Analysis of Alternative Splicing Events in The Transcriptome of Potato Plants

by

Atinuke E Ogungbayi

Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of

Science in the

Biological Sciences

Program

YOUNGSTOWN STATE UNIVERSITY

May 2022

Signature

Analysis of Alternative Splicing Events in The Transcriptome of Potato Plants

Atinuke E Ogungbayi

I hereby release this thesis to the public. I understand that this thesis will be made available from the Ohio LINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

_____

Atinuke E Ogungbayi, Graduate Student Date

Approvals:

_____

Dr. Xiangjia Min, Thesis Advisor Date

_____

Dr. David Asch, Committee Member Date

_____

Dr. Jonathan Caguiat, Committee Member Date

_____

Dr. Salvatore A. Sanders, Dean of Graduate Studies

ABSTRACT

Potato (Solanum tuberosum L.) is the third most important food crop in the world following rice and wheat. Due to its high calories, it is a good source of energy and nutrients such as proteins, vitamins and minerals. Thus, potato plays an important role in providing food, nutritional and economic security of the world. To further improve the quantity and quality of potato as a food crop, it is imperative to understand the transcriptome diversity, gene expression dynamics and associated developing methods. The aim of this study is to identify the alternative spliced events of genes and their differential expression in nitrogen-treated and drought-stressed potato plants. In nitrogen grown potato plants, alternative acceptor site was observed to be most dominant alternative spliced events while intron retention was dominant in drought-stressed potato plants. Additionally, differential gene and isoform expressions of four different varieties of drought-stressed leaf samples of potato plants were analyzed by cuffdiff module of cufflinks V2.2.1 package. Statistically significant ($P \leq 0.05$) differential gene expressions with corresponding transcript isoforms were identified such as 10 (up-regulated) and 13 (down-regulated) in Algeria cultivars, 24 (up-regulated) and 21 (down-regulated) in Desiree cultivars with 39 transcript isoforms generated, 33 (up-regulated) and 51 (down-regulated) with 62 generated transcript isoforms in Saturna cultivars, 43 (up-regulated) , 32 (down-regulated) and 55 transcript isoforms generated in Milva cultivars.

Overall, conserved alternative spliced genes and alternative spliced events were identified in both nitrogen-treated and drought-stressed potato plants. Furthermore, among the differentially expressed genes, PYR1-like (pyrabactin resistance 1) and heat shock protein families were most upregulated genes. These genes play a crucial role in enhancing development in potato plants under extreme conditions of drought.

ACKNOWLEDGEMENT

I would like to acknowledge and give my warmest thanks to my supervisor, Dr Xiangjia Min, for the continuous support of my Masters study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Master's degree.

I would also like to thank my thesis committee: Dr. Dave Asch, and Dr. Jonathan Caguiat, for their encouragement, insightful comments, and brilliant suggestions.

I would also like to give special thanks to my dearest husband, Babatunde Ibrahim Ojoawo for his continuous support and understanding when undertaking my research and writing my thesis. Your prayer for me was what sustained me this far.

Finally, I would like to thank God, for letting me through all the difficulties. I have experienced your guidance day by day. You are the one who let me finish my degree. I will keep on trusting you for my future.

Table of Content

Contents

List of Figures

List of Tables

CHAPTER I: INTRODUCTION

1.1 Background

Despite the challenges of ensuring food availability for a growing world population, the potato has remained one of the most important nongrain crops produced globally, with over 388 million tons in 2018 (http://www.fao.org; Lemke et al., 2020; Devaux et al., 2021). Potato (Solanum tuberosum L.) belongs to the Solanaceae family whose members contribute to economic development due to its important roles of high yield in food, nutritional composition, and a good source of bioenergy to combat food insecurity and provide environmentally friendly energy (Burlingame et al., 2009; Tiwari et al.,2020). However, the cultivation of potato plants for this economic advantage has been sabotaged by abiotic stresses such as extremes of temperature, high salinity, and drought. Thereby affecting potato growth, development, survival, and secondary metabolism (Pennisi, 2008; Sprenger et al., 2016; QingLi et al., 2020). However, various agricultural practices have been explored to enhance crop productivity, such as excessive use of nitrogen fertilizer to increase potato tuber yield (Tiwari et al.,2021). According to several researchers, approximately 50% of the nitrogen in applied fertilizer is utilized by the potato crop, whereas excessive use of fertilizer negatively impacts the environment and causes a decline in soil and water quality (Qu et al.,2020). In an attempt to adapt to these abiotic stresses, plants generally, have evolved several strategies one of such is alternative splicing of primary transcripts which contributes to protein diversity and regulates gene expression in plants.

1.2 Definition Of Alternative Splicing

Alternative Splicing (AS) is a process of generating more than one Messenger Ribonucleic Acid (mRNA) transcript from a precursor RNA (pre-mRNA) that increases the diversity of functional proteins. mRNA molecules serve as a template for protein synthesis in all species. Identification

and analysis of Alternative Splicing (AS) events are crucial for crop improvement and understanding regulatory mechanisms (Min, 2017).

Identifying and characterizing AS enables our understanding of the biological role of transcript isoform diversity (Mei et al., 2017). The presence of segments of non-coding DNA, called introns, interspaced with coding DNA segments, the exons, is a characteristic of eukaryotic protein-coding genes (Melo et al.,2020). The processing of the precursor mRNA leads to the identification of specific splice sites, the introns are removed, and the exons are joined together (or spliced). However, splice sites can be differentially recognized, leading to the inclusion or removal of different RNA segments, resulting in multiple transcripts and potentially proteins originating from the same gene. This mechanism of alternative splicing also represents an effective means of increasing transcriptome and proteome diversity and regulating gene expression by affecting the stability of the transcripts (Shang et al., 2017).

1.3 Alternative Splicing Events and Their Significance

Messenger RNA transcript isoforms are generally generated through four primary events in AS: intron retention (IR) in the mature mRNA; exon skipping (ES) resulting from alternative exon usage (AEU); alternative donor site (AltD), and alternative acceptor site (Alta) that are resulted from the use of cryptic splice sites that may elongate or shorten an exon (Reddy et al., 2013; G. Sablok et al., 2017; Staiger and Brown, 2013).

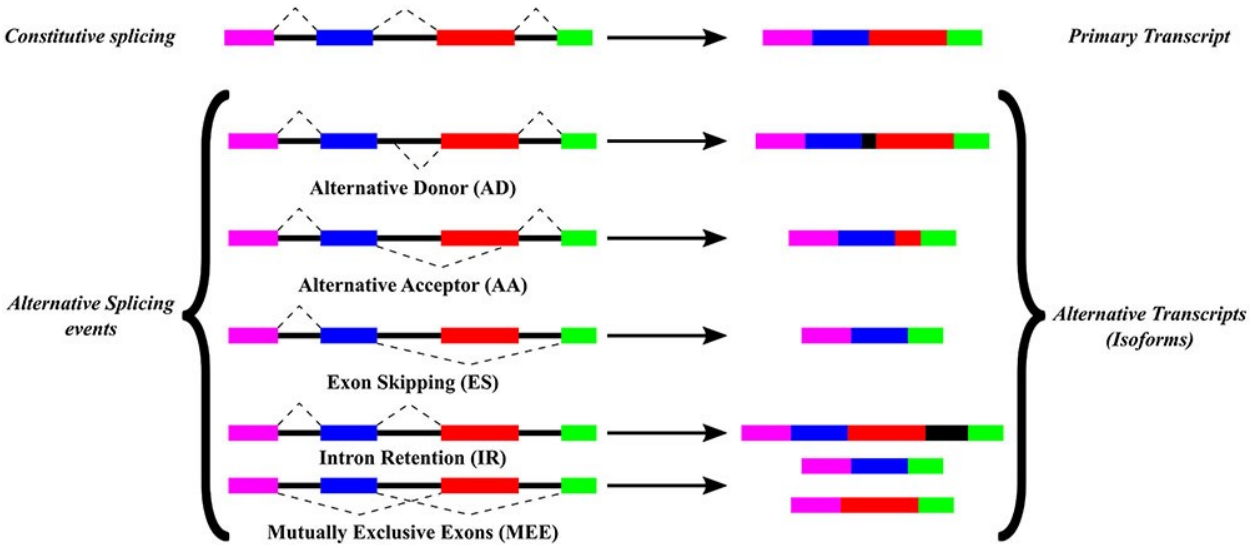


Figure 1 Formation of different types of AS events from primary transcripts. Íñiguez LP et al., 2017

Various complex types can be formed by combining basic events (Sablok et al., 2011). While these basic types can be found in all kingdoms of eukaryotes, ES is the most prevalent event in animals, including humans, and IR is the dominant event in plants (McGuire et al., 2008). A multiexon gene can produce a single mature mRNA by constitutive splicing (CS), where only one set of splice sites is used (Reddy et al., 2013).

Patterns of alternatively spliced transcripts have been widely observed, with reports suggesting that approximately 60–75% of AS events occur within the protein coding regions of mRNAs, resulting in changes in binding properties, intracellular localization, protein stability, enzymatic, and signaling activities (Stamm et al.,2005). In plants, IR is

the most dominant form, with reports suggesting the proportions of intron containing genes undergoing AS in plants ranged from ~30% to >60%, depending on the depth of available

transcriptome data (Reddy et al., 2013; Sablok et al., 2017). For example, as many as 60% of multi-exon genes undergoing AS in the model plant Arabidopsis thaliana (Reddy et al.,2013; Carvalho et al., 2013; Sablok et al., 2017). Over time, this number has been increasing as genome annotations improve and next-generation sequencing provides more and deeper data (Syed et al., 2012).

CHAPTER II

2.1 STATEMENT OF OBJECTIVE

1. To identify AS genes and conserved AS genes in drought-stressed potato plants.

2. This will further lead to analysis of alternatively spliced (AS) events in these AS genes in drought-stressed and also nitrogen treated potato plants.

3. To identify and analyze significant and most upregulated differentially expressed genes in drought-stressed potato plants.

2.2 Hypothesis

The treatment of potato plants with nitrogen or subjecting potato plants to the condition of drought should trigger a series of adaptive responses, which could serve as an escape, avoidance, and tolerance mechanism. This mechanism may result from differential gene expression; therefore, we expect to identify novel mRNA transcripts responsible for this gene expression through analyzing AS events.

