DIFFERENTIAL ALTERNATIVE SPLICING ANALYSIS UNDER ABIOTIC STRESS CONDITIONS IN BARLEY (*Hordeum vulgare*)

Project Work submitted to the Central University of Punjab

For the award of
Master of Science

In

Life Sciences
Specialization in Plant Sciences

BY
Diksha Bisht

Supervisor
Dr. Pankaj Bhardwaj

Department of Plant Sciences
School of Basic and Applied Sciences
Central University of Punjab, Bathinda

May, 2018

CERTIFICATE

I declare that the project work entitled "DIFFERENTIAL ALTERNATIVE SPLICING ANALYSIS UNDER ABIOTIC STRESS CONDITIONS IN BARLEY (*Hordeum vulgare*)" has been prepared by me under the guidance of Dr. Pankaj Bhardwaj, Assistant Professor, Department of Plant Sciences, School of Basic and Applied Sciences, Central University of Punjab. No part of this project work has formed the basis for the award of any degree or fellowship previously.

Diksha Bisht

Department of Plant Sciences
School of Basic and Applied Sciences
Central University of Punjab, Bathinda-151001
Date:

CERTIFICATE

I certify that Diksha Bisht has prepared her project work entitled "DIFFERENTIAL ALTERNATIVE SPLICING ANALYSIS UNDER ABIOTIC STRESS CONDITIONS IN BARLEY (*Hordeum vulgare*)", for the award of M.Sc. degree of the Central University of Punjab, under my guidance. She has carried out this work at the Department of Plant Sciences, School of Basic and Applied Sciences, Central University of Punjab.

Dr. Pankaj Bhardwaj

Department of Plant Sciences
School of Basic and Applied Sciences
Central University of Punjab, Bathinda-151001
Date:

ABSTRACT

Differential alternative splicing analysis under abiotic stress conditions in barley (*Hordeum vulgare*).

Name of student: Diksha Bisht

Registration number: 16mslsps05

Degree for which submitted: Master of Science

Name of supervisor: Dr. Pankaj Bhardwaj

Name of Department: Department of Plant Sciences

Name of School: School of Basic and Applied Sciences

Keywords: Transcriptomic data, differentially spliced genes, annotation, splicing factors.

Biotic and abiotic stress factors adversely affect the growth and development of plants, living in constantly changing environment. Plant abiotic stress responses seem to be frequently associated with the process of alternative splicing. Alternative splicing generates multiple products from a single gene by variable splicing of exons. RNA-seq utilizes NGS platforms to produce massive amount of sequence data. Using transcriptomic data, we tried to identify the differentially spliced genes in barley under heat and cadmium stress. Raw RNA seq data was pre-processed and aligned with reference barley genome. Then differential alternative splicing was analyzed on the basis of intron excision. In this study, a total of 196 significant differentially spliced clusters were found under two abiotic stress conditions, among them 47 and 149 clusters were associated with cadmium and heat stress condition, respectively. Overall, 133 fully annotated clusters and 63 unannotated clusters were observed. Further, annotation was performed for the significant genes to determine their functions. Some of these significant genes encode for splicing factors, stress tolerance protein, non snRNPs and antioxidants. From this study, it was concluded that barley undergo differential spliced events to maintain its homeostasis under different abiotic stress conditions.

(Diksha Bisht) (Dr. Pankaj Bhardwaj)

Signature of student Signature of supervisor

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Pankaj Bhardwaj, Assistant Professor, Department of Plant Sciences, Central University of Punjab for his continuous support, guidance, patience, motivation, enthusiasm, immense knowledge and constructive criticism throughout the course of investigation and during the preparation of this manuscript.

Besides my supervisor, I would like to thank all the Ph.D. Scholars of Department of Plant Sciences especially Aasim Majeed and Shruti Choudhary for their constant guidance, encouragement and support. One simply could not wish for a better or friendlier seniors. Aasim Majeed helped me a lot to get on the road to Bioinformatics and provided an experienced ear for my doubts about writing this project. In many ways I have learnt much from and because of them.

The cooperation of all the faculty members of Department of Plant Sciences is highly acknowledged. I am thankful to my lab mate and technical staff for their help and support. I would like to thank all my friends of DU and CUPB who helped me to get through two years of Masters.

Above all of them, I want to give my warmest regards to the all mighty god and my family for their blessings.

DIKSHA BISHT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

S.No.	**Full Form**	**Abbreviations**
1	Abscisic acid	ABA
2	Alternative splicing	AS
3	Base pair	bp
4	Basic Local Alignment Search Tool	BLAST
5	Branch binding point	BBP
6	Cadmium	Cd
7	Dehydration responsive factor	DRF
8	Kilo base	kb
9	messenger Ribonucleic Acid	mRNA
10	National Centre for Biotechnology Information	NCBI
11	Next Generation Sequence	NGS
12	Polypyrimidine tract	Py tract
13	RNA sequence	RNA-seq
14	Reverse transcriptase polymerase chain reaction	RT-PCR
15	Small nuclear RNAs	SnRNA
16	Small nuclear ribosomal proteins	SnRNPs
17	Short nucleotide archive	SRA
18	Splice site	SS

CHAPTER 1

INTRODUCTION

Hordeum vulgare (Barley) belongs to family Poaceae and is a very important cereal crop in the world. According to Food and Agriculture organization of the United States, it ranks 4th in world cereal production and also in the area of cultivation of cereal crops in the world. Annual barley production of India is 1.51 million tonnes. In temperate regions it is grown as summer crop whereas in tropical regions as winter crop. It has short growing seasons. Primarily barley is used for direct consumption for both human and animals. Apart from this it is also used for beer production, food processing, and malting industries (Zhang et al., 2016)

According to USDA (https://www.fas.usda.gov/) data, in India significant decrease in the barley production and area under cultivation has been observed in last 15 years. Many factors could be responsible for the declination in the barley production and restriction of its agriculture land. Abiotic stress is one of the major cause for it, particularly drought stress. Abiotic stress is defined as the adverse impact of non-living factors on plant growth and development (Zhu et al., 2016). It includes drought, flooding, metal toxicity, high temperature, mineral toxicity and salinity. Under different environmental conditions genes are expressed differently. mRNA isoforms and transcripts are also expressed variably in response to developmental and environmental stress conditions. Transcript level differential analysis enables us to study the variability across the replicates or samples and also the uncertainty in expression of abundant transcripts caused by reads-mapping ambiguity.

The high throughput RNA seq technology has revolutionized the today's biological research. RNA-Seq utilizes various next-generation sequencing platforms like Illumina, Roche, SOLiD, 454 (Martin et al., 2011). Compared to conventional Sanger Sequencing methods, it has allowed to produce sequence reads with size ranging from 100-1000 bp or even longer in a very short time scale (Martin et al., 2011). It generates high quality of robust data with low noise. NGS data is widely used in transcriptomic analysis, ribosome profiling, chip-seq, de-novo sequencing, meta-genomics, shotgun sequencing etc. Now a days it has become a powerful tool to address the challenges of transcriptomic study. Over microarrays RNA seq has provided an unbiased view of transcriptomics data. It enables rapid transcript profiling, small RNA analysis, targeted RNA expression and comparative study of relative abundant transcripts for any species. Bioinformatics tools has allowed us to analyze the transcriptomics data. After obtaining high quality reads, first step of analysis includes alignment and mapping to the reference genome (Wang et al., 2009). Further downstream analysis is based on the objective to study and convenient bioinformatics tools.

Progress in the sequencing and high-throughput technologies has resulted in massive amount of sequencing and transcriptomics data. It has thrown the light towards the study of transcriptional level mechanism and regulatory pathways. Alternative Splicing based research in plants had made a significant progress in past few years. Alternative splicing is a regulated process where pre-mRNA are spliced differently to form variable protein products. It play a major role to generate functional diversity of proteins. As a single gene is responsible for multiple proteins formation. It is making system economically and energetically more efficient. NGS data along with the advent of bioinformatics tools has allowed computational analysis to identify conservation of AS events across the species under different environmental conditions. It also enables to understand the mechanism, function and regulation of AS.

In view of this, the objective of the study is to analyze the differential alternative splicing under cadmium and heat stress condition in barley. This study of differential alternative splicing helps us to identify the functions of genes undergoing alternative splicing and are being differentially spliced to produce isoforms under different abiotic conditions in barley.

CHAPTER 2

REVIEW OF LITERATURE

Plants are living in a constantly changing environment that can be either favorable or stressful for its proper growth and development. The adverse conditions include both biotic and abiotic factors. Temperature, drought, heavy metals, salt are some of the major abiotic factors influencing the geographical distribution, high productivity in agriculture and also threatening food security. So improving plant resistance and tolerance to combat stress conditions has become a challenge to work with. Alternative splicing was first discovered in 1977 in adenovirus as a result of rearrangement in type 2 mRNA (Berget et al. 1977; Chow et al. 1977). Alternative splicing (AS) is the process in which exons of pre mRNAs are spliced with different arrangements to produce multiple products from a single gene. This results in excision of the introns. Introns are the intervening sequences constituting the non-coding region of the genes. The multiple products resulting from alternative splicing of a single gene are named as isoforms. They constitute structurally and functionally different mRNA and protein variants. In plants more than 60% of genes undergo alternative splicing process. It regulate the alterations in gene expressions during the growth, development and response to environmental stresses. Regulation in different cell conditions depends on the abundance and interactions of RNA-binding proteins (Naeem H. Syed et al. 2012). Generation of the huge amount of sequencing data and its analysis has revealed new pathways towards the study of alternative splicing and its regulatory role in differential gene expression. Recent updates show that in *Arabidopsis* the frequency of AS is more than 61%, when grown in normal growth conditions. On the other hand, it was increased variably when different tissues at developmental stages and growth conditions were analyzed. (Marquez, Y. et al. 2012). Alternative Splicing can be cell, tissue or condition specific process regulating the post transcriptional mechanism (Pan et al., 2008; Merkin et al., 2012).

