फलियों में सर्क्युलर आर.एन.ए. की पहचान एवं लक्षण वर्णन

IDENTIFICATION AND CHARACTERIZATION OF CIRCULAR RNAs IN LEGUMES

BY

TANWY DASMANDAL

MASTER OF SCIENCE
IN
BIOINFORMATICS

ICAR-INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE

ICAR-INDIAN AGRICULTURAL RESEARCH INSTITUTE
NEW DELHI – 110012

2019

फलियों में सर्क्युलर आर.एन.ए. की पहचान एवं लक्षण वर्णन

IDENTIFICATION AND CHARACTERIZATION OF CIRCULAR RNAs IN LEGUMES

BY

TANWY DASMANDAL

A thesis submitted to the Faculty of Post-Graduate School,
ICAR-Indian Agricultural Research Institute, New Delhi,
In partial fulfilment of the requirements for the award of the degree of

MASTER OF SCIENCE
IN
BIOINFORMATICS

Approved By:

Chairperson:	**Dr. A.R. Rao**	
Co-chairperson:	**Dr. D.C. Mishra**	
Member:	**Mr. Sanjeev Kumar**	
Member:	**Dr. Ankur Biswas**	
Member:	**Dr. Ganapati Mukri**	

ICAR - INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE
Library Avenue, New Delhi-110012

Dr. A. R. Rao
Principal Scientist & Professor (Bioinformatics)
Centre for Agricultural Bioinformatics (CABin)

CERTIFICATE

This is to certify that the thesis entitled **"Identification and characterization of circular RNAs in legumes"** submitted to the faculty of Post Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, in partial fulfillment of the requirement for the award of the degree of **Master of Science** in **Bioinformatics** by **Ms. Tanwy Dasmandal**, Roll No. 20956 embodies the results of *bona fide* research work carried out by her under my guidance and supervision and no part of the study reported here has so far been submitted for publication or for any other degree or diploma.

The assistance and help availed during the course of this investigation as well as source of information have been duly acknowledged by her.

New Delhi
Date:

(Dr. A. R. Rao)
Chairman
Advisory Committee

THIS THESIS
IS MOST RESPECTFULLY
DEDICATED TO MY BELOVED
PARENTS AND TEACHERS

Acknowledgements

"The highest education is that which does not merely gives us information but makes our life in harmony with all existence."

…….. Rabindranath Tagore

This thesis is the fruitful outcome of the knowledge gained over the entire period of my M.Sc. study at ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI), New Delhi during which I have been in touch with a great number of people whose contribution in varied ways led to the research and making of the thesis which deserves special mention. It is a pleasure to convey my gratitude to all of them by way of my humble acknowledgements.

First and foremost, I want to express deepest sense of gratitude to ***Dr.A.R Rao****, Principal Scientist and Professor, Division of Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI), New Delhi and Chairperson of my Advisory Committee for his initiative, benevolence, endurance, constructive criticism and constant monitoring during the period of my investigation and also during preparation of this thesis. Above all and most needed, he provided me constant support and encouragement in various ways. Indeed, I am very fortunate enough to have his guidance. I am really indebted to him.*

I am also equally indebted to ***Dr.D.C Mishra****, Scientist, Division of Bioinformatics, ICAR- IASRI and Co-Chairman of my Advisory Committee for his moral support during the course of my research work. He gave me constant encouragement from the time to time for completion of my research work. I am really grateful to him for the support and encouragement.*

I am extremely thankful to ***Dr. Sanjeev Kumar****, Scientist, ICAR-IASRI, New Delhi and member of my Advisory Committee for his useful suggestions and constant support during my research work.*

I am also highly grateful to ***Dr. Ankur Biswas****, Scientist, Division of Agricultural Statistics, ICAR- IASRI, New Delhi and member of my Advisory Committee for his valuable suggestions and help.*

I would like to convey my deep sense of gratitude and appreciation to ***Dr. Anil Rai****, Principal Scientist and Head, Division of Bioinformatics, ICAR-IASRI, New Delhi, for his help and support during my entire course work. I am indebted to him for his valuable advice and constructive criticism during the preparation of this thesis.*

I am also thankful to ***Dr. SeemaJaggi****, Professor, Division of Agricultural Statistics, ICAR-IASRI, New Delhi, for his help and support during my entire course work.*

I am indebted to ***Dr. A.K. Singh****, Director, ICAR-Indian Agricultural Research Institute (ICAR-IARI), New Delhi,* ***Dr. Rashmi Aggarwal*** *, Dean, ICAR-IARI, and* ***Dr. R.K. Jain****, Former Dean & Joint Director (Education), ICAR-IARI for providing facilities to carry out this research work.*

I am deeply indebted to ***Dr. L.M. Bhar****, Director, ICAR-IASRI for the help and infrastructure provided by him.*

I want to express my respect and gratitude to my family, especially ***Maa, Baba*** *and* ***my didi*** *for their sacrifices they made for me. They always provide strength, love and support to me.*

I am thankful to all my batchmates especially ***Mona, Bhattu, Anku, Shruti, Shilpa, Tanima, Naveen, Baibhav, Nitesh, Jutan, Krishna, Rohit, Abhishek, Rahul, Amit, Lishi, Jagdish, Vinay, Vinayaka, Debopam, Sona, Nandini,*** *for their friendly approach and moral support throughout my M.Sc course and* ***Sarathi, and Debo*** *for their constant support and encouragement throughout my B.Sc and M.Sc student life.*

I do not have any words to express my thanks for the help and support rendered by my seniors specially, ***Sutanwa di, Asha mam, Dipro da, Amit da, Sayantani di, Debdali di, Dipankar da, Rittika di, Tanuj da, Soumya mam, Sonica mam, Shweta mam, Sneha mam, Sapna mam, Priyanka di, Shaba mam, Sayanti di, and Sanchita di*** *and my juniors* ***Parinita, Sharan, Mailar, Shreyosi, Shreya, and Nandita.***

I take this opportunity to appreciate the help rendered by the staff of Training and Administrative Cell (TAC), TAC Lab, Bioinformatics Lab and Library, ICAR-IASRI. Special thanks to ***Sanjeev ji, Gaganji, and Usha Jain ma'am****.*

I would like to thank all the staff of the PG School for their helpful attitude and cooperation throughout the period of study. I would also like to thank IARI for providing scholarship throughout the period of study.

Finally, I would like to thank everyone who was imperative to the successful realization of this thesis as well as articulate my apology if I missed anyone.

Date:

Tanwy Dasmandal

Place: New Delhi-110012 ***(Tanwy Dasmandal)***

CONTENTS

LIST OF FIGURES

	stress conditions in soybean
Figure 27	Lengthwise distribution of intronic circRNAs under control and wilt stress conditions in soybean
Figure 28	Lengthwise distribution of intergenic circRNAs under control and wilt stress conditions in soybean
Figure 29	Distribution of circRNAs based on expression values under control and wilt stress conditions in soybean
Figure 30	GO Terms of differentially expressed circRNAs in soybean-wilt in terms of biological processes, cellular components and molecular functions
Figure 31	Classification of biological processes for mRNAs in soybean-wilt
Figure 32	Classification of molecular functions for mRNAs in soybean-wilt
Figure 33	Classification of cellular components for mRNAs in soybean-wilt

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

The present chapter provides a brief description of the leguminous crops and their importance. This is followed by the introduction to coding and non-coding RNAs, various types of non-coding RNAs along with their potential roles, a brief detailing about circular RNAs, their biogenesis, functions and computational prediction and finally the chapter ends with the motivation, objectives and scope of the thesis.

1.2 Leguminous crops and their importance

Legumes are the plants belonging to the family Fabaceae (previously known as leguminosae). The name 'Fabaceae' comes from the genus *Faba*, now included in *Vicia*. The term "faba" is a Latin word which simply means “bean”. Previously the name of this family was Leguminosae, which is still considered valid (International Code of Nomenclature for algae, fungi, and plants. Article 18.5) and the fruits of the plants belonging to this family are called legumes. This is one of the largest family of flowering plants and consists of about 690 genera and 18000 species (Jambunathan, *et al.,* 1991). The leguminous plants are rich source of proteins (especially rich in amino acids: lysine and methionine), dietary fibre, carbohydrates and different dietary minerals. In terms of number of species, they are widely distributed as the third-largest land plant family in the world. Based on the total cultivated area and production, legumes are the second most important crops, the first one being the cereals. Worldwide about 12-15% of the Earth’s arable surface is used in the production of grains and forage legumes (Morel, *et al.,* 2012). The major quality of leguminous crops is that they are capable to interact symbiotically to soil bacteria *Rhizobium sp.* that infect their root nodules through which they can fix atmospheric nitrogen into the soil and thus their production can increase the indigenous nitrogen availability of the soil. The leguminous crops also help in solubilising insoluble phosphorous in soil, improve the physical properties of the soil, increase microbial activity in soil, and can play a vital role in controlling weeds. There are several types of legumes like- grain legumes (pulses), forage legumes, oilseed legumes etc. One of the most widely cultivated legumes is the oilseed soybean which accounts for about 50% of the total world production of grain legumes, while chickpea provides about

7% of the production. But the production of the leguminous crops is highly affected by several biotic and abiotic stresses. Biotic stresses include the diseases caused by different pathogens like fungi, bacteria, viruses, etc as well as pests like insects, and nematodes. Whereas abiotic stresses mainly include drought stress, heat stress, salinity, frost, water logging etc. For legumes, drought stress has adverse effects on total biomass, pod number, seed number, seed weight and quality, and seed yield per plant (Toker *et al*., 2007; Charlson *et al*., 2009; Khan *et al*., 2010; Toker & Mutlu, 2011; Impa *et al*., 2012; Hasanuzzaman *et al*., 2013).

1.2.1 Taxonomy of family Fabaceae (Leguminosae)

As per United States Department of Agriculture (https://plants.usda.gov/), the taxonomic classification of Fabaceae is as follows:

Kingdom: Plantae – Plants
Subkingdom: Tracheobionta – Vascular plants
Superdivision: Spermatophyta – Seed plants
Division: Magnoliophyta – Flowering plants
Class: Magnoliopsida – Dicotyledons
Subclass: Rosidae
Order: Fabales
Family: Fabaceae – Pea family

Two most important crops under Fabaceae are chickpea and soybean. A brief introduction about these crops are described in the next few subsections.

1.2.2 Chickpea

Among the grain legumes, chickpea (*Cicer arietinum*), commonly known as Bengal gram is one of the most important pulse crops in India. India is the leading producer of chickpea. They are of two types- kabuli and desi. The seeds of the kabuli type are large and light coloured, whereas the seeds of the desi type are comparatively smaller than the kabuli type and are dark yellow-brown coloured. They can be consumed in different forms like as whole dehulled grains, immature pods, sprouted grains, mature green seeds or as daal. Chickpea is a good source of proteins and carbohydrates, accounting about 80% of the total dry seed mass (Geervani, *et al.,*1989). The whole chickpea grain contains about 20.8-25.9% proteins and the starch content of the seed may vary from 41% to 50%. Chickpea also contains about

6% fat, which is an important source of fat for the vegetarian diet, and rich in important minerals especially calcium and iron. The crop can be used in a variety of ways, one of which is rather unusual, like, the acid exudates from chickpea have medicinal value. The seeds of the crop are consumed at immature stage as well as in the form of dried grains. The yield of the crop is highly affected by several abiotic and biotic stresses. Chickpea is mainly grown in areas of tropical climate under conditions of residual soil moisture. Thus the crop greatly experiences conditions of drought and heat stress. Moreover, along with the abiotic stresses the crop is reported to be affected by more than 50 pathogens of which few are highly devastating like fusarium wilt, ascochyta blight, etc (Rheenen, 1991).

1.2.3 Soybean

Other than the pulse crops, there are legumes like soybean which have great industrial importance. It is one of the main crops cultivated for oil extraction. Soybean contains about 20% oil. The crop is also rich in protein content to a extent of 40%. Among the cultivated crops in the world, soybean has the highest protein content as well as the highest gross output of vegetable oil (Singh, 2010). The crop has several uses other than extracting oil like making many food products such as bean curd, soybean milk, bean curd stick etc. The fermented soybean products are soy paste, soybean cheese, fermented soybean, soybean sauce, etc. Moreover, livestock are fed with soybean cakes, which are high protein rich products. The crop has also great role in improving the soil properties through their deep tap root system, as green manuring crop as well as for conserving soil moisture and contribution to soil nitrogen enrichment. (Kulcheski, *et al*., 2011). However, soybean productivity is greatly affected by several abiotic and biotic stresses. Drought is the major abiotic stress factor that negatively affect soybean productivity around the world. The impact of water stress during the flower formation can cause shorter flowering period, and water stress during the later phases of reproductive development accelerate senescence thereby decreasing the duration of the seed-filling period. With regards to biotic stress, fusarium blight or wilt is an important disease that affects the yield of soybean.

1.3 Coding and non-coding RNAs

RNAs are the macromolecules that convert the genic information contained in the DNA of organisms into proteins which ultimately helps in the expression of

various characters in an organism. Broadly RNAs can be classified into two categories- coding RNAs and non-coding RNAs. Through transcription, messenger RNAs (mRNAs) are formed from the DNA which further gets translated into various proteins. But for the synthesis of the proteins, other than these mRNAs, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) also play vital roles. These three types of RNAs are known as the protein coding RNAs or simply coding RNAs. Earlier it was thought that the transcriptome of an organism is mainly composed of these coding RNAs. But with the advancement of high throughput sequencing technologies the complexity of the genomes of higher organisms have been revealed and it has been found that a minority of the genomes of multicellular organisms is composed of protein coding sequences and that also decreases with increase in the complexity of the organism, with an increase in the amount of intergenic and intronic non-coding sequences most of which are infact transcribed. Thus there is a shift in the transcriptional output of multicellular organisms as compared to microorganisms from mainly protein-coding RNAs to mainly non-coding RNAs (Mattick, *et al.,* 2006). So it would not be right anymore to conclude that protein coding gene expression in organisms is purely attributed to the proteins involved in transcription, RNA processing and translation. Rather very vital and potent role in regulation of gene expressions at various levels are carried out by these non-coding RNAs.

1.4 Types of non coding RNAs

Non-coding RNAs in organisms can be classified into different classes that have different modes of biogenesis as well as different modes of target regulation. Some important classes of non-coding RNAs along with their functions are as follows (Collins, *et al.,* 2011).

1.4.1 Short non-coding RNAs

Short non-coding RNAs is a class of non-coding RNAs which are shorter in length varying from 20-50 nucleotides. These short ncRNAs regulate the gene expression through RNAi (RNA interference) mechanism, where some nucleotides of the ncRNAs bind to either the coding region or the promoter region of the genes which interferes with the normal processing of the mRNAs and silences the mRNA expression. These short ncRNAs can be further classified into three major classes-

1.4.1.1 MicroRNAs (miRNAs)

miRNAs is a class of short non-coding RNAs that are 21-23 nucleotides in length. These are initially single-stranded RNA (ssRNA) molecules produced through transcription or through splicing, which fold into stem-loop structures forming imperfect double-stranded RNA molecules (dsRNAs). These dsRNAs are then processed by RNase III which is an endonuclease (generally Dicer) before being denatured. One of the RNA strands (usually the less stable one of the two) binds to the RNA-induced silencing complex (RISC), which then binds to a specific target mRNA containing sequence complementary to the miRNA, to induce either degradation or cleavage, or block translation.

1.4.1.2 Short-interfering RNAs (siRNAs)

Short interfering RNAs are the non-coding RNA molecules that are 20-25 nucleotides in length. These RNA molecules are produced as dsRNAs, and can enter the Post-Transcriptional Gene Silencing (PTGS) pathway, which leads to degradation of mRNAs in the cytoplasm, or the Transcriptional Gene Silencing (TGS) pathway involved in modification of the chromatin. In plants, siRNAs are reported to be involved in RNA-directed DNA Methylation (RdDM), that was first observed in viroid infected tobacco plants where the sequences that are similar in sequence to the viral genes got methylated.

1.4.1.3 PIWI-interacting RNAs (piRNAs)

PIWI-interacting RNAs are about 27-30 nucleotides long and unlike miRNAs and siRNAs, piRNAs are not produced by “Dicing”, but mainly by bi-directional promoters through the mechanism which is known as the “ping pong” cycle of biogenesis and amplification. These non-coding RNAs mainly function as chromatin regulators by interacting with PIWI proteins and are also found to play critical roles in transposon “control” i.e. preventing transposon activation and hence keeping the levels of transposons interrupting genes to a minimum.

1.4.2 Long non-coding RNAs (lncRNAs)

lncRNAs are non-coding RNAs that are more than 200 nucleotides long. The broad functional repertoire of lncRNAs includes roles in high-order chromosomal dynamics, telomere biology and subcellular structural organization. One major

emergent theme is the involvement of these ncRNAs in regulating the expression of neighbouring protein-coding genes (Mercer, *et al.,* 2009).

Other than these major classes of ncRNAs, a new class of ncRNAs called **circular RNAs (circRNAs)** also play an important role in regulating genes of several interests.

1.5 Circular RNAs

The existence of circular RNAs was first reported in 1970s in plant viroid (potato spindle tuber viroid) and yeast mitochondria and later on in higher eukaryotes. These are nothing but a special class of non-coding RNAs that are produced by the formation of covalent linkage between the 5' and 3' ends of an RNA molecule. They may vary in size from 100 nucleotides to several kilobases in length and may originate from coding as well as non-coding genes.

1.5.1 Biogenesis of circular RNAs

Circular RNAs were primarily thought to be produced as a byproduct of splicing. But as more and more circular RNAs were identified from different organisms, it was found that circular RNAs were also formed from the exonic regions of eukaryotic cells. Eventually it was found that based on their genomic origin, circRNAs are broadly of three kinds – exonic circRNAs, intronic circRNAs and intergenic circRNAs which are found in almost all eukaryotic clades like animals, insects, plants, fungi, etc. Though these different circRNAs originates through different ways of biogenesis, all are related directly or indirectly to the mechanism of splicing. During splicing the intronic lariat (excluding the 3' tail) that is generated forms a perfect intronic circRNA. Exonic circRNAs are formed mainly through a mechanism known as backsplicing. In this event the 3' hydroxyl end of an exon released during splicing get covalently linked to the 5' end of the same or different exon situated upstream. The covalent linkage between 3' end 5' end of the same exon results in the formation of single-exonic circRNAs. While if the linkage is between the 3' end of one exon and 5' end of another upstream exon, then the circRNA formed consists of several intermediate intronic and exonic regions; the introns may remain in between or may get spliced post-transcriptionally. Majority of backsplicing events are cis-ones i.e it occurs between the exons of the same gene. From different studies it has been observed that in case of animals, for circularization of exonic regions,

canonical splice signals are mostly required. However, plants may use non-canonical splice signals also. Circular RNAs are mostly found to be conserved across different plant species.

1.5.2 Functions of circular RNAs

i) The circular RNAs in both plants and animals mainly function as miRNA sponge to regulate the function of the miRNAs because of presence of miRNA binding sites in them. It has been found through several studies that a single circular RNA may have more than one miRNA binding sites. For example in rice, about 31 exonic circRNAs have two or more miRNA binding sites.

ii) CircRNAs may also regulate the function of RNA-binding proteins through direct interaction.

iii) They may regulate the transcription of their parental genes through their interaction with the RNA Pol II, U1 nuclear RNA and the promoter of their host gene.

iv) CircRNAs may have potential role in developmental/stress-specific biological processes in plants. The differential expression of circular RNAs in different stress conditions have been studied in several crops and it has been found that differentially expressed circular RNAs may act as important functional regulators involved in stress-specific biological processes in plants. For example, 27 rice exonic circRNAs differentially expressed under phosphate-sufficient or -starvation conditions have been identified. In response to cold and heat treatment, 163 and 1583 circRNAs were identified to be differentially expressed in tomato and Arabidopsis, respectively .

1.5.3 Identification of circular RNAs

From the time circular RNAs were first discovered, different biochemical methods have been developed to test the circularity of a segment of RNA as well as for validating the existence of different circular RNAs. Along with this several statistical and bioinformatic approaches have also been developed for identifying circular RNAs and to predict their potential role in gene expression. The first genome wide identification of circRNAs has been done in Arabidopsis, followed by rice and later on studies have been done in other major crops like maize, soybean, wheat, barley, etc.

1.5.3.1 Biochemical identification of circular RNAs

Since circular RNAs do not have 5' end and 3' end like linear non-coding RNAs, nor they contain poly(A) tail, traditional sequencing techniques used for identifying linear non-coding RNA molecules cannot be applied for identification of the circular RNAs. One of the most fundamental tools for validation of a circRNA is reverse transcription-PCR (RT-PCR). Another effective way of identify circularity of RNAs involves RNase H treatment which is an endonuclease capable of cleaving RNA at RNA-DNA hybrids and Polyacrylamide gel electrophoresis can be used to distinguish between the linear and the circular RNAs. Other than these two approaches, there are a few other techniques also like- northern blotting (Capel *et al*., 1993). RNase R treatment (an exoribonuclease which is capable of degrading RNA from its 3′ to 5′ end) (Suzuki *et al*., 2006; Vincent and Deutscher, 2006).

1.5.3.2 Computational identification of circular RNAs

Although there are a variety of experimental procedures for the identification of circular RNAs, but recently several approaches have been made for the computational prediction and detection of expression of the circular RNAs. Total RNA or ribosomal RNA-depleted libraries are usually used for circRNA profiling. From RNA-seq datasets the circular RNAs are predicted based on the principle that those reads are capable of forming circular RNAs which are 'chimeric' in the sense that with respect to transcription the 5′ sequence is downstream of the 3′ sequence in the read. Based on this principle a variety of algorithms have been developed for prediction of circular RNAs like- CIRI, circExplorer, findCirc, Mapsplice, segmehl etc.

1.6 Motivation

In order to cope up with the increasing demand of food supply, along with cereals the world demand for legumes is also expected to grow in the near future, in both developing countries as well as in the developed nations given the trend towards healthy dieting. As the therapeutic uses of legumes are better recognized (Duranti, 2006) and the health risk of consuming animal proteins is more widely understood, the demand for legume-based products is likely to retain its upward trajectory.

Quite often droughts negatively impact the yield of most cultivated crops, including the legumes (Pandey, *et al.,* 1984). The yield of food legumes grown in

dryland areas such as the Mediterranean regions are usually variable or low due to terminal droughts (Karrou and Oweis, 2012). Along with drought stress, the severity of loss in leguminous crops increases due to biotic stresses like infection by *Fusarium sp*.

Several studies have suggested that different non-coding RNAs play a vital role in regulating different stress tolerance in various leguminous crops (Arenas-Huerter, *et al.,* 2009). From studies, it has been found that circular RNAs which is a new class of non coding RNAs have significant role under dehydration stress in wheat (Wang, *et al.,* 2017).

To the best of our knowledge, till date, studies have not been done regarding the identification of circRNAs in leguminous crop species like chickpea, and soybean that too for traits like abiotic stress and biotic stress tolerance. So with the understanding of the great values of chickpea and soybean and realizing their loss in production due to drought stress and wilt disease infection, importance was felt to study the potential role of circular RNAs in regulating genes involved in drought and wilt stress tolerances in the two crops.

1.7 Objectives

Keeping the research gap given in section 1.6 in mind, a study on "**Identification and characterization of circular RNAs in legumes**" was done with the following objectives:

i) To identify and characterize circRNAs responsible for biotic and abiotic stress tolerance in the leguminous crops.

ii) To study the relationship between circRNA-miRNA-mRNA.

