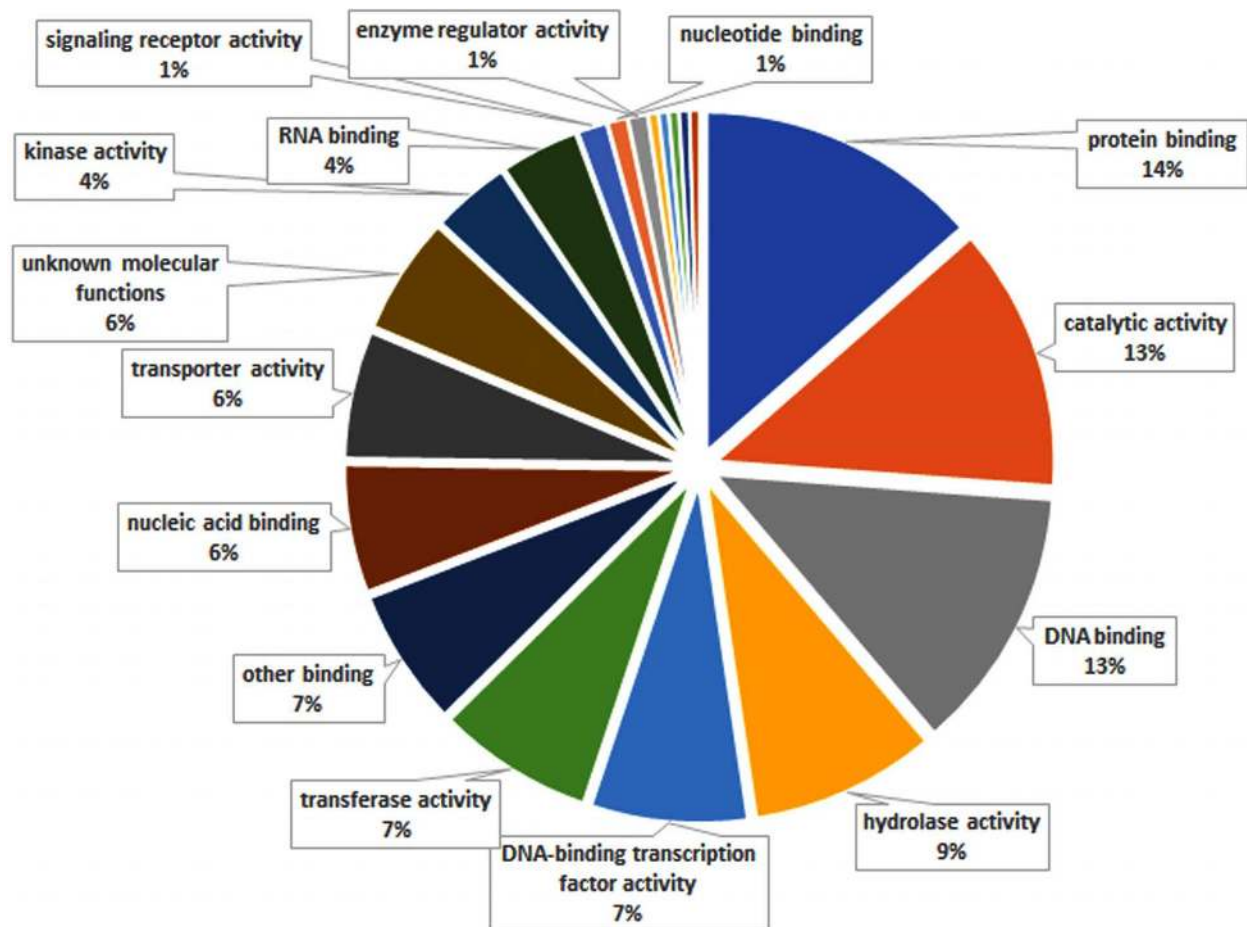


Fig. (7). Pie chart showing gene ontology classification of molecular functions of the target loci. (A higher resolution/colour version of this figure is available in the electronic copy of the article).