2.1) Mechanism of alternative splicing: spliceosome assembly

Alternative splicing is a process to generate isoforms of mature mRNA by the variable splicing of exons. Pre-mRNAs include both introns and exons. The borders of introns and exons are marked by specific sequences. 5’end of the intron is defined as the 5’splice site (SS) marking the exon-intron boundary at 5’ end. At

3’end of intron, intron-exon boundary is known as 3’ splice site. Branchpoint site is found to be located entirely within the intron followed by Polypyrimidine tract. Two transesterification reactions are involved in the removal of an intron. First reaction takes place between the 2-OH of branchpoint and phosphoryl group of 5’ splice site, 2-OH act as nucleophile to attack the phosphoryl group of 5’SS. The next transesterification reaction occurs between 3-OH of 5’ SS that now acts as nucleophile to phosphoryl of 3’SS. This resulted in the splicing of two exons and removal of an intron in the form of lariat. A properly regulated co-transcriptional pathway is needed for the alternative splicing (Lee et al., 2015). It is performed by the large molecular complex named as spliceosome comprising of five small nuclear RNAs, snRNAs (U1, U2, U4, U5 and U6) along with several proteins. The RNA-protein complexes are called small nuclear ribonuclear proteins (snRNPs). Besides these snRNPs it also requires recognition of splice sites, RNA-RNA, RNA-protein and protein-protein interactions (Reddy et al., 2013). 5’SS is first recognized by U1 snRNP and then U6 snRNP. U2snRNP binds to the branchpoint and displaces Branchpoint-binding protein (BBP). U2 snRNP recognizes branchpoint site. U4 and U5 snRNPs are loaded along with U6snRNP as tri-snRNP particle and U1snRNPS leaves the 5’SS. Interaction between U2 and U6 snRNPs bring 5’SS and branch site together. This drives the first splicing reaction. Some non-snRNPs are also involved like U2AF that recognize Py tract/3’SS. Other than this some RNA-annealing factors and DEAD-box helicase proteins are also required for second transesterification reaction. Second transesterification reaction is then aided by U5snRNP, it allows the splicing of two exons together and removal of intron forming a lariat. The mechanism of alternative splicing and its regulation is well elucidated in animals. However, development in the sequencing technologies has led to the availability of whole genome sequence data in many plants species also. This has thrown light to identify and study the protein orthologs and spliceosome machinery in plants. Plants have specific spilceosomal machinery and intronic sequences that differ from animals in many ways. Plants have much smaller intron then animals with an average size of 160bp, animals can have as large as 5kb of intronic sequences (Barbazuk et al., 2008). In Plants uridine rich sequences at 3'splice site is found to be more important for splicing efficiency. This difference may influence the splicing effect and RNA polymerase elongation rate. Although in plants, difference can also be seen between the species like dicotyledons have more than 60% of AU rich introns as compared to the monocotyledons (Barbazuk et al., 2008).

2.2) Types of alternative splicing

Alternative splicing can occur in various ways. Exon skipping, alternative 3' site, alternative 5' splice site, intron retention and constitutive splicing, are the five most common types of alternative splicing known (Laloum et al., 2017). Exon skipping type of alternative splicing either includes or excludes the exon. In alternative 3' and 5' site alternative splicing, variable mRNAs are generated based on usage of a proximal and distal 3' and 5' splice site respectively. Intron retention involves retention or excision of intron in mRNA. Adjacent exons are spliced in constitutive type of alternative splicing. More than one type of alternative splicing types could occur to generate multiple mRNA products from a single gene.

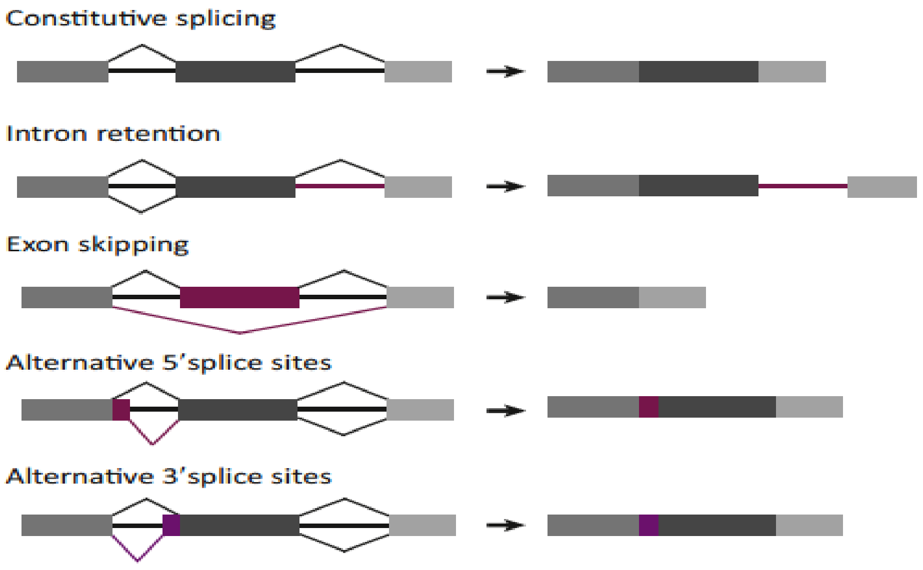

Figure 1: Most common five types of Alternative Splicing. Source: Laloum et al., 2017

In *Arabidopsis* a large fraction of intron retention is observed and confirmed by RT-PCR analysis. As the partially spliced transcripts tend to remain in nucleus, it may play a regulatory role in nuclear transport, RNA stability and translation efficiency (Ner-Gaon et al., 2004). For example, an intron retained in the 3' UTR in yeast HAC1 transcript inhibit the mRNA translation as a part of unfolded protein response (Ruegsegger et al., 2001). Such regulatory mechanisms not yet been discovered in plants but are becoming major area of research.

2.3) Role of alternative splicing in plant stress responses

Plant's exposure to various abiotic and biotic stresses alter the efficiency and patterns of alternative splicing. Plants stress related genes are more prone to undergo alternative splicing events. High throughput analysis using RNA-Seq data have confirmed the indications that abiotic stresses alters the alternative splicing events in plants ((Laloum et al., 2017). Recent studies shows that in grapes more than 1000 genes undergo differential splicing under high temperature stress (J et al., 2017), whereas in *Arabidopsis* salt stress induces alternative splicing events of more than 6000 genes (Feng et al., 2015). Under heat stress condition two alternatively spliced isoforms of heat shock transcription factors are generated in *Arabidopsis*. One isoform is involves in nonsense-mediated decay where as other binds to the promoter and activates its own transcription (Matsukura et al., 2010).

In Barley*,* dehydration-responsive factor 1 (HvDRF1) member of Apetala2/ethylene-responsive factor (AP2/ERF) is associated with abscisic acid (ABA) mediated gene regulation and produces three isoforms by alternative splicing. Two of them encoded for AP2 proteins. Expression of HvDRF1 gene is upregulated in barley under salt, drought, ABA stress condition and in embryos during seed maturation. Both isoforms of HvDRF1 gene are transcriptional activators and promotes the activity of HVA1s, an ABA inducible gene. (Xue et al., 2004). This suggests that HvDRF1 transcriptional activator along with ABA-responsive factors involves in upregulation of stress gene expressions. Also in some plants, studies report that alternative splicing in pre-mRNA of resistance R gene (Reddy et al., 2007) are responsible for various plant defense responses. In addition to alternative splicing process, splicing regulators act as mediators to regulate the plant stress response. Plants Serine/Arginine rich genes are regulated by stress at both transcriptional and translational levels (Tanabe et al., 2007 and Zhang et al., 2014). For example, during Cd stress, expression of SR34b gene is upregulated whereas loss of function of SR34b gene promotes the accumulation of Cd ions and toxicity in *Arabidopsis.* Cd interferes with many metabolic processes, causes chlorosis and necrosis, reduces chlorophyll production and photosynthetic efficiency, lipid peroxidation in different plant parts and increases the production of ROS. Different spilceosomal components also affects the plant responses to stress. In *Arabidopsis* loss of function of SNW, Ski-interacting protein led to the oversensitivity of plant towards salt and osmotic stress (Feng et al., 2015). SNW protein interacts with SR45 protein and induces alternative splicing through recognition of splice sites (Wang et al., 2012). Regulation of alternative splicing thus represents the transcriptional activation and inactivation of pre-mRNA. It allows plant to adapt and survive in adverse environmental conditions (Laloum et al., 2017). Alternative splicing analysis studies can pave the new strategies for selecting genotypes with higher productivity under unfavorable environmental conditions (Mastrangelo et al., 2012).

2.4) Differential alternative splicing in barley

Hordeum vulgare species was selected for the in silico analysis of differential alternative splicing. Genus Hordeum includes total 16 species and 26 accepted taxa, as per the Gramene database. It holds good historical record as a model plant for genetic studies (Schulte et al. 2009). This study aims at differential alternative splicing on the basis of intron usage during alternative splicing events. Besides increasing proteomic diversity, alternative splicing also plays a regulatory role in stress tolerance mechanism and regulates the level of functional transcripts RUST (Regulated Unproductive Splicing and Translation). Another reason to opt for Barley is that ample amount of transcriptomics data is freely available in various barley databases along with its whole genome information. Lastly with the best of my knowledge, this kind of intron usage based differential alternative splicing study in barley under abiotic stress condition has not been done earlier.

Differential alternative splicing was observed during seed germination in barley (Zhang et al., 2016). In between 552 to 669 common alternative splicing were found in germinating embryos of four varieties of barley. AS transcripts were found predominantly in regulatory pathways, protein and energy pathways. VP1 gene is involved in seed germination process undergo consistent alternative splicing in all four varieties of barley. But, it is differentially expressed at different time period. For example, it underwent Alternative 3' splice site alternative splicing at 24 hour and 48 hour, alternative 5' splice site and intron retention type of alternative splicing at 48 hour (Zhang et al., 2016). Differential alternative splicing analysis can be based upon numerous aspects such as exon usage, intron excision, splice junctions and quantifying differential isoforms within one or more than two conditions. Various tools are developed to study the differential alternative splicing like Cufflinks, Diffsplice, FDM, DEXSeq, rMATS etc. However, cufflinks estimates the uncertainty but statistic test does not differentiates the replicate's contribution with low or high degree of uncertainty (Trapnell et al., 2013). Diffsplice is unable to measure uncertainty of isoforms quantification in individual replicates (Hu et al., 2013). DEXSeq works with the different approach of testing for read counts deviation of individual exon as compared to whole gene but statistically it does not estimate the isoform proportion (Anders et al., 2012). rMATS requires the reads of equal length that might be responsible for loss of some information. In view of this, we have used leafcutter (Li et al., 2017), an annotation free software which identifies differential alternative splicing based on intron usage. It employs RNA-seq data to quantify RNA splicing variation. It have many advantages over the exon based quantification. It focus on what is removed during splicing events. Excised introns can be inferred directly from the reads spanning exon junctions. Thus it provides an accurate quantification usage ratio and is little ambiguous about the excised intron. Leafcutter allows easy detection of novel introns and modelling of complex splicing events. It also excludes the challenge of isoform abundance estimation. It emphases on all the alternative splicing events that can be detected by intron excision differences like exon skipping, 5' and 3' splice site usage (Li et al., 2017).