Comprehensive transcriptome analysis in potato plants by mapping the genome information of EST and mRNA with RNA-sequence reads from samples of potato plants should lead to the identification of AS genes and AS events from genomic loci generating several different mRNA transcripts. Like several other plants, intron retention is expected to be the most prevalent type of AS event at a higher estimated rate which further contributes to diversity in mRNA transcripts and proteins in potato plan

CHAPTER III: LITERATURE REVIEW

Several studies have revealed the gene expression profiles of salt-stressed potatoes (QingLi et al., 2020), and prior to these studies, many authors have reported mechanisms underlying salt- and drought- stress tolerance in plants (Zhu et al., 2016; Sprenger et al., 2015) and the pivotal role of genes involved in salt- and drought-stressed responses (Sun et al., 2015). Moreover, to curb excessive nitrogen use in potato cultivation, (Tiwari et al., 2021) identified candidate genes involved in enhancing nitrogen use efficiency in potatoes and a regulatory element connected with low and high nitrogen. Before this, several researchers have made advancements to improve the effective use of nitrogen in potato plants. The use of transgenic innovations through overexpression or manipulation of genes that encodes nitrogen metabolism pathway results in genetically modified organism (Beatty et al.,2016; Tiwari et al., 2020b). In addition, a breakthrough CRISPR/CAS9 innovation has sped up genomic-enhanced crop yield (Li et al., 2020). All these investigations were explored through transcriptomic analysis using RNA sequencing technologies.

In recent years, several well-planned pieces of research employed transcriptome analysis to achieve many outcomes in the Solanaceae family of the potato plant. One of such is amplifying Alternative Splicing (AS) identification and comprehensively classifying AS genes and AS events (Clark et al., 2019). This results in identifying conserved genes and differentially expressed in different tissues with comparative studies (Zhang et al., 2017) critical for crop quality and quantity yield, including understanding regulatory mechanisms (Min et al., 2015; Min et al., 2017). Therefore, this method has facilitated potato crop improvement in quantity and quality which is underlined by the dynamics of transcriptome diversity and gene expression (Clark et al., 2019).

An increase in the diversity of protein and mRNA transcripts (transcriptome) in eukaryotes is facilitated by AS, resulting in the modulation of protein function (Min et al., 2015) and regulating gene expression by affecting the stability of the transcripts (Shang et al., 2017). The variation in the number of transcripts in the cell is influenced by AS, which is fundamental to biotic and abiotic stress responses in plants and, thus, gene expression (Shang et al., 2017 & Laloum et al., 2018). According to researchers, several of these genes were expressed differently under certain conditions of stress (Galvez et al., 2016 & Tiwari et al., 2020), indicating that these genes could be manipulated to play a regulatory role in response to a specific condition of stress (Li et al., 2020).

While investigating gene expression level of potato plants under a different type of abiotic stress, Tiwari et al. (2020), in his experiment, discovered that a group of stress-responsive genes under low Nitrogen concentration were upregulated in the leaf and root tissues of potato plants while another group was subsequently down-regulated in the leaf under the same condition (Wanniarachch et al.,2018). In addition, the gene expression profile of salt-stressed potato was examined (by QingLi et al., 2020); there were more upregulated genes than down-regulated ones. Following the experiment Heike Sprenger and his cohorts carried out in 2016 on drought-responsive transcripts of the potato reference cultivars, more drought-responsive genes were more upregulated than down-regulated.

Thus, the study of potatoes by these researchers under different abiotic stress conditions indicates the implication of stress-responsive genes in response to low nitrogen, salinity, and drought at the early growth stage in various tissues of potato plants. (Tiwari et al.,2020; Abu-Romman and Shatnawi, 2011; QingLi et al., 2020 & Sprenger et al., 2015). Furthermore, Inherent in the ability to regulate biotic and abiotic responses is Alternative Splicing (AS) mechanism producing protein

diversity and also acts to modulate multiple gene expressions (Staiger et al., 2013; Chaudhary et al., 2019).

From the literature, it is clear that AS increases protein diversity; however, the degree to which AS is improving this diversity requires further clarification (Tress et al., 2017). Therefore, more understanding of AS in conjunction with protein diversity and gene expression in different tissues in potato plants needs to be studied through comprehensive AS analyses.

The thesis may provide a basic resource for further research into differentially expressed genes and alternatively spliced transcripts in nitrogen treated and drought-stressed potato plants, which may increase our understanding of how potato plants cope with different stress conditions, thereby developing a sustainable agricultural production.

CHAPTER IV: MATERIALS AND METHODS

4.1 Sample Description

The reads for the RNA sequencing data were obtained from two different published papers (H. Sprenger et al., 2016; Tiwari et al., 2020). The overall design of the first paper is based on a transcriptome profiling of 48 leaf samples from 4 potato cultivars grown under control or drought stress conditions in 6 independent experiments by RNA sequencing method (H.Sprenger et al.,2016). The study involves metabolic and transcriptomic responses of leaves from European potato reference (Solanum tuberosum) cultivars with differential tolerance to long-term drought. The experiment was conducted based on independently repeated field and greenhouse trials using four potato cultivars- Alegria, Milva, Desiree, and Saturna. The classifications of each experiment were in triplicate and designated F1, F3, & F4, and G1, G2, & G3 for field and greenhouse, respectively. The control plants were optimally irrigated throughout the experiment, whereas the drought-stressed was induced by withholding water. These are the two conditions in which each class of the experiment was carried out. The drought-stressed potato plants (treatment) were compared with sufficiently irrigated control grown together.

The second study involves transcriptome sequencing of potato cv. Kufri Gaurav for nitrogen use efficiency at the early stage (Tiwari et al., 2020). The Indian potato cultivar was grown in an aeroponic condition with low(0.75mM) and high Nitrogen (7.5mM) concentrations designated as treatment and control, respectively. The leaf and root samples were harvested at an early stage of growth, whereas the stolon was harvested when tuberization was seen.

4.2 Data Description

The raw sequencing read data was downloaded as Short Reads Archive, SRA format from National Centre for Biotechnology Institute (NCBI) database. The two data groups from the published papers are available under the Bioproject ID # PRJNA311702 and PRJN565618, respectively.

The RNA-seq data drought-stressed potato libraries were sequenced using Illumina HIseq 2000, producing 49 base pairs with single end reads. While in nitrogen treated potato, the libraries were sequenced using ion_torrent proton producing single end reads with 200 base pairs.

Potato genome sequence, predicted transcripts and annotation GFF3 files (version 4.03) were downloaded from the Phytozome database (https://phytozome-next.jgi.doe.gov/).

4.3 RNA Sequence Data Mapping to the Genome

With the use of the SRA toolkit, a total of 48 and 6 samples of RNA sequence (RNA-seq) data of drought-stressed and nitrogen-treated Potato "Solanum tuberosum" cultivars were downloaded from NCBI short read archives (SRA) database. The RNA-seq data were retrieved from 2 published papers (Tiwari et al., 2020; Heike Springer et al., 2016). The data from each published paper were solely analyzed.

The RNA-seq reads were mapped to the annotated potato reference genome (Potato Genome Sequencing Consortium, 2011) using TopHat2 with the reads in FastQ format. The resulting transcript alignment file with Stuberosum_448_v4.03. gene. gff3 annotation were used as an input in Cufflinks module to assemble the transcripts (Trapnell et al., 2010). The output files contain gene and isoform information, and the "transcript" contains the mapping information.

This resultant transcript gene transfer file (GTF) was used to classify the AS events using AStalavista (http://astalavista.sammeth.net/) (Foissac et al.,2007) The output file, a landscape.gtf was further processed using different Perl programs to estimate each type of events.

The transcript gtf file generated after cufflinks from each RNA seq datasets was further merged using Cuffmerg script within the Cufflink package (Trapnell et al., 2010) the GTF file "merge.gtf" generated from Cuffmerg script were analyzed to estimate the number of each event and to detect AS genes using 'gtf2event.pl' and 'gtf2gene2transcript. Pl' Perl script program.

4.4 Analysis of Differentially Expressed Genes in Each Drought Stressed Potato Cultivars

The BAM file 'accepted_hit.bam' produced by cufflinks program was combined Stuberosum_448_v4.03. gene. gff3 annotation using cuffdiff module. The cuffdiff analysis provides the estimated expression level of genes and transcript isoforms in drought-stressed potato plants. These expression levels were estimated as fragments per kilobase of exon per million mapped reads. The parameters of differential expression of genes and isoform transcripts in each cultivar of drought-stressed potato plants were $P \leq 0.05$, $Q \leq 0.05$, and log2 fold change. These parameters were used to estimate the number of significantly expressed genes and isoform transcripts and identify the number of upregulated and down-regulated genes in each cultivar.

The Venn diagram software (Venny2.1 tool) detected and identified the common and unique differentially expressed genes and isoforms. The annotation file that contains information on proteins encoded by these genes was downloaded, and the proteins were identified.

4.5 Summary of the Tools Employed to Identify and Analyze Alternative Splicing Events in Potato Plants

- The SRA (short read archive) toolkit downloads the raw reads and converts the SRA format to FASTQ format.

- Tophat2 (v2.2.6) combines short read with genome index and were mapped to the reference genome to produce an aligned transcript.

- Cufflinks module within cufflink v2.2.1 package assembles the resulting aligned transcript isoform along with an annotation file to produce an assembled file. These files are “transcript.gtf”, “skipped.gtf”, genes. fpkm_tracking, isoform.fpkm_tracking.BAM file.

- Cuffmerge module within cufflink v2.2.1 package merges groups of assembled transcripts into a single consensus assembly.

- Cuffdiff module also within cufflink v2.2.1 package analyses the differential expression of genes and transcripts using the assembled BAM file “accepted_hits.bam”. Thus, aiding in investigating their transcriptional and post-transcriptional regulation under different conditions.

- AStalavista classifies the assembled transcripts “transcript.gtf” in the form of gene transfer format(gtf). The GTF file can be used as input to the AStalavista server to analyze AS events. The resulting “Landscape. Gtf” contains the classified AS events quantitatively summarized. Astalavista server (http://astalavista.sammeth.net/)

- Perl Script Programs: different perl script programs were written to analyze the “landscape.gtf “quantitatively with different AS events types. The “gtfASevents.pl” was used to differentiate each type of event, while “countASevent.pl was written to count each of the event types.

- Venny 2.1 (Chen et al.,2018) carries out Venn diagram analysis to identify common and unique differential gene expressions.

CHAPTER V: RESULTS AND DISCUSSIONS

5.1 Mapping Downloaded RNA-Sequence Datasets

5.11 Nitrogen- treated Potato Plants

A total input of 25-37 million reads were generated from each of the 6 samples of nitrogen-grown potato plants. These reads were mapped between 37.40%- 56.60% with the annotated reference potato genome Solanum tuberosum (PGSC DM V3.4) (Potato Genome Sequencing Consortium, 2011) The range of multiple aligned reads of the total mapped reads varied between 13.20% - 18.70% as shown in Table 1 below.