3.4.1. Transcription Factors/ Co-Factors

Many miRNAs predicted in *A. deserti* targeted loci coding for transcription factors/ co-factors or DNA binding proteins involved in transcriptional regulation. Gene coding for Squamosa Promoter-binding protein (SPL10) was found to be a common target of miRNAs belonging to *A. deserti* MI156 family i.e. ade-miR156, ade-miR156c, ade-miR156g and ade-miR156x. Plant specific transcription factors encoded by Squamosa Promoter binding protein-Like (SPL) genes are important for phase transition, flower and fruit development, hormone signaling, sporogenesis and response to toxins [30, 31]. Likewise, other genes encoding for proteins having transcription factor activity were found to be targeted by different miRNAs such as: Auxin response factor-ARF17 (ade-miR160f), GRAS family transcription factor-HAM1/scarecrow-like protein27 (ade-miR171h), Nuclear Factor-YA8 (ade-miR169b-5p, ade-miR169c and ade-miR169l), polymerase associated factor- PAF1 (ade-miR172a-5p), transcription factor RAX3 (ade-miR172c-3p), transcription factor- MYB68 (ade-miR172c-3p) and RING/U-box superfamily protein (ade-miR11451). From the Gene Ontology results, these transcription factors were

indicated to be associated mainly with biological functions of plant growth and development, response to hormones and stimuli, and signaling pathways (Table S1).

3.4.2. Nucleotide (ATP/ADP/GTP) Binding Proteins

ATP binding proteins have diverse functions and are engaged in performing specific functions related to normal functioning of the cells. Examples of such genes targeted by *A. deserti* miRNA includes: LRR receptor-like serine/threonine-protein kinase (ade-miR156 and ade-miR156c), AHA11, H(+)-ATPase 11 (ade-miR156b-3p), Phragmoplast orienting kinesin 2 (ade-miR167a-3p), P-loop containing nucleoside triphosphate hydrolase (ade-miR167a-3p) and Receptor-like kinase (ade-miR396a-5p and ade-miR396e-5p).

3.4.3. Disease Resistance/ Response Proteins

Disease resistance proteins have ATP binding activity and function in signal transduction and defense response. Micro-RNA ade-miR482b targeted the locus coding for disease resistance protein of TIR-NBS-LRR class while and CC-NBS-LRR class protein gene is targeted by ade-miR482b and ade-miR482e-3p. Zhu, *et al.* [32] showed that the release of