CHAPTER 3

METHODOLOGY

3.1) Data collection

Barley reference genome assembly and annotation file ((Mascher et al., 2017) were downloaded from Plant Genome System Biology database (http://pgsb.helmholtz-muenchen.de/plant/plantsdb.jsp). Raw RNA seq data of *Hordeum vulgare* under Cd (80 µ mol) and heat stress (35℃) conditions were retrieved from SRA-NCBI database. Detailed list of experiments runs along with accession number of RNA seq data is given in Table1.

Table1: Details of transcriptomics data of barley.

S.No	*Hordeum vulgare* Condition	Runs (accession number)	Tissue
1	Cd stress control	i) SRX2740785 ii) SRX2740783 iii) SRX2740781	Root
2	Cd stress treated (80 µ mol)	i) SRX2740779 ii) SRX2740777 iii) SRX2740775	Root
3	Cd stress control	i) SRX2740786 ii) SRX2740784 iii) SRX2740782	Shoot
4	Cd stress treated (80 µ mol)	i) SRX2740780 ii) SRX2740778 iii) SRX2740776	Shoot
5	Heat stress control (22℃)	i) SRX1815080 ii) SRX1815079 iii) SRX1815078	Shoot
6	Heat stress treated (35.5℃)	i) SRX1815083 ii) SRX1815082 iii) SRX1815081	Shoot

3.2) Data Analysis

The raw reads were first pre-processed and quality reports were generated by using FastQC (Andrew, 2010). Cleaning of raw reads was done. The low-quality bases and adapter sequences were then trimmed by TRIM-GALORE. The pre-processed reads were then mapped to reference genome assembly using STAR (Dobin et al., 2013) software version 2.5.1a. Using STAR, first, genome indexes were generated followed by 2nd pass mapping. 2nd pass mapping was done for most sensitive novel junctions. Differential alternative splicing was performed using software leafcutter (Li et al., 2017). Leafcutter is an open source software used to quantify variable intron splicing events from RNA seq data across samples. It incapacitates the challenges to estimate the exon usage and estimating relative isoforms in complex splicing events. STAR mapped sam files were then converted to bam files using samtools. These bam files were converted to junction files using Leafcutter. It pooled all the mapped reads and found overlapping introns which were demarcated by split reads. These junction files were then used to generate count files containing intron clusters together. It was followed by differential intron excision analysis, using the counts from clustering step.

The plotting of differentially spliced clusters and junctions was done using R functions of leafcutter. Visualization was then performed by using Leafcutter Shiny App. Then differentially spliced clusters with significant p value and delta PSI value were obtained.

Sequences of the genes having clusters with significant p value < 0.01 and delta PSI value > 0.1, were extracted from genome FASTA file using EMBOSS. BLAST was performed for individual gene against the *Hordeum vulgare* genome using BLAST search tool in Plant Genome System Biology. The locus id with the highest score was used for the next step, annotation. Since the barley annotation file lacks gene level annotation, so we functionally annotated the significant genes. Annotation was done online using the gene annotation tool in PGSB database.

The flowchart of pre-processing and differential splicing analysis is represented in figure 1.

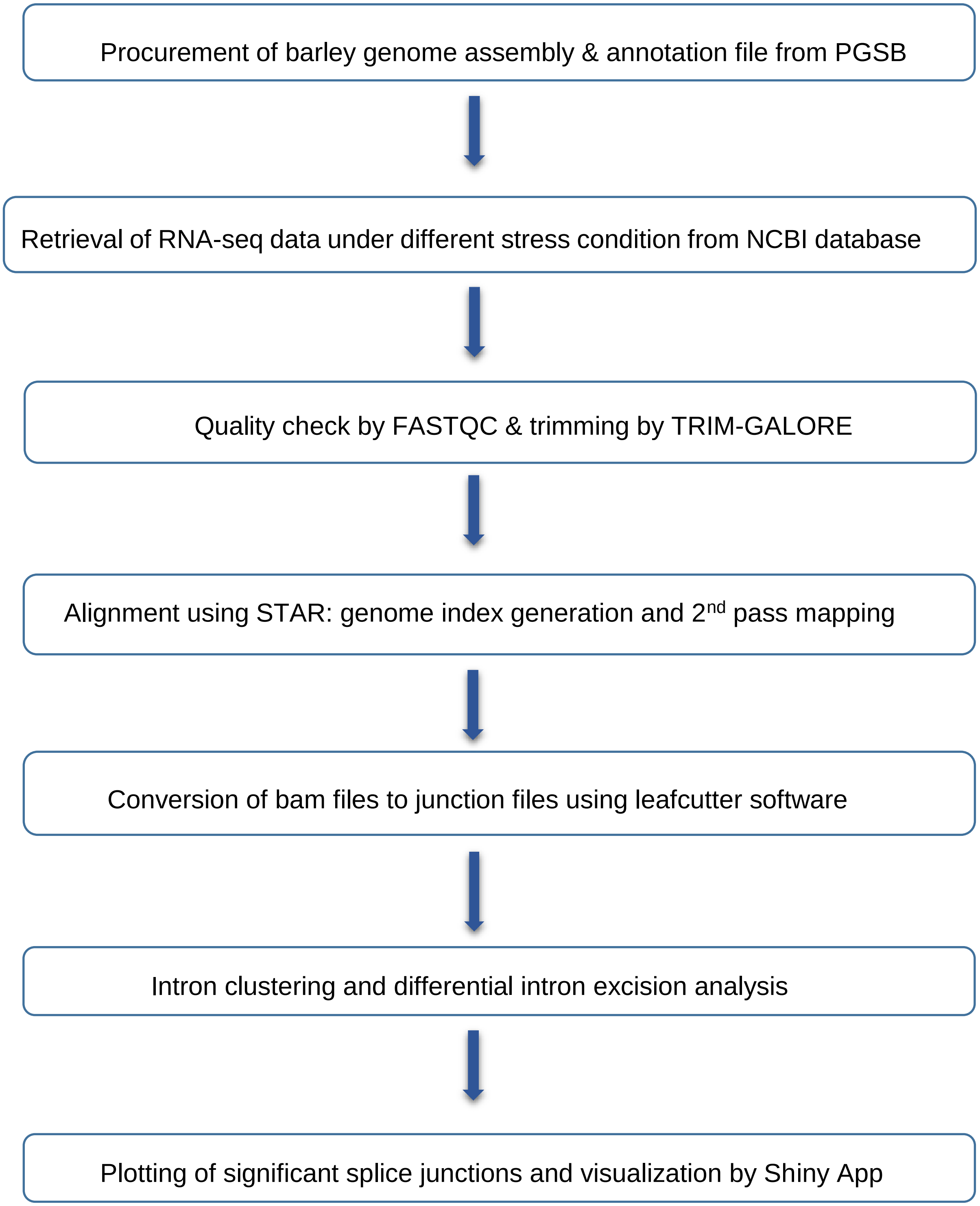

Figure 2: Workflow of RNA seq data processing and differential alternative splicing analysis

CHAPTER 4

RESULTS

The pre-processed reads of barley under Cd stress (root/shoot) and heat stress (root) were aligned against reference genome. Further analysis was carried out to obtain differentially spliced clusters and genes undergoing alternative splicing process. Differential intron usage across the conditions was studied using leafcutter. A cluster is defined as set of overlapping introns or spliced junctions. A number of differentially spliced clusters were identified. These clusters were filtered based on their statistical significance (≤ 0.01). Under cadmium stress (shoot), 465 clusters (figure 3) were generated as success clusters out of which 5 clusters (1.07%) were highly significant which involve 4 genes and 13 introns in alternative splicing events. Among these significant clusters, only 1 cluster with 5 splice junctions were found to be fully annotated (figure 4). Remaining 4 clusters had unannotated junctions. In roots, 433 success clusters and 42 (9.69%) highly significant clusters were obtained (figure 3). These highly significant clusters involved 34 genes and 103 introns. In addition to this, 29 clusters and 81 splice junctions were fully annotated whereas 13 clusters (figure 4) were identified with unannotated junctions. Under heat stress (shoot), 884 success clusters were found out of which 149 (16.85%) were highly significant which include 105 genes and 395 introns participating in alternative splicing events (figure 3). Amongst highly significant clusters, 103 clusters and 278 junctions were fully annotated while 46 highly significant clusters had unannotated junctions (figure 4).

In total, 133 fully annotated clusters with 63 clusters containing novel junctions. Further, 364 fully annotated junctions and 19 novel junctions were obtained in two different abiotic stress conditions. Also, 23 junctions with cryptic 5’ splice site, 7 junctions with cryptic 3’ splice site and 24 junctions with both cryptic sites were found. None of the clusters were found to be common in the stress conditions studied. Detailed report for the clusters and junctions obtained by leafcutter analysis are presented in Table 2.

Table 2: Details of highly significant clusters and junctions as obtained by differential analysis of Barley under Cd and heat stress conditions.