Table 1. The RNA-Sequence Mapping Summary of Different Tissues of Potato Plants Grown in Aeroponic Nitrogen (N) Conditions. The percentage of multiple alignment is based on the total mapped reads.

ACCESSION	N.TREATMENT	INPUT READS	MAPPED READS(%)	M. ALIGNMENT(%)
SRR10135769	Low N.Stolon	32,473,725	56.6	15.3
SRR10135770	High N.Stolon	37,217,935	51.3	14.7
SRR10135771	Low N.Root	26,246,911	37.4	13.2
SRR10135772	High N.Root	25,364,583	41.7	13.4
SRR10135773	Low N.Leaf	29,052,428	51.9	18.7
SRR10135774	High N.Leaf	28,412,698	39.7	18.2

Data in Table 1 was visualized in Figure 1 below. It can be seen that the sample with low nitrogen at the stolon (SRR10135769) has the highest mapping rate of 56.60%, hence, best aligned transcripts while the sample with low nitrogen at the root tissue has the lowest aligned transcript at mapping rate of 37.40%. Multiple alignment or multi-mapping is a result of reads mapping to more than one gene locus and a low percentage of multiple alignment results in increased accuracy in gene/transcript analysis. Therefore, all the reads from each of the nitrogen- grown samples have a very low percentage of multiple alignment, approximately below 20%, meaning a good percentage of the reads were accurately mapped to one gene location on the reference genome as

seen in the input reads in Table 1 above. It should be noted that variation in mapping percentage is caused by sample preparation and sequencing procedure, not an effect of treatments.

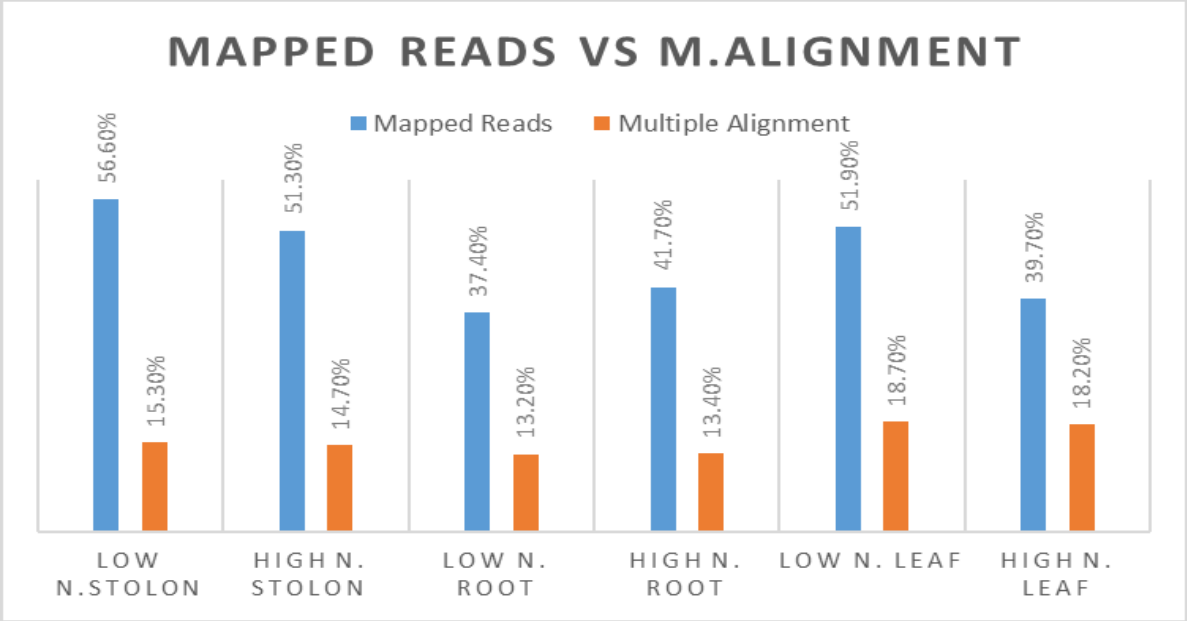


Figure 2 Visual representation of percent mapped reads and reads mapped to more than one locus (multiple alignments)

5.12 Drought-treated Potato Plants

A total input of 10.32 – 22.22 million reads were generated from each of the 48 samples of control and drought stressed potato cultivars. The mapping of reads to potato reference genome (Potato Genome Consortium, 2011) is approximately between 84%-92%, while the percentage multiple alignment was in a range of 5.6% -8% as shown in Table 2. The relatively much higher percentage of mapping of RNA-seq reads in this experiment indicate the quality of data much better than the data obtained from nitrogen treatment experiments

ACCESSION	TREATMENT	INPUT READS	MAPPED READS(%)	M. ALIGNMENT(%)
SRR3161990	Leaf Control	12,019,653	89.6	6.8
SRR3161991	Leaf Control	10,317,924	88.6	7.4
SRR3161992	Leaf Control	11,403,051	91.1	7.3
SRR3161993	Leaf Control	11,930,144	91.4	6.7
SRR3161994	Leaf Treatment	11,694,058	90.7	6.9
SRR3161995	Leaf Treatment	11,742,842	90.3	7.1
SRR3161996	Leaf Treatment	11,706,105	90.5	7.3
SRR3161997	Leaf Treatment	12,529,686	91.4	7
SRR3161998	Leaf Control	12,197,816	90.8	6
SRR3161999	Leaf Control	11,835,090	91	6.8
SRR3162000	Leaf Control	11,258,480	90.8	6.8
SRR3162001	Leaf Control	11,930,898	90.9	5.9
SRR3162002	Leaf Treatment	12,392,609	89.8	5.8
SRR3162003	Leaf Treatment	12,066,501	90.6	6.1
SRR3162004	Leaf Treatment	11,787,042	91.1	6.4
SRR3162005	Leaf Treatment	12,471,557	90.8	6
SRR3162006	Leaf Control	12,341,619	91.1	6.2
SRR3162007	Leaf Control	12,412,385	90.9	6.6
SRR3162008	Leaf Control	12,596,439	90.7	7.1
SRR3162009	Leaf Control	11,783,906	91.6	6.2
SRR3162010	Leaf Treatment	11,742,409	90.9	6.2
SRR3162011	Leaf Treatment	11,869,495	91.5	6.5
SRR3162012	Leaf Treatment	12,592,720	90.8	6.2
SRR3162013	Leaf Treatment	12,288,675	91.3	6
SRR3162014	Leaf Control	11,930,331	90.5	6.2
SRR3162015	Leaf Control	12.319733	91.1	6.9
SRR3162016	Leaf Control	12,403,548	90.9	7.2
SRR3162017	Leaf Control	12,063,024	91.6	6.6
SRR3162018	Leaf Treatment	12,177,348	91.1	6.2
SRR3162019	Leaf Treatment	11,823,475	91.3	7
SRR3162020	Leaf Treatment	12,442,088	91.6	6.6
SRR3162021	Leaf Treatment	12,301,068	92	6.5
SRR3162022	Leaf Control	21,380,309	83.6	6.4
SRR3162023	Leaf Control	18,275,400	91.9	7.1
SRR3162024	Leaf Control	12,403,548	90.9	7.2
SRR3162025	Leaf Control	20,352,632	91.8	6.6
SRR3162026	Leaf Treatment	22,215,641	88.2	5.6
SRR3162027	Leaf Treatment	19,163,175	91.4	6.2
SRR3162028	Leaf Treatment	18,498,595	91.1	6.3
SRR3162029	Leaf Treatment	20,525,443	90.9	6.9
SRR3162030	Leaf Control	19,969,325	88.6	6.7
SRR3162031	Leaf Control	20,086,368	91	7.7
SRR3162032	Leaf Control	17,428,812	90.7	7.4
SRR3162033	Leaf Control	19,893,587	91.5	7.2
SRR3162034	Leaf Treatment	20,008,844	86.5	6.7
SRR3162035	Leaf Treatment	19,487,715	90.3	7.2
SRR3162036	Leaf Treatment	17,537,040	90.6	7.6
SRR3162037	Leaf Treatment	15,995,317	91.7	6.6

Table 2. The RNA-Sequence Mapping Summary of Drought-Treated and Control of Potato Leaf samples.

The mapping data is also shown in Figure 2 below. Overall, the drought-treated potato has a considerable higher mapping rate, this implies majority of the reads being aligned to the reference genome at a rate between ~84% – 92%. Therefore, a low percentage of multiple alignment of reads improves accurate quantification of genes with little or no ambiguity (Deschamps-Francoeur et al,.2020).

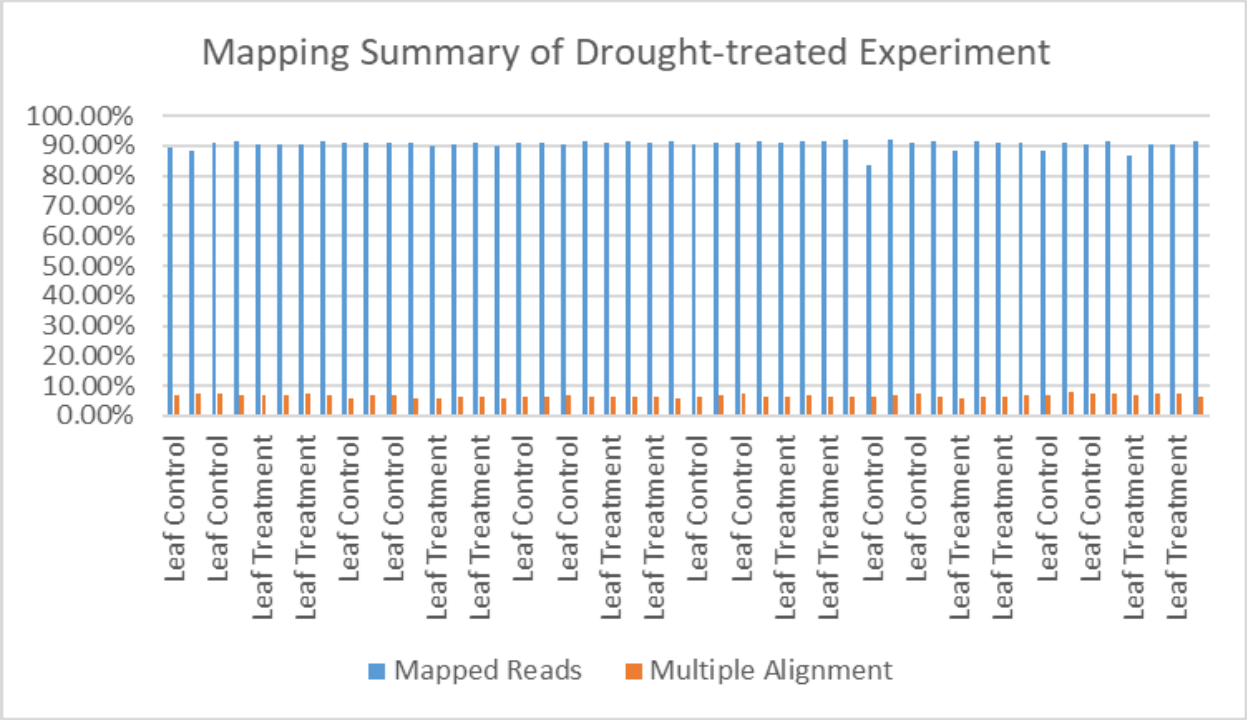


Figure 3 Graphical view of mapped reads and corresponding multiple alignment in each experimental condition.