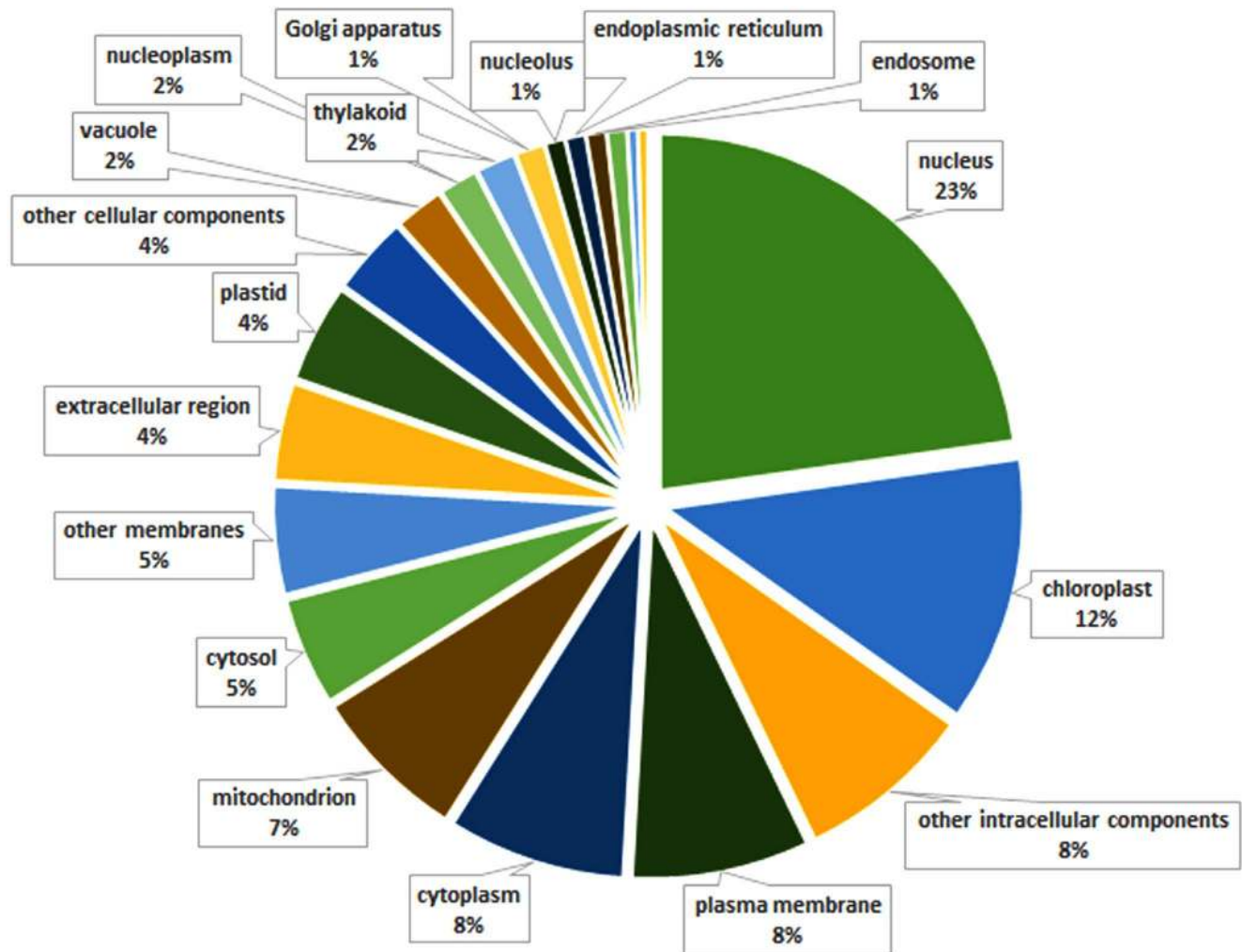


Fig. (8). Pie chart showing gene ontology classification of cellular components of the target loci. (A higher resolution/colour version of this figure is available in the electronic copy of the article).

the suppression of miRNA (miRN482 family)-mediated gene silencing pathway upon fungal attack triggers the expression of NBS-LRR genes in cotton to stimulate the defense against disease. Probably, L-type lectin-domain containing receptor kinase I.6 plays important roles in defense response to bacterium, defense response to oomycetes and protein phosphorylation; the gene for this protein is targeted by ade-miR482b. Concanavalin A-like lectin protein kinase family gene is the target of ade-miR169c and ade-miR169l; this protein helps in defense response to bacterium and oomycetes, and takes part in protein phosphorylation.

3.4.4. Stress Response Proteins

DEA(D/H)-box RNA helicase family gene STRS2 is targeted by ade-miR156, ade-miR156c and ade-miR156x; and the gene has a potential role in the regulation of tolerance to environmental stresses (cold, salinity and drought) [33]. RNA helicases of the DEAD box family exists in all eukary-

otic cells, and are needed for RNA metabolism (from transcription to degradation) and hence play an important role in gene expression. They use ATP to carry out unwinding of short duplex RNA in a unique fashion, and also act as ATP-dependent RNA clamps to provide nucleation centers that establish larger RNA-protein complexes [34].