S.N	**SAMPLES**	**Cd stress, Shoot**	**Cd stress, Root**	**Heat stress, Shoot**
i	Control	3 reps	3 reps	3 reps
ii	Treated	3 reps	3 reps	3 reps
	CLUSTERS			
i	Number of differentially spliced clusters	5	42	149
ii	Fully annotated	1	29	103
iii	Contain unannotated junctions	4	13	46
	JUNCTIONS			
i	Number of fully annotated junctions	5	81	278
ii	Number of junction with cryptic 5'splice site	1	6	16
iii	Number of junction with cryptic 3'splice site	2	4	1
iv	Number of junction with two cryptic splice site	1	0	23
v	Number of novel junctions that connect two annotated splice sites	1	3	15

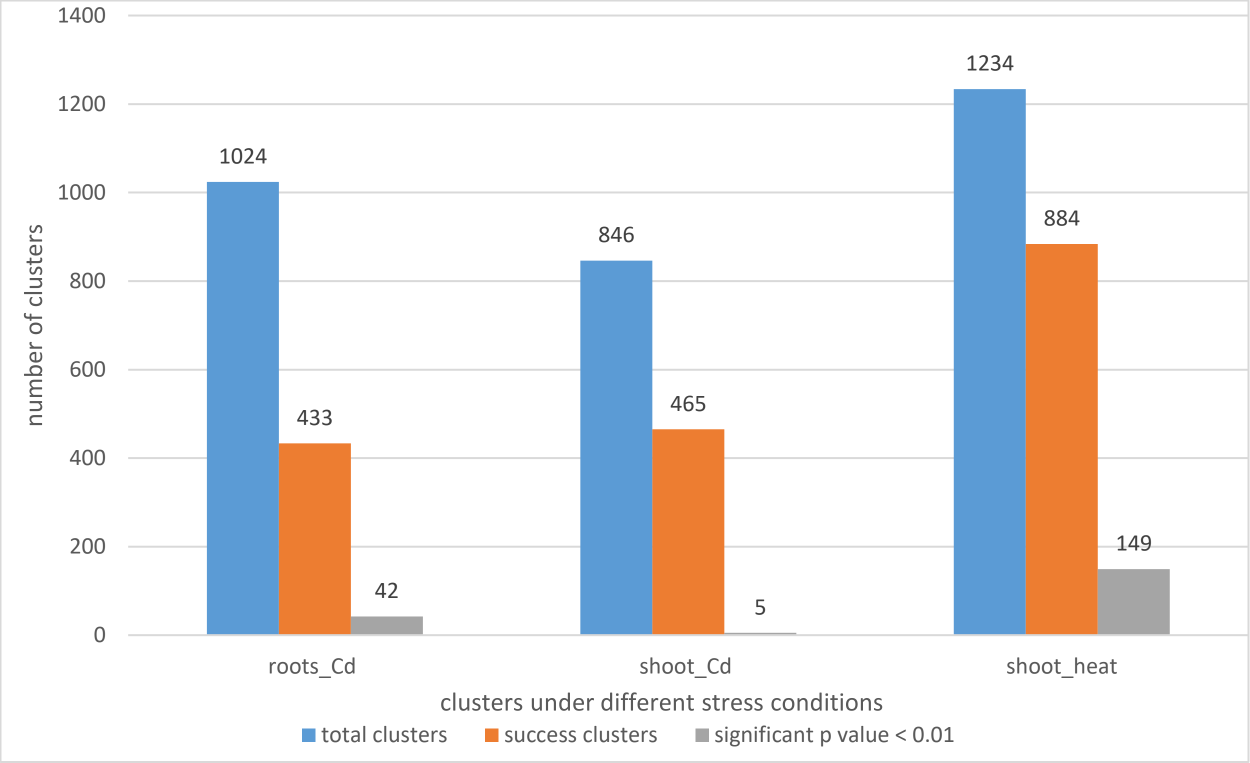

Figure3. A graphical representation of significant differentially spliced clusters under two different abiotic stress conditions.

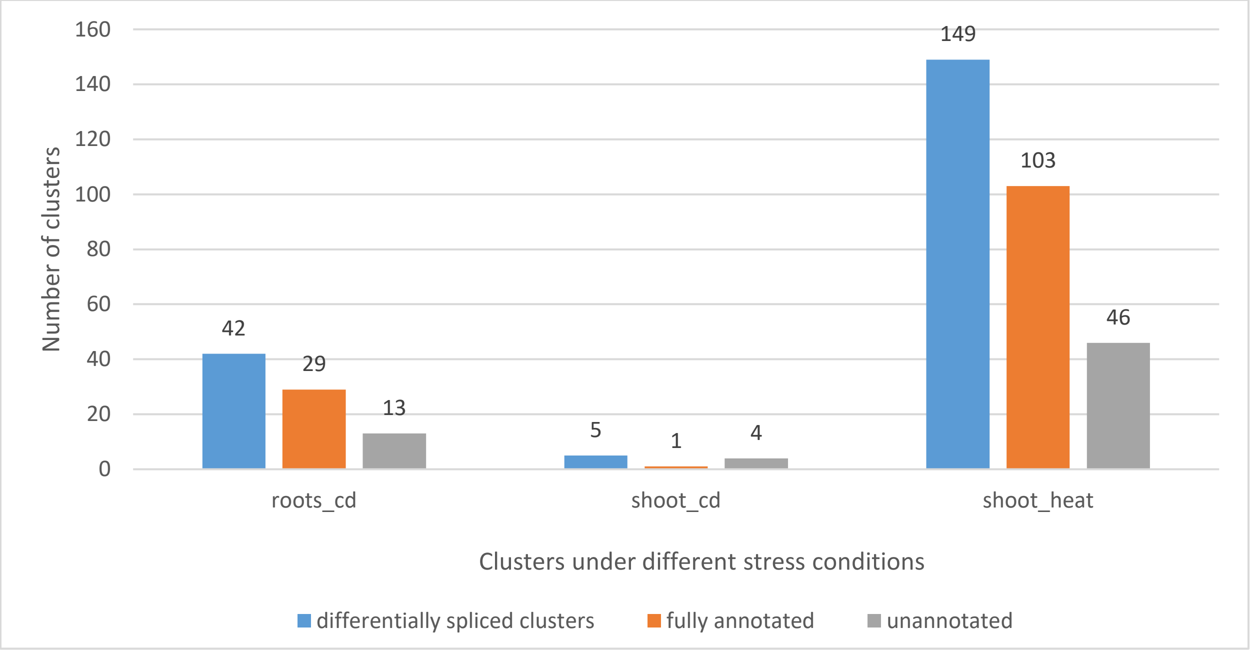

Figure 4. A graphical representation of fully annotated and unannotated highly significant differentially spliced clusters.

Differentially spliced genes were identified based on the combination of both, p value (≤0.01) and delta PSI value (≥0.1). A total of 73 genes were found to be significant (Table 2), among which 2 and 17 genes were involved in cadmium stress in shoot and root respectively while 54 genes belonged to heat stress in shoot of barley.

Gene HORVU7Hr1G103170 encodes for Polyubiquitin under Cd stress condition in shoot of barley. Splicing event visualizations signified that in treated conditions 'intron a' underwent more significant spliced events with the ΔPSI value of 0.384 involving cluster 657_+ (figure 7.1).

In roots, genes HORVU2Hr1G019410 and HORVU4Hr1G029030 encode for Arginine/serine-rich splicing factor and Splicing factor U2af large subunit A respectively. Both of these splicing factors are involved in alternative splicing machinery. In HORVU2Hr1G019410, intron b had more supported spliced junctions in treated conditions where as HORVU4Hr1G029030 had intron c (figure 5.1, 5.2). Genes identified in shoot under heat stress conditions includes HORVU1Hr1G030790, HORVU2Hr1G074280 and HORVU7Hr1G085150 which encodes for Chaperone DnaK, PAP fibrillin family protein and Lactoylglutathione lyase respectively (figure5.3, 5.4, 5.5). In all these three genes 'intron a' underwent more number of splicing events indicating its more participation in splicing process.

In addition to these, three genes from each of the studied conditions, HORVU7Hr1G103170 (Cd stress in shoots), HORVU7Hr1G120030 (Cd stress in roots) and HORVU1Hr1G043930 (heat stress in shoots) were found to encode for Polyubiquitin protein (figure 7). Further, HORVU3Hr1G115540 is found common in two stress conditions, encoding for telomere binding proteins (figure 6.1, 6.2).

Table 3: List of annotated significant genes with their p-value, ΔPSI value and their function