1.8 Scope of the thesis

The thesis starts with **Chapter I** on introduction, which describes a brief description of the leguminous crops, coding and non-coding RNAs, types of non-coding RNAs, circular RNAs, their biogenesis, functions and computational prediction. The **Chapter II** contains the literature related to biotic and abiotic stresses in leguminous crops with special reference to soybean, and chickpea and non-coding RNAs, with main emphasis given on circular RNAs of various crops and different tools developed for their identification and characterization. The details on the materials and methodology followed for achieving the objectives set for the study are

discussed in **Chapter III.** Whereas a details on the results obtained from the study are given in **Chapter IV**. **Chapter V** includes discussion on the results obtained from the study as well as in the context of findings available in the literature .

The thesis ends with a summary of the salient findings of the study, abstract, bibliography and annexure.

The previous work done on non-coding RNAs in general and circular RNAs in particular along with their roles in abiotic and biotic stress tolerance mechanisms in legumes is presented in the next chapter.

CHAPTER II

REVIEW OF LITERATURE

2.1 Background

Literatures have been reviewed regarding the topic of the study and is compiled in this chapter. The review relates to general overview of biotic and abiotic stresses in leguminous crops with special reference to the implication of non-coding RNAs in various biotic and abiotic stress mechanisms in soybean, and chickpea. To be more particular, the main emphasis is given on the effect of circRNAs in regulating the genes responsible for stress mechanisms. Therefore, the review of literature on previous works done on circular RNAs of various crops and different tools developed for their identification and characterization are provided in the present chapter.

2.2 Biotic and abiotic stresses in leguminous crops

Reddy, *et al.,* (2012) made a thorough discussion about the various omics approaches along with their applications with respect to abiotic stresses in important legume crops. The recent advances in omics technologies and availability of the genome sequences of model legumes provide great opportunity to study different aspects related to stress tolerance in the legume crops. Different studies have been carried out regarding the genome-wide expression profiling in various leguminous crops to identify the potential genes involved in the response mechanism of different abiotic stresses. Among various grain legumes, soybean has been intensively studied and recently, characterization of different tolerant and sensitive varieties of chickpea and peanut under abiotic stress conditions have been carried out.

2.2.1 Soybean

Le*, et al.* (2012) identified the most suitable reference genes related to abiotic stresses in soybean. For this they have collected 13 candidate genes from literature and have evaluated them for stability of expression under several abiotic stresses like-dehydration, high salinity, cold and ABA (abscisic acid) treatments. Among those genes, they have found 60 genes to be the best reference gene in different tissues under the said stress conditions. It was also observed that those genes identified to be the best from their study were also able to detect subtle variations in expression rates that otherwise would not be possible using a less stable reference gene. Their study revealed that a single reference gene cannot be best for all conditions as well as for all the tissues. Rather, they are tissue and stress specific in nature.

Deshmukh, *et al.* (2014) provided a review on the advances in the development of omic tools used for dissecting abiotic stress tolerance in soybean. The review also provides a comprehensive catalog of the different omic resources available online for dealing with stress tolerance mechanisms in soybean. Besides, they addressed the importance of phenomics in different approaches related to the abiotic stresses and identified high-throughput multi-dimensional phenotyping as a major limiting factor for the improvement of abiotic stress tolerance in soybean.

2.2.2 Chickpea

Mantri, *et al.,* (2010) employed a 768-featured microarray for comparing the genes expressed by chickpea in response to drought, cold, high salinity and the fungal pathogen Ascochyta rabiei. From the study, a totl of 46, 54, 266 and 51 differentially expressed transcripts were identified for drought, cold, high salinity and the fungal pathogen Ascochyta rabiei respectively. Along with this the genes commonly expressed in biotic and abiotic stress conditions were also identified. Previously several studies have been conducted regarding the transcriptional profiling of plant responses to several biotic and abiotic stresses. But majority of them had dealt with either biotic or abiotic stresses, thus making it difficult to find the genes that were common to both biotic as well as abiotic stresses.

Garg, *et al.,* (2016) performed the analysis of whole transcriptome of chickpea to detect the genes responsible for drought and salinity stress responses. They considered the RNA-seq of the root tissues under drought and salinity stress conditions at both vegetative as well as reproductive stages and through transcriptomics they found divergent gene expressions at different developmental stages. They also identified the key enzymes involved in metabolic pathways (carbohydrate metabolism, photosynthesis, lipid metabolism, etc.), which were also affected by drought and salinity stresses.

2.3 Small and long non-coding RNAs

Reinhart, *et al.,* (2002) studied the presence of miRNAs in plants, indicating that this class of noncoding RNA arose early in eukaryotic evolution. In their study, 16 miRNAs of *Arabidopsis* were described and observed that many of them have differential expression patterns in development. Out of which 8 were found to be absolutely conserved in the rice genome.

Zhu, *et al.,* (2012) reviewed the current understanding about the functions of long ncRNA in plants and have discussed the different approaches of in-silico and de novo identification of lnc RNAs in plants. The review also contains the molecular functions carried out by the lnc RNAs in plants as well as the challenges faced during their functional characterization.

Kulcheski, *et al.,* (2011) developed libraries from drought-sensitive and tolerant seedlings and rust-susceptible and resistant soybeans, sequenced the libraries and analyzed the data and detected 256 miRNAs. From the study 24 families of novel miRNAs, 2 families of conserved miRNAs were found in soybean when compared with other plant species. Several iso miRNAs were also observed. On validation, 11 miRNAs showed differential expression profiles during biotic and abiotic stresses in soybean. The majority of miRNAs were also found to be up-regulated during water deficit stress in the sensitive plants. However, for the tolerant genotype, most of the miRNAs were found to be down regulated.

2.4 Characterization of circular RNAs

Ye *et al.* (2015) performed genome-wide identification of circRNAs in *Oryza sativa* and *Arabidopsis thaliana* using publically available RNA-Seq data. And the results demonstrated that circRNAs are widespread in plants, revealing the common and distinct features of circRNAs between plants and animals, and suggested that circRNAs could be a critical class of noncoding regulators in plants.

Lu *et al.* (2015) reported 2354 circRNAs in **rice** that were identified through deep sequencing and computational analysis of ssRNA-seq data and suggested that circRNA and its linear form might have acted as a negative regulator of its parental gene. Overall, the study revealed the prevalence of circRNAs in rice and provided new biological insights into rice circRNAs.

Darbani *et al.* (2016) used the software CIRI (CircRNA Identifier) to analyze RNA-Seq reads for existence of circular RNAs in **barley** and also demonstrated that the levels of circular RNAs vary across tissues and stages. An interesting discovery in their study was the identification of microRNAs and putative long non-coding/microprotein coding RNAs as the transcripts of origin for circular RNAs.

Wang *et al* (2017) extracted RNA from leaves of **wheat** seedlings under dehydration-stressed and well-watered conditions and identified the circRNAs using bioinformatics tool CIRI. This study revealed a possible connection between the

regulation of circRNAs with the expressions of functional genes in wheat leaves associated with dehydration resistance.

Zhao, *et al.* (2017) used deep sequencing technology coupled with RNase R enrichment strategy and bioinformatic approach CIRI to uncover circRNAs in **soybean** and identified 5,372 circRNAs, approximately 80% of which were paralogous circRNAs but have different expression patterns. They also predicted that among the identified circRNAs, 2134 circRNAs were the targets of 92 miRNAs.

Chen *et al* (2018) hypothesized that transposons in maize may be involved in the formation of circRNAs and further modulated phenotypic variation and performed circRNA-Seq on B73 seedling leaves and uncovered 2804 high-confidence **maize** circRNAs, which showed distinct genomic features. This first study of maize circRNAs uncovers a potential new way for transposons to modulate transcriptomic and phenotypic variations.

Chu *et al*.(2018) summerized the progress achieved so far on plant circRNAs, including identification and functional characterization, and compared the similarities and differences of circRNAs between plant and animal species, and discussed the challenges for detection of plant circRNAs .

2.5 Prediction of circularRNAs

Zhang, *et al.* (2014) developed a computational pipeline named **CIRCexplorer** to precisely identify back-spliced junction reads for circular RNAs and characterized circular RNA formation by both bioinformatic and biochemical lines of evidence, demonstrating that flanking complementary sequences, including both repetitive and nonrepetitive sequences, play important roles in exon circularization.

Fu, *et al.* (2014) introduced **circRNAFinder** as a software tool for identifying circRNAs using RNA deep sequencing data, as well as annotating and visualizing circRNAs and comparing their expression in different tissues or conditions.

Gao *et al.* (2015) developed a novel chiastic clipping signal-based algorithm, **CIRI**, to unbiasedly and accurately detect circRNAs from transcriptome data by employing multiple filtration strategies. By applying CIRI to encode RNA-seq data, they have identified and experimentally validated the prevalence of intronic/intergenic circRNAs as well as fragments specific to them in the human transcriptome.

Chu, *et al.* (2017) build a database of plant circRNAs, termed as **PlantcircBase** (http://ibi.zju.edu.cn/plantcircbase/) for plant research community. In this database, circRNAs identified in recent years by prediction and/or experimental validation from *O. sativa, Triticum aestivum, A. thaliana, Zea mays, Solanum lycopersicum, Hordeum vulgare, Solanum tuberosum, Glycine max, Gossypium sp.* were collected and made publicly available. It includes a total of 77,595 unique circRNAs from diverse tissues of the crops.

Gao *et al* (2018) attempted to illustrate the key steps and summarized the trade off between different strategies, covering all popular algorithms for circRNA detection (like CIRI, CircExplorer, KNIFE, etc) and various downstream analyses.

Based on the literature review and keeping in view the objectives of the present study, the methodology along with the materials used for the present study is presented in the next chapter.

CHAPTER III

MATERIALS AND METHODS

This chapter explains various materials and methods used for identification of circular RNAs under various stress conditions of different leguminous crops. The identification is followed by the characterization of those identified circular RNAs, predicting the differentially expressed circular RNAs in various stress situations and observing the interaction between the circular RNAs, their target miRNAs and mRNAs.

There are a number of bioinformatics tools available for identification of circular RNAs like CIRCexplorer, findcirc, circRNAfinder, CIRI and so on. Every tool has its own advantages and disadvantages. But from the literature survey it has been observed that the most used tool for identification of circular RNAs especially in case of plants is CIRI, because of its high sensitivity as compared to the other tools. Apart from this another advantage of CIRI is that submission of genome annotation data is not compulsory in the tool, thus extending the scope of identification of circular RNAs from crops not having genome annotations.

Since much work has not been done on leguminous crops regarding the role of circular RNAs in various abiotic and biotic stress conditions, this work aims at identifying the circular RNAs in chickpea and soybean under drought stress (abiotic stress) and fusarium wilt (biotic stress) affected condition and predicting the interaction between the circRNAs-miRNAs-mRNAs, i.e., network of circRNAs-miRNAs-mRNAs in leguminous crops.

3.1 Source and extent of data

RNA-seq data were taken for chickpea and soybean crops under various stress conditions which are given below:

i) For chickpea and soybean, RNA-seq data under control and drought conditions (abiotic stress condition) were downloaded from the public domain (https://www.ncbi.nlm.nih.gov/) with the following accession numbers:
   Chickpea control data- SRR5927135
   Chickpea drought data- SRR5927136
   Soybean control data- SRR2545896

Soybean drought data- SRR2545900

ii) RNA-seq data for wilt (biotic stress) under control and wilt affected chickpea and soybean were downloaded from the public domain (https://www.ncbi.nlm.nih.gov/) with the following accession numbers:
Chickpea control data- SRR2961640
Chickpea wilt data- SRR2961641
Soybean control data- SRR4095539
Soybean wilt data- SRR4095541

iii) The whole genome data for chickpea was downloaded from the public domain (https://www.ncbi.nlm.nih.gov/).

iv) The whole genome data for soybean was downloaded from the public domain (http://plants.ensembl.org/index.html).

v) Data of all the miRNAs for chickpea were downloaded from the **Supplementary table S3** given in Jain, *et al.,* (2014).

vi) Data for all the miRNAs for soybean were downloaded from the public database (http://www.mirbase.org/).

3.2 Methodology

3.2.1 Data preparation

The raw data downloaded from the public databases were processed to get quality reads. For this at first the report on the quality of the reads was obtained using **FastQC** software which provides a way to do quality control checks on raw sequence data obtained from high throughput sequencing pipelines. It outputs summary graphs and tables to quickly assess the data based on which poor quality reads can be discarded as well as poor quality segments of the reads can be trimmed to obtain quality reads for downstream analysis.

Next, based on the report obtained from FastQC software the poor quality segments of the reads were trimmed using the **Trimmomatic** tool. This tool performs useful trimming tasks for paired-ended and single ended RNA-Seq data. Trimming the reads using Trimmomatic removes the adapters, “leading low quality or N bases (below quality 3)”, “trailing low quality or N bases (below quality 3)”. It also scan the reads with a 4-base wide sliding window, thereby cutting them when the average quality per base drops below 15 and drop reads below the 36 bases long. The trimmed

clean reads were then used as input data during identification of circular RNAs in CIRI.

3.2.2 Identification and characterization of circular RNAs

For the identification of circular RNAs and their characterization, following steps were followed:

Step1: The mapping of the reads to the reference genome was done by using **BWA-MEM**. BWA-MEM is a tool that aligns single end reads or paired-end reads to specific reference genome using the BWA-MEM algorithm. The reads have to be supplied in FASTQ format as input. For mapping two modules were executed

i) An index file of the reference genome was built by 'bwa index' module.
ii) Mapping of the clean good quality fastq reads (may be paired end or single end) to the index file was done by 'bwa mem' module. BWA-MEM generated SAM file as output.

Step2: After the mapping of the reads, identification was done using CIRI. It employs a de novo algorithm for identifying circular RNAs in an annotation independent manner. For the identification and characterization of the circular RNAs, CIRI requires two files as input- one is the SAM file generated by BWA-MEM mapping and the other is the reference genome FASTA file. CIRI basically involves three major steps (Gao, *et al.,* 2015):

i) Balanced junction reads detection:
ii) Filtering of the predicted junction reads
iii) Unbalanced junction reads detection and final filtering

The major advantages of identification of circRNAs with CIRI are that it neither requires RNA-seq data generated through circRNA enrichment step, such as RNase treatment, nor an annotation file as input, and is applicable to almost all commonly generated read lengths in various sequencing platforms. Moreover, the systematic filtering in the algorithm ensures a quite low false positive rate without sacrificing the sensitivity of detecting small circRNAs and non-exonic circRNAs.

Step3: The final output of CIRI contained the IDs of the identified circular RNAs along with their start and end positions on the chromosomes. Those information were used to fetch the nucleotide sequences of each circular RNA in FASTA format using

bedtools. Bedtools is a software suite for the comparison, manipulation and annotation of genomic features. It supports a wide range of operations for interrogating and manipulating genomic features like finding overlaps between two BED files, removing the portion of an interval that is overlapped by another feature, merging overlapping features into a single feature, converting BAM records to FASTQ records, and so on. Among those tools, the **'getfasta'** tool was used to extract the sequences from the reference FASTA file using intervals.

Step4: After fetching the sequences of the circRNAs, the next step performed was the calculation of the isoform expression levels of the identified circular RNAs from the clean RNA-seq data using **RSEM**. RSEM is an user-friendly software package used for quantifying gene and isoform abundances from single-end or paired-end RNA-Seq data. In contrast to other existing tools, a reference genome is not required by the software. Thus, in combination with a de novo transcriptome assembler, RSEM enables accurate transcript quantification for species without sequenced genomes. A typical run of the RSEM software consists of two steps. First, a set of reference transcript sequences are to be generated. Second, a set of RNA-Seq reads are to be aligned to the reference transcripts and the resulting alignments are then used to estimate abundances and their credibility intervals. Thus two scripts were executed for calculating the isoform expression levels of the predicted circular RNAs.

i) Firstly, the 'rsem-prepare-reference' module was executed using the FASTA file containing the identified circRNA sequences obtained in Step3 as the reference_fasta_file to built the RSEM references.
ii) Secondly, the 'rsem-calculate-expression' module was executed using the (a) trimmed RNA-seq data as upstream_read_file for single-end reads and both upstream_read_file and downstream_read_file for paired-end reads (b) the output file of the previous executed module as the reference file.

Step5: From the output of Step4, the file containing isoform level expression estimates were used for detection of the differentially expressed circular RNAs using the R package **DESeq2**. The package DESeq2 provides methods to test for differential expression by use of negative binomial generalized linear models. As input, the DESeq2 package expects count data as obtained from RNA-seq or another high-throughput sequencing experiment in the form of a matrix of integer values.

Results tables are generated using the function *results*, which extracts a results table with log2 fold changes, *p* values and adjusted *p* values.

i) At first, the transcript IDs from the file containing the isoform level expression estimates of both control and treatment data were used as input in software **Venny** to find out the transcript IDs common in both of them.
ii) Next, a csv (comma separated values) file was generated containing the common transcript IDs along with their FPKM values obtained in the file containing the isoform level expression estimates for both control and treatment data using the **VLOOKUP** function in Excel.
iii) That csv file was used in the **DESeq2** package in R software to detect the differentially expressed circular RNAs and **bedtools** to fetch the sequences of those differentially expressed circRNAs.

3.2.3 CircRNA-miRNA-mRNA interaction

The main function of circular RNAs as reported till date is that they act as miRNA sponges. The circRNAs bind to specific miRNAs having sequence complementarity and restrict the miRNAs from regulating the expression of mRNAs. Therefore, after identification of the differentially expressed circRNAs the next objective was to predict the target miRNAs of the identified circular RNAs and also the mRNAs which may be regulated by those miRNAs. For carrying out this objective the steps followed were-

Step1: The miRNA targets of the differentially expressed circRNAs were predicted using **TargetFinder**. It is a plant small RNA target prediction tool. TargetFinder computationally predicts small RNA binding sites on target transcripts from a sequence database. This is done by aligning the input small RNA sequences against all transcripts, followed by site scoring using a position-weighted scoring matrix. The script 'targetfinder_threads.pl' was run using FASTA file containing the miRNA sequences downloaded for the specific crop as input for small RNA sequences and the FASTA file containing the sequences of the differentially expressed circRNAs as the target sequence database. The result can be obtained in different formats, viz., table form, tab-delimited form, GFF format etc.

Step2: From the results of TargetFinder, the miRNAs that were predicted to be targets of different circRNAs were sorted out along with their sequences to generate a

FASTA file of those predicted miRNAs. Those miRNA sequences were then used to predict their mRNA targets using the online tool **psRNATarget** (http://plantgrn.noble.org/psRNATarget/). psRNATarget is a plant small RNA target analysis server which features reverse complementary matching between small RNA and target transcript, and target site accessibility evaluation by calculating unpaired energy (UPE) required to 'open' secondary structure around small RNA's target site on mRNA. The FASTA file containing the miRNA sequences were then uploaded into psRNATarget by searching the preloaded transcript of the specific crop. The tool outputs a comprehensive list of small RNA-target pairs.

Step3: The next step involves the annotation of the predicted mRNAs. For this, the predicted mRNAs from psRNATarget were subjected to blast, mapped and annotated using **BLAST2GO** software with the default parameters (https://www.blast2go.com/). Blast2GO is a bioinformatics platform for high-quality functional annotation and analysis of genomic datasets (Conesa and Götz, 2008). It provides bioinformatics solution for functional annotation of (novel) sequences and the analysis of annotation data. Its main function is to assign information about the biological function of gene or protein sequences. Thus the functions of the genes that may get regulated by the circRNAs through interaction with miRNAs were known through this software.

By using the described materials and methodology the analysis was carried out and the results obtained there under are presented in the next chapter.

CHAPTER IV

RESULTS

In this chapter, a brief description of the results obtained under the study of prediction of circular RNAs in chickpea under drought and wilt stress as well as in soybean under drought and wilt stress are given. Besides, the results of the interaction of the predicted circular RNAs-miRNAs-mRNAs are also given.

4.1 Prediction of circular RNAs in chickpea under drought stress

4.1.1 Identification of circular RNAs

The clean trimmed reads of chickpea under control condition and drought stress condition were given separately as input for mapping in BWA-MEM and the SAM files generated were given further as input in CIRI. Based on the algorithm followed in CIRI, as discussed in materials and methods, 200 circular RNAs were identified in control condition (CT) and 285 circular RNAs were identified in drought stress condition (DR).

The result files obtained from control and stress conditions contain the identified circular RNA IDs, the chromosome IDs on which the circular RNAs are present, the start and end coordinates of the circular RNAs, the lengths of the circRNAs as well as the type of circular RNAs like exonic, intronic or intergenic.

In chickpea under control condition out of the 200 identified circular RNAs, 164 (82%) were intergenic circRNAs, 28 (14%) were exonic circRNAs and 8 (4%) were intronic circRNAs (**Fig. 1a**).

Whereas, under drought stress condition out of the 285 identified circular RNAs, 228 (80%) were intergenic circRNAs, 46 (16%) were exonic circRNAs and 11 (4%) were intronic circRNAs (**Fig. 1b**). Thus it can be observed that in both the conditions, highest number of circular RNAs are of intergenic type followed by exonic type.