5.2 Abundance Estimation

The aligned or mapped transcript and the reference annotation file (GFF file) were assembled with the use of cufflinks (v2.2.1) (Trap Nell et al, 2010) This results in generation of four files: 'transcript.gtf ', 'isoforms. fpkm_tracking', 'genes. fpkm_tracking', and 'accepted_hits.BAM' files. The 'genes. fpkm_tracking', and 'isoforms. fpkm_tracking' files contain the estimated level of expression of genes and isoforms respectively. This is measured in Fragment per Kilobase of transcript per Million mapped reads, FPKM.

In drought-treated potato, the total FPKM values obtained from genes. fpkm_tracking and isoform. fpkm_tracking was 2,369,848 and 3,429,826 respectively for gene and isoform expression level while in nitrogen-treated potato, the values are 272,435 gene level and 415,285 isoforms level.

5.3 Detection and Analysis of AS genes in Drought- and Nitrogen-Treated Potato Plants.

The transcripts assembled by cufflinks "transcript.gtf" were further merged with cuffmerge program to detect the genes that were alternatively splice. The Perl script program was further employed to identify the conserved AS genes between drought-stressed, and control potato plants in each cultivar.

In drought-treated potato plant, a total of 18,225 genes were alternatively splice out of 50,418 genes merged. When compared with nitrogen- treated counterpart, 14,929 genes out of 45,666 genes were alternatively spliced. In both nitrogen- & drought-treated, the non- AS genes are twice as much as the AS genes.

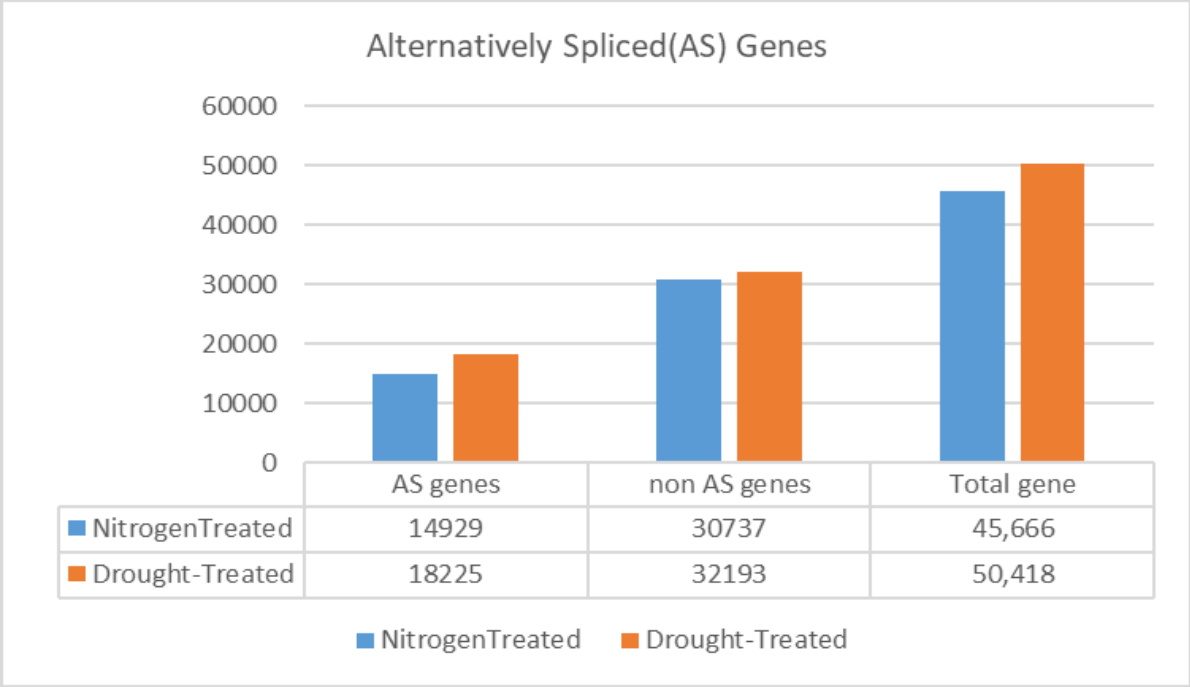


Figure 4 Comparative visualization of alternatively spliced genes in nitrogen- and drought-treated potato plants.

From Table 3 and the corresponding Figure 5 below, approximately the difference in number of alternatively spliced genes across all control varieties and drought- stressed varieties are comparatively small, however, the difference is large between drought-stressed and control sample.

Table 3. Comparative Number of Alternatively Spliced Genes in Drought- Treated Potato Cultivars.

Cultivars	Treatment	AS genes	Non-AS genes
Alegria	control	14,484	32,842
	Drought-Treated	14,817	33,732
Desiree	control	14,556	33,325
	Drought-Treated	14,792	33,847
Milva	control	14,425	32,946
	Drought-Treated	14,743	33,086
Saturna	control	14,597	32,687
	Drought-Treated	14,803	33,230

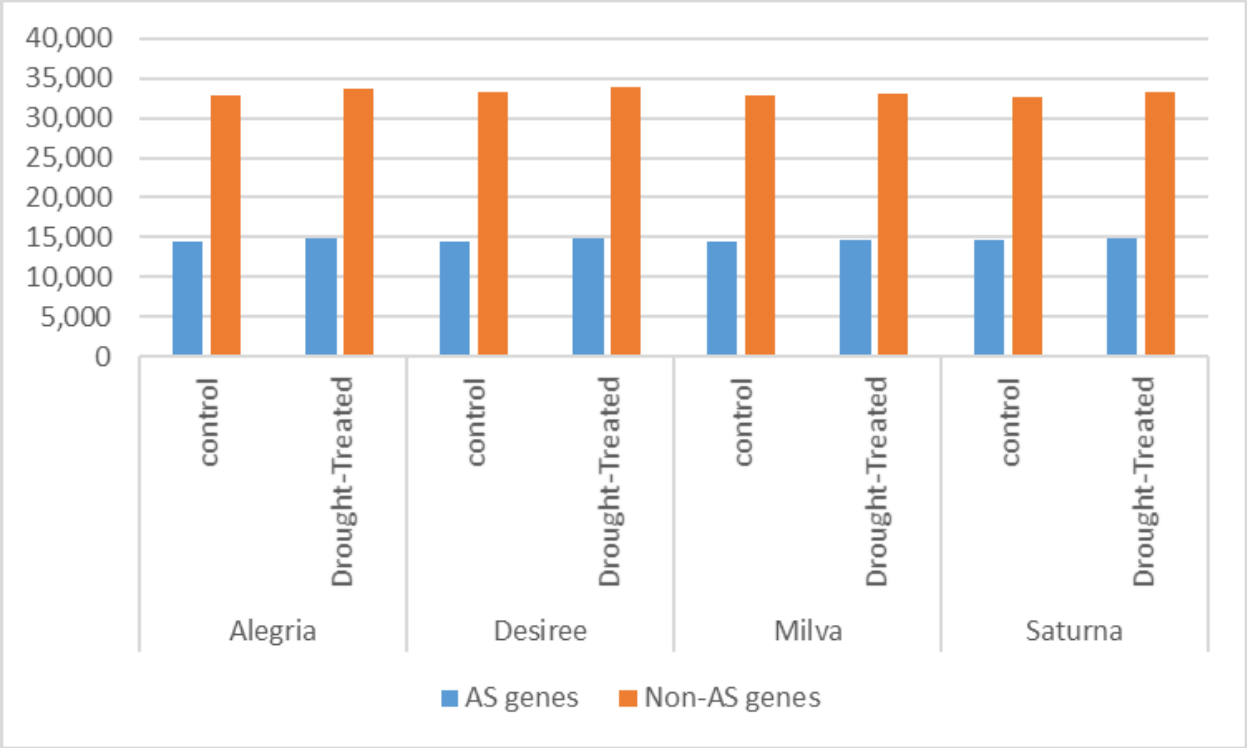


Figure 5 Graphical visualization of comparative analysis of alternatively spliced genes of drought-treated potato cultivars.

While in nitrogen- treated potato plants, the control (low N.,7.5mM) has more alternatively spliced genes than the treatments(0.75mM) potato plants as seen in the Figure 6 and Table 4. Out of all the expressed genes that were analyzed from all merged data, about one-third of the genes were found to be alternatively spliced.