S.N	Gene	Clusters	ΔPSI	p value	Function
1	HORVU6Hr1G050160	Chr6H:clu_344_+	0.426	6.55E-10	Retrotransposon protein, putative, unclassified
2	HORVU7Hr1G103170	Chr7H:clu_657_+	0.384	1.28E-08	Polyubiquitin
3	HORVU1Hr1G079380	chr1H:clu_340_+	0.17	0.002	Hydrolase, alpha/beta fold family protein
4	HORVU2Hr1G019410	chr2H:clu_98_+	0.155	0.001	Arginine/serine-rich splicing factor
5	HORVU2Hr1G109120	chr2H:clu_554_-	0.347	0.0004	beta-fructofuranosidase
6	HORVU2Hr1G082800	chr2H:clu_542_-	0.182	0.002	Isocitrate dehydrogenase, putative
7	HORVU3Hr1G115540	chr3H:clu_93_-	0.357	8.56E-05	Telomere-binding protein 1
8	HORVU3Hr1G077800	chr3H:clu_76_-	0.248	0.001	unknown protein
9	HORVU3Hr1G053490	Chr3H:clu_833_+	0.114	0.001	N-alpha-acetyltransferase 50
10	HORVU4Hr1G029030	chr4H:clu_900_+	0.261	0.001	Splicing factor U2af large subunit A
11	HORVU4Hr1G026340	chr4H:clu_895_+	0.333	0.0003	Major facilitator superfamily antiporter
12	HORVU4Hr1G052170	Chr4H:clu_667_-	0.127	0.001	Alpha-L-arabinofuranosidase
13	HORVU4Hr1G057650	chr4H:clu_672_-	0.205	0.003	5-hydroxyisourate hydrolase
14	HORVU5Hr1G085070	chr5H:clu_456_-	0.163	7.56E-05	Histone H1.1, putative
15	HORVU5Hr1G103130	chr5H:clu_1012_+	0.147	0.002,	1-aminocyclopropane-1-carboxylate synthase
16	HORVU5Hr1G091970	chr5H:clu_461_-	0.189	0.003	Protein kinase superfamily protein
17	HORVU7Hr1G120030	chr7H:clu_781_+	0.501	8.33E-06	Polyubiquitin
18	HORVU7Hr1G075790	chr7H:clu_220_	0.253	0.001	Lysine-specific demethylase 5B
19	HORVU7Hr1G033230	chr7H:clu_182_-	0.269	0.003	Sucrose synthase
20	HORVU1Hr1G063320	chr1H:clu_431_+	0.133	1.79E-05	Cysteine proteinase inhibitor
21	HORVU1Hr1G030790	chr1H:clu_333_-	0.181	6.38E-05	Chaperone DnaK
22	HORVU1Hr1G035720	chr1H:clu_342_-	0.243	6.69E-06	Ribulose bisphosphate carboxylase small chain
23	HORVU1Hr1G073170	chr1H:clu_438_+	0.154	6.28E-05	Zinc finger protein LSD1
24	HORVU1Hr1G083730	chr1H:clu_446_+	0.225	0.001	Protein disulfide isomerase family
25	HORVU1Hr1G028030	chr1H:clu_330_-	0.155	0.005	Malic enzyme
26	HORVU1Hr1G037460	chr1H:clu_343_-	0.188	0.00643	Ganglioside induced differentiation associated protein
27	HORVU1Hr1G043930	chr1H:clu_346_-	0.249	0.00709	Ubiquitin
28	HORVU2Hr1G017960	chr2H:clu_128_+, chr2H:clu_614_-	0.18,0.41	0.001,1.22E-06	Retrotransposon protein, putative, unclassified
29	HORVU2Hr1G060770	chr2H:clu_163_+	0.221	2.13E-05	Endonuclease-reverse transcriptase
30	HORVU2Hr1G045230	chr2H:clu_638_-	0.219	0.0008	Oral cancer overexpressed protein 1
31	HORVU2Hr1G013400	chr2H:clu_613_-	0.199	0.001	Pseudo-response regulator
32	HORVU2Hr1G028210	chr2H:clu_625_-	0.122	0.001	Transcription antitermination protein nusG
33	HORVU2Hr1G074280	chr2H:clu_663_-	0.198	0.002	PAP fibrillin family protein
34	HORVU3Hr1G113740	chr3H:clu_108_-	0.122	2.63E-05	Receptor-like kinase
35	HORVU3Hr1G074850	chr3H:clu_80_-	0.261	0.0003	Plastid transcriptionally active 12

37	HORVU3Hr1G074210	chr3H:clu_79_-	0.21	0.0003	Hemimethylated DNA binding protein
38	HORVU3Hr1G030250	Chr3H:clu_978_+	0.119	0.0044	Unknown
39	HORVU3Hr1G115540	chr3H:clu_1065_+, clu113_-	0.2,0.128	0.0004	Telomere-binding protein 1
40	HORVU3Hr1G068840	chr3H:clu_73_-	0.36	0.0003	WD-40 repeat family protein
41	HORVU3Hr1G116880	Chr3H: clu_1069_+	,0.13	0.004	Vacuolar fusion protein MON1-like protein
42	HORVU3Hr1G030580	chr3H:clu_979_+	0.17	0.0006	Enoyl-CoA hydratase/isomerase family protein
43	HORVU3Hr1G089190	chr3H:clu_1047_+	0.272	0.0006	Pentatricopeptide repeat (PPR) superfamily protein
44	HORVU3Hr1G006250	chr3H:clu_12_-	0.416	0.0007	Phosphatidylserine decarboxylase proenzyme 2
45	HORVU3Hr1G085480	chr3H:clu_87_-	0.146	0.004	Fructokinase, putative
46	HORVU3Hr1G098170	chr3H:clu_1056_+,clu_100_-	0.24,0.20	0.0009	Mitochondrial carrier protein-like
47	HORVU4Hr1G010270	chr4H:clu_790_-	0.127	1.27E-07	Acetylglucosaminyltransferase family protein
48	HORVU4Hr1G060940	chr4H:clu_1129_+, chr4H:clu_844_-	0.3,0.279	2.04E-05,6.37E 05	RNA-binding protein
49	HORVU4Hr1G034300	chr4H:clu_818_-	0.14	2.86E-05	Inositol monophosphatase 3, putative, expressed
50	HORVU4Hr1G040290	chr4H:clu_1112_+, chr4H:clu_824_-	0.18,0.152	0.001,0.001	cDNA clone:J033145I11, full insert sequence
51	HORVU4Hr1G047340	chr4H:clu_1115_+,clu_827_-	0.128,0.11	0.0001	Transmembrane protein
52	HORVU4Hr1G011120	chr4H:clu_1083_+, chr4H:clu_792_-	0.15,0.17	0.0001,0.001	Methyltransferase
53	HORVU4Hr1G071070	chr4H:clu_1137_+, chr4H:clu_852_-	0.221,0.19	0.007,0.0003	Nogo-B receptor
54	HORVU4Hr1G051620	chr4H:clu_830_-,clu1120_+	0.19,0.16	0.0004	Uncharacterised conserved protein UCP022280
55	HORVU5Hr1G044410	chr5H:clu_1182_+	0.114	5.82E-06	Hydroxymethylglutaryl-CoA lyase
56	HORVU5Hr1G014170	chr5H:clu_544_-	0.207	4.79E-05	bZIP transcription factor family protein
57	HORVU5Hr1G051230	chr5H:clu_563_-	0.102	0.0001	L-rich repeat receptor-like protein kinase family
58	HORVU5Hr1G054450	chr5H:clu_570_-	0.177	0.002	Myb family transcription factor
59	HORVU5Hr1G111530	chr5H:clu_604_-	0.115	0.002	RING/U-box superfamily protein
60	HORVU5Hr1G119790	chr5H:clu_1228_+	0.18	0.003	Carotenoid isomerase 1
61	HORVU5Hr1G066790	chr5H:clu_578_-	0.19	0.003	Unknown protein
62	HORVU6Hr1G016240	chr6H:clu_705_-	0.23	0.0005	G-box-binding factor
63	HORVU6Hr1G037500	chr6H:clu_730_-,clu484_+	0.31,0.243	0.001	Peptidyl-prolyl cis-trans isomerase
64	HORVU6Hr1G036640	chr6H:clu_482_+	0.197	0.004	ARF-GAP domain 5
65	HORVU7Hr1G085150	chr7H:clu_280_-, chr7H:clu_937_+	0.272,0.11	0.001	Lactoylglutathione lyase
66	HORVU7Hr1G070870	chr7H:clu_924_+, chr7H:clu_264_-	0.21,0.25	6.11E-09	Circadian clock associated 1
67	HORVU7Hr1G121210	Chr7H:clu_955_+	0.14	3.24E-05	At5g37260-like protein
68	HORVU7Hr1G074890	chr7H:clu_927_+	0.114	0.0002	Succinate dehydrogenase iron sulfur subunit B
69	HORVU7Hr1G017880	chr7H:clu_212_-, chr7H:clu_878_+	0.107,0.11	0.0002, 0.0004	Ribosome-binding factor A
70	HORVU7Hr1G088210	chr7H:clu_938_+	0.264	0.0005	Multidrug resistance protein ABC transporter family
71	HORVU7Hr1G044150	chr7H:clu_236_-	0.239	0.001	Major facilitator superfamily protein
72	HORVU7Hr1G120660	chr7H:clu_311_-, chr7H:clu_954_+	0.20,0.242	0.005,0.001	Phytoene synthase 7A
73	HORVU7Hr1G037600	chr7H:clu_229_-	0.18	0.003	Calcium-binding EF-hand family protein