Other important genes targeted by predicted ade-miRNAs include glutathione transferase (ade-miR168 and ade-miR168a), gene for heat shock protein DnaJ (ade-miR11451), Tocopherol cyclase VTE1 gene (ade-miR528), gene for phloem protein 2-A8 (ade-miR396g-3p), gene encoding Probable cysteine protease RD21B (ade-miR396a-5p and ade-miR396e-5p), Auxin signaling F-box 3 gene (ade-miR393c-5p), phosphoinositide 4-kinase gamma 4 gene (ade-miR172c-3p), NDR1/HIN1-like protein 6 gene (ade-miR172c-3p), arginine decarboxylase 1 gene (ade-miR172a-5p) and BTB and TAZ domain protein 4 gene (ade-miR156b-3p).

3.4.5. Translation Factors

Elongation factor 1- α promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis [35]. From our analysis, Elongation factor 1- α gene was targeted by ade-miR156g, while another gene coding for elongation factor 1- α was predicted target of ade-miR482e-3p. Eukaryotic initiation factor 4A-III (EIF4A-III) gene helps in mRNA processing and transport; this gene was predicted to be targeted by ade-miR172a-5p. Moreover, a gene regulating the process of the translation and functioning in mitochondrial protein synthesis, small GTP-binding protein gene, was targeted by ade-miR171h.

3.4.6. Photosynthetic Proteins

Photosystem II reaction center PSB28 gene was targeted by ade-miR156 and ade-miR156c. Psb28 is reported to have several roles, including guiding PSII biogenesis and assembly, stabilizing dimeric PSII, and facilitating PSII repair after photo-inhibition [36, 37]. Micro-RNAs ade-miR156b-3p was predicted to target gene encoding accumulation of photosystem I (DUF794) protein gene, which plays an important role in chloroplast organization and facilitates photosynthesis [38].

3.4.7. Other Proteins

A diversity of other protein coding genes participating in vital cell processes were also predicted as potential targets. Floral homeotic protein APETALA 2, which participates in floral organ development [39]; this gene was targeted by ade-miR172c-3p. In *A. thaliana*, miRNA172 is involved in the regulation of flowering time by down-regulation of AP2-like target genes and when over-expressed, caused disruption in the specification of floral organs' identity [40].

Embryo-specific protein 3 gene having biological role in embryo development, ending in seed dormancy, was the target of ade-miR172c-3p. ALC-interacting protein 1 gene was targeted by ade-miR396a-5p and ade-miR396e-5p functions in protein binding and has a biological role in fruit dehiscence. Likewise, copper chaperone for SOD1 gene which participates in cellular copper ion homeostasis and removal of superoxide radicals is targeted by ade-miR398a-3p.

CONCLUSION

Contemporary methods involving high-throughput sequencing technologies generate huge amount of genomic/transcriptomic data, which act as a rich resource for prediction, and analysis of miRNAs in plant species. Agave species have tremendous potential for commercial applications such as food, fiber, ornamentals and insecticides.

The target of this study was to identify possible miRNAs encoded in *Agave deserti* by using a computational approach that harnesses the available next generation sequencing (RNA-seq) data on its transcriptome. A local database of *A. deserti* transcriptome was created with 128,869 contigs followed by standalone Blast-N against 6,028 non-redundant

plant miRNAs retrieved from miRBase with defined parameters. Protein coding transcripts were removed from the target dataset using standalone BLASTx followed by the selection of stable pre-miRNA structures by the implementation of stringent criteria keeping in consideration %G+C content, ΔG values (minimum free energy) kcal/mol of stem-loop structures, and mismatches and bulge size in the mature miRNA region. The resultant 30 conserved miRNAs were identified named: ade-miR156, ade-miR156b-3p, ade-miR156c, ade-miR156g, ade-miR156x, ade-miR160f, ade-miR167, ade-miR167a-3p, ade-miR167b-3p, ade-miR168, ade-miR168a, ade-miR168b-3p, ade-miR169b-5p, ade-miR169c, ade-miR169l, ade-miR171h, ade-miR172a-5p, ade-miR172c-3p, ade-miR393c-5p, ade-miR396-3p, ade-miR396a-5p, ade-miR396b-3p, ade-miR396e-5p, ade-miR396g-3p, ade-miR398a-3p, ade-miR482b, ade-miR482e-3p, ade-miR528, ade-miR828 and ade-miR11451. Further, multiple sequence alignment and phylogenetic trees gained insights into the evolutionary relationship between the identified pre-miRNAs and homologue pre-miRNAs of other species. Finally, we predicted the targets of identified miRNAs and identified the target gene functions by Gene Ontology analysis.