Table 4. Estimate of AS genes analysis in nitrogen-treated potato

TREATMENT	% AS gene	%Non AS genes
CONTROL	31.63	68.37
TREATMENT	27.13	72.87

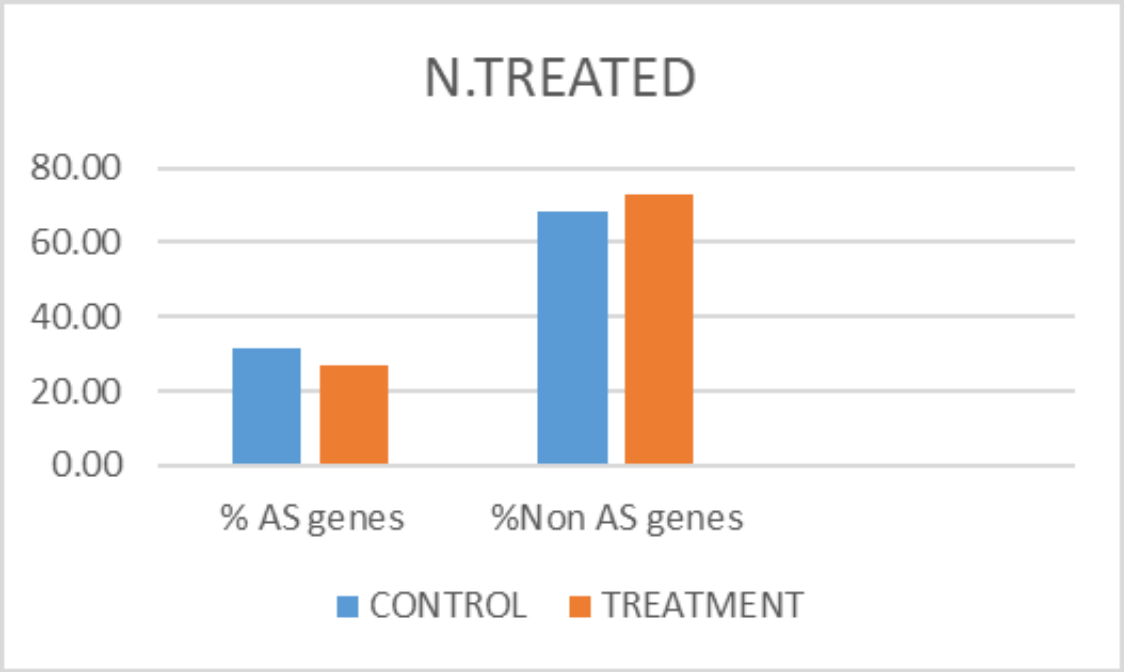


Figure 6 Data Representation of alternative spliced (AS) genes in nitrogen grown potato plants

5.4 Detection and Classification of AS events.

The assembled transcripts produced by the cufflinks in the 'transcript.gtf' file was used for the analysis of the alternative splicing events. The AS events in drought- and nitrogen- treated potato plants were detected and classified using AStalavista software.

5.4.1 AS events in nitrogen-treated plants

In Nitrogen-treated potato plants, a sum total of 26100 AS events from 'Landscape.gtf' file was identified by AStalavista from the assembled transcript. The landscaped AS events were visualized with the percentage estimate of each type of AS events in the figure below.

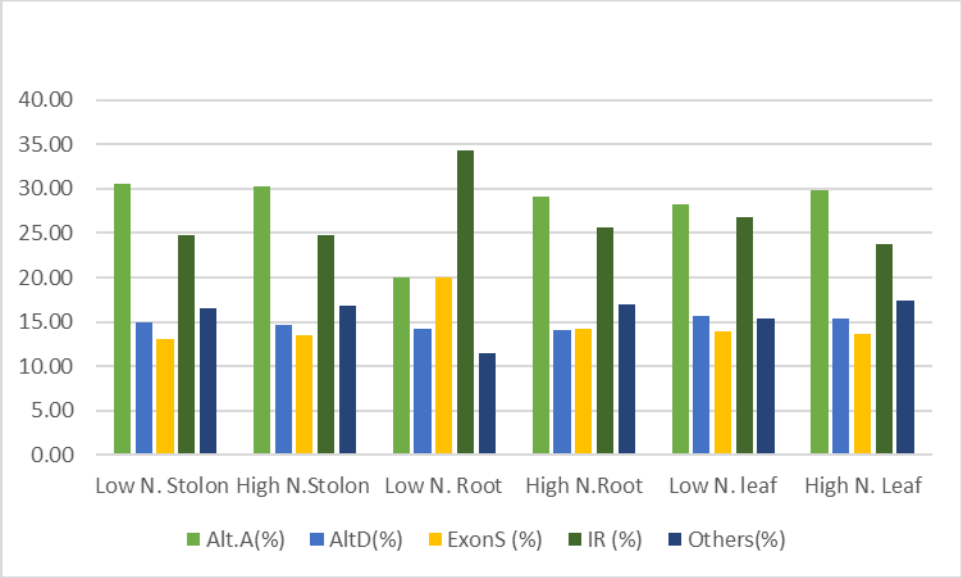


Figure 7 The bar chart indicates the relative percentage of each AS events in different parts of the plant at different concentration of nitrogen. Control (High Nitrogen, 7.5mM), Treatment (Low Nitrogen, 0.75mM)

From the chart, the percentage of alternative acceptor (AltA) is highest in the stolon, and leaf of potato plant tissue irrespective of whether it is high or low nitrogen-treated. However, the percentage of intron retention (IR or Intron.R) is higher than AltA in both control and treatment of the root of potato plants. The stolon samples (6492 and 6470 for control and treatment respectively)

seems to have highest number of alternatively spliced events as seen in the Table 4 below. Also, the root treatment has the lowest AS events (35) overall. Such a low number of AS events in root and leaf samples of the low nitrogen treatment is caused by relatively low mapping rate or by low nitrogen treatment needs to be further investigated, because unfortunately there were no replicates collected in the samples in this experiment.

Table 5. Alternative splicing events in Nitrogen- treated potato plants. Treatment (Low Nitrogen,0.75mM) Control (High Nitrogen, 7.5mM)

ACCESSION	TREATMENT	AltA	Alt.A(%)	AltD	AltD(%)	ExonS	ExonS (%)	IntronR	IR (%)	Others	Others(%)	Total
SRR10135769	Low N. Stolon	1979	30.59	968	14.96	845	13.06	1606	24.82	1072	16.57	6470
SRR10135770	High N.Stolon	1968	30.31	956	14.73	873	13.45	1604	24.71	1091	16.81	6492
SRR10135771	Low N. Root	7	20.00	5	14.29	7	20.00	12	34.29	4	11.43	35
SRR10135772	High N.Root	1851	29.14	894	14.07	904	14.23	1630	25.66	1074	16.91	6353
SRR10135773	Low N. leaf	221	28.26	122	15.60	109	13.94	210	26.85	120	15.35	782
SRR10135774	High N. Leaf	1778	29.79	921	15.43	813	13.62	1420	23.79	1036	17.36	5968

The above result is different from other previous studies on AS event in plants such as Arabidopsis, rice, maize and sacred lotus, since IR was found to be the predominant AS type in these plants (Min et al.2015; VanBuren et al. 2013)

5.4.2 AS events in drought-treated potato plants

A total of 258,786 AS events were identified from all the 48 samples consisting of four different cultivars of drought- treated potato using AStalavista to predict alternative splicing events in a pairwise fashion by comparing all overlapping transcript as shown below.

Table 6. The estimated number of each AS events in each experiment under conditions drought and irrigation (control). The types of AS events are indicated by red color, these are: Alternative acceptor site (Alt.A), Alternative donor site (Alt.D), Exon skipping/ExonS (ES), Intron retention/IntronR (IR), and Others (these are referred to complex events). The total of each event

is at the last row. The highest is indicated in the blue cell, followed by the light red, and the lowest is the dark red cell. Lastly, the grand total indicated in the green cell.

ACCESSION	Treatment	Experiment	AltA	Alt.A(%)	AltD	AltD(%)	ExonS	ExonS (%)	IntronR	IR (%)	Others	Others(%)	Total
SRR3161990	control	F1	1285	25.16	630	12.33	732	14.33	1533	30.01	928	18.17	5108
SRR3161991	control	F1	1301	25.71	648	12.80	725	14.33	1489	29.42	898	17.74	5061
SRR3161992	control	F1	1260	25.19	626	12.51	745	14.89	1474	29.47	897	17.93	5002
SRR3161993	control	F1	1329	25.13	651	12.31	738	13.95	1626	30.74	945	17.87	5289
SRR3161994	drought-stressed	F1	1337	25.03	671	12.56	742	13.89	1660	31.08	931	17.43	5341
SRR3161995	drought-stressed	F1	1358	25.82	648	12.32	731	13.90	1605	30.51	918	17.45	5260
SRR3161996	drought-stressed	F1	1341	25.41	642	12.16	747	14.15	1613	30.56	935	17.72	5278
SRR3161997	drought-stressed	F1	1361	24.81	663	12.09	746	13.60	1761	32.11	954	17.39	5485
SRR3161998	control	G3	1392	25.59	671	12.34	729	13.40	1675	30.80	972	17.87	5439
SRR3161999	control	G3	1327	25.36	641	12.25	729	13.93	1599	30.56	937	17.91	5233
SRR3162000	control	G3	1269	25.15	648	12.84	697	13.82	1518	30.09	913	18.10	5045
SRR3162001	control	G3	1388	25.12	694	12.56	772	13.97	1735	31.40	936	16.94	5525
SRR3162002	drought stressed	G3	1465	25.56	706	12.32	774	13.51	1781	31.08	1005	17.54	5731
SRR3162003	drought stressed	G3	1415	25.25	705	12.58	754	13.46	1761	31.43	968	17.28	5603
SRR3162004	drought stressed	G3	1416	25.25	680	12.13	763	13.61	1772	31.60	976	17.41	5607
SRR3162005	drought stressed	G3	1426	25.50	683	12.21	737	13.18	1796	32.11	951	17.00	5593
SRR3162006	control	G1	594	39.79	217	14.53	157	10.52	364	24.38	161	10.78	1493
SRR3162007	control	G1	1365	26.46	655	12.70	742	14.39	1496	29.00	900	17.45	5158
SRR3162008	control	G1	1326	24.96	667	12.56	721	13.57	1642	30.91	956	18.00	5312
SRR3162009	control	G1	1362	25.97	662	12.62	723	13.79	1541	29.39	956	18.23	5244
SRR3162010	drought -stressed	G1	1303	25.78	616	12.19	726	14.36	1506	29.80	903	17.87	5054
SRR3162011	drought -stressed	G1	1247	24.48	632	12.41	737	14.47	1537	30.18	940	18.46	5093
SRR3162012	drought -stressed	G1	1391	25.91	646	12.03	754	14.05	1608	29.96	969	18.05	5368
SRR3162013	drought -stressed	G1	1382	25.46	696	12.82	751	13.83	1662	30.61	938	17.28	5429
SRR3162014	control	G2	1374	25.66	675	12.61	713	13.31	1631	30.46	962	17.96	5355
SRR3162015	control	G2	1350	25.89	639	12.26	718	13.77	1582	30.34	925	17.74	5214
SRR3162016	control	G2	1382	26.16	663	12.55	712	13.48	1564	29.61	961	18.19	5282
SRR3162017	control	G2	133	26.49	71	14.14	63	12.55	150	29.88	85	16.93	502
SRR3162018	drought stressed	G2	1377	25.84	667	12.52	753	14.13	1598	29.99	934	17.53	5329
SRR3162019	drought stressed	G2	1293	24.91	657	12.66	747	14.39	1556	29.97	938	18.07	5191
SRR3162020	drought stressed	G2	1383	25.49	677	12.48	788	14.52	1606	29.60	972	17.91	5426
SRR3162021	drought stressed	G2	1404	26.23	661	12.35	749	13.99	1611	30.10	927	17.32	5352
SRR3162022	**control**	**F4**	1552	25.73	760	12.60	848	14.06	1858	30.80	1014	16.81	6032
SRR3162023	**control**	**F4**	1533	25.81	733	12.34	822	13.84	1835	30.90	1016	17.11	5939
SRR3162024	**control**	**F4**	1566	25.80	747	12.31	850	14.00	1861	30.66	1046	17.23	6070
SRR3162025	**control**	**F4**	1652	25.97	793	12.46	877	13.78	1945	30.57	1095	17.21	6362
SRR3162026	drought-stressed	**F4**	1669	27.16	783	12.74	848	13.80	1831	29.79	1015	16.51	6146
SRR3162027	drought-stressed	**F4**	1599	26.06	755	12.31	860	14.02	1859	30.30	1062	17.31	6135
SRR3162028	drought-stressed	**F4**	1631	26.22	766	12.32	828	13.31	1950	31.35	1045	16.80	6220
SRR3162029	drought-stressed	**F4**	1633	25.52	795	12.42	861	13.46	2031	31.74	1079	16.86	6399
SRR3162030	**control**	**F3**	1544	26.42	742	12.70	829	14.19	1755	30.04	973	16.65	5843
SRR3162031	**control**	**F3**	1546	26.27	727	12.35	854	14.51	1725	29.31	1033	17.55	5885
SRR3162032	**control**	**F3**	1496	25.44	745	12.67	853	14.51	1755	29.85	1031	17.53	5880
SRR3162033	**control**	**F3**	1606	25.80	786	12.63	865	13.90	1894	30.43	1074	17.25	6225
SRR3162034	drought stressed	**F3**	1593	25.74	779	12.58	863	13.94	1876	30.31	1079	17.43	6190
SRR3162035	drought stressed	**F3**	1601	25.95	774	12.55	862	13.97	1880	30.47	1052	17.05	6169
SRR3162036	drought stressed	**F3**	1557	25.78	738	12.22	857	14.19	1826	30.23	1062	17.58	6040
SRR3162037	drought stressed	**F3**	1508	25.78	761	13.01	828	14.16	1747	29.87	1005	17.18	5849
Total			66622		32292		36020		78680		45172		258786