Fig. 1a

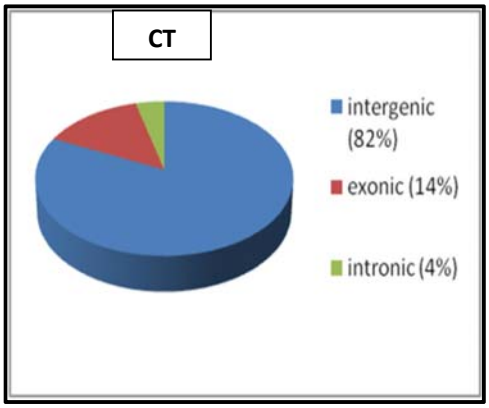

Fig. 1b

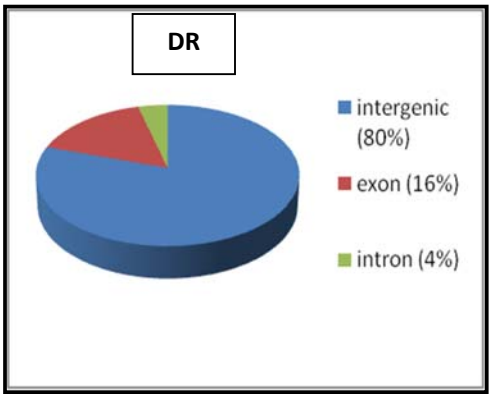

Figure1: Distribution of various types of circRNAs (**a**) under control and (**b**) under drought stress conditions in chickpea

A sample of the circular RNAs identified under control condition is given in **Table 1** and that of under drought stress condition is given in **Table 2**

Table 1: List (sample) of identified circRNAs and their length, chromosomal distribution under control condition of drought in chickpea

circRNA_ID	chr	circRNA_start	circRNA_end	circRNA_type	circRNA_length
NC_021160.1:882975|883213	NC_021160.1	882975	883213	intergenic	239
NC_021160.1:3357697|3357915	NC_021160.1	3357697	3357915	intergenic	219
NC_021160.1:3735039|3735350	NC_021160.1	3735039	3735350	intergenic	312
NC_021160.1:4923970|4924192	NC_021160.1	4923970	4924192	intergenic	223
NC_021160.1:6097199|6097435	NC_021160.1	6097199	6097435	intergenic	237
NC_021160.1:6781654|6781794	NC_021160.1	6781654	6781794	intergenic	141
NC_021160.1:6878045|6878272	NC_021160.1	6878045	6878272	intergenic	228
NC_021160.1:9029955|9030582	NC_021160.1	9029955	9030582	intergenic	628
NC_021160.1:9485873|9495992	NC_021160.1	9485873	9495992	intergenic	10120
NC_021160.1:9487726|9497857	NC_021160.1	9487726	9497857	intergenic	10132
NC_021160.1:9489520|9499731	NC_021160.1	9489520	9499731	intergenic	10212
NC_021160.1:9608392|9608917	NC_021160.1	9608392	9608917	intergenic	526
NC_021160.1:10603715|10625867	NC_021160.1	10603715	10625867	intergenic	22153
NC_021160.1:12227003|12227161	NC_021160.1	12227003	12227161	intergenic	159
NC_021160.1:13966080|14003288	NC_021160.1	13966080	14003288	intergenic	37209
NC_021160.1:13966080|14003928	NC_021160.1	13966080	14003928	intergenic	37849
NC_021160.1:14023144|14029972	NC_021160.1	14023144	14029972	intergenic	6829
NC_021160.1:18083094|18084439	NC_021160.1	18083094	18084439	intergenic	1346
NC_021160.1:21924746|21924945	NC_021160.1	21924746	21924945	intergenic	200
NC_021160.1:21928803|21930191	NC_021160.1	21928803	21930191	exon	1389
NC_021160.1:22723020|22746757	NC_021160.1	22723020	22746757	intergenic	23738
NC_021160.1:26129538|26129799	NC_021160.1	26129538	26129799	intergenic	262
NC_021160.1:34325825|34326682	NC_021160.1	34325825	34326682	intergenic	858
NC_021160.1:39060913|39075419	NC_021160.1	39060913	39075419	intergenic	14507

Table 2: List (sample) of identified circRNAs and their length, chromosomal distribution under drought stress condition in chickpea

circRNA_ID	chr	circRNA_start	circRNA_end	circRNA_length	circRNA_type
NC_021160.1:882975|883213	NC_021160.1	882975	883213	239	intergenic
NC_021160.1:3357697|3357915	NC_021160.1	3357697	3357915	219	intergenic
NC_021160.1:3735039|3735350	NC_021160.1	3735039	3735350	312	intergenic
NC_021160.1:4923970|4924192	NC_021160.1	4923970	4924192	223	intergenic
NC_021160.1:6097199|6097435	NC_021160.1	6097199	6097435	237	intergenic
NC_021160.1:6781654|6781794	NC_021160.1	6781654	6781794	141	exon
NC_021160.1:6878045|6878272	NC_021160.1	6878045	6878272	228	intergenic
NC_021160.1:9029955|9030582	NC_021160.1	9029955	9030582	628	intergenic
NC_021160.1:9485873|9495992	NC_021160.1	9485873	9495992	10120	intergenic
NC_021160.1:9487726|9497857	NC_021160.1	9487726	9497857	10132	intergenic
NC_021160.1:9489520|9499731	NC_021160.1	9489520	9499731	10212	intergenic
NC_021160.1:9608392|9608917	NC_021160.1	9608392	9608917	526	intergenic
NC_021160.1:10603715|10625867	NC_021160.1	10603715	10625867	22153	intergenic
NC_021160.1:12227003|12227161	NC_021160.1	12227003	12227161	159	exon
NC_021160.1:13966080|14003288	NC_021160.1	13966080	14003288	37209	intergenic
NC_021160.1:13966080|14003928	NC_021160.1	13966080	14003928	37849	intergenic
NC_021160.1:14023144|14029972	NC_021160.1	14023144	14029972	6829	intergenic
NC_021160.1:18083094|18084439	NC_021160.1	18083094	18084439	1346	intergenic
NC_021160.1:21924746|21924945	NC_021160.1	21924746	21924945	200	exon
NC_021160.1:21928803|21930191	NC_021160.1	21928803	21930191	1389	intergenic
NC_021160.1:22723020|22746757	NC_021160.1	22723020	22746757	23738	intergenic
NC_021160.1:26129538|26129799	NC_021160.1	26129538	26129799	262	intron
NC_021160.1:34325825|34326682	NC_021160.1	34325825	34326682	858	intergenic

4.1.2 Circular RNA sequences

Based on the circRNA_IDs of the identified circular RNAs under control and drought stress conditions, the sequences of the predicted circular RNAs were fetched from the reference genome of chickpea and a sample of the sequences of the circular RNAs identified under control condition is given in **Fig. 2**

Figure 2: FASTA sequences of circRNAs of chickpea under control condition

Similarly a sample of the sequences of the circular RNAs identified under drought stress condition is given in **Fig. 3**

Figure 3: FASTA sequences of circRNAs of chickpea under drought stress condition

4.1.3 Lengthwise distribution of the circular RNAs

The circular RNAs identified were found to be of different lengths having a range of 135 nucleotides to 308 nucleotides for exonic circRNAs, 146 nucleotides to 420 nucleotides for intronic circRNAs and 135 nucleotides to 173732 nucleotides for intergenic circRNAs under control condition; whereas under drought stress condition the length ranges from 135 nucleotides to 427 nucleotides for exonic circRNAs, 142 nucleotides to 463 nucleotides for intron nucleotides, and 136 nucleotides to 137434 nucleotides for intergenic circRNAs. The lengthwise distribution of different types of circular RNAs under both the conditions are given in **Fig. 4-6** (where x-axis denotes the length of circRNAs in terms of the number of nucleotides and y-axis denotes the number of circRNAs)

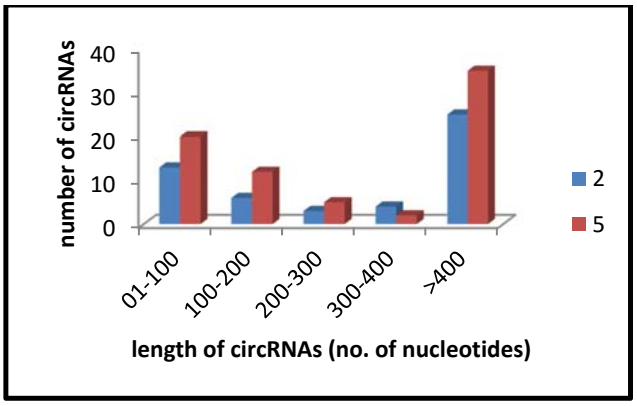

Figure 4: Lengthwise distribution of exonic circRNAs under control and drought stress condition in chickpea

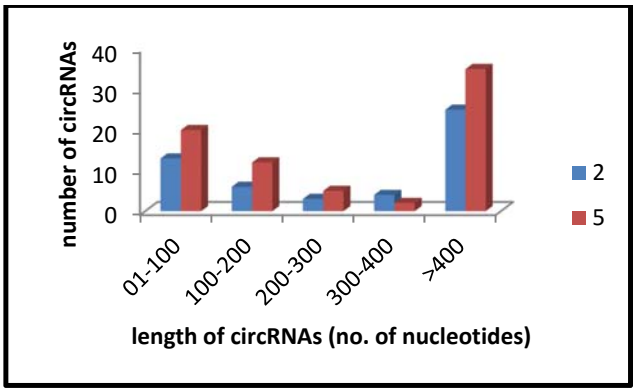

Figure 5: Lengthwise distribution of intronic circRNAs under control and drought stress condition in chickpea

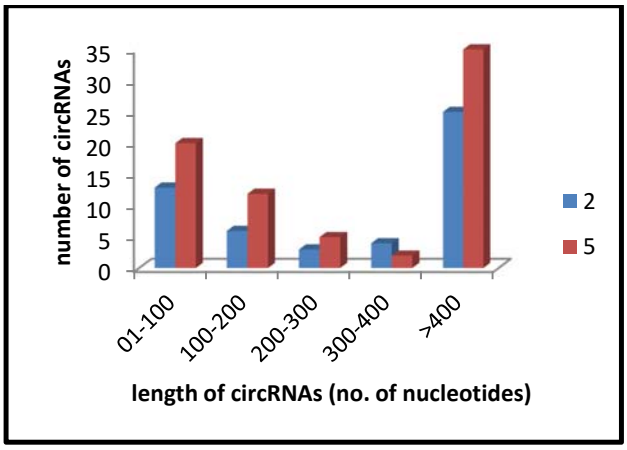

Figure 6: Lengthwise distribution of intergenic circRNAs under control and drought stress condition in chickpea

4.1.4 Distribution of the circular RNAs based on expression values (FPKM)

The expression values of the circular RNAs under control (CT) and drought (DR) conditions were obtained using the RSEM tool and were divided into six classes having values 0, 1-100, 100-200, 200-300, 300-400 and >400 and the distribution of the circular RNAs based on their expression values were observed. The expression values for most of the circular RNAs under both the conditions were more than 400 with a maximum of 341838.3 under control condition and 383558.9 under drought stress condition. 90 circular RNAs under control condition have expression value greater than 400; whereas in case of drought stress condition 130 circular RNAs have expression value greater than 400. The distribution of the circular RNAs based on the expression values as obtained through RSEM is shown in **Fig. 7** (where x-axis denotes the expression value of circRNAs in terms of FPKM value and y-axis denotes the number of circRNAs)

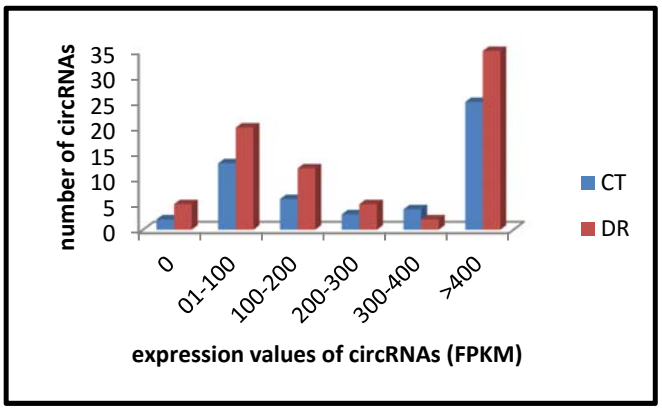

Figure 7: Distribution of circRNAs based on expression values under control and drought stress conditions in chickpea

4.1.5 Differentially expressed circular RNAs

From the identified circular RNAs under control and drought conditions, through the pipeline discussed in materials and methods, DESeq2 package of R software was used to get the differentially expressed circular RNAs. 44 differentially expressed circular RNAs were identified under control and drought stress condition of chickpea with a p-value ranging from 0.13 to 0.97.

A sample of the result is given in **Table 3**

Table 3: List (sample) of differentially expressed circRNAs in chickpea-drought

circRNA_ID	baseMean	log2FoldChange	lfcSE	stat	pvalue
NC_021160.1:13966080-14003928	189.0466949	-7.35E-06	0.000217157	-0.03383909	0.973005464
NC_021160.1:43850404-43854172	123.7624791	0.000111876	0.00016602	0.673870595	0.500393588
NC_021160.1:6878045-6878272	57141.59495	-6.28E-05	6.92E-05	-0.907551147	0.364115415
NC_021160.1:882975-883213	166.3531636	-0.000115118	7.97E-05	-1.445212822	0.148398116
NC_021160.1:9487726-9497857	1400.076463	-8.36E-06	1.52E-05	-0.551660277	0.58118113
NC_021161.1:3642891-3643045	98816.95218	-1.59E-05	2.12E-05	-0.751402778	0.452410291
NC_021161.1:6381136-6381276	304571.5174	0.003622166	0.004052836	0.893736147	0.37146308
NC_021162.1:15669142-15674247	254.0844465	0.000168561	0.00017872	0.943157085	0.34560056
NC_021162.1:26213382-26221978	33.47677008	7.96E-05	0.000110338	0.721442322	0.470637415
NC_021162.1:32406943-32431029	35.80398058	9.26E-05	7.26E-05	1.276149431	0.201902701
NC_021162.1:37942779-37956143	99.6631421	0.000126918	0.000132146	0.960442028	0.336832794
NC_021162.1:451283-452967	22.39784889	7.23E-05	9.74E-05	0.742709359	0.457657662
NC_021162.1:8535399-8536263	12.81683663	-7.33E-05	7.28E-05	-1.00735243	0.313765443
NC_021163.1:14092578-14092810	98433.48638	-8.85E-05	0.000258191	-0.342620368	0.731884077

Among the 44 identified circular RNAs, the log_2FC was less than 0 in 22 circular RNAs indicating down regulation (log2FoldChange<0) (**Table 4**), whereas 22 circular RNAs were found to be upregulated in terms of $log_2FC>0$ (**Table 5**)

Table 4: List of downregulated circRNAs in chickpea-drought

circRNA_ID	baseMean	log2FoldChange	lfcSE	stat	pvalue
NC_021160.1:882975-883213	166.3531636	-0.000115118	7.97E-05	-1.445212822	0.148398116
NC_021163.1:48301112-48330263	415.6803647	-0.000170241	0.000119474	-1.424917169	0.154181146
NC_021163.1:3742204-3742359	918.3656595	-0.000125897	9.69E-05	-1.299062867	0.193922355
NC_021163.1:8647041-8666318	229.663721	-0.000162285	0.00014092	-1.15161171	0.249480667
NC_021162.1:8535399-8536263	12.81683663	-7.33E-05	7.28E-05	-1.00735243	0.313765443
NC_021166.1:4382294-4382476	33799.89856	-3.06E-05	3.17E-05	-0.966874521	0.333606765
NC_021165.1:15094820-15095896	83.94052114	-0.000118885	0.000129763	-0.916163811	0.359580983
NC_021160.1:6878045-6878272	57141.59495	-6.28E-05	6.92E-05	-0.907551147	0.364115415
NC_021166.1:4382447-4382608	10892.64248	-2.51E-05	2.81E-05	-0.891714799	0.372545819
NC_021163.1:38962677-38965979	1100.880642	-0.000114849	0.000133905	-0.857692017	0.391062549
NC_021163.1:3057553-3057762	19063.87889	-2.74E-05	3.24E-05	-0.844175268	0.398571473
NC_021164.1:37736425-37736621	23219.93681	-2.91E-05	3.51E-05	-0.827741124	0.407817124
NC_021161.1:3642891-3643045	98816.95218	-1.59E-05	2.12E-05	-0.751402778	0.452410291
NC_021166.1:4382369-4382608	26613.37174	-1.31E-05	1.90E-05	-0.690668185	0.489774087
NW_004516369.1:383261-383416	1365.07972	-8.48E-05	0.000124395	-0.681503367	0.495553037
NC_021164.1:17855071-17863795	153.9263065	-0.000119225	0.000175293	-0.680146458	0.49641173
NC_021160.1:9487726-9497857	1400.076463	-8.36E-06	1.52E-05	-0.551660277	0.58118113
NC_021165.1:135464-141284	1037.38356	-6.90E-05	0.000187885	-0.367146452	0.71350978
NC_021163.1:14092578-14092810	98433.48638	-8.85E-05	0.000258191	-0.342620368	0.731884077
NC_021165.1:17714024-17738537	91.98719713	-3.42E-05	0.000171502	-0.199667926	0.8417403
NW_004516020.1:4076-15829	177.8477704	-2.87E-05	0.00021125	-0.135735453	0.89203042
NC_021160.1:13966080-14003928	189.0466949	-7.35E-06	0.000217157	-0.03383909	0.973005464

Table 5: List of upregulated circRNAs in chickpea-drought

circRNA_ID	baseMean	log2FoldChange	lfcSE	stat	pvalue
NC_021165.1:41831549-41836044	24097.56294	5.52E-06	5.02E-05	0.109840045	0.912536232
NC_021165.1:17714024-17738439	435.982358	6.51E-05	0.000289968	0.224570843	0.822313149
NC_021162.1:451283-452967	22.39784889	7.23E-05	9.74E-05	0.742709359	0.457657662
NC_021162.1:26213382-26221978	33.47677008	7.96E-05	0.000110338	0.721442322	0.470637415
NC_021165.1:58830645-58842270	1526.392373	8.50E-05	9.81E-05	0.866062503	0.386455887
NC_021165.1:4441168-4458979	1074.687519	8.96E-05	0.000170558	0.525501502	0.599234609
NC_021162.1:32406943-32431029	35.80398058	9.26E-05	7.26E-05	1.276149431	0.201902701
NW_004515823.1:299669-350369	46.24645813	9.54E-05	0.000115832	0.823965412	0.409959211
NC_021167.1:9169457-9169972	67.33925603	0.00010378	7.75E-05	1.339806662	0.180308209
NC_021163.1:25254569-25286975	57.80539236	0.000110559	0.000110372	1.001696536	0.31649018
NC_021160.1:43850404-43854172	123.7624791	0.000111876	0.00016602	0.673870595	0.500393588
NC_021164.1:10826352-10915704	66.5463818	0.000115632	0.000111675	1.035441098	0.300462961
NC_021164.1:41677974-41701939	196.5747365	0.000117335	7.96E-05	1.474940469	0.140228549
NC_021165.1:58265166-58265346	86895.44799	0.000119288	0.000161804	0.737238858	0.460977107
NC_021162.1:37942779-37956143	99.6631421	0.000126918	0.000132146	0.960442028	0.336832794
NC_021165.1:58831729-58843385	845.9173024	0.000131367	9.75E-05	1.347510686	0.177815814
NC_021165.1:59150102-59150852	383.1913964	0.000134068	0.000263064	0.50963928	0.6103042
NC_021166.1:38655845-38656106	229.483654	0.000156711	0.000124478	1.258947525	0.208049286
NC_021162.1:15669142-15674247	254.0844465	0.000168561	0.00017872	0.943157085	0.34560056
NC_021164.1:41798703-41814114	323.423067	0.000182974	0.000202511	0.903525142	0.366247251
NC_021165.1:23492486-23492644	138713.2403	0.000303976	0.000199731	1.52192806	0.128027107
NC_021161.1:6381136-6381276	304571.5174	0.003622166	0.004052836	0.893736147	0.37146308

4.2 Circular RNAs-miRNAs-mRNAs network in chickpea for drought stress trait

The identified differentially expressed circular RNAs of chickpea under control and drought stress conditions were used for identification of the miRNAs for which those circular RNAs act as sponges i.e the miRNAs that target the differentially expressed circular RNAs, using TargetFinder. In TargetFinder, the FASTA file containing the miRNA sequences downloaded for chickpea were used as input for small RNA sequences and the FASTA file containing the sequences of the differentially expressed circRNAs as the target sequence database. The result was retrieved from TargetFinder in 'table' format.

The results were sorted to get the miRNAs targeting one or more identified differentially expressed circular RNAs (**Table 6**).

Table 6: List (sample) of miRNAs targeting circRNAs in chickpea-drought

miRNA_ID	circRNA_ID	Coordinates		strand	Score	target_sequence	base_pairing	miRNA_sequence
Cat-miR1509a	NC_021163.1:8647041-8666318	4276	4295	+	4	GAUGUGAUUUCUUUGAUUAG	. ::::::::..:::::.	UGACACUAAAGGGACUAAUU
Cat-miR1509a	NC_021164.1:10826352-10915704	46000	46019	+	4	GUUCUUAUUUCCUUGAUUAA	..: : :::::.:::::::	UGACACUAAAGGGACUAAUU
Cat-miR1520e	NC_021165.1:17714024-17738537	14236	14255	+	2	GUAUCAUAUGUUACGUUAUU	::::::::::.:::::::	GAUAGUAUACAGUGCAAUAA
Cat-miR1520e	NC_021165.1:17714024-17738439	14236	14255	+	2	GUAUCAUAUGUUACGUUAUU	::::::::::.:::::::	GAUAGUAUACAGUGCAAUAA
Cat-miR166g-5p	NC_021162.1:32406943-32431029	194	214	+	3.5	CCACGGUCCAAGCAACAUUCC	:: ::. :::.::::::::	GGAGCUCGGUUUGUUGUAAG
Cat-miR171c-5p.2	NC_021165.1:17714024-17738537	20122	20142	+	3.5	UAUUGAGAUGGACCAAUGUCC	:::::. .:::::::::.::	AUAACUUGGCCUGGUUAUAG
Cat-miR171c-5p.2	NC_021165.1:17714024-17738439	20122	20142	+	3.5	UAUUGAGAUGGACCAAUGUCC	:::::. .:::::::::.::	AUAACUUGGCCUGGUUAUAG
Cat-miR172h-5p.1	NC_021163.1:48301112-48330263	24933	24951	+	4	GUGGAA-UUGAUGAUGUUCU	::::: ::::::::::.:.	CACUUAGAACUNCNACGAGG
Cat-miR1878	NC_021164.1:41677974-41701939	21286	21305	+	4	UUUAUUCUGAACGAGAUAAA	..:: :::::::.:::.:::	GAGUUAGACUUGUUCUGUUU
Cat-miR2081	NC_021165.1:17714024-17738537	1840	1858	+	4	AUCA-ACAUAUAACUCUAGC	:::: :::.::::::::.::	UAGUCUGUGUAUUGAGGUCA
Cat-miR2081	NC_021165.1:17714024-17738439	1840	1858	+	4	AUCA-ACAUAUAACUCUAGC	:::: :::.::::::::.::	UAGUCUGUGUAUUGAGGUCA
Cat-miR2661	NC_021164.1:10826352-10915704	39053	39074	+	4	UUGACUCAUUUUCUCAAAUUAA	.:: :.:::::::::::::.	GAC-GGGUAAAAGAGUUUAGU
Cat-miR395h	NC_021162.1:451283-452967	672	691	+	4	AUUUCACCAAAACACAUAAU	: ::: :::::::::: ::	UCAAGAGGUUUUGUGUAGUA
Cat-miR408-3p	NC_021164.1:37736425-37736621	141	162	+	4	GCCAGGACAGAGGCAGUGCUAU	:::::: :::::::::: :	CGGUCCCUUCUCCGUCACG-UA
Cat-miR408-5p.2	NC_021160.1:9487726-9497857	6703	6722	+	4	CAU-CGCAGCCUGUUUCUUGU	::: : :::::::::.:.:::	GUACGAGUCGGACAAGGGACA
Cat-miR408-5p.3	NC_021160.1:9487726-9497857	6703	6721	+	4	CAU-CGCAGCCUGUUUCUUG	::: : :::::::::.:.::	GUACGAGUCGGACAAGGGAC
Cat-miR408b	NC_021160.1:9487726-9497857	6703	6722	+	4	CAU-CGCAGCCUGUUUCUUGU	::: : :::::::::.:.:::	GUACGAGUCGGACAAGGGACA
Cat-miR419	NC_021163.1:48301112-48330263	26069	26088	+	4	AAACAUCAACAUCAUGCAUC	: :::::: ::::::.::::	UAUGUAGUAGUAGUAUGUAG
Cat-miR529	NW_004515823.1:299669-350369	34860	34879	+	3	ACCUGAGCUUUCUAUCUUCU	: ::: :::.::::::::::	UAGACACGAGAGAUAGAAGA
Cat-miR530a.1	NC_021165.1:17714024-17738537	16342	16362	+	4	ACAGGUGCAGGUGCAGGUGCA	::::::::::::::..::::	AUUCCACGUCCACGUUUACGU
Cat-miR530a.1	NC_021165.1:17714024-17738439	16342	16362	+	4	ACAGGUGCAGGUGCAGGUGCA	::::::::::::::..::::	AUUCCACGUCCACGUUUACGU
Cat-miR5554a-5p	NC_021164.1:10826352-10915704	79297	79315	+	4	AUCAU-GUUCAAAAGGUAUG	:.::: ::::::::::.:...	UGGUAACAAGUUUUCCGUGU
Cat-miR774b-3p	NC_021165.1:4441168-4458979	4667	4686	+	4	UAGAAGAUAAAAAAUAUGAAG	: :::::::::::::::	AAGCUCUAUUUUUUAUACUUA
Cat-NovmiR100a	NC_021165.1:4441168-4458979	15255	15273	+	3	GUUUUAUUCCAUCA-UCAAA	.:::::.::::::: :::::	UAAAAUGAGGUAGUUAGUUU

The total number of target miRNAs predicted were 54, but the number of unique miRNAs targeting the differentially expressed circRNAs were 40. This is because there are several miRNAs which target more than one differentially expressed circular RNA. Moreover, out of the total 44 differentially expressed circular RNAs, only 21 circular RNAs have the miRNA targets. Most of those 21 circular RNAs have more than one target miRNA sites. The maximum number of miRNA target sites is found to be 9 in circular RNA having ID as NC_021165.1:4441168-4458979. Only 9 circular RNAs of those 21 have a single miRNA target site, 3 circRNAs having 2 miRNA target sites, 5 circRNAs having 3 miRNA target sites, 2 circRNAs having 4 miRNA target sites, and 1 circRNA having 7 miRNA target sites.