Scrutinizing the framework of miRNA loci in the species of interest is a rudimentary step that can further unleash ways to improve productivity, and for enhancing biomass and stress tolerance in crops. To the best of our knowledge, this is the first study that reports micro-RNAs of *A. deserti* based on the transcriptome data. Further validation through wet-lab experimentation and functional analyses, can lead to understand the role of miRNAs and regulatory functions.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIAL

The data that support the findings of this study are available within the manuscript.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

- [1] Iwakawa, H.O.; Tomari, Y. The functions of microRNAs: mRNA decay and translational repression. *Trends Cell Biol.*, **2015**, *25*(11), 651-665.
<http://dx.doi.org/10.1016/j.tcb.2015.07.011> PMID: 26437588
- [2] Beauclair, L.; Yu, A.; Bouché, N. microRNA-directed cleavage and translational repression of the copper chaperone for superoxide dismutase mRNA in Arabidopsis. *Plant J.*, **2010**, *62*(3), 454-462.
<http://dx.doi.org/10.1111/j.1365-313X.2010.04162.x> PMID: 20128885
- [3] Park, W.; Li, J.; Song, R.; Messing, J.; Chen, X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr. Biol.*, **2002**, *12*(17), 1484-1495.
[http://dx.doi.org/10.1016/S0960-9822\(02\)01017-5](http://dx.doi.org/10.1016/S0960-9822(02)01017-5) PMID: 12225663
- [4] Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *cell*, **2004**, *2*(116), 281-297.
- [5] He, L.; Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.*, **2004**, *5*(7), 522-531.
<http://dx.doi.org/10.1038/nrg1379> PMID: 15211354
- [6] Jeong, D-H.; Park, S.; Zhai, J.; Gurazada, S.G.R.; De Paoli, E.; Meyers, B.C.; Green, P.J. Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell*, **2011**, *23*(12), 4185-4207.
<http://dx.doi.org/10.1105/tpc.111.089045> PMID: 22158467
- [7] Guo, H-S.; Xie, Q.; Fei, J-F.; Chua, N-H. MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *Plant Cell*, **2005**, *17*(5), 1376-1386.
<http://dx.doi.org/10.1105/tpc.105.030841> PMID: 15829603
- [8] Ling, L-Z.; Zhang, S-D.; Zhao, F.; Yang, J-L.; Song, W-H.; Guan, S-M.; Li, X-S.; Huang, Z-J.; Cheng, L. Transcriptome-wide identification and prediction of mirnas and their targets in paris polyphylla var. Yunnanensis by high-throughput sequencing analysis. *Int. J. Mol. Sci.*, **2017**, *18*(1), 219.
<http://dx.doi.org/10.3390/ijms18010219> PMID: 28117746
- [9] Zhang, B.; Pan, X.; Cobb, G.P.; Anderson, T.A. Plant microRNA: a small regulatory molecule with big impact. *Dev. Biol.*, **2006**, *289*(1), 3-16.
<http://dx.doi.org/10.1016/j.ydbio.2005.10.036> PMID: 16325172
- [10] Lee, R.C.; Feinbaum, R.L.; Ambros, V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *cell*, **1993**, *75*(5), 843-854.
- [11] Reinhart, B.J.; Slack, F.J.; Basson, M.; Pasquinelli, A.E.; Bettinger, J.C.; Rougvie, A.E.; Horvitz, H.R.; Ruvkun, G. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *nature*, **2000**, *403*(6772), 901.
- [12] Berezikov, E.; Cuppen, E.; Plasterk, R.H. Approaches to microRNA discovery. *Nat. Genet.*, **2006**, *38*(Suppl.), S2-S7.
<http://dx.doi.org/10.1038/ng1794> PMID: 16736019
- [13] Wu, F.; Shu, J.; Jin, W. Identification and validation of miRNAs associated with the resistance of maize (*Zea mays* L.) to Exserohilum turcicum. *PLoS One*, **2014**, *9*(1), e87251.
<http://dx.doi.org/10.1371/journal.pone.0087251> PMID: 24489881
- [14] Luis Lauro, E-T. Potential of plants from the genus *Agave* as bioenergy crops. *BioEnergy Res.*, **2012**, *5*(1), 1-9.
- [15] Sarwar, M.B.; Ahmad, Z.; Rashid, B.; Hassan, S.; Gregersen, P.L.; Leyva, M.O.; Nagy, I.; Asp, T.; Husnain, T. *De novo* assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Sci. Rep.*, **2019**, *9*(1), 396.