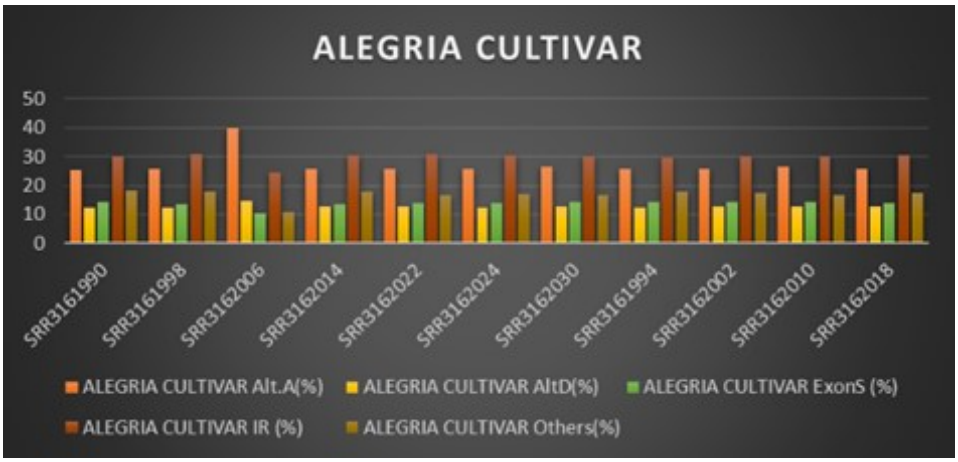


Figure 8 The percentage of AS events in Algeria cultivars. The dark brown indicates intron-retention (IR) with highest percentage.

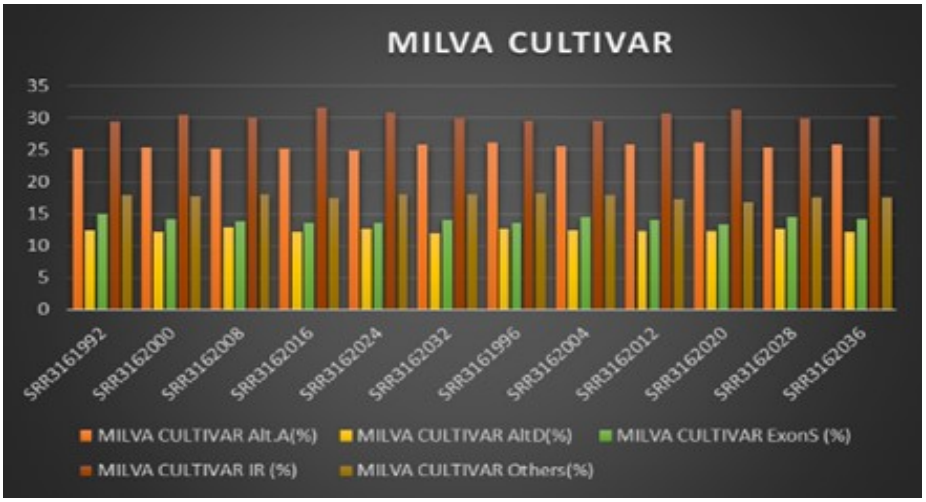


Figure 9 The percentage of AS events in Milva cultivars. The dark brown indicates intron-retention (IR) with highest percentage.

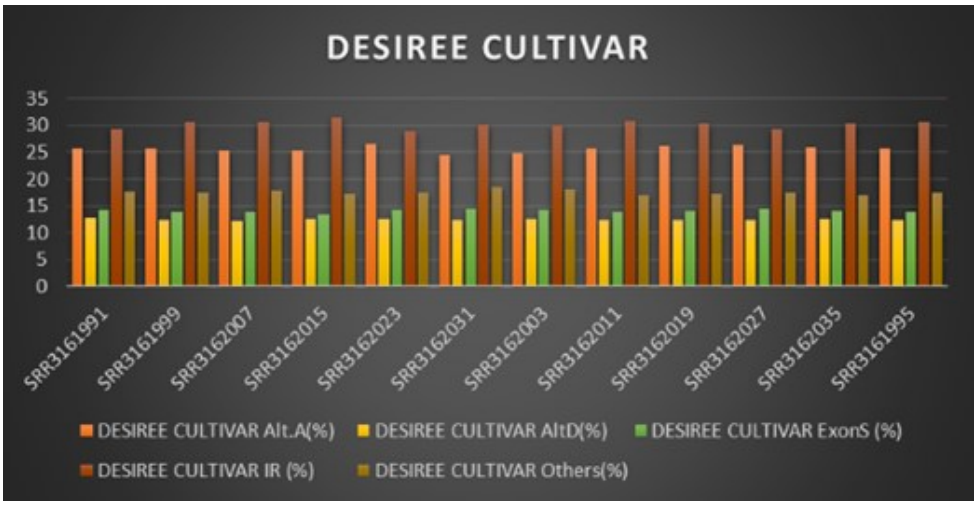


Figure 10 The percentage of AS events in Desiree cultivars. The dark brown indicates intron-retention (IR) with highest percentage.

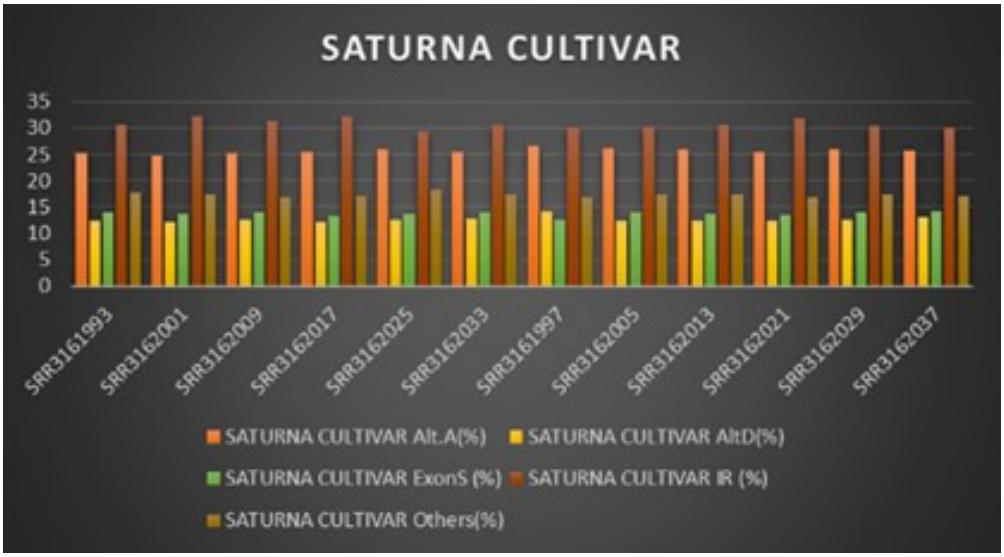


Figure 11. The percentage of AS events in Saturna cultivars. The dark brown indicates intron-retention (IR) with highest percentage.

Among different AS events, Intron retention (IR) is the most dominant (30.4%) followed by alternative acceptor site, AltA (25.74%), exon skipping (13.92%) and alternative donor, AltD, is the least occurring AS event type with 12.48%. Others are various complex events are formed by combination of basic events and 17.46% are complex types of AS events.

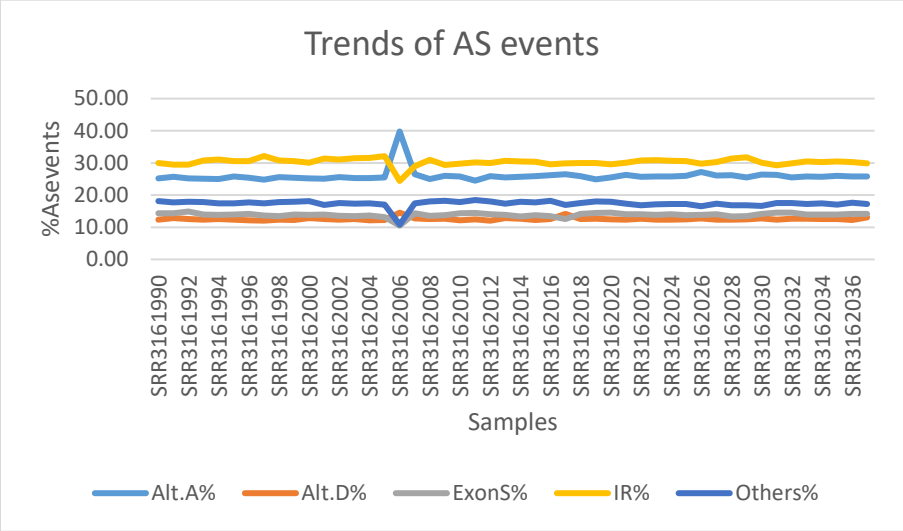


Figure 12. Displays the trends of percent AS events types in each sample according to the experiment.