The identification of the targeting miRNAs was followed by the identification of the mRNAs in chickpea that were targeted by those miRNAs in order to understand the whole interaction between the circRNAs, miRNAs and mRNAs. When the miRNAs get bind to their target circRNAs (sponges), these miRNAs are no more available to regulate the function of their target mRNAs. That's why identification of the target mRNAs of the predicted 40 miRNAs was done using psRNATarget with Expectation value <2. Out of 40 miRNAs, 30 miRNAs were found to target 145 unique mRNAs. The result file contains the IDs of both miRNAs and their target mRNAs, expectation value, unpaired potential energy, start and end coordinates of the alignment of both miRNAs and mRNAs, as well as the inhibition type**.** It was observed that the potential energy for all the interactions between the miRNAs and target mRNAs is -1. It was also found that there was only one translation inhibition and the rest 153 interactions results in cleavage inhibition. A sample list of mRNAs targeted by miRNAs is given in **Table 7.**

Table 7: List of mRNAs targeted by miRNAs in chickpea-drought

miRNA_Acc.	Target_Acc.	Expectation	UPE$	miRNA_start	miRNA_end	Target_start	Target_end	miRNA_aligned_fragment	Target_aligned_fragment	Inhibition
Cat-NovmiR107a	XM_004494079.3	1.5	-1	1	21	1115	1135	CUUUAUAUUCGUUUUUAUUAU	ACAAUAGAAAUGGAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_012716994.2	1.5	-1	1	21	106	126	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_027335624.1	1.5	-1	1	21	106	126	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_012716993.2	1.5	-1	1	21	106	126	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_004504586.3	1.5	-1	1	21	105	125	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_012716992.2	1.5	-1	1	21	106	126	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_027335623.1	1.5	-1	1	21	106	126	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_027335622.1	1.5	-1	1	21	106	126	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XM_027335621.1	1.5	-1	1	21	106	126	CUUUAUAUUCGUUUUUAUUAU	UAAAUAAACACGAAUAUAAAG	Cleavage
Cat-NovmiR107a	XR_003473317.1	1.5	-1	1	21	99	119	CUUUAUAUUCGUUUUUAUUAU	UGAAUGAAAAUGAAUAUGAAG	Cleavage
Cat-NovmiR107h	XM_004516942.3	1.5	-1	1	21	210	230	CUUUAUUUUCGUUUUUAUCAU	AAGAUCAAAACGAAGAUAAAG	Cleavage
Cat-NovmiR107h	XM_004498752.3	1.5	-1	1	21	511	531	CUUUAUUUUCGUUUUUAUCAU	AAGAAAAAAACGAAAAUGAAG	Cleavage
Cat-NovmiR107i	XM_004506168.3	1.5	-1	1	21	1112	1132	CUUUAUAUUUGUUUUUAUCAU	AAGAUGAAGACAAAUAUGAAG	Cleavage
Cat-NovmiR107j	XM_004506168.3	1.5	-1	1	21	1112	1132	CUUUAUAUUUGUUUUUAUCAU	AAGAUGAAGACAAAUAUGAAG	Cleavage
Cat-NovmiR107k	XM_004509994.3	1.5	-1	1	21	358	378	CUUUAUAUUCUUUUUAUCAU	AAGAUAAGAAAGAAUAUGAGG	Cleavage
Cat-NovmiR16	XM_004508782.3	1.5	-1	1	20	342	361	AAGAAGAGAGAAGUAAAUAC	GUAGUUACUUCUUUCUUCUU	Cleavage
Cat-NovmiR16	XM_004488672.3	1.5	-1	1	20	1452	1471	AAGAAGAGAGAAGUAAAUAC	GUAUUUUUUUCUCUCUUCUU	Cleavage
Cat-NovmiR16	XM_027336693.1	1.5	-1	1	20	345	364	AAGAAGAGAGAAGUAAAUAC	GUAGUUACUUCUUUCUUCUU	Cleavage
Cat-NovmiR16	XM_004491220.3	1.5	-1	1	20	51	70	AAGAAGAGAGAAGUAAAUAC	AUAUUUGCUUCUCUCUUCUG	Cleavage
Cat-NovmiR16	XM_004493632.3	1.5	-1	1	20	1761	1780	AAGAAGAGAGAAGUAAAUAC	UUAUUUAUUUUUCUCUUUUU	Cleavage
Cat-NovmiR16	XM_004497042.3	1.5	-1	1	20	935	954	AAGAAGAGAGAAGUAAAUAC	UUAUUUAUUUCUUUCUUUUU	Cleavage
Cat-NovmiR16	XM_004509176.3	1.5	-1	1	20	988	1007	AAGAAGAGAGAAGUAAAUAC	AUAUUUCCUUUUCUCUUCUU	Cleavage
Cat-NovmiR16	XM_004514250.3	1.5	-1	1	20	267	286	AAGAAGAGAGAAGUAAAUAC	AUGUUUUCUUCUCUCUUCUU	Cleavage

Blast2Go results:

To understand the functional role of mRNAs involved in the network, the GO analysis was performed by employing the BLAST2GO software. The GO functional categorization generated 136 annotations from the 145 predicted mRNAs that were targeted by miRNAs. In that, a total of 57, 64, and 15 mRNAs were classified as the first level classification of biological processes, molecular functions, and cellular components, respectively (**Fig. 8**). Among the genes involved in biological process, 38 and 42 mRNAs were classified into the categories of "metabolic process (GO: 0008152)" and "cellular process (GO: 0009987)", respectively. Interestingly, 13 and 20 mRNAs were categorized as the categories of "biological regulation (GO: 0065007)" and "response to stimulus (GO: 0050896)", respectively (**Fig. 9**). In the classification of molecular functions, two main classes were "binding (GO: 0005488)" and "catalytic activity (GO: 0003824", which had 54 and 34 predicted mRNAs, respectively (**Fig. 10**). When the predicted mRNAs were classified according to the cellular component classification, categories "cell (GO: 0005623)" and "cell part (GO: 0044464)" both made up the largest proportion of 16 predicted mRNAs, followed by "organelle (GO: 0043226)" that had 6 predicted mRNAs (**Fig. 11**). The GO analysis on predicted mRNAs showed that the targets of differentially expressed circRNAs under drought stress were associated with various functions involving in different cellular components, biological process, and molecular functions.

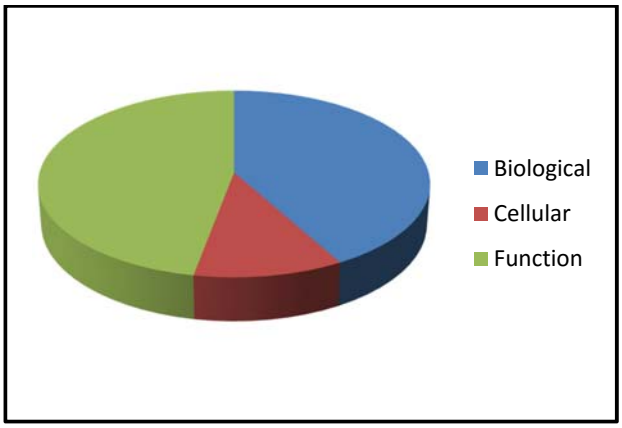

Figure 8: Differentially expressed circRNAs of chickpea under drought stress condition in GO terms of biological processes, cellular component and molecular functions

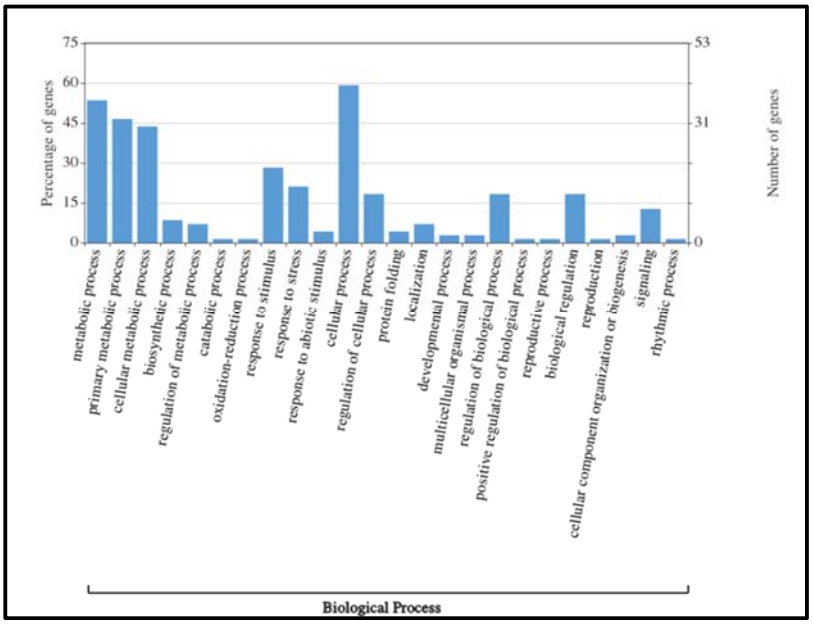

Figure 9: Classification of biological processes for mRNAs in chickpea-drought

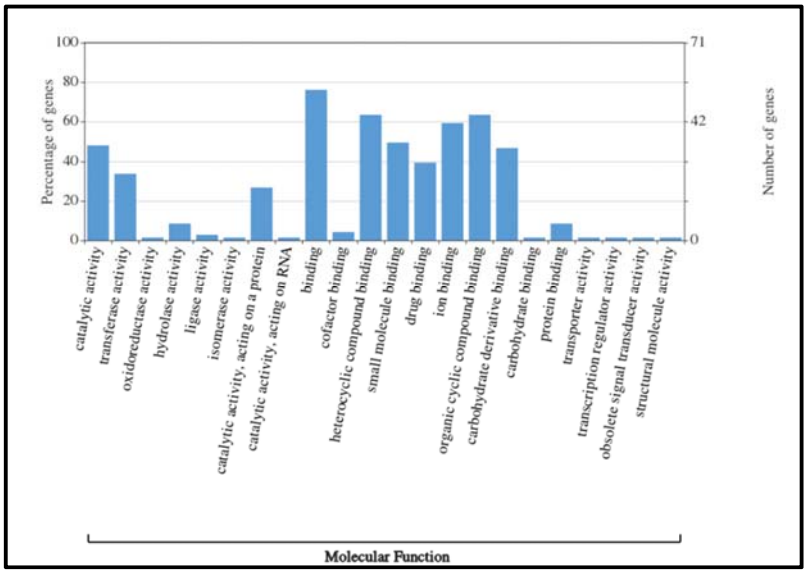

Figure 10: Classification of molecular functions for mRNAs in chickpea-drought

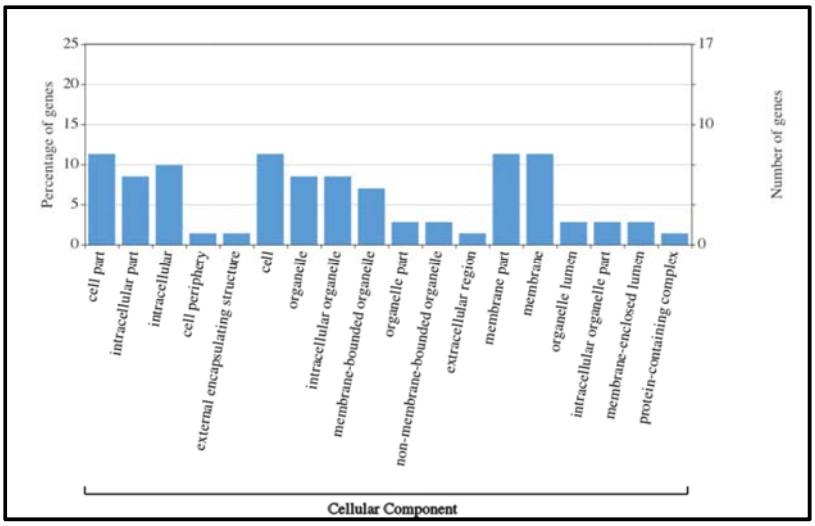

Figure 11: Classification of cellular components for mRNAs in chickpea-drought

The potential functions of the mRNAs categorized in the GO terms were also found and listed in **Table 8**

Table 8: GO based functions of the mRNAs in chickpea-drought
integral component of membrane, structural constituent of ribosome
signal transduction
response to stress, defense response, response to auxin
protein kinase activity, oxidoreductase activity
glycogen phosphorylase activity
zinc ion binding, hydrolase activity
transcription factor activity

These functions carried out by the mRNAs are involved in regulation of drought stress responsive mechanisms in chickpea. That's why when the miRNAs that are responsible for silencing these mRNAs get bind to the circRNAs, they remain no more available to regulate the mRNAs and so the genes become free to carry out their functions and help the chickpea plants in tolerating the drought stress conditions.

4.3 Prediction of circular RNAs in soybean under drought stress

4.3.1 Identified circular RNAs

The reads of soybean under control condition and drought stress conditions obtained after trimming were given as input for mapping in BWA-MEM and the SAM files generated were given further as input in CIRI. 57 circular RNAs were identified in control condition (CT) and 66 circular RNAs were identified in drought stress condition (DR) in CIRI.

The result files obtained from CIRI contain the identified circular RNA IDs, the chromosome IDs on which the circular RNAs are present, the start and end coordinates of the circular RNAs, the lengths of the circRNAs as well as the type of circular RNAs like exonic, intronic or intergenic.

In soybean under control condition out of the 57 identified circular RNAs, 50 (87.7%) were intergenic circRNAs, 5 (8.7%) were exonic circRNAs and 2 (3.5%) were intronic circRNAs (**Fig. 12a**). Whereas, under drought stress condition out of the 66 identified circular RNAs, 53 (80.3) were intergenic circRNAs, 11 (16.6%) were exonic circRNAs and 2 (3%) were intronic circRNAs (**Fig. 12b**). Thus it can be observed that in both the conditions most of the circular RNAs were of intergenic type followed by exonic type.

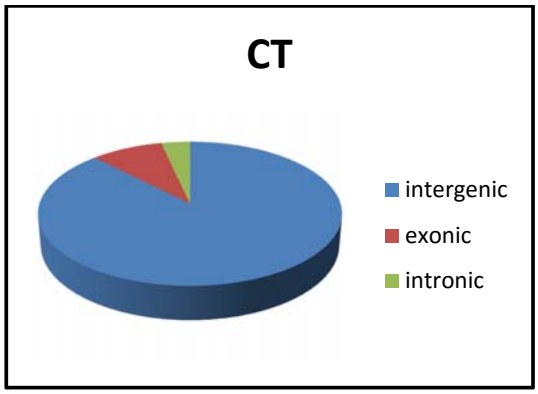

Fig. 12a

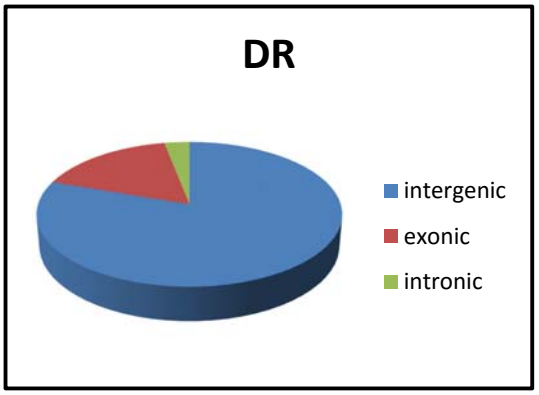

Fig. 12b

Figure 12 : Distribution of different types of circRNAs (**a**) under control and (**b**) under drought stress conditions in soybean

A sample of the circular RNAs identified under control condition is given in **Table 9** and that of under drought stress condition is given in **Table 10**

Table 9: List (sample) of identified circRNAs and their length, chromosomal distribution under control condition of drought in soybean

circRNA_ID	chr	circRNA_start	circRNA_end	circRNA_length	circRNA_type
11:8927977|8931149	11	8927977	8931149	3173	intergenic_region
11:8928133|8933250	11	8928133	8933250	5118	intergenic_region
11:26728832|26733600	11	26728832	26733600	4769	intergenic_region
11:30284970|30311997	11	30284970	30311997	27028	intergenic_region
11:30286028|30315056	11	30286028	30315056	29029	intergenic_region
12:977423|983081	12	977423	983081	5659	intergenic_region
12:3921774|3930152	12	3921774	3930152	8379	intergenic_region
13:16981593|16986402	13	16981593	16986402	4810	intergenic_region
15:12635940|12644708	15	12635940	12644708	8769	intergenic_region
15:12636344|12645124	15	12636344	12645124	8781	intergenic_region
15:42308943|42500339	15	42308943	42500339	191397	intergenic_region
16:5841009|5847710	16	5841009	5847710	6702	intergenic_region
16:6072989|6088828	16	6072989	6088828	15840	intergenic_region
16:34067902|34075561	16	34067902	34075561	7660	intergenic_region
16:36344782|36387037	16	36344782	36387037	42256	intergenic_region
16:36720161|36724310	16	36720161	36724310	4150	intergenic_region
16:37161171|37180330	16	37161171	37180330	19160	intergenic_region
16:37167407|37262135	16	37167407	37262135	94729	intergenic_region
18:1099021|1104695	18	1099021	1104695	5675	intergenic_region
18:6856493|6858812	18	6856493	6858812	2320	intergenic_region
18:7390770|7391226	18	7390770	7391226	457	exon
18:8118704|8189069	18	8118704	8189069	70366	intergenic_region
18:9125163|9149112	18	9125163	9149112	23950	intergenic_region

Table 10: List (sample) of identified circRNAs and their length, chromosomal distribution under drought stress condition in soybean

circRNA_ID	chr	circRNA_start	circRNA_end	circRNA_length	circRNA_type
10:44644870|44656965	10	44644870	44656965	12096	intergenic_region
10:49351762|49393985	10	49351762	49393985	42224	intergenic_region
11:406172|414451	11	406172	414451	8280	intergenic_region
11:8927977|8931149	11	8927977	8931149	3173	intergenic_region
11:8928133|8933250	11	8928133	8933250	5118	intergenic_region
11:8933378|8940805	11	8933378	8940805	7428	intergenic_region
11:9832652|9890747	11	9832652	9890747	58096	intergenic_region
11:9860230|9873017	11	9860230	9873017	12788	intergenic_region
11:9860230|9881467	11	9860230	9881467	21238	intergenic_region
11:9860230|9891731	11	9860230	9891731	31502	intergenic_region
11:9866857|9892608	11	9866857	9892608	25752	intergenic_region
11:9873816|9892608	11	9873816	9892608	18793	intergenic_region
12:3902873|3920705	12	3902873	3920705	17833	intergenic_region
12:3902873|3928991	12	3902873	3928991	26119	intergenic_region
12:3921774|3930152	12	3921774	3930152	8379	intergenic_region
12:4759886|4768531	12	4759886	4768531	8646	intergenic_region
14:3731889|3757602	14	3731889	3757602	25714	intergenic_region
14:46987313|47005711	14	46987313	47005711	18399	intergenic_region
15:919572|942558	15	919572	942558	22987	intergenic_region
15:1715649|1731295	15	1715649	1731295	15647	intergenic_region
15:12636344|12645124	15	12636344	12645124	8781	intergenic_region
15:48531659|48546480	15	48531659	48546480	14822	intergenic_region
16:5841009|5847710	16	5841009	5847710	6702	intergenic_region

4.3.2 Circular RNA sequences

The circRNA_IDs of the identified circular RNAs under control and drought stress conditions were used to fetch their sequences from the reference genome of soybean and a sample of the sequences of the circular RNAs under control condition is given in **Fig. 13**