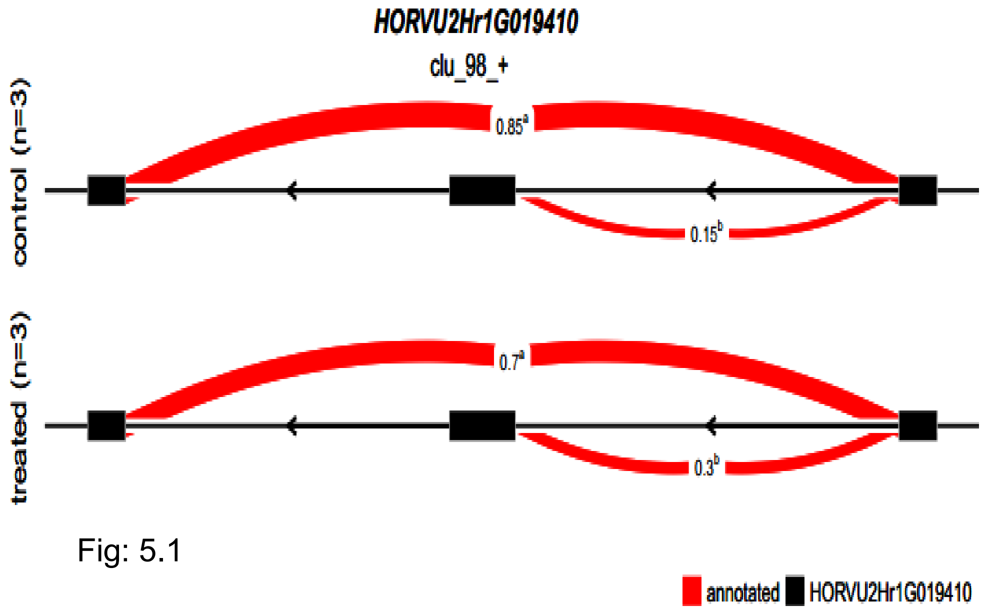

Fig: 5.1

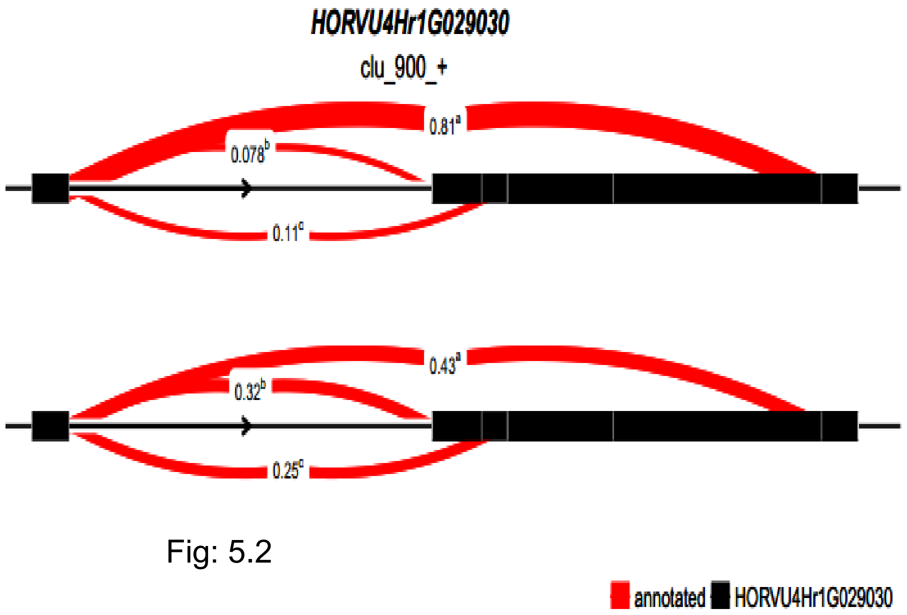

Fig: 5.2

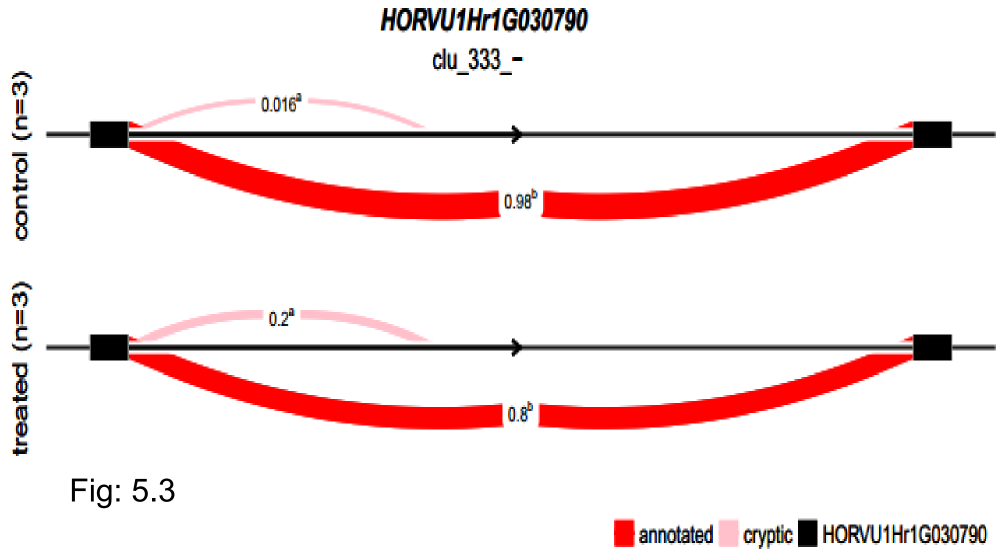

Fig: 5.3

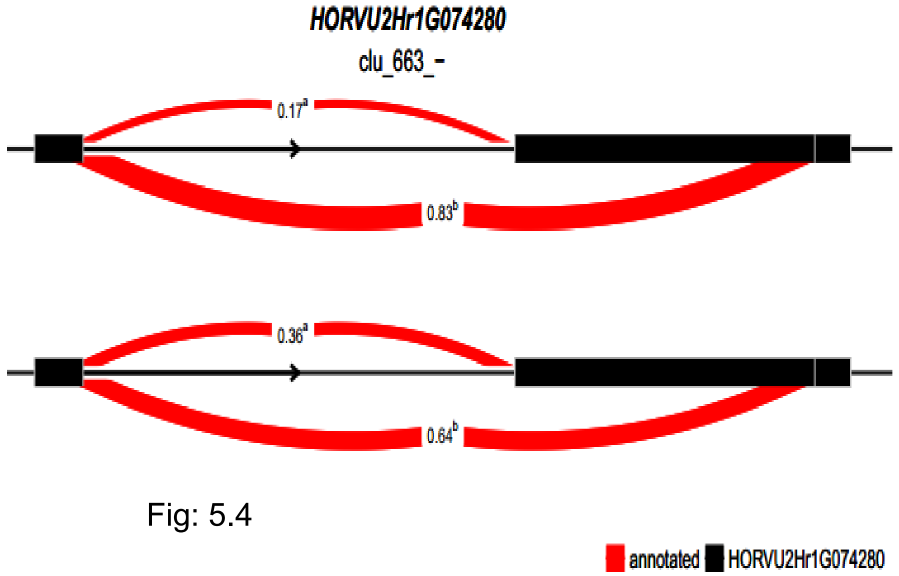

Fig: 5.4

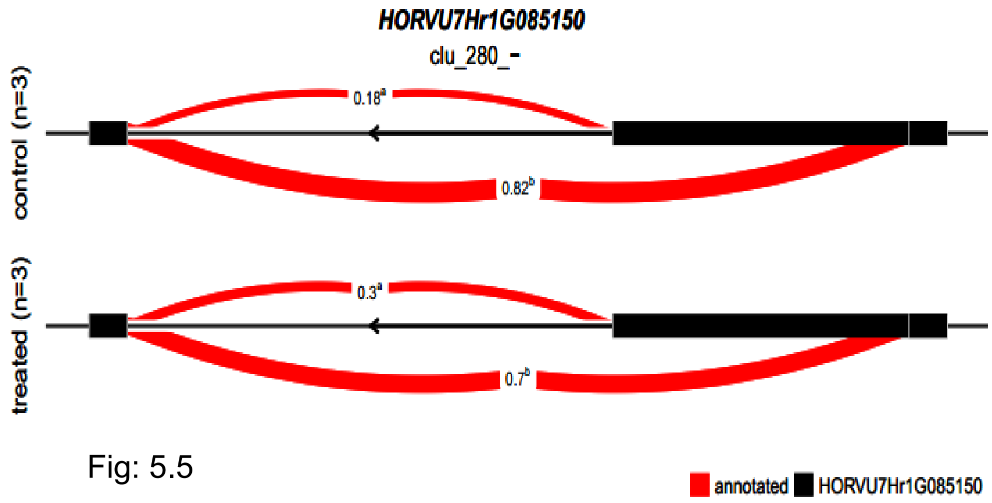

Fig: 5.5

Figure 5: Splicing event visualization of stress related genes. **5.1, 5.2** genes encoding splicing factors under Cd stress (roots). **5.3, 5.4, 5.5** genes encoding DnaK, PAP fibrillin and lyase proteins respectively under stress condition (shoot).

The line connecting the start and end coordinates represent an intron with a 'letter value'. Each intron is represented with a normalized count value, which is the mean number of splice junctions supporting that particular intron and delta PSI (percent spliced in) value. Black color rectangles represent exons. The red color intron indicates that it is annotated whereas pink color indicates unannotated intron.

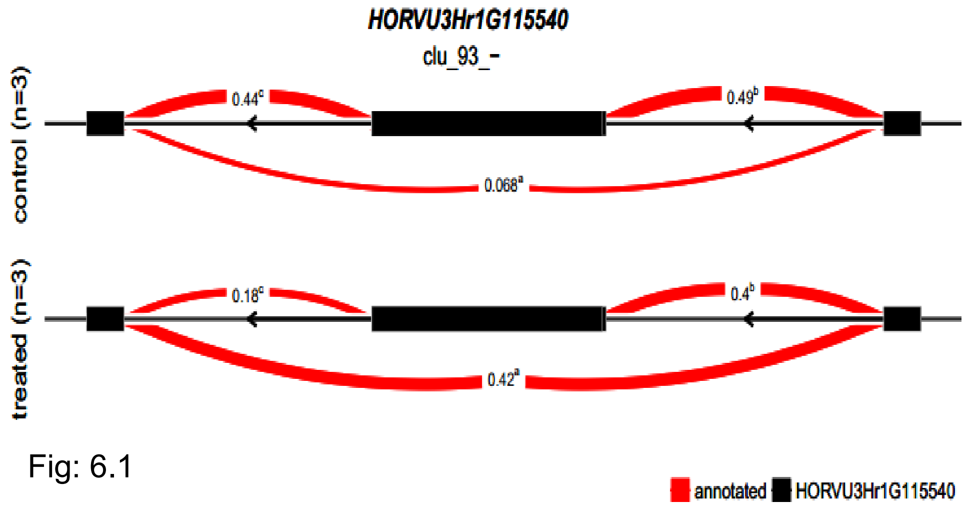

Fig: 6.1

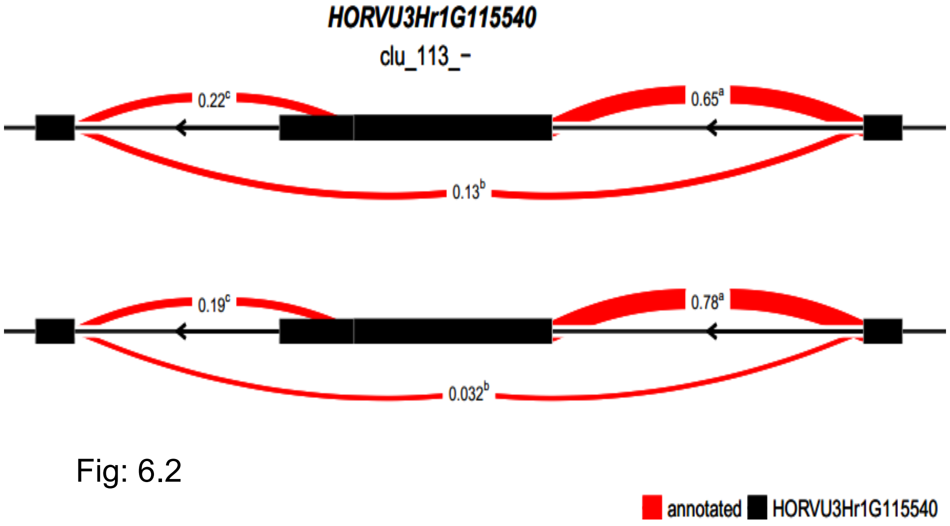

Fig: 6.2

Figure 6: Variation in AS in common gene HORVU3Hr1G115540. **6.1** Cd stress (root) **6.2** Heat stress (shoot)