<http://dx.doi.org/10.1038/s41598-018-35891-6> PMID: 30674899
- [16] Gross, S.M.; Martin, J.A.; Simpson, J.; Abraham-Juarez, M.J.; Wang, Z.; Visel, A. *De novo* transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics*, **2013**, *14*(1), 563.
<http://dx.doi.org/10.1186/1471-2164-14-563> PMID: 23957668
- [17] Martin, J.; Bruno, V.M.; Fang, Z.; Meng, X.; Blow, M.; Zhang, T.; Sherlock, G.; Snyder, M.; Wang, Z. Rnnotator: an automated *de novo* transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics*, **2010**, *11*(1), 663.
<http://dx.doi.org/10.1186/1471-2164-11-663> PMID: 21106091
- [18] Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **1980**, *16*(2), 111-120.
<http://dx.doi.org/10.1007/BF01731581> PMID: 7463489
- [19] Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, **2018**, *35*(6), 1547-1549.
<http://dx.doi.org/10.1093/molbev/msy096> PMID: 29722887
- [20] Dai, X.; Zhao, P.X. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.*, **2011**, *39*(suppl_2), W155-W159.
- [21] Binns, D.; Dimmer, E.; Huntley, R.; Barrell, D.; O'Donovan, C.; Apweiler, R.; Quick, G.O. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*, **2009**, *25*(22), 3045-3046.
<http://dx.doi.org/10.1093/bioinformatics/btp536> PMID: 19744993
- [22] Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M. UniProt: the universal protein knowledgebase. *Nucleic acids research*, **2004**, *32*(suppl_1), D115-D119.
<http://dx.doi.org/10.1093/nar/gkh131>
- [23] Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L. InterPro: the integrative protein signature database. *Nucleic acids research*, **2008**, *37*(suppl_1), D211-D215.
- [24] Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **2003**, *31*(13), 3406-3415.
<http://dx.doi.org/10.1093/nar/gkg595> PMID: 12824337
- [25] Zhang, B.H.; Pan, X.P.; Cox, S.B.; Cobb, G.P.; Anderson, T.A. Evidence that miRNAs are different from other RNAs. *Cell. Mol. Life Sci.*, **2006**, *63*(2), 246-254.
<http://dx.doi.org/10.1007/s00018-005-5467-7> PMID: 16395542
- [26] Yang, H.; Zhang, H.; Zhu, L.; Zhang, C.; Li, D. Identification and characterization of microRNAs in Macaca fascicularis by EST analysis. *Comparative Funct. Genom.*, **2012**, *2012*
- [27] Zhang, B.H.; Pan, X.P.; Wang, Q.L.; Cobb, G.P.; Anderson, T.A. Identification and characterization of new plant microRNAs using EST analysis. *Cell Res.*, **2005**, *15*(5), 336-360.
<http://dx.doi.org/10.1038/sj.cr.7290302> PMID: 15916721
- [28] Panda, D.; Dehury, B.; Sahu, J.; Barooah, M.; Sen, P.; Modi, M.K. Computational identification and characterization of conserved miRNAs and their target genes in garlic (*Allium sativum* L.) expressed sequence tags. *Gene*, **2014**, *537*(2), 333-342.
<http://dx.doi.org/10.1016/j.gene.2014.01.010> PMID: 24434367
- [29] Nodine, M.D.; Bartel, D.P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.*, **2010**, *24*(23), 2678-2692.
<http://dx.doi.org/10.1101/gad.1986710> PMID: 21123653
- [30] Chen, X.; Zhang, Z.; Liu, D.; Zhang, K.; Li, A.; Mao, L. SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development. *J. Integr. Plant Biol.*, **2010**, *52*(11), 946-951.
<http://dx.doi.org/10.1111/j.1744-7909.2010.00987.x> PMID: 20977652
- [31] Huijser, P.; Schmid, M. The control of developmental phase transitions in plants. *Development*, **2011**, *138*(19), 4117-4129.
<http://dx.doi.org/10.1242/dev.063511> PMID: 21896627
- [32] Zhu, Q-H.; Fan, L.; Liu, Y.; Xu, H.; Llewellyn, D.; Wilson, I. miR482 regulation of NBS-LRR defense genes during fungal pathogen infection in cotton. *PLoS One*, **2013**, *8*(12), e84390.
<http://dx.doi.org/10.1371/journal.pone.0084390> PMID: 24391949
- [33] Stief, A.; Altmann, S.; Hoffmann, K.; Pant, B.D.; Scheible, W-R.; Bäurle, I. Arabidopsis miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. *Plant Cell*, **2014**, *26*(4), 1792-1807.