Figure 13 : FASTA sequences of circRNAs of soybean under control condition

```
>11 dna:chromosome chromosome:Glycine_max_v2.1:11:1:34766867:1 REF:8927977-8931149
TATTGGCGTGAGAGTAATAAAAGAATCATATAGTAACACAAGTATCATAAGATCAAGTAACGCAAACAGATCAGAGATCACAGATCAGAAAGGGATCGAGAAAATTCCGAATCCGAATCGTGTGAGATCAC
>11 dna:chromosome chromosome:Glycine_max_v2.1:11:1:34766867:1 REF:8928133-8933250
TTCGTTTCAACGATTTCGCCGTACTGAGAGAAGGCTCTCTCGAGAGCTTGGTCGTCAGTGGCCCAAGCGAGCCCACCGACAAAGCAACGGTACTCAACATCTGAAGAAGCCATCGAAAATTGTATAAACCCTA
>11 dna:chromosome chromosome:Glycine_max_v2.1:11:1:34766867:1 REF:26728832-26733600
GAATTCAAGGTGGGTACATGCAATGAAGTGGTCAGGCAAAAAGGCCTTTGGAAAATCTCCTACAGTGAAATTTGTGGTTGATGGTTCAAAAGCAGGGTCTCTTAATAGCTATGGACCTCTTTCTTTCCTCAAGG
>11 dna:chromosome chromosome:Glycine_max_v2.1:11:1:34766867:1 REF:30284970-30311997
GTGACAATTTTAATCCAAAGGCAAGAAGATGTTGCTCCTGCCAAGAGGGCAACTCCAAAGGCCATTTCAAAAAGTGAAACACAAATCCAGAATACCTGACAATGATTAGTCAGGAACAACCAAAGGTCATG
>11 dna:chromosome chromosome:Glycine_max_v2.1:11:1:34766867:1 REF:30286028-30315056
TGCATGTGAAGGAAAAATGTAATTGGAAATATAATTGTCAAAGTAAGAAAAGTGCACATTGCTGCAAGTAGTGGATGTCTCTTCCATCTCAACAACGGTACCTGAAAAGTTTCAAAACAACGCTATCACCATC
>12 dna:chromosome chromosome:Glycine_max_v2.1:12:1:40091314:1 REF:977423-983081
TTGCTACAATGTATGCAGCCATGCGAGATACAGACAGACTACATGCCAAAAGATAGGTTCAAGGAAAATTTCTGGAATGACTGGCGAAAGTTGCTTGGAAGAAATCATGTTATCCAGAACTTGAAAGATTGT
>12 dna:chromosome chromosome:Glycine_max_v2.1:12:1:40091314:1 REF:3921774-3930152
AATGGCGCAAATGTAAAGGGATATCATGTATGGTCTTTGTTCGATAATTTTGAGTGGTCTTCAGGTTATACAGTACGGTTCGGAATGATCTATATAGATTACAAAAATGATATGAAGAGACACAAAAAACTCT
>13 dna:chromosome chromosome:Glycine_max_v2.1:13:1:45874162:1 REF:16981593-16986402
TGCATATCTGGTGTCAATCCTAGTGCTGCCCACAATTGCAAGCAGATACTTGCGCGTAGTGAATTAGAATGAGTCAAGTCCTGCCCCCGTGGGATAAAGTCCTTGCAGTGGTGTGAGGATTCAATTTTCTCGGA
>15 dna:chromosome chromosome:Glycine_max_v2.1:15:1:51756343:1 REF:12635940-12644708
GTACTAAGGCCATTAGAAGCATATTGATTCACTTGCCAACATTCATGAAGCAAGAGTTAGGCCCTCACATATTTGGAAAGATGAACAGACTACAATTTCTGGAAATTTCTGGGAAATGTGAAGAAGATAGCTT
>15 dna:chromosome chromosome:Glycine_max_v2.1:15:1:51756343:1 REF:12636344-12645124
ATCTGGTGAATTTAAAAGAACTTCACCTCACTGACTCCAAGATGTTAGAGGAGCTGCCAGATTTATCGAATGCCACAAATCTTGAAGTACTGGTTCTCGAGGGTTGTTCTATGTTAACTACGGTGCATCCATCTA
>15 dna:chromosome chromosome:Glycine_max_v2.1:15:1:51756343:1 REF:42308943-42500339
TCCATTCCACGAAAGAAGGGACCAAGAAGGCCCCCGAACTGATGAGTGAGTCAAAGGTTTCCGATTGTGCATGCGTCAAATGGTTTAACTTTGACTCGATCTCTACCACCACAACATCACGATATGTTGTTACA
AGCAAGAAAATTTTAATTAGAAAATAACAGAATTAAAACATGTTGTTTCCCATTGATTCACAAGAAAGTCTTTTATCCTAGGTTATGAGGATTTATCCCTAACCAGTTCAACCACTTAATCCAACCCTAAATTAA
TTGGAACGAGAAACCAAGAAGGCCCGAAAGGAAGAACACAACCAAAGCAAAGTTTTGAGGGGCTTTATAGGGCAGAAATAGTGAGCTCAAGCTCCGAAGAGGTGAAAGGAATCATCATGGGTCAAAGGC
ATTCAACCCCCCCTTCTTAATTATTTCGAGGCCATTTGATCCAACACTCCAATATAGTAGTGTCAAGGTGGGTAAAGAAGCCCAGTTCTCCTTTCTACAGCATGCGAACAATGCTATAAATGTAGGGGAGGTGTC
```

Similarly a sample of the sequences of the circular RNAs identified under drought stress condition is given in **Fig. 14**

Figure 14: FASTA sequences of circRNAs of soybean under drought condition

4.3.3 Lengthwise distribution of the circular RNAs

When the lengths of the circRNAs were observed it was found that they have varying lengths ranging from 162 nucleotides to 641 nucleotides for exonic circRNAs, 652 nucleotides to 670 nucleotides for intronic circRNAs and 2320 nucleotides to 191397 nucleotides for intergenic circRNAs under control condition; whereas under drought stress condition the length ranges from 147 nucleotides to 651 nucleotides for exonic circRNAs, 620 nucleotides to 677 nucleotides for intron nucleotides, and 2320 nucleotides to 174240 nucleotides for intergenic circRNAs. The lengthwise distribution of the circular RNAs under both the conditions are given in **Fig. 15-17** (where x-axis denotes the length of circRNAs in terms of the number of nucleotides and y-axis denotes the number of circRNAs)

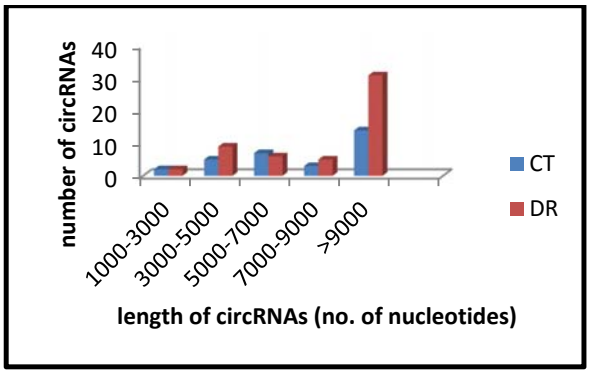

Figure 15: Lengthwise distribution of exonic circRNAs under control and drought stress conditions in soybean

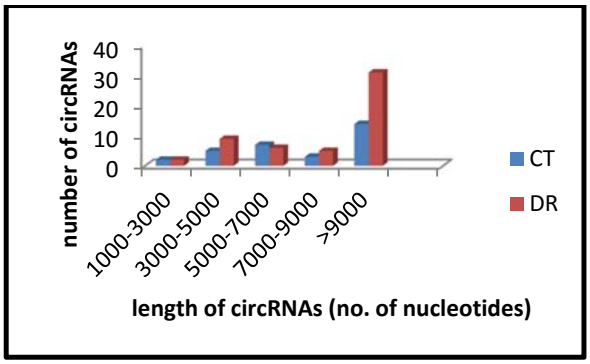

Figure 16: Lengthwise distribution of intronic circRNAs under control and drought stress conditions in soybean

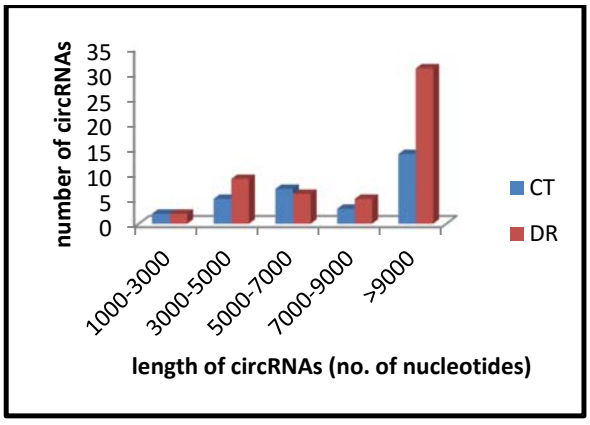

Figure 17: Lengthwise distribution of intronic circRNAs under control and drought stress conditions in soybean

4.3.4 Distribution of the circular RNAs based on expression values (FPKM)

To observe the distribution of the expression values of the circular RNAs under control and drought conditions obtained from RSEM, they were divided into six classes having values 0, 1-100, 100-200, 200-300, 300-400 and >400. The expression values for most of the circular RNAs under both the conditions is more than 400 with a maximum of 833914.4 under control condition and 554353.8 under drought stress condition. 35 circular RNA under control condition have expression value greater than 400, whereas in case of drought stress condition 38 circular RNAs have expression value greater than 400. The distribution of the circular RNAs based on the expression values as obtained through RSEM is shown in **Fig. 18** (where x-axis denotes the expression value of circRNAs in terms of FPKM value and y-axis denotes the number of circRNAs)

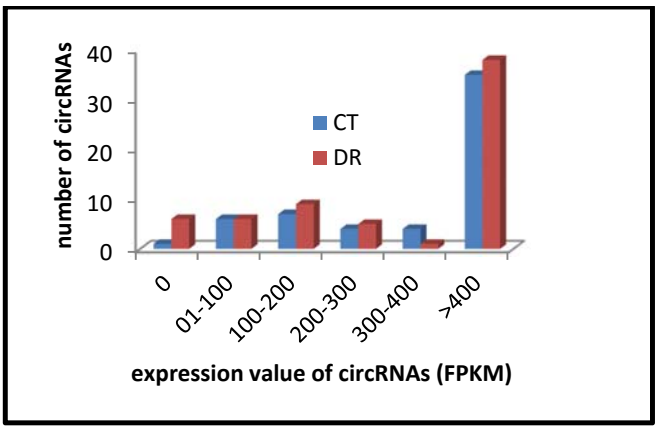

Figure 18: Distribution of circRNAs based on expression values under control and drought stress conditions in soybean

4.3.5 Differentially expressed circular RNAs

The identified circular RNAs under the two conditions (control and drought stress) were used to find out the differentially expressed circRNAs using R package DESeq2. 23 differentially expressed circular RNAs were identified under control and drought stress condition of soybean with a p-value ranging from 0.07 to 1. Only one circular RNA was identified as differentially expressed with p-value less than 0.1, originated from the intergenic region and is upregulated.

A sample of the result is given in **Table 11**

Table 11: List (sample) of differentially expressed circRNAs in soybean-drought

circRNA_ID	baseMean	log2FoldChange	lfcSE	stat	pvalue
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:49532142-49539485	147.8680179	0.846455303	0.47461189	1.783468389	0.074510089
9_dna:chromosome_chromosome:Glycine_max_v2.1:9:1:50189764:1_REF:1827313-1832264	9529.181292	-0.832494619	0.539896489	-1.541952274	0.123085192
8_dna:chromosome_chromosome:Glycine_max_v2.1:8:1:47837940:1_REF:7434887-7443884	255.5519428	0.682062764	0.472055786	1.444877457	0.148492309
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:45697097-45728257	561.502906	0.634257297	0.569995781	1.112740335	0.265819967
9_dna:chromosome_chromosome:Glycine_max_v2.1:9:1:50189764:1_REF:41497262-41497902	2225.065073	0.596924443	0.554284616	1.076927676	0.281512578
11_dna:chromosome_chromosome:Glycine_max_v2.1:11:1:34766867:1_REF:8928133-8933250	1819.102368	-0.57176334	0.572327382	-0.999014477	0.317787678
12_dna:chromosome_chromosome:Glycine_max_v2.1:12:1:40091314:1_REF:3921774-3930152	604.6443325	0.555032052	0.572763757	0.969041852	0.332524308
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:45830034-45836851	7037.381348	-0.31656812	0.333069854	-0.950455635	0.341880786
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:1099021-1104695	1178.889366	0.402468073	0.425408729	0.946073848	0.344110922
15_dna:chromosome_chromosome:Glycine_max_v2.1:15:1:51756343:1_REF:12636344-12645124	984.7336515	0.447290266	0.481950182	0.928084026	0.353363985
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:50841293-50844818	769.5785904	0.436836954	0.560569534	0.779273449	0.435818652
11_dna:chromosome_chromosome:Glycine_max_v2.1:11:1:34766867:1_REF:8927977-8931149	10669.9713	-0.436041834	0.568929274	-0.766425378	0.443423237
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:44630646:1_REF:43200672-43217184	50.57804171	-0.371806739	0.521193108	-0.713376161	0.47561302
3_dna:chromosome_chromosome:Glycine_max_v2.1:3:1:45779781:1_REF:5941334-5980395	291.7359305	-0.211563905	0.457823205	-0.462108303	0.644003656
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:44630646:1_REF:4027348-4042367	112.2722809	0.202399978	0.441479591	0.45845829	0.646623222
3_dna:chromosome_chromosome:Glycine_max_v2.1:3:1:45779781:1_REF:4794675-4894600	148.8625502	-0.191737061	0.43224672	-0.443582455	0.657344493
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:6856493-6858812	2644.338057	0.080553027	0.264120491	0.304985905	0.760376888
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:4487693-4497730	251.9423074	-0.100815679	0.439660414	-0.229303517	0.81863302
16_dna:chromosome_chromosome:Glycine_max_v2.1:16:1:37887014:1_REF:5841009-5847710	37.48778728	-0.112873053	0.551928718	-0.204506577	0.837957653
8_dna:chromosome_chromosome:Glycine_max_v2.1:8:1:47837940:1_REF:47391302-47397126	5290.867515	-0.032367874	0.226942964	-0.14262559	0.886585886
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:8118704-8189069	199.8348138	0.02937779	0.426543735	0.06887404	0.945089882
16_dna:chromosome_chromosome:Glycine_max_v2.1:16:1:37887014:1_REF:36720161-36724310	1536.754063	-0.000116711	0.515586682	-0.000226366	0.999819386
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:44630646:1_REF:495938-513138	3731.401881	-1.47E-09	0.202502704	-7.28E-09	0.999999994

Among the 23 identified circular RNAs, 12 circular RNAs were found to be down regulated in terms of $log_2FC<0$ (**Table 12)**, whereas 11 were found to be upregulated in terms of $log_2FC>0$ (**Table 13**)

Table 12: List of downregulated circRNAs in soybean-drought

circRNA_ID	baseMean	log2FoldChange	lfcSE	stat	pvalue
9_dna:chromosome_chromosome:Glycine_max_v2.1:9:1:501	9529.181292	-0.832494619	0.539896	-1.54195	0.123085
11_dna:chromosome_chromosome:Glycine_max_v2.1:11:1:3	1819.102368	-0.57176334	0.572327	-0.99901	0.317788
11_dna:chromosome_chromosome:Glycine_max_v2.1:11:1:3	10669.9713	-0.436041834	0.568929	-0.76643	0.443423
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:446	50.57804171	-0.371806739	0.521193	-0.71338	0.475613
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:514	7037.381348	-0.31656812	0.33307	-0.95046	0.341881
3_dna:chromosome_chromosome:Glycine_max_v2.1:3:1:457	291.7359305	-0.211563905	0.457823	-0.46211	0.644004
3_dna:chromosome_chromosome:Glycine_max_v2.1:3:1:457	148.8625502	-0.191737061	0.432247	-0.44358	0.657344
16_dna:chromosome_chromosome:Glycine_max_v2.1:16:1:3	37.48778728	-0.112873053	0.551929	-0.20451	0.837958
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:514	251.9423074	-0.100815679	0.43966	-0.2293	0.818633
8_dna:chromosome_chromosome:Glycine_max_v2.1:8:1:478	5290.867515	-0.032367874	0.226943	-0.14263	0.886586
16_dna:chromosome_chromosome:Glycine_max_v2.1:16:1:3	1536.754063	-0.000116711	0.515587	-0.00023	0.999819
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:446	3731.401881	-1.47E-09	0.202503	-7.28E-09	1

Table 13: List of upregulated circRNAs in soybean-drought

circRNA_ID	baseMean	log2FoldChang	lfcSE	stat	pvalue
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58(	199.834814	0.02937779	0.426544	0.068874	0.94509
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58(	2644.33806	0.080553027	0.26412	0.304986	0.760377
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:4463(	112.272281	0.202399978	0.44148	0.458458	0.646623
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58(	1178.88937	0.402468073	0.425409	0.946074	0.344111
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416	769.57859	0.436836954	0.56057	0.779273	0.435819
15_dna:chromosome_chromosome:Glycine_max_v2.1:15:1:517	984.733652	0.447290266	0.48195	0.928084	0.353364
12_dna:chromosome_chromosome:Glycine_max_v2.1:12:1:40(	604.644333	0.555032052	0.572764	0.969042	0.332524
9_dna:chromosome_chromosome:Glycine_max_v2.1:9:1:50185	2225.06507	0.596924443	0.554285	1.076928	0.281513
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416	561.502906	0.634257297	0.569996	1.11274	0.26582
8_dna:chromosome_chromosome:Glycine_max_v2.1:8:1:47837	255.551943	0.682062764	0.472056	1.444877	0.148492
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58(	147.868018	0.846455303	0.474612	1.783468	0.07451

4.4 CircRNA-miRNA-mRNA network in soybean for drought stress trait

The identified differentially expressed circular RNAs of soybean under control and drought stress condition were used for identification of the miRNAs for which those circular RNAs act as sponges i.e the miRNAs targeting those differentially expressed circular RNAs, using psRNATarget. In psRNATarget the FASTA file containing the miRNA sequences downloaded for soybean were used as input for small RNA sequences and the FASTA file containing the sequences of the differentially expressed circRNAs as the target sequence database. A sample list of circRNAs targeted by miRNAs is given in **Table 14**

Table 14: List (sample) of miRNAs targeting circRNAs in soybean-drought

miRNA_Acc.	Target_Acc.	Expectation	UPE$	miRNA_start	miRNA_end	Target_start	Target_end	miRNA_aligned_f	Target_aligned	Inhibition
gma-miR1514a-5p	7_dna:chromoson	0.5	-1	1	21	1184	1204	UUCAUUUUUAAA	AAUGCCUAUU(	Cleavage
gma-miR4378b	18_dna:chromoso	0.5	-1	1	24	54833	54856	UAGAACUGUCUU	GACACACAUU(	Cleavage
gma-miR1514b-5p	7_dna:chromoson	1.5	-1	1	21	1184	1204	UUCAUUUUUAAA	AAUGCCUAUU(	Cleavage
gma-miR10193a	18_dna:chromoso	2	-1	1	22	26899	26920	UCUGAAACCGAU(	AUUAGUUAAC	Cleavage
gma-miR10193b	18_dna:chromoso	2	-1	1	22	26899	26920	UCUGAAACCGAU(	AUUAGUUAAC	Cleavage
gma-miR10193c	18_dna:chromoso	2	-1	1	22	26899	26920	UCUGAAACCGAU(	AUUAGUUAAC	Cleavage
gma-miR10193d	18_dna:chromoso	2	-1	1	22	26899	26920	UCUGAAACCGAU(	AUUAGUUAAC	Cleavage
gma-miR10200	18_dna:chromoso	2	-1	1	21	26544	26564	AGGUUUUAAAGA	CUUGUAAUUU(	Cleavage
gma-miR10407a	3_dna:chromoson	2	-1	1	24	16858	16881	AGUUAACGGAUG	GAUAAAUUUA	Cleavage
gma-miR10407a	3_dna:chromoson	2	-1	1	24	93765	93788	AGUUAACGGAUG	GAUAAAUUUA	Cleavage
gma-miR10407b	3_dna:chromoson	2	-1	1	24	16858	16881	AGUUAACGGAUG	GAUAAAUUUA	Cleavage
gma-miR10407b	3_dna:chromoson	2	-1	1	24	93765	93788	AGUUAACGGAUG	GAUAAAUUUA	Cleavage
gma-miR1513a-5p	7_dna:chromoson	2	-1	1	21	4198	4218	UGAGAGAAAGCC	GUAAGUCAUG(	Cleavage
gma-miR1513b	7_dna:chromoson	2	-1	1	21	4198	4218	UGAGAGAAAGCC	GUAAGUCAUG(	Cleavage
gma-miR1533	3_dna:chromoson	2	-1	1	19	54784	54802	AUAAUAAAAAUA	UUAUUAUUAU	Cleavage
gma-miR1533	3_dna:chromoson	2	-1	1	19	54805	54823	AUAAUAAAAAUA	UUAUUAUUAU	Cleavage
gma-miR1533	3_dna:chromoson	2	-1	1	19	88405	88423	AUAAUAAAAAUA	UUAUUAUUAU	Cleavage
gma-miR1533	8_dna:chromoson	2	-1	1	19	4107	4125	AUAAUAAAAAUA	UUAUUAUUAU	Cleavage
gma-miR169l-3p	18_dna:chromoso	2	-1	1	22	53270	53291	CGGGCAAGUUGU(	GAUGUCAAAA	Cleavage
gma-miR6299	3_dna:chromoson	2	-1	1	22	10325	10346	AUUUAAAAUUAU	UAUCAAAUUA	Translation
gma-miR9722	18_dna:chromoso	2	-1	1	21	2901	2921	UAAUAGAGGGAA	GACAUUUUUU(	Cleavage

The total number of miRNAs targeting the differentially expressed circRNAs were predicted to be 21, but the number of unique miRNAs targeted were 17. This is because there are several miRNAs which target more than one differentially expressed circular RNA. Moreover, out of the total 23 differentially expressed circular RNAs, only 7 circular RNAs have miRNA targets. Most of those 7 circular RNAs have more than one target miRNA sites. The maximum number of miRNA target sites is found to be 7 in 2 circular RNAs. Other than these 2, only 3 circular RNAs of those 7 have a single miRNA target site, and 2 circRNAs have 2 miRNA target sites. It was observed that the potential energy of interaction between the miRNAs and the circRNAs is -1 in all the cases with 20 cleavage type of inhibition and a single translation type of inhibition.

The identification of the miRNAs was followed by the identification of the mRNAs in soybean that were targeted by those miRNAs in order to understand the whole interaction between the circRNAs, miRNAs and mRNAs. The identification of the target mRNAs of the predicted 17 miRNAs was done using psRNATarget with Expectation value ≤ 2. Out of 17 miRNAs, 12 miRNAs were found to target 281 unique mRNAs and total 366 interactions. It was observed that the potential energy for all the interactions between the miRNAs and target mRNAs is -1. It was also found that there was only one translation inhibition and the rest 365 interactions results in cleavage inhibition. A sample list of mRNAs targeted by miRNAs is given in **Table 15**

Table 15: List (sample) of miRNAs targeting mRNAs in soybean-drought

miRNA_Acc.	Target_Acc.	Expectatic	UPE$	miRNA_sta	miRNA_end	Target_start	Target_end	miRNA_align	Target_align	Inhibition
gma-miR1040	Glyma.11G1	0	-1	1	24	2605	2628	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1040	Glyma.11G1	0	-1	1	24	2624	2647	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1040	Glyma.11G1	0	-1	1	24	2669	2692	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1040	Glyma.11G1	0	-1	1	24	3258	3281	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1040	Glyma.11G1	0	-1	1	24	2605	2628	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1040	Glyma.11G1	0	-1	1	24	2624	2647	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1040	Glyma.11G1	0	-1	1	24	2669	2692	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1040	Glyma.11G1	0	-1	1	24	3258	3281	AGUUAACGG	GACAAAUU	Cleavage
gma-miR1533	Glyma.14G0	0	-1	1	19	267	285	AUAAUAAAA	UCAUUAUU	Cleavage
gma-miR1533	Glyma.14G0	0	-1	1	19	267	285	AUAAUAAAA	UCAUUAUU	Cleavage
gma-miR1533	Glyma.14G0	0	-1	1	19	267	285	AUAAUAAAA	UCAUUAUU	Cleavage
gma-miR1533	Glyma.14G0	0	-1	1	19	267	285	AUAAUAAAA	UCAUUAUU	Cleavage
gma-miR1533	Glyma.14G0	0	-1	1	19	267	285	AUAAUAAAA	UCAUUAUU	Cleavage
gma-miR1020	Glyma.13G1	0.5	-1	1	21	621	641	AGGUUUUAA	AUUUUAAU	Cleavage
gma-miR1513	Glyma.07G1	0.5	-1	1	21	104	124	UGAGAGAAA	GUAAGUCA	Cleavage
gma-miR1513	Glyma.07G1	0.5	-1	1	21	89	109	UGAGAGAAA	GCAAGUCA	Cleavage
gma-miR1513	Glyma.07G1	0.5	-1	1	21	104	124	UGAGAGAAA	GUAAGUCA	Cleavage

Blast2Go results

To understand the functional role of the mRNAs involved in the network in soybean under drought stress condition, the GO analysis was performed by employing the BLAST2GO software. The GO functional categorization generated 16 annotations from the 281 predicted mRNAs. In that a total of 6, 5, and 5 mRNAs were classified as the first level classification of biological processes, molecular functions, and cellular components, respectively (**Fig. 19**). Among the genes involved in biological processes, 3 mRNAs were classified into each of the categories of "metabolic process (GO: 0008152)" and "cellular process (GO: 0009987)", respectively. Interestingly, 2 mRNAs were categorized in each of the categories of "biological regulation (GO: 0065007)" and "response to stimulus (GO: 0050896)", respectively (**Fig. 20**). In the classification of molecular functions, two main categories were "binding (GO: 0005488)" and "catalytic activity (GO: 0003824)", which had 3 and 4 predicted mRNAs, respectively (**Fig. 21**). When the predicted mRNAs were classified according to the cellular component classification, categories "cell (GO: 0005623)" and "cell part (GO: 0044464)" both made up the largest proportion of 8 predicted mRNAs (**Fig. 22**). The GO analysis on predicted mRNAs showed that the targets of differentially expressed circRNAs under drought stress in soybean were associated with various functions involving in different cellular components, biological process, and molecular functions.