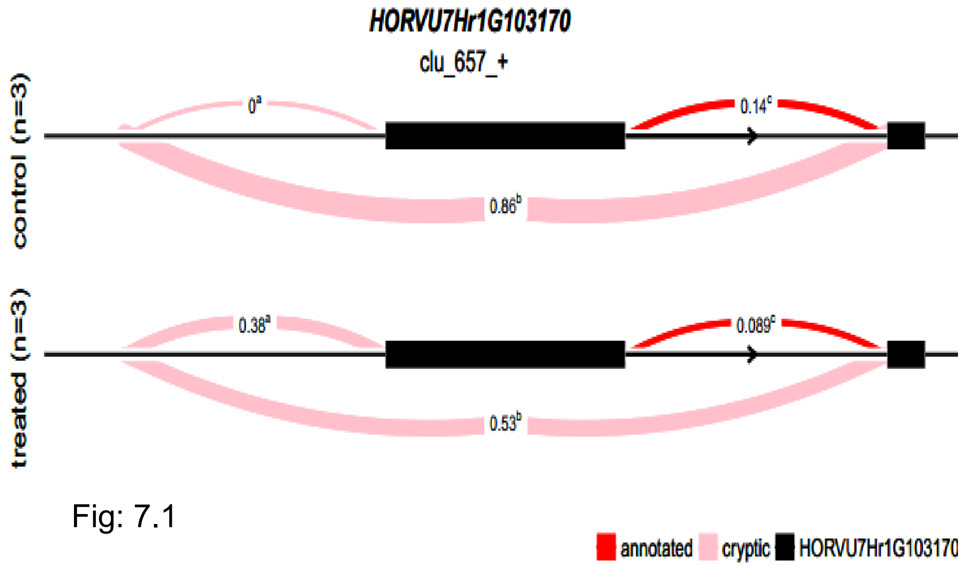

Fig: 7.1

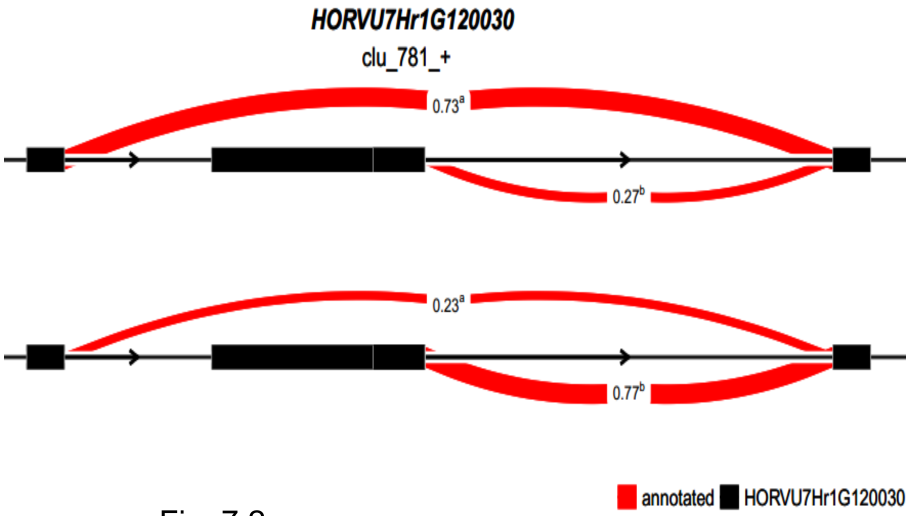

Fig: 7.2

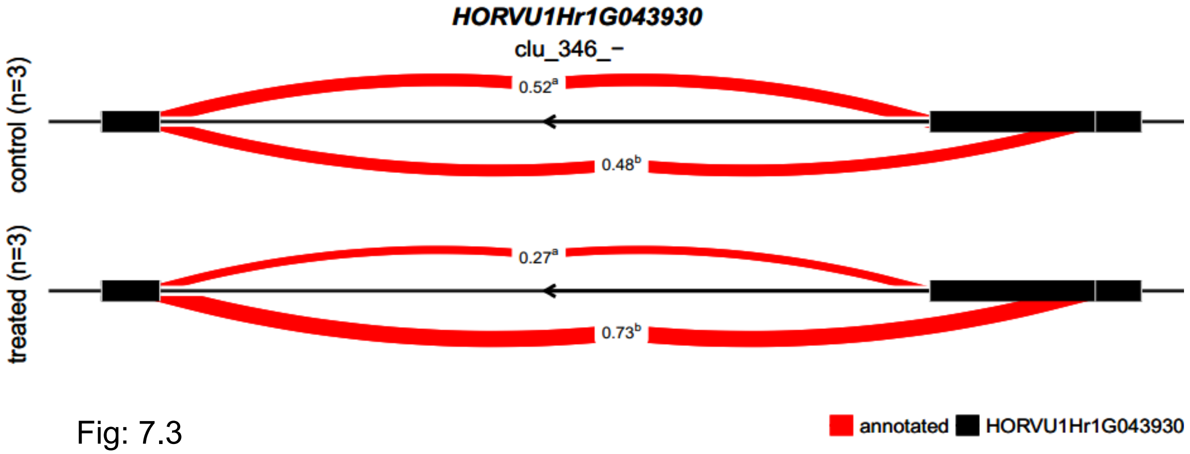

Fig: 7.3

Figure 7: Splicing event visualizations of genes encoding for ubiquitin. **7.1** Cd stress (shoot). **7.2** Cd stress (root). **7.3** heat stress (shoot)

CHAPTER 5

DISCUSSION

Alternative splicing emerges as a powerful mechanism for regulating plant stress responses. Plant responses to any stress is based on the interactions of various mechanisms involved in gene expression regulation at different stages of transcriptional and post translational processes (Mastrangelo et al., 2012). With the advent of sequencing technologies, transcriptomic sequences have been used for computational analysis and aimed to describe the extent of alternative splicing. The rate of alternative splicing were similar among different genotypes of same species, particularly dicotyledons species (Ner-Gaon et al., 2007). This could be due to very recent origins of cultivars. Further, lower differences were found in cereals crops like barley (20%) and maize. This might be due to selection pressure exerted by domestication and plant breeding technologies. In barley, rate of alternative splicing in expressed genes was found to be 26-31% (Ner-Gaon et al., 2007). In this study we identified 73 significant genes that are involved in various metabolic pathways like signalling, proteolysis, nucleotide excision repair pathways, flavonoid synthetic pathways, splicing pathways, respiration, yang cycle, calvin cycle, protein folding, glycolysis, circadian oscillator and stress related responses.

In accordance with this study alternative splicing events were found to be more prominent in roots under Cd stress. This might be possible because roots of plants are more prone to cadmium stress, which ultimately affects the uptake and translocation of minerals and nutrients. The more AS in the roots depicts the regulation occurring in response to stress. Cd is a non-essential toxic heavy metal, can cause oxidative stress and affect the uptake and translocation of minerals and nutrients (Zhang et al., 2012).

Even the splicing factors were found to be spliced alternatively under stress as HORVU2Hr1G019410 and HORVU4Hr1G029030 genes encode for Arginine/serine-rich splicing factor and Splicing factor U2af large subunit A respectively. Arginine/serine-rich splicing factor is an essential protein for splicing process. It binds to the exonic splicing enhancers (ESE) within the exons and regulates the alternative splicing. It ensure the efficiency and accuracy of the process and some of them are controlled by physiological signals. Apart from these proteins, U2AF splicing factors also play a key role during AS process. These are the non snRNPs known to recognize the branch point near the 3'splice site in spliceosome machinery. So both of the genes coding for splicing factors were found to be differentially spliced in response to the stress condition in barley.

In addition to Cd stress, heat stress also impairs photosynthetic efficiency, reduces the membrane stability and reduce plant water content. Heat stress disturbs homeostasis and retard plant growth and development. It also induces decrease in plant height and spike length and reduces tillering in barley (Abou-Elwafa and Amein, 2016). Plants have evolved various heat shock proteins to combat heat stress. We got 54 significant genes involved in alternative splicing events under heat stress condition in shoots of barley. We also found genes HORVU1Hr1G035720 and HORVU1Hr1G028030 underwent differential splicing which encode for enzymes involved in photosynthesis. Here, it might be possible that during photosynthesis, alternative splicing of these genes were responsible for heat stress responses. HORVU1Hr1G030790 encodes for Chaperone DnaK, homolog of Heat shock protein 70. Heat shock proteins are generally produced during stress conditions, they act as chaperones and prevent the aggregations of newly synthesized proteins and fold them properly during their translocation (Sung et al., 2001; Su and Li, 2008). It also forms the part of translocon, guidance complex import in organelles like chloroplast (Jackson-Constan et al., 2001; Soll, 2002). Differential alternative splicing in HORVU1Hr1G030790 encoding DnaK chaperone might play a regulatory role to provide heat stress tolerance in barley. HORVU2Hr1G074280 encodes Plastid-lipid Associated Protein, PAP fibrillin family protein, a stress tolerant protein. Plants have antioxidants to cope up with the heavy metal stress conditions (Demirevska et al., 2006). Heavy metal contamination of soils and crop plants are the issues of major concern due to their deleterious health effects on both humans and animals (Nan Z et al., 2002). HORVU7Hr1G085150 gene found to encode for Lactoylglutathione lyase, an enzyme act as antioxidant in which 'intron a' underwent more number of splicing events. It catalyzes the detoxification of methylglyoxal (MG). MG is highly toxic and produced during the uncontrolled influx of glycolytic substrates, lipid peroxidation and amino acid metabolism. It generates mutations in DNA base pairs and proteins. Thus its detoxification is very crucial and necessary (Chakraborty et al., 2015). It could be inferred that alternative splicing in HORVU7Hr1G085150 was involved in heat stress related responses.