- [34] <http://dx.doi.org/10.1105/tpc.114.123851> PMID: 24769482
Linder, P.; Jankowsky, E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.*, **2011**, *12*(8), 505-516.
- [35] <http://dx.doi.org/10.1038/nrm3154> PMID: 21779027
Ransom-Hodgkins, W.D. The application of expression analysis in elucidating the eukaryotic elongation factor one alpha gene family in *Arabidopsis thaliana*. *Mol. Genet. Genomics*, **2009**, *281*(4), 391-405.
- [36] <http://dx.doi.org/10.1007/s00438-008-0418-2> PMID: 19132394
Shi, L.-X.; Lorković, Z.J.; Oelmüller, R.; Schröder, W.P. The low molecular mass PsbW protein is involved in the stabilization of the dimeric photosystem II complex in *Arabidopsis thaliana*. *J. Biol. Chem.*, **2000**, *275*(48), 37945-37950.
- [37] <http://dx.doi.org/10.1074/jbc.M006300200> PMID: 10950961
Hagman, A.; Shi, L.-X.; Rintamäki, E.; Andersson, B.; Schröder, W.P. The nuclear-encoded PsbW protein subunit of photosystem II undergoes light-induced proteolysis. *Biochemistry*, **1997**, *36*(42), 12666-12671.
- [38] <http://dx.doi.org/10.1021/bi970685o> PMID: 9335523
Amann, K.; Lezhneva, L.; Wanner, G.; Herrmann, R.G.; Meurer, J. Accumulation of photosystem one1, a member of a novel gene family, is required for accumulation of [4Fe-4S] cluster-containing chloroplast complexes and antenna proteins. *Plant Cell*, **2004**, *16*(11), 3084-3097.
- [39] <http://dx.doi.org/10.1105/tpc.104.024935> PMID: 15494558
Jung, J.-H.; Lee, S.; Yun, J.; Lee, M.; Park, C.-M. The miR172 target TOE3 represses AGAMOUS expression during Arabidopsis floral patterning. *Plant Sci.*, **2014**, *215-216*, 29-38.
- [40] <http://dx.doi.org/10.1016/j.plantsci.2013.10.010> PMID: 24388512
Aukerman, M.J.; Sakai, H. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell*, **2003**, *15*(11), 2730-2741.
- <http://dx.doi.org/10.1105/tpc.016238> PMID: 14555699