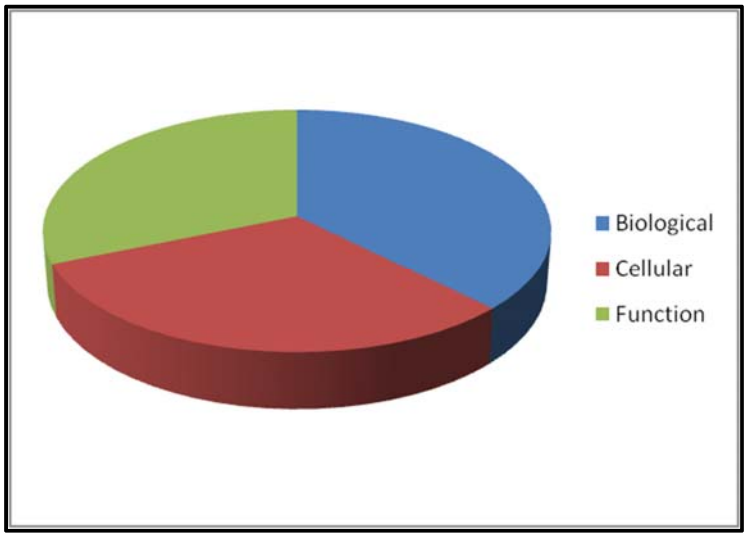

Figure 19: Differentially expressed circRNAs of soybean under drought stress condition in GO terms of biological processes, cellular components and molecular functions

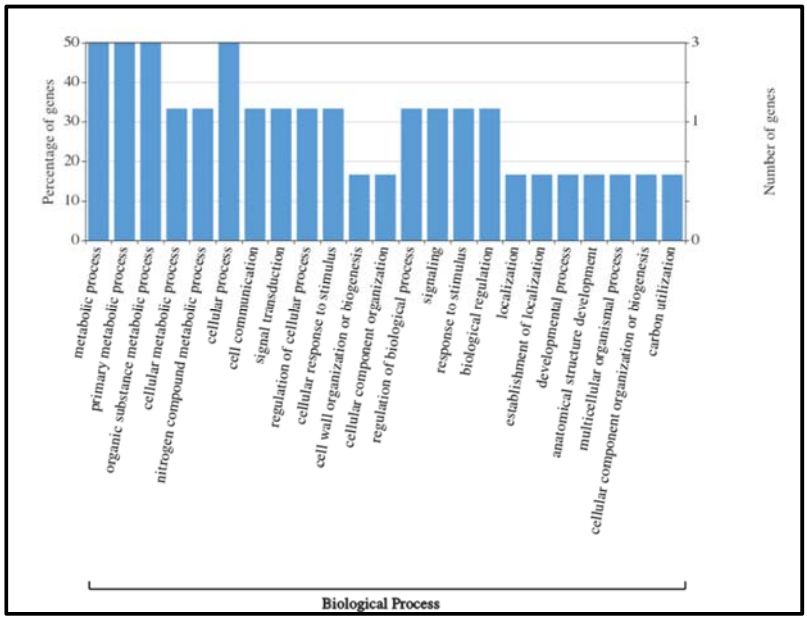

Figure 20: Classification of biological processes for mRNAs in soybean-drought

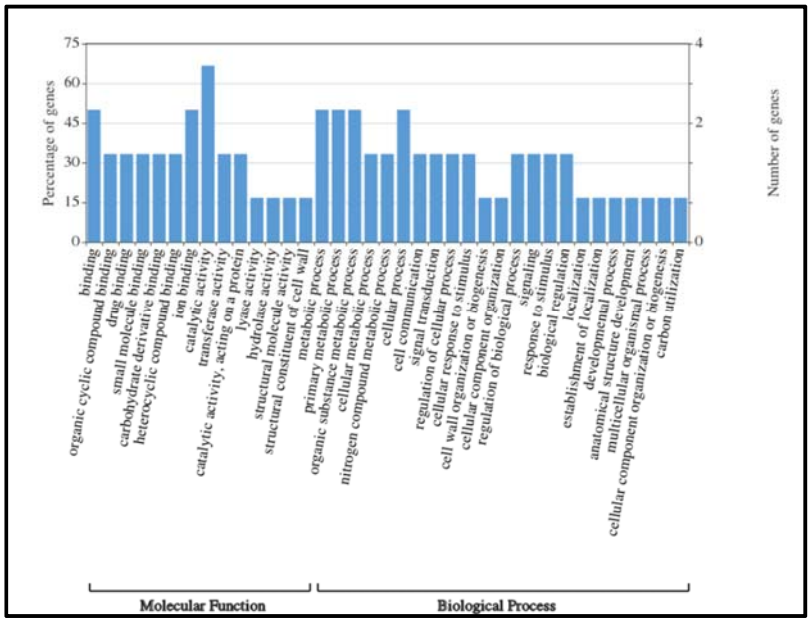

Figure 21: Classification of molecular functions for mRNAs in soybean-drought

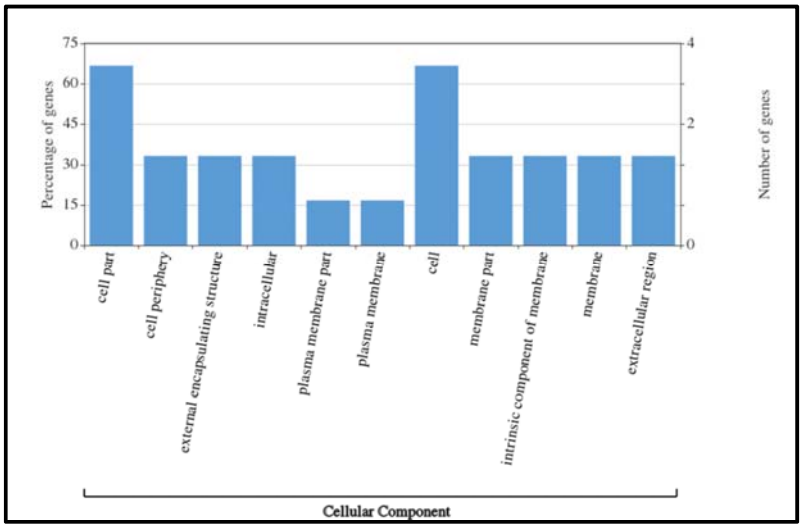

Figure 22: Classification of cellular components for mRNAs in soybean-drought

The potential functions of the mRNAs categorized in the GO terms were also found and listed in **Table 16**

Table 16: GO based functions of mRNAs in soybean-drought
Transcription initiation factor TFIID subunit 1
Probable polygalacturonase At3g15720
itogen-activated kinase 3
Sugar transport 14-like
Carbonic anhydrase

These functions carried out by the mRNAs are involved in regulation of drought stress responsive mechanisms in soybean. That's why when the miRNAs that are responsible for silencing these mRNAs get bind to the circRNAs, they remain no more available to regulate the mRNAs and so the genes become free to carry out their functions and help the soybean plants in tolerating the drought stress conditions.

4.5 Prediction of circular RNAs in soybean under wilt stress

4.5.1 Identified circular RNAs

For soybean under wilt stress trait, the trimmed reads for control condition and wilt stress conditions were given as input for mapping in BWA-MEM and the SAM files generated were further given as input in CIRI. As a result 48 circular RNAs were identified in control condition and 75 circular RNAs were identified in wilt stress condition.

The result files obtained under control and wilt stress conditions contain the identified circular RNA IDs, the chromosome IDs on which the circular RNAs are present, the start and end coordinates of the circular RNAs, the lengths of the circRNAs as well as the type of circular RNAs like exonic, intronic or intergenic.

In soybean under control condition out of the 48 identified circular RNAs, 31 (64.5%) were intergenic circRNAs, 15 (31.25%) were exonic circRNAs and 2 (4.16%) were intronic circRNAs. Whereas, under wilt stress condition out of the 75 identified circular RNAs, 53 (70.66%) were intergenic circRNAs, 18 (24%) were exonic circRNAs and 3 (4%) were intronic circRNAs. Thus it can be observed that in both the conditions highest number of circular RNAs are formed from the intergenic regions of the genome which is followed by circRNAs formed from the exonic regions. The distribution of different types of circular RNAs under control and wilt stress conditions are given in **Fig. 23**.

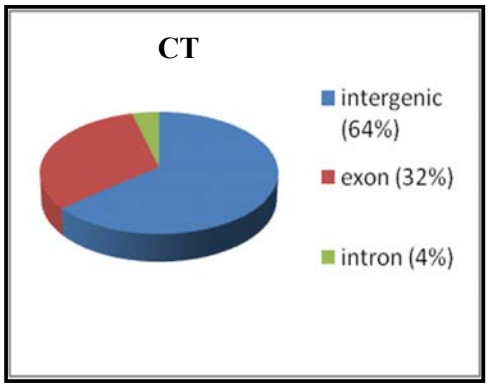

Fig. 23a

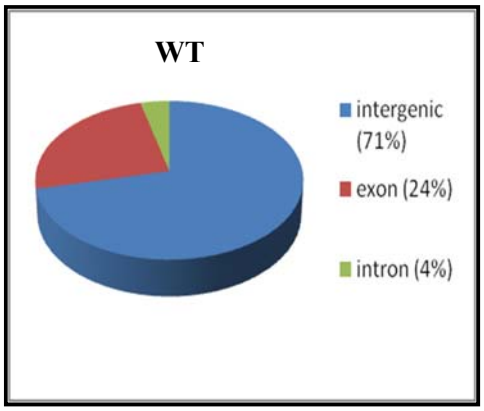

Fig. 23b

Figure 23: Distribution of different types of circular RNAs (**a**) under control and (**b**) under wilt stress conditions in soybean

A sample of the circular RNAs identified under control condition is given in **Table 17** and that of under wilt stress condition is given in **Table 18.**

Table 17: List (sample) of identified circRNAs and their length, chromosomal distribution under control condition of wilt in soybean

circRNA_ID	chr	circRNA_start	circRNA_end	circRNA_length	circRNA_type
10:44644870|44656965	10	44644870	44656965	12096	intergenic_region
11:8928133|8933250	11	8928133	8933250	5118	intergenic_region
11:9042743|9042952	11	9042743	9042952	210	exon
11:9861138|9866965	11	9861138	9866965	5828	intergenic_region
11:9866857|9892608	11	9866857	9892608	25752	intergenic_region
12:3920868|3928991	12	3920868	3928991	8124	intergenic_region
12:3921774|3930152	12	3921774	3930152	8379	intergenic_region
13:37978908|37983760	13	37978908	37983760	4853	intergenic_region
14:2855219|2860398	14	2855219	2860398	5180	intergenic_region
14:3801742|3808153	14	3801742	3808153	6412	exon
14:9302918|9389342	14	9302918	9389342	86425	intergenic_region
15:3011238|3011548	15	3011238	3011548	311	exon
16:5841009|5847710	16	5841009	5847710	6702	intergenic_region
16:33010788|33010934	16	33010788	33010934	147	exon
16:33924501|33930298	16	33924501	33930298	5798	intergenic_region
17:2892531|2895545	17	2892531	2895545	3015	intergenic_region
17:9232942|9234837	17	9232942	9234837	1896	intergenic_region
18:9125163|9149112	18	9125163	9149112	23950	intergenic_region
18:50306502|50306660	18	50306502	50306660	159	exon
19:275127|276513	19	275127	276513	1387	exon
2:44596302|44602019	2	44596302	44602019	5718	intergenic_region

Table 18: List (sample) of identified circRNAs and their length, chromosomal distribution under wilt stress condition in soybean

circRNA_ID	chr	circRNA_start	circRNA_end	circRNA_length	circRNA_type
1:1745463|1745696	1	1745463	1745696	234	exon
1:2601841|2601999	1	2601841	2601999	159	exon
10:4510843|4514869	10	4510843	4514869	4027	intergenic_region
10:45713078|45721325	10	45713078	45721325	8248	intergenic_region
10:45713078|45721328	10	45713078	45721328	8251	intergenic_region
11:8928133|8933250	11	8928133	8933250	5118	intergenic_region
11:9866857|9892608	11	9866857	9892608	25752	intergenic_region
12:3920868|3928991	12	3920868	3928991	8124	intergenic_region
12:3921774|3930152	12	3921774	3930152	8379	intergenic_region
13:10492941|10541030	13	10492941	10541030	48090	intergenic_region
13:14430357|14448409	13	14430357	14448409	18053	intergenic_region
13:16981593|16986402	13	16981593	16986402	4810	intergenic_region
13:37978908|37983760	13	37978908	37983760	4853	intergenic_region
14:3801742|3808153	14	3801742	3808153	6412	exon
14:9302918|9389342	14	9302918	9389342	86425	intergenic_region
15:915884|937311	15	915884	937311	21428	intergenic_region
15:4800797|4800966	15	4800797	4800966	170	exon
15:10277367|10311098	15	10277367	10311098	33732	intergenic_region
16:5841009|5847710	16	5841009	5847710	6702	intergenic_region
16:5957190|5962053	16	5957190	5962053	4864	intergenic_region

4.5.2 Circular RNA sequences

Based on the circRNA_ID of the identified circular RNAs under control and drought stress conditions, the sequences of the predicted circular RNAs were fetched from the reference genome of chickpea and a sample of the sequences of the circular RNAs identified under control condition is given in **Fig. 24.** Similarly a sample of the sequences of the circular RNAs identified under wilt stress condition is given in **Fig. 25**

Figure 24: FASTA sequences of circRNAs under control condition of soybean-wilt

Figure 25: FASTA sequences of circRNAs under wilt stress condition of soybean

4.5.3 Lengthwise distribution of the circular RNAs

The circular RNAs identified were found to be of different lengths having a range of 147 nucleotides to 642 nucleotides for exonic circRNAs, 352 nucleotides to

387 nucleotides for intronic circRNAs and 1896 nucleotides to 121747 nucleotides for intergenic circRNAs under control condition; whereas under wilt stress condition the length ranges from 135 nucleotides to 674 nucleotides for exonic circRNAs, 582 nucleotides to 903 nucleotides for intronic circRNAs and 2320 nucleotides to 173741 nucleotides for intergenic circRNAs. The lengthwise distribution of the circular RNAs under both the conditions are given in **Fig. 26-28** (where x-axis denotes the length of circRNAs in terms of the number of nucleotides and y-axis denotes the number of circRNAs).

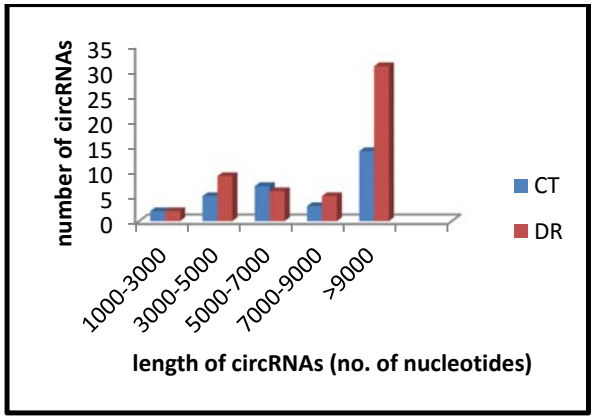

Figure 26: Lengthwise distribution of exonic circRNAs under control and wilt stress conditions in soybean

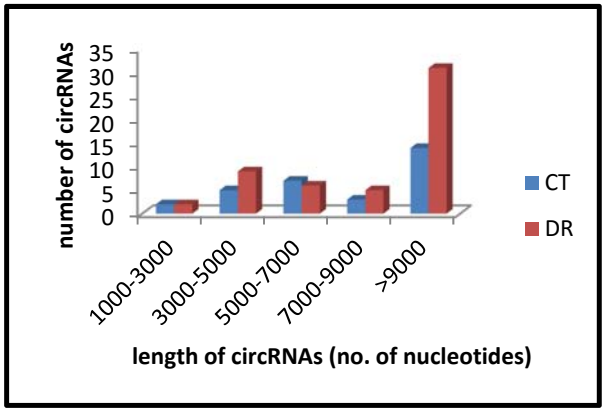

Figure 27: Lengthwise distribution of intronic circRNAs under control and wilt stress conditions in soybean

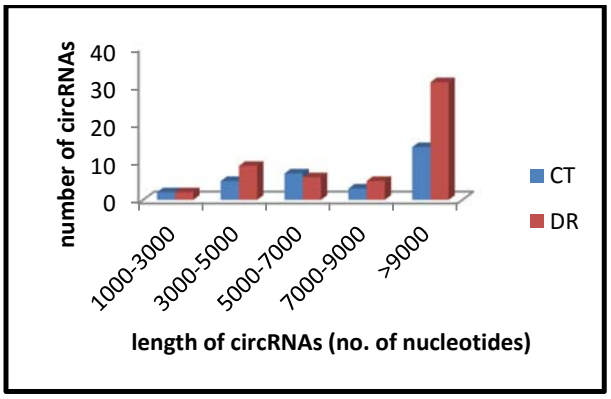

Figure 28: Lengthwise distribution of intergenic circRNAs under control and wilt stress conditions in soybean

4.5.4 Distribution of the circular RNAs based on expression values (FPKM)

The expression values of the circular RNAs under control and drought conditions were obtained using the RSEM tool and were divided into six classes having values 0, 1-100, 100-200, 200-300, 300-400 and >400. The expression values for most of the circular RNAs under both the conditions is more than 400. 25 circular RNA under control condition has expression value greater than 400, whereas in case of wilt stress condition 35 circular RNAs have expression value greater than 400. The distribution of the circular RNAs based on the expression values as obtained through RSEM is shown in **Fig. 29** (where x-axis denotes the expression value of circRNAs in terms of FPKM value and y-axis denotes the number of circRNAs)

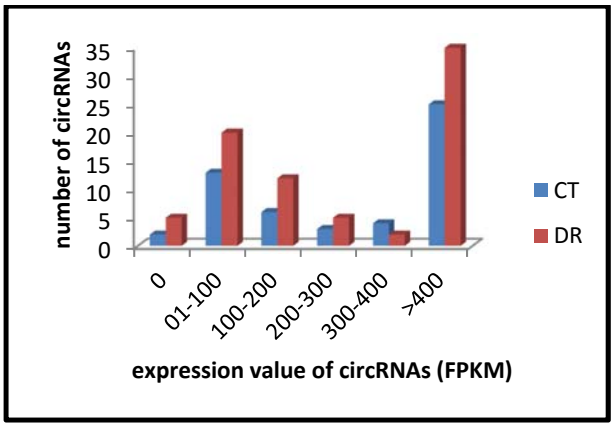

Figure 29: Distribution of circRNAs based on expression values under control and wilt stress conditions in soybean

4.5.5 Differentially expressed circular RNAs

From the identified circular RNAs under control and wilt stress conditions, the differentially expressed circular RNAs were identified using DESeq2 package of R software. 24 differentially expressed circular RNAs were identified under control and wilt stress condition of soybean with a p-value ranging from 0.089 to 1. Only one circular RNA was identified as differentially expressed with p-value less than 0.1, originated from the intergenic region and is downregulated.

A sample of the result is given in **Table 19**

Table 19: List (sample) of differentially expressed circRNAs in soybean-wilt

circRNA_ID	baseMean	log2FoldChange	lfcSE	stat	pvalue
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:49532142-49539485	147.8680179	0.846455303	0.47461189	1.783468389	0.074510089
9_dna:chromosome_chromosome:Glycine_max_v2.1:9:1:50189764:1_REF:1827313-1832264	9529.181292	-0.832494619	0.539896489	-1.541952274	0.123085192
8_dna:chromosome_chromosome:Glycine_max_v2.1:8:1:47837940:1_REF:7434887-7443884	255.5519428	0.682062764	0.472055786	1.444877457	0.148492309
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:45697097-45728257	561.502906	0.634257297	0.569995781	1.112740335	0.265819967
9_dna:chromosome_chromosome:Glycine_max_v2.1:9:1:50189764:1_REF:41497262-41497902	2225.065073	0.596924443	0.554284616	1.076927676	0.281512578
11_dna:chromosome_chromosome:Glycine_max_v2.1:11:1:34766867:1_REF:8928133-8933250	1819.102368	-0.57176334	0.572327382	-0.999014477	0.317787678
12_dna:chromosome_chromosome:Glycine_max_v2.1:12:1:40091314:1_REF:3921774-3930152	604.6443325	0.555032052	0.572763757	0.969041852	0.332524308
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:45830034-45836851	7037.381348	-0.31656812	0.333069854	-0.950455635	0.341880786
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:1099021-1104695	1178.889366	0.402468073	0.425408729	0.946073848	0.344110922
15_dna:chromosome_chromosome:Glycine_max_v2.1:15:1:51756343:1_REF:12636344-12645124	984.7336515	0.447290266	0.481950182	0.928084026	0.353363985
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:50841293-50844818	769.5785904	0.436836954	0.560569534	0.779273449	0.435818652
11_dna:chromosome_chromosome:Glycine_max_v2.1:11:1:34766867:1_REF:8927977-8931149	10669.9713	-0.436041834	0.568929274	-0.766425378	0.443423237
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:44630646:1_REF:43200672-43217184	50.57804171	-0.371806739	0.521193108	-0.713376161	0.47561302
3_dna:chromosome_chromosome:Glycine_max_v2.1:3:1:45779781:1_REF:5941334-5980395	291.7359305	-0.211563905	0.457823205	-0.462108303	0.644003656
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:44630646:1_REF:4027348-4042367	112.2722809	0.202399978	0.441479591	0.45845829	0.646623222
3_dna:chromosome_chromosome:Glycine_max_v2.1:3:1:45779781:1_REF:4794675-4894600	148.8625502	-0.191737061	0.43224672	-0.443582455	0.657344493
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:6856493-6858812	2644.338057	0.080553027	0.264120491	0.304985905	0.760376888
6_dna:chromosome_chromosome:Glycine_max_v2.1:6:1:51416486:1_REF:4487693-4497730	251.9423074	-0.100815679	0.439660414	-0.229303517	0.81863302
16_dna:chromosome_chromosome:Glycine_max_v2.1:16:1:37887014:1_REF:5841009-5847710	37.48778728	-0.112873053	0.551928718	-0.204506577	0.837957653
8_dna:chromosome_chromosome:Glycine_max_v2.1:8:1:47837940:1_REF:47391302-47397126	5290.867515	-0.032367874	0.226942964	-0.14262559	0.886585886
18_dna:chromosome_chromosome:Glycine_max_v2.1:18:1:58018742:1_REF:8118704-8189069	199.8348138	0.02937779	0.426543735	0.06887404	0.945089882
16_dna:chromosome_chromosome:Glycine_max_v2.1:16:1:37887014:1_REF:36720161-36724310	1536.754063	-0.000116711	0.515586682	-0.000226366	0.999819386
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:44630646:1_REF:495938-513138	3731.401881	-1.47E-09	0.202502704	-7.28E-09	0.999999994
7_dna:chromosome_chromosome:Glycine_max_v2.1:7:1:44630646:1_REF:502415-511575	0	NA	NA	NA	NA

Among the 24 identified circular RNAs, 12 circRNAs were found to have $log_2FC<0$ indicating downregulation (**Table 20**), whereas 12 circRNAs were found to have $log_2FC>0$ indicating upregulation (**Table 21**).