Overall, the percentage of intron retention is higher than all other class of AS events regardless of the treatment category of drought in the experiment as seen from figure5. However, the percentage of alternative acceptor site (~40%) in one of the control sample (SRR3162007) is higher than every other AS event. This makes the result less consistent from previous investigations in other plant species such as pineapple, tomatoes and rice. Therefore, it is important to compare the AS events between the control and drought-treated potato and to estimate the conserved AS events in all four cultivars.

5.5 Comparative Identification of Unique and conserved AS events between control and drought-treated in each Cultivars.

Perl program was employed to compare the AS events between the control and drought stressed sample among the four cultivars. The purpose is to identify some AS events that are unique and conserved in control and drought stressed potatoes among the cultivars. “Gtf2event.pl” and “uniqpair.pl” are the Perl program used for this identification.

Prior to Perl analysis, the samples in each cultivar were merged as control and drought-specific assemblies into a single consensus assembly (i.e. control and drought) with cuff merge module in cufflink package. Each of these assemblies were used to identify the common AS events in each of the cultivars.

In Alegria and Milva cultivars, comparative analysis of AS events in drought stressed and control identified a total of 5550 and 5497 conserved AS events with alternative acceptor site being most dominant event in the two drought- sensitive cultivars. While in Desiree and Saturna cultivars, 5512 and 5491 AS events were conserved in total with alternative acceptor site as most dominating event in the two drought-tolerant cultivars. The significance of this, according to Yan Wang et al., 2014, is that conserved AS events plays an important role in species differences and genome evolution.

Also, the estimate of total AS events is higher in the control than in the drought-stressed in Milva and Saturna while the reverse occurs in Algeria and Saturna cultivars.

Table 7 Comparative estimate of conserved AS events between control and drought-stressed potato plants in each cultivar. The estimate of conserved AS events is indicated in red, and the white color within the red column is the highest conserved AS events, while the green indicates highest events among the control and drought followed by light green.

ALGERIA CULTIVAR				DESIREE CULTIVAR			
AS EVENTS	Control	Conserved	Drought	AS EVENTS	Control	Conserved	Drought
ExonS	1579	879	1704	ExonS	1579	897	1706
Alt.D	1525	783	1576	Alt.D	1494	761	1583
Alt.A	3497	1801	3794	Alt.A	3547	1782	3665
IntronR	1754	921	1878	IntronR	1632	897	1847
complex events	2326	1166	2369	complex events	2259	1175	2388
Total	10681	5550	11321	Total AS events	10511	5512	11189
MILVA CULTIVAR				SATURNA CULTIVAR			
AS EVENTS	Control	Conserved	Drought	AS EVENTS	Control	Conserved	Drought
ExonS	1633	908	1699	ExonS	1599	868	1718
Alt.D	1501	737	1513	Alt.D	1531	756	1648
Alt.A	3411	1776	3706	Alt.A	3543	1801	3813
IntronR	1670	873	1835	IntronR	1695	914	1957
complex events	2238	1203	2448	complex events	2278	1152	2403
Total AS events	10453	5497	5663	Total AS events	10646	5491	6048

Figure 13. A cross section of visual estimate of conserved AS events and treatment specific AS events in each cultivar.

Figure 13. shows that alternative acceptor site is higher in the drought stressed (green) than control(orange) and in all the cultivar. However, alternative donor site (Alt.D), and intron-retention seems to be the least conserved AS events in both the drought-treated samples and the control samples in all the cultivars.

5.6 Identification of Unique and Common AS events among cultivars

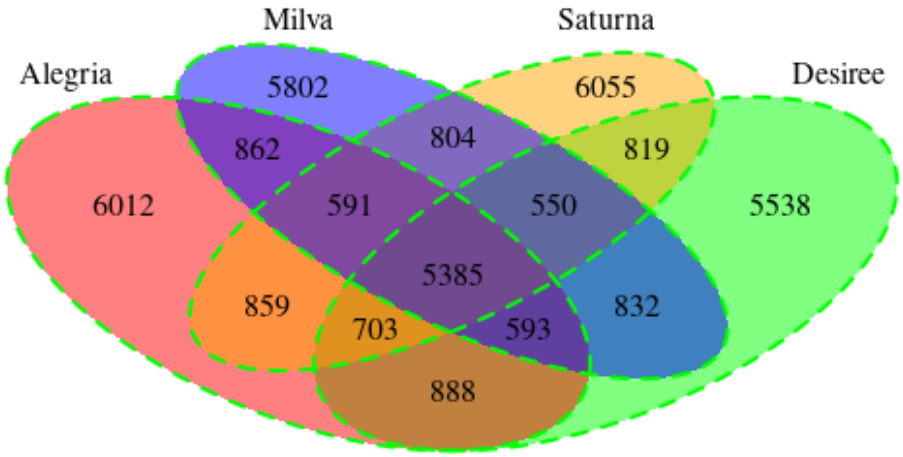


Figure 14. Venn diagram showing estimates of unique AS events in all the four cultivars

Figure 14 represents the number of unique AS events in each of the cultivars, there are 862 and 819 AS events are conserved in both Alegria/Milva cultivars and Desiree/Saturna cultivars respectively as listed below

Table 8 Comparative estimates of Common AS events in Algeria and Milva cultivars, Desiree and Saturna cultivars and among all cultivars.

AS EVENTS	ES (%)	ALT.D(%)	ALT.A(%)	IR(%)	COMPLEX EVENTS(%)	TOTAL
All Cutivars	1070(19.87)	785(14.58)	1757(32.63)	419(7.78)	1354(25.14)	5385
Algeria_Milva Cultivars	1508(20.29)	1055(14.20)	2445(32.90)	642(8.64)	1781(23.97)	7431
Desiree_Saturna Cultivars	1476(19.79)	1066(14.30)	2520(33.79)	635(8.52)	1760(23.60)	7457
Total AS events	4054	2906	6722	1696	4895	20273

From Table.7 alternative acceptor site (Alt.A) is the most dominant AS event common in all the merged cultivars followed by complex events while intron retention is the least conserved event in both cultivars. This may be due differences in the acceptor site defining exon-intron boundaries in highly conserved sequences that may change the interaction between pre-mRNA and spliceosomes complex (Riolo et al., 2021).

5.7 Differential Gene Expression And Alternative Spliced Genes

Differential expression of genes and alternative splicing events were analyzed simultaneously in drought-treated potato plants in each cultivar. This is done in other to identify and compare the level of gene expression and AS genes that occur in response to changing environmental conditions like drought.

In Alegria cultivars, 63,946 gene loci were identified with approximately 39,000 differentially expressed genes and 53,131 isoform transcripts. Majority of this expression were formed from intron-retention, however, there were some samples among this cultivar whose expressions were formed as a result of complex AS event types. While some of these expressed genes showed some

levels of significance (as judged by the p-value ≤0.05), none of the isoform's expressions were significant in this cultivar.

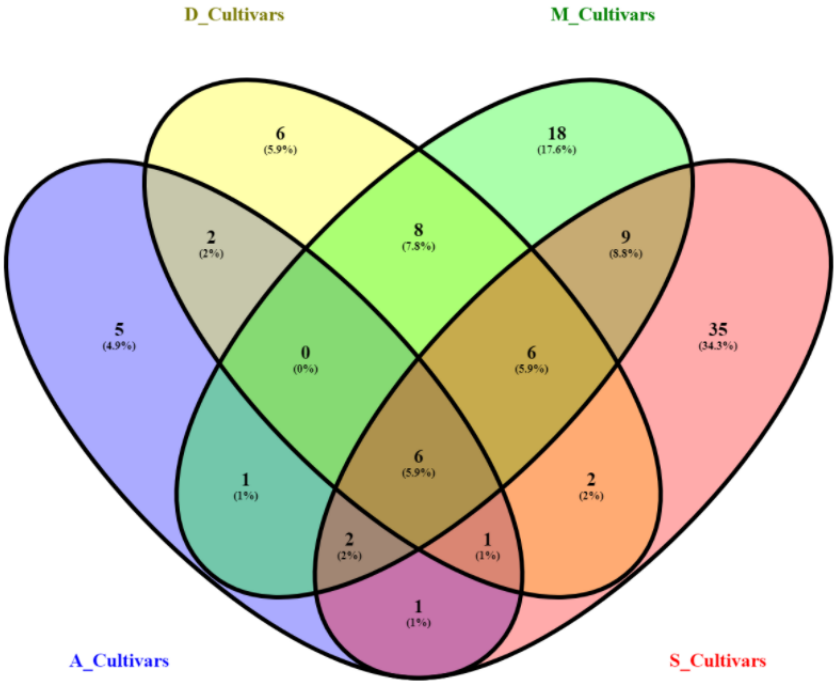


Figure 15. Represents the Venn Diagram Representation of Significant Differentially Expressed Genes in the Four Cultivars. A-Alegria, D-Desiree, M- Milva, and S- Saturna

However, out of the average 65,247 gene loci in the remaining three cultivars, 8 isoforms, 18 isoforms and 28 isoforms were significantly expressed in respectively in Desiree, Milva, and Saturna cultivars (Figure 15). About 39,000 gene loci and 53,131 isoform transcripts were expressed in all the three cultivars.