Further, under cd stress and heat stress in root and shoot, HORVU7Hr1G103170, HORVU7Hr1G120030 and HORVU1Hr1G043930 genes encode for Polyubiquitin protein were alternatively spliced. Ubiquitin is involved in proteolysis and play a key role in stress response pathways (Flick et al., 2012). It allows the degradation of damaged proteins and eliminate their accumulation. We also found that these genes have introns with variable start and end points participating in splicing process. So, these genes might be important for survival under stress conditions. In addition to this, HORVU3Hr1G115540 was found to be common in both stress conditions and also underwent differential splicing process. It encodes for telomere binding proteins. Telomere are generally associated with replicative senescence and oxidative stress. It was found that telomere shortening rates were higher in cells with low antioxidants capacity whereas it is lower in cells with high antioxidants (Saretzki et al., 2002). In *Arabidosps* ten1 mutants (telomere-binding proteins) are highly sensitive to thermal stress (Lee et al., 2010). So, stress responsive genes undergo frequent alternative splicing events so as to cope up with the stress conditions and regulates the gene expression during stress. In addition to these stress related genes we found 63 clusters containing novel junctions associated with the differential alternative splicing. These novel clusters can be used for further studies to identify the novel transcripts and novel genes regulating alternative splicing mechanism.

CHAPTER 6

SUMMARY

NGS has provided a platform to produce huge amount of transcriptomics data. RNA seq is used quantitatively to examine splicing diversity using the splice junctions and clusters. Alternative splicing is a regulatory process that produces multiple products from a single gene. During stress, plants undergo differential alternative spliced events so as to tolerate or combat the stress conditions. In view of this, we had studied differential alternative splicing in barley under different abiotic conditions. We had found annotated and unannotated differentially significant spliced clusters contributing to form multiple gene products under stress conditions. We had identified different genes with significant clusters undergoing splicing events with variable intron usage and participating in stress related enzymes. Various genes were found to be involved in many metabolic pathways like photosynthesis, respiration, DNA repair, stress related responses indicating their role in stress responses. Different genes and clusters were found in all the different stress condition. Only one gene is found to be similar in heat stress and cadmium stress conditions in shoots and roots of barley respectively. This study depicts that under abiotic stress conditions, barley undergoes differential splicing process in order to enhance its survival possibilities.

The future perspective of this study could include identification of various types of alternative splicing along with the PCR mediated studies to confirm and validate the splicing events. It might involve the study of genes regulating the growth and development of barley during stress conditions. Also effects of alternative splicing in regulatory pathways could be studied.

CHAPTER 7

REFERENCES

Abou-Elwafa, S. F., and Amein, K. A. (2016). Genetic diversity and potential high temperature tolerance in barley (*Hordeum vulgare*). *World J. Agric. Res.* 4, 1–8

Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNAseq data. Genome Res 22(10):2008–2017.

Andrews, S., & FastQC, A. (2015). A quality control tool for high throughput sequence data. 2010. *Google Scholar*.

Su, P. H., & Li, H. M. (2008). Arabidopsis stromal 70-kD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds. *Plant Physiology*, *146*(3), 1231-1241.

Barbazuk, W. B., Fu, Y., & McGinnis, K. M. (2008). Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome research*, *18*(9), 1381-1392.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 59 terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci 74: 3171–3175.

Bradley, R. K., Merkin, J., Lambert, N. J., & Burge, C. B. (2012). Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS biology*, *10*(1), e1001229.

Chakraborty, S., Gogoi, M., & Chakravortty, D. (2015). Lactoylglutathione lyase, a critical enzyme in methylglyoxal detoxification, contributes to survival of Salmonella in the nutrient rich environment. *Virulence*, *6*(1), 50-65.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 59 ends of adenovirus 2 messenger RNA. Cell 12: 1–8. Colwill K, Feng LL, Yeakley JM, Gish GD, Caceres JF, Pawson T, Fu XD

D. Jackson-Constan, M. Akita, K. KeegstraMolecular chaperones involved in chloroplast protein import Biochim. Biophys. Acta, 1541 (2001), pp. 102-113

Sung, D. Y., Kaplan, F., & Guy, C. L. (2001). Plant Hsp70 molecular chaperones: protein structure, gene family, expression and function. *Physiologia plantarum*, *113*(4), 443-451.

Demirevska-Kepova, K., Simova-Stoilova, L., Stoyanova, Z. P., & Feller, U. (2006). Cadmium stress in barley: growth, leaf pigment, and protein composition and detoxification of reactive oxygen species. *Journal of plant nutrition*, *29*(3), 451-468.

Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Current protocols in bioinformatics*, 11-14.).

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21.

Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., ... & Zhu, Z. (2015). SKIP confers osmotic tolerance during salt stress by controlling alternative gene splicing in Arabidopsis. *Molecular plant*, *8*(7), 1038-1052.

Flick, K., & Kaiser, P. (2009). Proteomic revelation: SUMO changes partners when the heat is on. *Sci. Signal.*, *2*(81), pe45-pe45.

Flick, K., & Kaiser, P. (2012, July). Protein degradation and the stress response. In *Seminars in cell & developmental biology* (Vol. 23, No. 5, pp. 515-522). Academic Press.

H. Ner-Gaon, N. Leviatan, E. Rubin, R. Fluhr, Comparative cross-species alternative splicing in plants, Plant Physiol. 144 (2007) 1632–1641).

Hu Y, et al. (2013) DiffSplice: The genome-wide detection of differential splicing events with RNA-seq. Nucleic Acids Res 41(2):e39.)

Jiang, J., Liu, X., Liu, G., Liu, C., Li, S., & Wang, L. (2017). Integrating omics and alternative splicing reveals insights into grape response to high temperature. *Plant physiology*, pp-01305.

Kocheva K., Lambrev P., Georgiev G., Goltsev V., Karabaliev M., Evaluation of chlorophyll fluorescence and membrane injury in the leaves of barley cultivars under osmotic stress, Bioelectrochemistry, 2004, 63, 121-124)

Lachman, J., Kotíková, Z., Zámečníková, B., Miholová, D., Száková, J., & Vodičková, H. (2015). Effect of cadmium stress on barley tissue damage and essential metal transport into plant. *Open Life Sciences*, *10*(1).

Laloum, T., Martín, G., & Duque, P. (2017). Alternative Splicing Control of Abiotic Stress Responses. *Trends in plant science*.)

Lee, J. R., Xie, X., Yang, K., Zhang, J., Lee, S. Y., & Shippen, D. E. (2016). Dynamic interactions of Arabidopsis TEN1: stabilizing telomeres in response to heat stress. *The Plant Cell*, *28*(9), 2212-2224.

Lee, Y., & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annual review of biochemistry*, *84*, 291-323.

Li, Y. I., Knowles, D. A., Humphrey, J., Barbeira, A. N., Dickinson, S. P., Im, H. K., & Pritchard, J. K. (2017). LeafCutter: Annotation-free quantification of RNA splicing. *bioRxiv*, 044107.

Marquez, Y. et al. (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res. 22, 1184–1195).

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, *12*(10), 671.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T. & Bayer, M. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature*, *544*(7651), 427.).

Mastrangelo, A. M., Marone, D., Laidò, G., De Leonardis, A. M., & De Vita, P. (2012). Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Science*, *185*, 40-49.

Matsukura S, Mizoi J, Yoshida T, Todaka D, Ito Y, Maruyama K, et al. (2010) Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stressresponsive genes. Molecular Genetics and Genomics 283: 185–196. doi: 10.1007/s00438-009-0506-y PMID: 20049613.

Nan Z., Li J., Zhang J., Cheng G., Cadmium and zinc interactions and their transfer in soil-crop system under actual field conditions, Sci. Total Environ., 2002, 285, 187- 195)

Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., & Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *The Plant Journal*, *39*(6), 877-885.).

Ner-Gaon, H., Leviatan, N., Rubin, E., & Fluhr, R. (2007). Comparative cross-species alternative splicing in plants. *Plant physiology*, *144*(3), 1632-1641.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, *40*(12), 1413.

Panahi, B., Mohammadi, S. A., Khaksefidi, R. E., Fallah Mehrabadi, J., & Ebrahimie, E. (2015). Genome-wide analysis of alternative splicing events in Hordeum vulgare: Highlighting retention of intron-based splicing and its possible function through network analysis. *FEBS letters*, *589*(23), 3564-3575.

Patel, R. K., & Jain, M. (2015). NGS QC toolkit: A platform for quality control of next-generation sequencing data. In *Encyclopedia of Metagenomics* (pp. 544-548). Springer US.

Protein import into chloroplasts Curr. Opin. Plant Biol., 5 (2002), pp. 529-535

Reddy, A. S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.*, *58*, 267-294.)

Reddy, A. S., Marquez, Y., Kalyna, M., & Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *The Plant Cell*, *25*(10), 3657-3683.

Rüegsegger, U., Leber, J. H., & Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. *Cell*, *107*(1), 103-114.

Saretzki, G., & Zglinicki, T. (2002). Replicative aging, telomeres, and oxidative stress. *Annals of the New York Academy of Sciences*, *959*(1), 24-29.

Schulte, D., Close, T. J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G., & Stein, N. (2009). The international barley sequencing consortium—at the threshold of efficient access to the barley genome. *Plant physiology*, *149*(1), 142-147.

Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., & Brown, J. W. (2012). Alternative splicing in plants–coming of age. *Trends in plant science*, *17*(10), 616-623.

Tanabe, N., Yoshimura, K., Kimura, A., Yabuta, Y., & Shigeoka, S. (2007). Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress. *Plant and Cell Physiology*, *48*(7), 1036-1049.

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology*, *31*(1), 46.

Wang Z., Gerstein M., & Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 10(*1*), 57-63. doi:10.1038/nrg2484.

Wang, X. et al. (2012) SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in Arabidopsis. Plant Cell 24, 3278–3295

Xue, G. P., & Loveridge, C. W. (2004). HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. *The Plant Journal*, *37*(3), 326-339.).

Zhang L., Chen Z., Zhu C., Endogenous nitric oxide mediates alleviation of cadmium toxicity induced by calcium in rice seedlings, J. Environ. Sci., 2012, 24, 940-948).

Zhang, Q., Zhang, X., Wang, S., Tan, C., Zhou, G., & Li, C. (2016). Involvement of alternative splicing in barley seed germination. *PloS one*, *11*(3), e0152824.

Zhang, W. et al. (2014) Splicing factor SR34b mutation reduces cadmium tolerance in Arabidopsis by regulating iron-regulated transporter 1 gene. Biochem. Biophys. Res. Commun. 455, 312–317

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell*, *167*(2), 313-324.