Table 20: List of downregulated circRNAs in soybean-wilt

circRNA_ID	baseMean	log2FoldChang	lfcSE	stat	pvalue
7_dna:chromosome_chromosome:Glycine_max_v2.1	709.3314047	-2.290626513	1.350556	-1.69606	0.089874
KZ847214_dna:supercontig_supercontig:Glycine_max	38.78358759	-1.34508724	1.109218	-1.21264	0.225266
11_dna:chromosome_chromosome:Glycine_max_v2.	356.1966334	-0.875846168	0.757436	-1.15633	0.247546
2_dna:chromosome_chromosome:Glycine_max_v2.1	710.5561496	-0.743598755	0.670172	-1.10956	0.267187
9_dna:chromosome_chromosome:Glycine_max_v2.1	319.6584114	-0.572228948	1.067934	-0.53583	0.592078
16_dna:chromosome_chromosome:Glycine_max_v2.	130.0270805	-0.445013204	0.531682	-0.83699	0.402598
18_dna:chromosome_chromosome:Glycine_max_v2.	145.7446397	-0.402091836	0.499557	-0.8049	0.420879
8_dna:chromosome_chromosome:Glycine_max_v2.1	52.86815362	-0.375586025	0.580686	-0.6468	0.517764
11_dna:chromosome_chromosome:Glycine_max_v2.	2636.46746	-0.269143363	0.354095	-0.76009	0.447202
7_dna:chromosome_chromosome:Glycine_max_v2.1	58.78775383	-0.233145689	0.510527	-0.45668	0.647904
13_dna:chromosome_chromosome:Glycine_max_v2.	42.25369806	-0.201202276	0.559047	-0.3599	0.71892
2_dna:chromosome_chromosome:Glycine_max_v2.1	234.5386429	-0.081902688	0.312221	-0.26232	0.793072

Table 21: List of upregulated circRNAs in soybean-wilt

circRNA_ID	baseMean	log2FoldChange	lfcSE	stat	pvalue
12_dna:chromosome_chromosome:Glycine_max_v	41.64132563	5.29E-08	0.540154	9.79E-08	1
6_dna:chromosome_chromosome:Glycine_max_v2	48.37742242	0.035331766	0.511035	0.069138	0.94488
4_dna:chromosome_chromosome:Glycine_max_v2	1509.702178	0.071321199	0.233514	0.305426	0.760041
3_dna:chromosome_chromosome:Glycine_max_v2	34.49698054	0.114331325	0.588745	0.194195	0.846023
14_dna:chromosome_chromosome:Glycine_max_v	9.185586535	0.16754938	0.966981	0.173271	0.862439
12_dna:chromosome_chromosome:Glycine_max_v	445.8071332	0.281126461	0.382938	0.73413	0.462869
14_dna:chromosome_chromosome:Glycine_max_v	433.5596845	0.368008202	0.440086	0.836218	0.403032
3_dna:chromosome_chromosome:Glycine_max_v2	82.46615467	0.387958513	0.536982	0.72248	0.469999
6_dna:chromosome_chromosome:Glycine_max_v2	2620.341652	0.511997786	0.518077	0.988266	0.323022
3_dna:chromosome_chromosome:Glycine_max_v2	740.1541506	0.586750648	0.573614	1.022902	0.306354
8_dna:chromosome_chromosome:Glycine_max_v2	1722.603662	0.866810737	0.736806	1.176443	0.239418
9_dna:chromosome_chromosome:Glycine_max_v2	1900.395792	1.843348513	1.436851	1.282908	0.199524

4.6 CircRNA-miRNA-mRNA network in soybean under wilt stress trait

The identified differentially expressed circular RNAs of chickpea under control and drought stress condition are used for identification of the miRNAs for which those circular RNAs act as sponges, using psRNATarget. In psRNATarget the FASTA file containing the miRNA sequences downloaded for soybean were used as input for small RNA sequences and the FASTA file containing the sequences of the differentially expressed circRNAs as the target sequence database. A sample list of circRNAs targeted by miRNAs is given in **Table 22.**

Table 22: List (sample) of miRNAs targeting circRNAs in soybean-wilt

miRNA_Acc.	Target_Ac	UPE$	miRNA_sta	miRNA_end	Target_star	Target_end	miRNA_alig	Target_aligr	Inhibition
gma-miR4415a	11	-1	1	21	11402	11422	UUGAUUCU(	UGAUGUUC(	Cleavage
gma-miR4415b	11	-1	1	21	11402	11422	UUGAUUCU(	UGAUGUUC(	Cleavage
gma-miR10196	8	-1	1	22	12408	12429	UGAUUGUG(	UCAAAGUG(	Cleavage
gma-miR10407	3	-1	1	24	16858	16881	AGUUAACG(	GAUAAAUU	Cleavage
gma-miR10407	3	-1	1	24	93765	93788	AGUUAACG(	GAUAAAUU	Cleavage
gma-miR10407	3	-1	1	24	16858	16881	AGUUAACG(	GAUAAAUU	Cleavage
gma-miR10407	3	-1	1	24	93765	93788	AGUUAACG(	GAUAAAUU	Cleavage
gma-miR1513a	7	-1	1	21	4198	4218	UGAGAGAA	GUAAGUCA(	Cleavage
gma-miR1513b	7	-1	1	21	4198	4218	UGAGAGAA	GUAAGUCA(	Cleavage
gma-miR1533	8	-1	1	19	14453	14471	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	8	-1	1	19	14474	14492	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	14	-1	1	19	32864	32882	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	14	-1	1	19	32885	32903	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	3	-1	1	19	54784	54802	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	3	-1	1	19	54805	54823	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	3	-1	1	19	88405	88423	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	3	-1	1	19	115094	115112	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	3	-1	1	19	116797	116815	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	11	-1	1	19	4628	4646	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR1533	11	-1	1	19	13078	13096	AUAAUAAA	UUAUUAUU	Cleavage
gma-miR4405	11	-1	1	24	17099	17122	AUUCUAAG	GGUUAUAC(	Cleavage
gma-miR6299	3	-1	1	22	1979	2000	AUUUAAAA	UAUCAAAU	Translatio

The total number of target miRNAs predicted were 22, but the number of unique miRNAs targeting the differentially expressed circRNAs were 10. This is because there are several miRNAs which target more than one differentially expressed circular RNA. Moreover, out of the total 24 differentially expressed circular RNAs, only 18 circular RNAs have the miRNAs targets. Most of those 18 circular RNAs have more than one target miRNA sites. The maximum number of miRNA target sites is found to be 2.

The identification of the targeting miRNAs was followed by the identification of the mRNAs in soybean that were targeted by those miRNAs in order to understand the whole interaction between the circRNAs, miRNAs and mRNAs. When the miRNAs get bind to their target circRNAs, that miRNAs are no more available to regulate the function of their target mRNAs. That's why identification of the target mRNAs of the predicted 10 miRNAs was done using psRNATarget with Expectation value <2. Out of 10 miRNAs, all were found to target 275 unique mRNAs with a total of 366 interactions. The result file contains the IDs of both miRNAs and their target mRNAs, expectation value, unpaired potential energy, start and end coordinates of the alignment of both miRNAs and mRNAs, as well as the inhibition type**.** It was observed that the potential energy for all the interactions between the miRNAs and target mRNAs is -1. It was also found that there was only one translation inhibition and the rest 365 interactions results in cleavage inhibition. A sample of the mRNAs targeted by miRNAs is given in **Table 23**

Table 23: List (sample) of mRNAs targeted by miRNAs in soybean-wilt

miRNA_Acc.	Target_Acc.	UPE$	miRNA_align	Target_aligne	Inhibition
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR104	Glyma.11G1	-1	AGUUAACGG	GACAAAUUC	Cleavage
gma-miR153	Glyma.14G0	-1	AUAAUAAAA	UCAUUAUUA	Cleavage
gma-miR153	Glyma.14G0	-1	AUAAUAAAA	UCAUUAUUA	Cleavage
gma-miR153	Glyma.14G0	-1	AUAAUAAAA	UCAUUAUUA	Cleavage
gma-miR153	Glyma.14G0	-1	AUAAUAAAA	UCAUUAUUA	Cleavage
gma-miR153	Glyma.14G0	-1	AUAAUAAAA	UCAUUAUUA	Cleavage
gma-miR151	Glyma.07G1	-1	UGAGAGAAA	GUAAGUCAU	Cleavage
gma-miR151	Glyma.07G1	-1	UGAGAGAAA	GCAAGUCAU	Cleavage
gma-miR151	Glyma.07G1	-1	UGAGAGAAA	GUAAGUCAU	Cleavage
gma-miR151	Glyma.07G1	-1	UGAGAGAAA	GCAAGUCAU	Cleavage
gma-miR153	Glyma.05G1	-1	AUAAUAAAA	UUAUUAUUA	Cleavage
gma-miR151	Glyma.06G1	-1	UGAGAGAAA	GUAAGUCGU	Cleavage
gma-miR151	Glyma.01G2	-1	UGAGAGAAA	GUAAGUCAU	Cleavage
gma-miR151	Glyma.16G2	-1	UGAGAGAAA	GUAAGUCAU	Cleavage
gma-miR151	Glyma.16G2	-1	UGAGAGAAA	GUAAGUCAU	Cleavage
gma-miR151	Glyma.01G2	-1	UGAGAGAAA	GUAAGUCAU	Cleavage

Blast2Go results

To understand the functional role of mRNAS involved in the circRNA-miRNA-mRNA network, the GO analysis was performed by employing the BLAST2GO software. The GO functional categorization generated 14 annotations from the 275 predicted mRNAs. In that, a total of 5, 5 and 4 mRNAs were classified as the first level classifications of biological processes, molecular functions, and cellular components, respectively (**Fig. 30**). Among the biological process classification, 3 mRNAs were classified into each of the categories of "metabolic process (GO: 0008152)" and "cellular process (GO: 0009987)", respectively (**Fig. 31**). In the molecular functions classification, the main category was "catalytic activity (GO: 0003824)", which had 3 mRNAs, (**Fig. 32**). When the predicted mRNAs were classified according to the cellular component classification, categories "cell (GO: 0005623)" and "cell part (GO: 0044464)" both made up the largest proportion of 8 predicted mRNAs (**Fig. 33**). The GO analysis on predicted mRNAs showed that the targets of differentially expressed circRNAs under wilt stress were associated with various functions involving in different cellular components, biological process, and molecular functions.

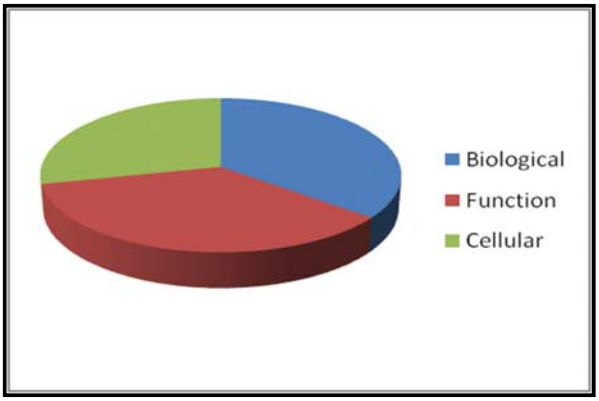

Figure 30: Differentially expressed circRNAs in soybean-wilt in GO terms of biological processes, cellular components and molecular functions

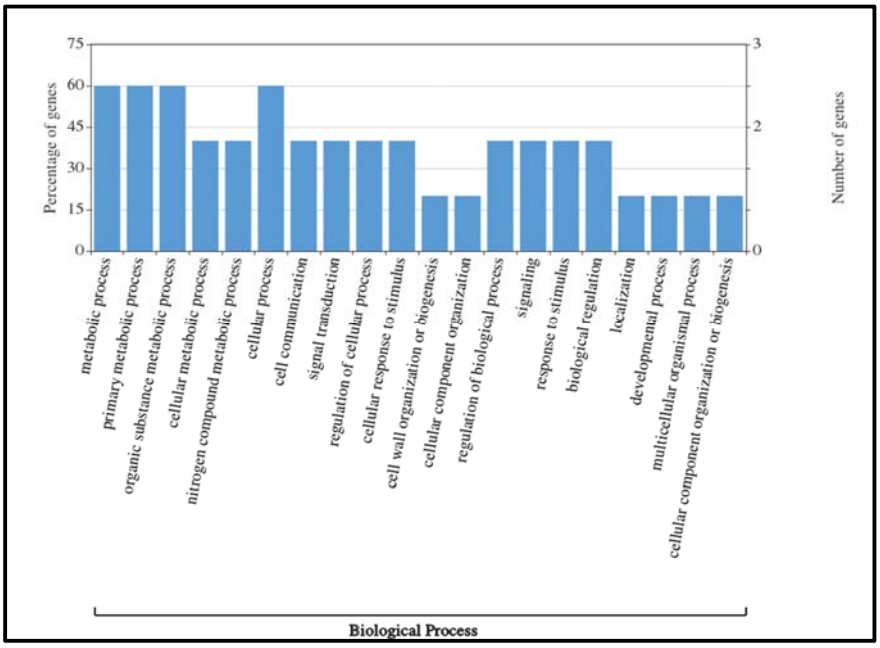

Figure 31: Classification of biological processes for mRNAs in soybean-wilt

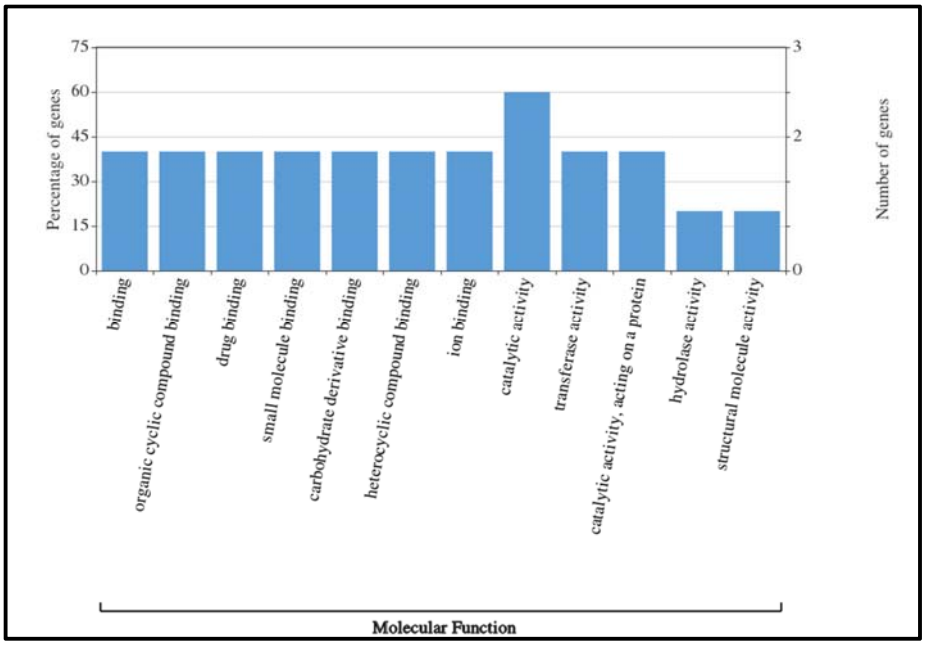

Figure 32: Classification of molecular functions for mRNAs in soybean-wilt

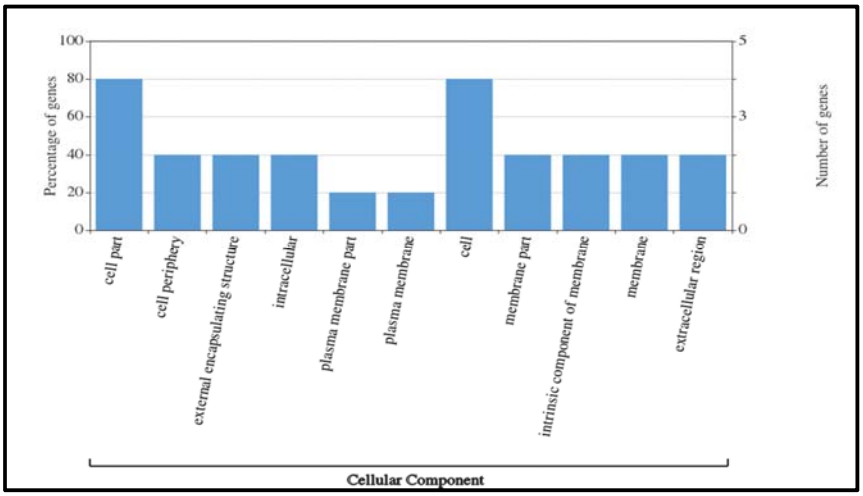

Figure 33: Classification of cellular components for mRNAs in soybean-wilt

The potential functions of the mRNAs categorized in the GO terms were also found and listed **Table 24**

Table 24: GO based functions of mRNAs in soybean-wilt

Structural constituent of cell wall
Polygalacturonase activity;carbohydrate metabolic process
Intracellular;MAP kinaseactivity; ATP binding; MAPK cascade
Integral component of membrane
Carbonate dehydratase activity;zinc ion binding;carbon utilization
Substrate-specific transmembrane transporter activity

These functions carried out by the mRNAs are involved in regulation of wilt stress responsive mechanisms of soybean. That's why when the miRNAs that are responsible for silencing these mRNAs get bind to the circRNAs, they remain no more available to regulate the mRNAs and so the genes become free to carry out their functions and help the soybean plants in tolerating the wilt stress conditions.

In case of wilt stress in chickpea, the RNA-seq data collected from NCBI, described in material section, was analysed for the identification and characterization of circular RNAs. However, no results on circular RNAs were found, probably, due to non availability of exhaustive transcriptome data in chickpea.

Based on the results obtained in this chapter, a detailed discussion on the role of circRNAs in drought and wilt stress tolerance mechanisms in chickpea and soybean is made in the next chapter.

CHAPTER V

Discussion

In this chapter, a brief discussion about the results obtained from identification and characterization of circular RNAs in chickpea and soybean under drought and wilt stress conditions is given. Besides, the discussion is made in the context of findings reported in literature with regard to the role of circular RNAs under biotic and abiotic stress tolerance mechanisms in chickpea and soybean.

5.1 Identification of circRNAs

Although the role of circRNAs have been reported in some plant species like *Arabidopsis thaliana* (Sablok et al., 2016), rice, etc., circRNAs in legumes under abiotic and biotic stress conditions is yet to be fully explored. In the present study, circular RNAs in chickpea as well as soybean under drought and wilt stress conditions were identified. Previously, by analyzing the publically available RNA-seq data 12,037 (6074exons) and 6012 (5152exons) circRNAs were identified from *Oryza sativa* and *Arabidopsis thaliana*, respectively (Ye et al., 2015). While 2354 circRNAs (1356 exons) were identified in rice through deep sequencing and computational analysis of ssRNA-seq data (Lu et al., 2015). Recently, deep sequencing research in tomato identified 854 circRNAs (615 exons), in which 163 exhibiting chilling responsive expression (Zuo et al., 2016), and 88 circular RNAs were identified in wheat under dehydration stress condition. In our study the number of circRNAs identified were lesser in number as compared to *Oryza sativa* and *Arabidopsis thaliana* (200 and 285 in chickpea under control and drought stress conditions respectively, 57 and 66 in soybean under control and drought stress conditions respectively and 48 and 75 in soybean under control and wilt stress conditions respectively). The percentages of exons (< 20% in both the crops under drought stress condition and about 30% under wilt stress condition in soybean) were also not as many as reported earlier in other plants. The results may be attributed to the following possible reasons: (i) The sequencing data is one of the most important factor influencing the number of identified circRNAs, because more number of reads denotes not only the high detection rate of circRNAs, but also could eliminate the false positives (Szabo and Salzman, 2016). The sequencing data size in present study is only 42M and 47M reads (chickpea-drought), 31M and 35M reads (chickpea-wilt), 26.5M and 28M reads (soybean-drought) and 23M and 31M reads (soybean-wilt),

while it was 710 million paired-end reads sized 100 bp in rice (Lu et al., 2015); (ii)The available chickpea genome sequences are limited in public database (http://plants.ensembl.org and http://ncbi.nlm.nih); (iii) The software used for the circRNAs prediction. We identified the circRNAs in legumes using CIRI algorithm (Gao et al., 2015). It is an efficient and unbiased algorithm for the *denovo* identification of circRNAs, which has already been used in organisms like animals, humans, and plants (Zuo et al., 2016). CIRI has been proven to be more useful as compared to other algorithms when conducting more unbiased circRNA analyses or in poorly annotated organisms (Hansen et al., 2015). However, it is known that the differences between organization of genomes might influence the results on circRNA identification (Chen*, et al.*, 2016), as well as non availability of specific algorithm for legume circRNAs detection.

5.2 The Differential Expression Patterns of circRNAs

To investigate whether the circRNAs expressed in a specific manner in chickpea and soybean under drought and wilt stress conditions, we compared the circRNA expression patterns between the control condition and stress conditions in these crops. The results showed differential expression in 44 circRNAs of which 22 upreglated and 22 downregulated in chickpea under drought stress, 23 circRNAs of which 11 upregulated and 12 downregulated in soybean under drought stress and 24 circRNAs of which 12 upregulated and 12 downregulated in soybean under wilt stress condition. The differential expressions were found to be significantly different at higher α level of significance suggesting that they might play roles in responding to the different stress conditions under study. In previous studies on wheat under dehydration stress, 66 circRNAs showed significant difference out of 88 identified circRNAs which indicated their specific roles in the anti-dehydration stress regulation, whereas in tomato under chilling stress condition 163 circRNAs out of 854 identified circRNAs had significant difference between control and chilling injury conditions, indicating their specific role in chilling injury regulation.

5.3 Prediction of circRNAs acting as miRNA Sponges in legumes

To evaluate whether circRNAs in legumes could affect post-transcriptional regulation of functional genes by binding to miRNAs, the bioinformatics methods were employed to identify the circRNA-originating target mimics in chickpea and

soybean based on the differentially expressed circRNAs. A total of 17 miRNAs, 40 miRNAs and 10 miRNAs were found to target 7, 21 and 18 differentially expressed circRNAs respectively in soybean-drought, chickpea-drought and soybean-wilt stress conditions. Moreover, most of the differentially expressed circRNAs had two to nine miRNA- binding sites, which was significantly higher than that reported in rice (Lu et al., 2015) and tomato (Zuo et al., 2016).

5.4 Interaction between circRNA-miRNA-mRNA

The functions of the predicted mRNAs as reported by BLAST2GO in chickpea and soybean under drought and wilt stress conditions showed that the mRNAs were involved in plant hormone signal transduction, response to stress, defence response mechanism, transcription factor activity, response to auxin as well as in various enzymatic activities like oxidoreductase activity, GTPase activity, hydrolase activity and so on in chickpea under drought stress conditions. While in soybean under drought stress condition the predicted mRNAs were found to be involved in activity of transcription initiation factor, polygalacturonase activity, mitogen-activated kinase activity and carbonic anhydrase activity. Those plant enzymes were found to participate in drought stress tolerance by mediating growth, development, nutrient allocation and gene expression. Recently, the role of auxins in drought tolerance was postulated by Peleg and Blumwald (2011). This supports our present finding on expression of genes/mRNAs involved in response to auxins, as miRNAs fail to regulate them due to their binding with circRNAs.

Wang, *et al.* (2017) identified circRNAs along with their targets in wheat leaves under dehydration stress. They also explained the involvement of these predicted circRNAs in plant hormone signal transduction involving auxins etc under dehydration stress. In a similar way we have also found the indirect involvement of circRNAs in drought stress mechanism due to their action as sponges for miRNAs that no more remains available to regulate the mRNAs.

Zuo, *et al.* (2016) deciphered the importance of circRNAs in regulating chilling responsive expression in tomato. Further they found circRNAs acting as miRNA sponges in tomato. In similar lines around 24 differentially expressed circRNAs to be involved in wilt stress responsive process such as, structural constituent of cell wall, polygalacturonase activity, carbohydrate metabolic process,

ATP binding, NAPK cascade, membrane activity, carbonate dehydratase activity, zinc ion binding, carbon utilization and substrate specific transmembrane activity are identified.

Keeping in view the objectives of the study, material methods used, results obtained and discussion made in the chapters 1-5, a summary of the findings, conclusion and future scope of work are given in the next chapter.

CHAPTER VI

Summary and Conclusion

This chapter presents a summary of findings and conclusion drawn from the research work carried out in this thesis. The whole thesis has been divided into 5 chapters followed by summary, abstract in English and Hindi and bibliography.

With the rapid increase in population the importance of growing more and more leguminous crops along with cereals are being realized, among which soybean and chickpea are two important crops. But the productivity of these crops is highly affected by adverse environmental conditions like drought stress and some biotic stress conditions like wilt disease. Several studies have shown the role of different non-coding RNAs in regulating those stress responsive mechanisms in plants. But circular RNAs being a new star of non-coding RNAs remained unexplored in leguminous crops specific to drought and wilt stress responses.

Keeping in mind the above research gap, the study on "**Identification and characterization of circular RNAs in legumes**" was taken up with the following objectives:

i) To identify and characterize circRNAs responsible for biotic and abiotic stress tolerance in the leguminous crops

ii) To study the relationship between circRNA-miRNA-mRNA

In chapter 1, a brief introduction of the leguminous crops and their importance, followed by the introduction to coding and non-coding RNAs, various types of non-coding RNAs along with their potential roles, a brief detailing about circular RNAs, along with their biogenesis, functions and identification was given. This chapter gives an overview of the context and the scope of the research work conducted.

Chapter 2, discusses the literature related to general overview of biotic and abiotic stresses in leguminous crops with special reference to soybean, and chickpea and non-coding RNAs, with main emphasis given on circular RNAs of various crops and different tools developed for their identification and characterization.

Chapter 3, deals with description of methodology, that has been used for the present study. A well established tool CIRI was used for the prediction of the circular RNAs in soybean and chickpea under drought and wilt stress conditions. Their identification

was followed by the identification of the differentially expressed circular RNAs under control and drought stress conditions as well as control and wilt stress conditions. Further, those differentially expressed circRNAs were used to detect their role as miRNA sponges using TargetFinder and their interaction with mRNAs through psRNATarget. Finally, the functional characterization of the mRNAs were done using BLAST2GO.

Chapter 4, deals with the results of the research work. The results were written based on the circular RNAs identified, their length distribution, the differentially expressed circular RNAs identified, their role as miRNA sponges as well as the interaction of the circRNAs-miRNAs-mRNAs. The results in brief were: the number of identified circRNAs were 57 (control) and 66 (drought stress) in soybean-drought; 200 (control) and 285 (drought stress) in chickpea-drought, and 48 (control) and 75 (wilt stress) in soybean-wilt conditions, whereas in chickpea-wilt condition no circRNA was identified by CIRI. The number of miRNAs targeting the DEcRNAs under drought stress condition in soybean is 17, whereas in chickpea it is 40 and in soybean under wilt stress condition is 10. The mRNAs targeted by the miRNAs under drought stress condition in soybean is 281, in chickpea under drought stress condition is 145 and in soybean under wilt stress is 275. The mRNAs targeted by the miRNAs are involved in intensifying the metabolic activities (polygalacturonase), enzymatic activities (MAP kinase, Carbonate dehydratase), and retaining the cell structure. In addition transcription initiation factor TFIID, sugar transport 14 are also involved in drought and wilt stress tolerance mechanisms.

Chapter 5, is devoted towards general discussion drawn based on the results obtained from the study. Here discussion was also made in the light of findings reported by a galaxy of earlier workers under various biotic and abiotic stress conditions in crops such as rice, *Arabidopsis thaliana* , tomato, wheat. The differential expression patterns of circRNAs and their role in regulating the genes responsible for drought and wilt stress conditions was discussed. Also the activity of circRNAs as sponges for miRNAs that target a number of genes involved in stress tolerance mechanisms was discussed. Besides the functional role of genes involved in the interaction of circRNA-miRNA-mRNA was also discussed.

Finally, the thesis ended with summary and conclusion followed by future scope of work.

Future Scope of work

In the present study, among various abiotic and biotic stresses affecting chickpea and soybean, drought stress condition and wilt infected conditions were only taken for identification of the circRNAs and the potential role played by them in regulation of genes involved in stress responsive mechanisms. Such attempts can be further extended to other important stress conditions like heat stress, salinity stress, root rot disease, bacterial blight disease etc that affect the productivity of leguminous crops.

In addition the identification of circular RNAs was done using the CIRI tool. Though this tool/algorithm was reported as an efficient tool, several other algorithms like CIRCExplorer, find_circ, Mapsplice etc can also be applied for circRNA identification followed by further downstream analysis.

Identification and characterization of circular RNAs in legumes

ABSTRACT

Leguminous crops are important crops next to cereals. The production and productivity of these crops has a major impact on country's economy as well as the economic status of the stakeholders-farmers. Moreover, the nutritional security can be ensured through growing of these leguminous crops. In this context, two leguminous crops, chickpea and soybean along with their transcriptome data were considered in the present study. With the advent of NGS technologies it has become feasible to unravel the underlying complex mechanisms at genome level. Even though the coding genes play a major role in various abiotic and biotic stress mechanisms, there are several non-coding RNAs which regulate the genes responsible for stress tolerance mechanisms. However, the identification and characterization of circular RNAs, a type of non-coding RNAs as well as the circRNA-miRNA-mRNA networks in legumes has not been fully explored. Hence, the present study on 'Identification and characterization of circular RNAs in legumes' has been taken up with the objectives: (i) to identify and characterize circRNAs responsible for biotic and abiotic stress tolerance in the leguminous crops and (ii) to study the relationship between circRNA-miRNA-mRNA. Here the transcriptome data of the said crops were collected from public domain (NCBI, Ensemble, etc) and the algorithm given in CIRI (CircRNA Identifier) was used to identify the circRNAs under drought (abiotic) and wilt (biotic) stress conditions. The characterization of circRNAs was done through their differential expression in both drought and wilt stress conditions. The differentially expressed (DE) circRNAs were further probed for their role in circRNA-miRNA-mRNA interaction. Finally, the identified genes from the network were studied for their functionality in stress tolerance mechanisms. The results revealed identification of 200 and 285 circRNAs under control and drought stress conditions in chickpea, 57 and 66 circRNAs under control and drought stress conditions in soybean and 48 and 75 circRNAs under control and wilt stress conditions in soybean. The number of DEcircRNAs were 44, 23 and 24 in chickpea-drought, soybean-drought and soybean-wilt respectively. These DEcircRNAs were found to act as sponges for 40 (chickpea-drought), 17 (soybean-drought) and 10 (soybean-wilt) miRNAs. Besides, these miRNAs were found to target 145 (chickpea-drought), 281 (soybean-drought) and 275 (soybean-wilt) mRNAs. GO study was carried out for the mRNAs and found that they are involved in biological processes like "metabolic process" and "cellular process", molecular functions like "binding activity" and "catalytic activity" and cellular components like "cell", "cell part" and "membrane part". Thus circRNA-miRNA-mRNA network play a vital role in stress responsive mechanisms through their activities in hormone signal transduction, response to stress, response to auxin and transcription factor activity.

Keywords : Leguminous, circRNAs, RNAs, Catalytic, Hormone

फलियों में सर्क्युलर आर.एन.ए. की पहचान एवं लक्षण वर्णन

सार

अनाज के बाद लेग्युमिनस (फलीदार) फसलें महत्वपूर्ण फसलें हैं । इन फसलों के उत्पादन और उत्पादक का देष की अर्थव्यवस्था के साथ–साथ स्टेकहोल्डर किसानों की आर्थिक स्थिति पर बहुत प्रभाव पड़ता है । इसके अतिरिक्त, इन फलीदार फसलों को उगाने के माध्यम से पोषण सुरक्षा सुनिष्चित की जा सकती है । इस संदर्भ में, प्रस्तुत अध्ययन में दो फलीदार फसलों, चना और सोयाबीन के साथ–साथ उनके ट्रांसक्रिप्टोम डाटा पर विचार किया गया । एन.जी.एस. प्रौद्योगिकियों के आगमन के साथ जीनोम स्तर पर अन्तर्निहित जटिल तंत्र को उजागर करने के लिए यह संभव हो गया है । भले ही, कोडिंग जीन विभिन्न अजैविक एवं जैविक तनाव तन्त्रों में एक प्रमुख भूमिका निभाते हैं लेकिन कई गैर–कोडिंग आर.एन.ए. हैं जो तनाव सहिष्णुता तंत्र के लिए जिम्मेदार जीन को रेगुलेट करते हैं । हालांकि, सर्क्युलर RNAs की पहचान एवं लक्षण वर्णन, गैर–कोडिंग RNAs का प्रकार, के साथ–साथ फलियों में circRNA-miRNA-mRNA नेटवर्क का पूरी तरह एक्सप्लोर नही किया गया है । अतः, ''फलियों में सर्क्युलर आर.एन.ए. की पहचान एवं लक्षण वर्णन'' पर प्रस्तुत अध्ययन निम्नलिखित उद्देष्यों से किया गया : (i) फलीदार फसलों में जैविक एवं अजैविक तनाव सहिष्णुता के लिए जिम्मेदार circRNAs की पहचान एवं वर्गीकरण (ii) circRNA-miRNA-mRNA के बीच संबंध का अध्ययन। यहां पर उक्त फसलों के ट्रांस्क्रिप्टोम डाटा पब्लिक डोमेन (एन.सी.बी.आई, एनसेम्बल, इत्यादि) से एकत्रित किए गए हैं तथा सी.आई.आर.आई. (CircRNA Identifier) में दिये गये एल्गोरिथ्म का उपयोग सूखे (अजैविक) एवं विल्ट (जैविक) स्ट्रैस दषाओं के अन्तर्गत पहचान के लिए किया गया । circRNAs का लक्ष्ण वर्णन सूखे एवं विल्ट स्ट्रैस दोनों दषाओं में उनके डिफ्रेन्षियल एक्सप्रैषन के माध्यम से किया गया । विभेदित रूप से अभिव्यक्त (डी.ई.) circRNAs की आगे circRNA-miRNA-mRNA इन्टरेक्षन में भूमिका के लिए जॉच की गयी । अन्त में, नेटवर्क से पहचान किए गए जीन्स की तनाव सहिष्णुता तंत्र में उनकी कार्यक्षमता के लिए अध्ययन किया गया । परिणामों से चने में नियंत्रित एवं सूखा तनाव दषाओं के अन्तर्गत 200 एवं 285 circRNAs की, सोयाबीन में नियंत्रित एवं सूखा तनाव दषाओं के अन्तर्गत 57 एवं 66 circRNAs की तथा सोयाबीन में नियंत्रित एवं विल्ट तनाव दषाओं के अन्तर्गत 48 एवं 75 circRNAs की पहचान हुई । चना–सूखा, सोयाबीन–सूखा एवं सोयाबीन–विल्ट में DEcircRNAs की संख्या क्रमषः 44, 23 तथा 24 थी । इन DEcircRNAs को 40 (चना–सूखा), 17 (सोयाबीन–सूखा) तथा 10 (सोयाबीन–विल्ट) miRNAs के लिए स्पॉन्ज के रूप में कार्य करते हुए पाया गया । इसके अतिरिक्त, इन miRNAs को 145 (चना–सूखा), 281 (सोयाबीन–सूखा) तथा 275 (सोयाबीन–विल्ट) को टारगेट करते हुए पाया गया । mRNAs के लिए GO अध्ययन किया गया तथा यह पाया कि ये जैविक क्रियाओं जैसे ''मैटाबॉलिक प्रोसेस'' तथा ''सेल्यूलर प्रोसेस'', मॉलीक्यूलर क्रियाएँ जैसे ''बाइडिंग एक्टिविटि'' तथा ''कैटालिटिक एक्टिविटि'' तथा सेल्यूलर घटक जैसे ''सैल'', ''सैल पार्ट'' और ''मेमब्रेन पार्ट'' में शामिल हैं । अतः, ट्रांस्क्रिप्षन फेक्टर एक्टिविटि तथा हॉरमोन सिंगनल ट्रांस्डक्षन, स्ट्रैस के प्रति रिस्पॉन्स, ऑक्सिन के प्रति रिस्पॉन्स में अपनी एक्टिविटिज के माध्यम से स्ट्रैस रिस्पॉन्सिव मैकेनिज्म्स में circRNA-miRNA-mRNA नेटवर्क की महत्वपूर्ण भूमिका है ।

Bibliography

Arenas-Huertero, C., Pérez, B., Rabanal, F., Blanco-Melo, D., De la Rosa, C., Estrada-Navarrete, G.,and Reyes, J. L. (2009). Conserved and novel miRNAs in the legume Phaseolus vulgaris in response to stress. *Plant molecular biology*, **70**(4), 385-401.

Barrett, S. P., Wang, P. L., and Salzman, J. (2015). Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *elife*, **4**, 7540-7558.

Barrett, S. P., and Salzman, J. (2016). Circular RNAs: analysis, expression and potential functions. *Development*, **143**(11), 1838-1847.

Bogard, B., Francastel, C., and Hubé, F. (2018). A new method for the identification of thousands of circular RNAs. *Non-coding RNA Investigation*, **2**(1), 116-121.

Charlson, D. V., Bhatnagar, S., King, C. A., Ray, J. D., Sneller, C. H., Carter, T. E., & Purcell, L. C. (2009). Polygenic inheritance of canopy wilting in soybean [Glycine max (L.) Merr.]. *Theoretical and Applied Genetics*, **119**(4), 587-594.

Chen, L., Yu, Y., Zhang, X., Liu, C., Ye, C., and Fan, L. (2016). PcircRNA_finder: a software for circRNA prediction in plants. *Bioinformatics,* **32**, 3528–3529.

Chen, L., Zhang, P., Fan, Y., Lu, Q., Li, Q., Yan, J., and Li, L. (2018). Circular RNAs mediated by transposons are associated with transcriptomic and phenotypic variation in maize. *New Phytologist*, **217**(3)**,** 1292-1306.

Chu, Q., Zhang, X., Zhu, X., Liu, C., Mao, L., Ye, C.,and Fan, L. (2017). PlantcircBase: a database for plant circular RNAs. *Molecular plant*, **10**(8)**,** 1126-1128.

Chu, Q., Shen, E., Ye, C. Y., Fan, L., and Zhu, Q. H. (2018). Emerging roles of plant circular RNAs. *Journal of Plant Cell Development*, **1**(1)**,** 1-14.

Collins, L. J., Schönfeld, B., and Chen, X. S. (2011). The epigenetics of non-coding RNA. *Handbook of Epigenetics*, 49-61, Academic Press.

Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International journal of plant genomics*, 619832.

Darbani, B., Noeparvar, S., and Borg, S. (2016). Identification of circular RNAs from the parental genes involved in multiple aspects of cellular metabolism in barley. *Frontiers in plant science*, **7,** 776-784.

Deshmukh, R., Sonah, H., Patil, G., Chen, W., Prince, S., Mutava, R.,and Nguyen, H. T. (2014). Integrating omic approaches for abiotic stress tolerance in soybean. *Frontiers in Plant science*, **5**, 244-256.

Duranti, M. (2006). Grain legume proteins and nutraceutical properties. *Fitoterapia*, **77**(2), 67-82.

Fu, X., and Liu, R. (2014). CircRNAFinder: a tool for identifying circular RNAs using RNA-Seq data. In *Proceedings of the 6th International Conference on Bioinformatics and Computational Biology, BICOB*.

Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome biology*, **16**(1)**,** 4-20.

Gao, Y., and Zhao, F. (2018). Computational strategies for exploring circular RNAs. *Trends in Genetics*, **34**(5)**,** 389-400.

Garg, R., Shankar, R., Thakkar, B., Kudapa, H., Krishnamurthy, L., Mantri, N.,and Jain, M. (2016). Transcriptome analyses reveal genotype-and developmental stage-specific molecular responses to drought and salinity stresses in chickpea. *Scientific reports*, **6**, 19228-19243.

Geervani, P. (1989). Utilization of chickpea in India and scope for novel and alternative uses. *Uses of Tropical Grain Legumes*, **27**, 47-54.

Hansen, T. B., Venø, M. T., Damgaard, C. K., & Kjems, J. (2015). Comparison of circular RNA prediction tools. *Nucleic acids research*, **44(**6), e58.

Hasanuzzaman, M., Nahar, K., Gill, S. S., & Fujita, M. (2013). Drought stress responses in plants, oxidative stress, and antioxidant defense. *Climate change and plant abiotic stress tolerance*, 209-250.

Impa, S. M., Nadaradjan, S., & Jagadish, S. V. K. (2012). Drought stress induced reactive oxygen species and anti-oxidants in plants. *Abiotic Stress Responses in Plants.* Springer, New York, NY, 131-147.

Jain, M., Chevala, V. N., and Garg, R. (2014). Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. *Journal of experimental botany*, **65**(20), 5945-5958.

Jambunathan, R., Hall, S. D., Sudhir, P., Rajan, V., & Sadhana, V. (1991). Uses of Tropical Grain Legumes *Proceedings of a Consultants Meeting*, 27-30 Mar 1989.

Karrou, M., and Oweis, T. (2012). Water and land productivities of wheat and food legumes with deficit supplemental irrigation in a Mediterranean environment. *Agricultural Water Management*, **107**, 94-103.

Khan, H. R., Paull, J. G., Siddique, K. H. M., & Stoddard, F. L. (2010). Faba bean breeding for drought-affected environments: A physiological and agronomic perspective. *Field Crops Research*, **115**(3), 279-286.

Kulcheski, F. R., de Oliveira, L. F., Molina, L. G., Almerão, M. P., Rodrigues, F. A., Marcolino, J.,and Abdelnoor, R. V. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC genomics*, **12**(1), 307-334

Le, D. T., Aldrich, D. L., Valliyodan, B., Watanabe, Y., Van Ha, C., Nishiyama, R., and Nguyen, H. T. (2012). Evaluation of candidate reference genes for normalization of quantitative RT-PCR in soybean tissues under various abiotic stress conditions. *PloS one*, **7**(9), 46487-46497.

Lu, T., Cui, L., Zhou, Y., Zhu, C., Fan, D., Gong, H., et al. (2015). Transcriptome-wide investigation of circular RNAs in rice. *RNA,* **21**, 2076–2087.

Mantri, N. L., Ford, R., Coram, T. E., and Pang, E. C. (2010). Evidence of unique and shared responses to major biotic and abiotic stresses in chickpea. *Environmental and experimental botany*, **69**(3), 286-292.

Mattick, J. S., and Makunin, I. V. (2006). Non-coding RNA. *Human molecular genetics*, **15**(1), 17-29.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A.,and Loewer, A. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**(7441)**,** 333-338.

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature reviews genetics*, **10**(3), 155-159

Morel, M. A., Braña, V., and Castro-Sowinski, S. (2012). Legume crops, importance and use of bacterial inoculation to increase production. In *Crop plant*. IntechOpen. 217-240.

Pandey, R. K., Herrera, W. A. T., and Pendleton, J. W. (1984). Drought Response of Grain Legumes Under Irrigation Gradient: I. Yield and Yield Components 1. *Agronomy Journal*, **76**(4), 549-553.

Peleg, Z., & Blumwald, E. (2011). Hormone balance and abiotic stress tolerance in crop plants. *Current opinion in plant biology*, **14**(3), 290-295.

Reddy, D. S., Bhatnagar, M., P., Vadez,V. and Sharma, K. K. (2012). *Grain Legumes (Soybean, Chickpea, and Peanut): Omics Approaches to Enhance Abiotic Stress Tolerance*. In: Improving Crop Resistance to Abiotic Stress. *Wiley-VCH Verlag, Singapore*. ISBN 978-3-527-32840-6

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. (2002). MicroRNAs in plants. *Genes and development*, **16**(13), 1616-1626.

Rheenen, van H. A., 1991. Production aspects and prospects of chickpea. In: *Proceedings of the consultancy meeting on uses of grain legumes* (ed. Jambunathan, R.), 31–35 ICRISAT, Pantancheru, India.

Sablok, G., Zhao, H., and Sun, X. (2016). Plant circular RNAs (circRNAs): transcriptional regulation beyond miRNAs in plants. *Molecular Plant,* **9**, 192–194.

Singh, G. (Ed.). (2010). *The soybean: botany, production and uses*. CABI, 161-165.

Szabo, L., and Salzman, J. (2016). Detecting circular RNAs: bioinformatic and experimental challenges. *Nature Reviews Genetics,* **17**, 679–692.

Toker, C., & Mutlu, N. (2011). Breeding for abiotic stresses. *Biology and breeding of food legumes*, 241-261.

Toker, C., Lluch, C., Tejera, N. A., Serraj, R., & Siddique, K. H. M. (2007). 23 Abiotic Stresses. *Chickpea breeding and management*, 474-478.

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L.,and MacLeod, J. N. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic acids research,* **38**(18), 178-192

Wang, Y., Yang, M., Wei, S., Qin, F., Zhao, H., and Suo, B. (2017). Identification of circular RNAs and their targets in leaves of *Triticum aestivum* L. under dehydration stress. *Frontiers in plant science*, **7**, 2024-2034.

Ye, C. Y., Chen, L., Liu, C., Zhu, Q. H., and Fan, L. (2015). Widespread noncoding circular RNAs in plants. *New Phytologist*, **208**(1)**,** 88-95.

Yin, J., Liu, M., Ma, D., Wu, J., Li, S., Zhu, Y., and Han, B. (2018). Identification of circular RNAs and their targets during tomato fruit ripening. *Postharvest Biology and Technology*, **136,** 90-98.

Zhang, X. O., Wang, H. B., Zhang, Y., Lu, X., Chen, L. L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell*, **159**(1)**,** 134-147.

Zhao, W., Cheng, Y., Zhang, C., You, Q., Shen, X., Guo, W., and Jiao, Y. (2017). Genome-wide identification and characterization of circular RNAs by high throughput sequencing in soybean. *Scientific reports*, **7**(1)**,** 5636-1547.

Zhu, Q. H., and Wang, M. B. (2012). Molecular functions of long non-coding RNAs in plants. *Genes*, **3**(1), 176-190.

Zuo, J., Wang, Q., Zhu, B., Luo, Y., & Gao, L. (2016). Deciphering the roles of circRNAs on chilling injury in tomato. *Biochemical and biophysical research communications*, **479**(2), 132-138.