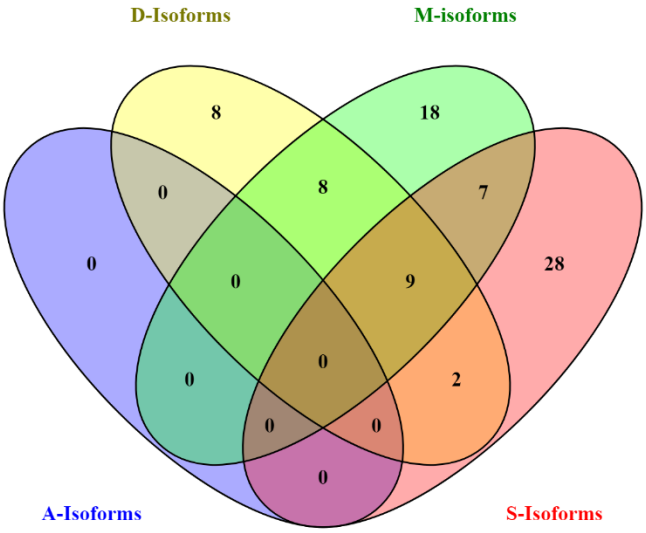
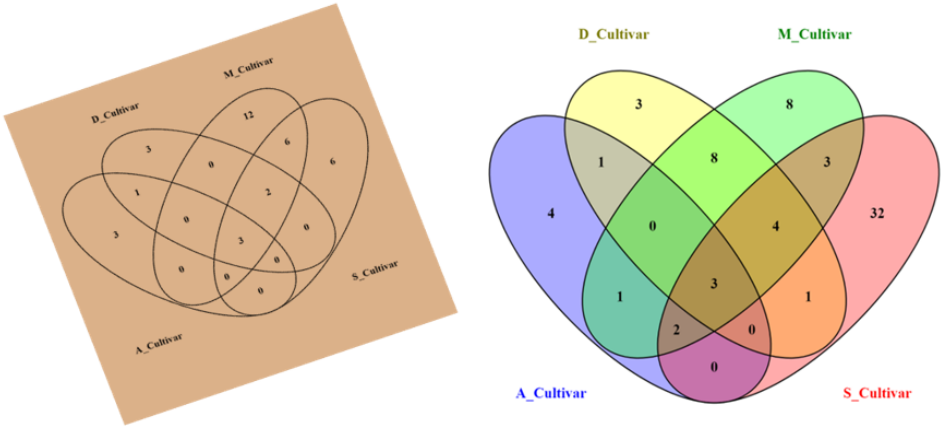


Figure 16. Represents the Venn Diagram Representation of Significantly Expressed isoforms in the Four Cultivars. A-Alegria, D-Desiree, M- Milva, and S- Saturna

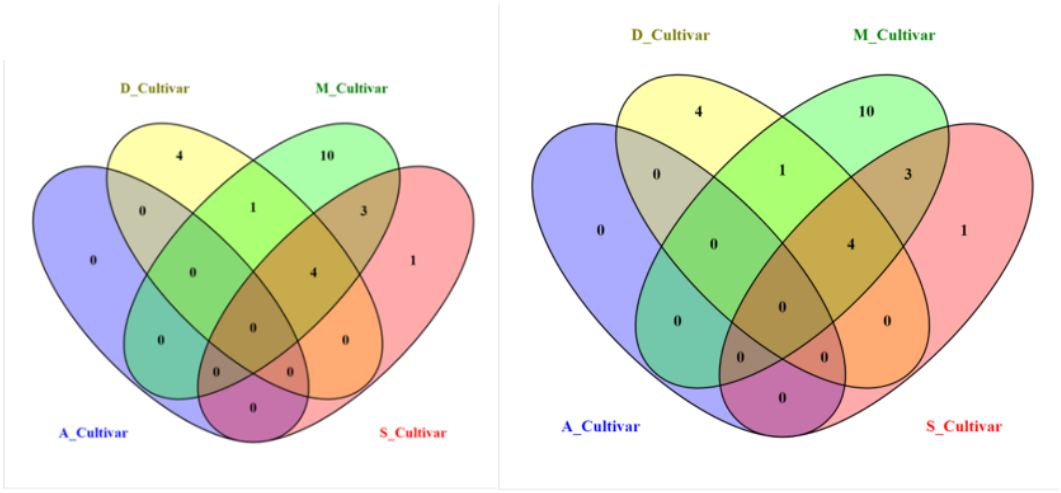
The expression levels of genes and generated transcript isoforms were estimated with cuffdiff module of cufflinks v2.2.1 programmed. A total of 228 and 167 of differentially expressed genes (DEGs) and differential isoforms respectively from each of the four cultivars were statistically significant ($P \leq 0.05$). These DEGs and isoforms were categorized as up-regulated and down-regulated and analyzed in each cultivar.



Up-regulated genes *Down-regulated genes*

Figure 17. Venn diagram of commonly and uniquely up-regulated and down-regulated genes in the four cultivars

Up-regulated Isoforms



Down-regulated Isoforms

Figure 15. Venn diagram analysis of differentially expressed genes with upregulated and down-regulated isoforms in the leaves of these four cultivars showed some common and unique genes and isoforms.

In Alegria cultivar, out of total 23 statistically significant DEGs, 10 (up-regulated) and 13 (down-regulated). Some highly upregulated genes were Hypoxia-responsive family protein, Glycine-rich protein family, and peroxidase superfamily protein etc.; whereas highly down-regulated genes were oxidoreductase, Di-glucose binding protein with Leucine-rich repeat domain.

In Desiree cultivar, out of total 45 statistically significant DEGs, 24 up-regulated and 21 down-regulated while 21 and 18 differential isoforms were upregulated and down-regulated respectively. Some highly upregulated genes in Desiree cultivars were heat shock protein -like chaperone and ribosomal proteins, etc.; whereas highly down-regulated genes were NmrA-like negative transcriptional regulator family protein and homocysteine S-methyltransferase 3.

In Saturna, out of total 84 statistically significant DEGs, (33) up-regulated and (51) down-regulated while 22 and 40 differential isoforms were upregulated and down-regulated respectively. Some highly upregulated genes were heat shock proteins, Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein etc.; whereas highly down-regulated genes were Glycosyl hydrolase superfamily protein, Oxidoreductase family protein and lipoxygenase etc.

In Milva cultivar, out of total 75 statistically significant DEGs, (43) up-regulated and 32 down-regulated while 28 and 27 differential isoforms were upregulated and down-regulated respectively under conditions of drought. Some highly upregulated genes were chloroplast heat shock protein and heat shock protein 90.1 etc. Whereas highly down-regulated genes were Leucine-rich repeat transmembrane protein kinase, and Glycosyl hydrolases family 31 protein.

Figure 18. Differentially Expressed Genes& Isoforms.

Table 9 Shows protein identification and the description of the 39 common differential expression of genes and their transcript isoforms.

GENE ID	PEPTIDE NAME	PROTEIN FAMILY	GENE DESCRIPTION
PGSC0003DMG400006448	PGSC0003DMP400011444	PF13578	caffeoyl-CoA 3-O-methyltransferase
PGSC0003DMG400009867	PGSC0003DMP400017393	PF11961,PF05003	Protein of unknown function (DUF668)
PGSC0003DMG400014325	PGSC0003DMP400025230	N/A	Sugar isomerase (SIS) family protein
PGSC0003DMG400006448	PGSC0003DMP400033458	PF01095	methylesterase PCR A
PGSC0003DMG400009867	PGSC0003DMP400041655	N/A	oxidative stress 3
PGSC0003DMG400014325	PGSC0003DMP400008119	PF00182	chitinase-like protein 2
PGSC0003DMG400006448	PGSC0003DMP400008672	PF00010	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
PGSC0003DMG400009867	PGSC0003DMP400014394	PF00011	17.6 kDa class II heat shock protein
PGSC0003DMG400014325	PGSC0003DMP400014501	PF00067	cytochrome P450, family 72, subfamily A, polypeptide 15
PGSC0003DMG400006448	PGSC0003DMP400015381	PF01734	Acyl transferase/acyl hydrolase/lysophospholipase superfamily protein
PGSC0003DMG400009867	PGSC0003DMP400022725	PF00234	lipid transfer protein 1
PGSC0003DMG400014325	PGSC0003DMP400034202	PF08263,PF00560	leucine-rich repeat transmembrane protein kinase family protein
PGSC0003DMG400006448	PGSC0003DMP400053102	PF04819	Family of unknown function (DUF716)
PGSC0003DMG400009867	PGSC0003DMP400042657	PF02861,PF07728,PF1043	heat shock protein 101
PGSC0003DMG400014325	PGSC0003DMP400003664	PF02458	HXXXD-type acyl-transferase family protein
PGSC0003DMG400006448	PGSC0003DMP400050837	PF00210	ferritin 4
PGSC0003DMG400009867	PGSC0003DMP400003107	PF00141	ascorbate peroxidase 2
PGSC0003DMG400014325	PGSC0003DMP400004546	PF00230	delta tonoplast integral protein
PGSC0003DMG400006448	PGSC0003DMP400004578	PF00400	Coatomer, beta\' subunit
PGSC0003DMG400009867	PGSC0003DMP400009082	PF00188	basic pathogenesis-related protein 1
PGSC0003DMG400022013	PGSC0003DMP400038079	PF00188,PF07887	CAP (Cysteine-rich secretory proteins, Antigen 5,
			and Pathogenesis-related 1 protein) superfamily protein
PGSC0003DMG400024475	PGSC0003DMP400042323	PF12695	Calmodulin binding protein-like
PGSC0003DMG400027960	PGSC0003DMP400048606	PF03760	carboxyesterase 20
PGSC0003DMG400011437	PGSC0003DMP400020263	PF13561	Late Embryogenesis Abundant 4-5
PGSC0003DMG400011601	PGSC0003DMP400020593	PF03106	NAD(P)-binding Rossmann-fold superfamily protein
PGSC0003DMG400012160	PGSC0003DMP400021486	N/A	WRKY family transcription factor
PGSC0003DMG400012182	PGSC0003DMP400021528	PF01490	Transducin/WD40 repeat-like superfamily protein
PGSC0003DMG400015198	PGSC0003DMP400026660	PF02153	lysine histidine transporter 1
PGSC0003DMG400020334	PGSC0003DMP400035312	PF00067	arogenate dehydrogenase
PGSC0003DMG401020453	PGSC0003DMP400035524	PF03168	cytochrome P450, family 71, subfamily B, polypeptide 23
PGSC0003DMG400020863	PGSC0003DMP400036253	PF00646,PF14299	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein fa
PGSC0003DMG400021065	PGSC0003DMP400036524	PF03107,PF07649	F-box family protein
PGSC0003DMG400025162	PGSC0003DMP400043675	PF07690	Cysteine/Histidine-rich C1 domain family protein
PGSC0003DMG400027425	PGSC0003DMP400047700	PF01053	sugar transporter 14
PGSC0003DMG400029022	PGSC0003DMP400050560	PF03407	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
PGSC0003DMG400029570	PGSC0003DMP400051495	PF13193,PF00501	Nucleotide-diphospho-sugar transferase family protein
PGSC0003DMG400029915	PGSC0003DMP400052109	PF11960,PF00487	AMP-dependent synthetase and ligase family protein
PGSC0003DMG400032204	PGSC0003DMP400055412	N/A	fatty acid desaturase 2
PGSC0003DMG400033125	PGSC0003DMP400055897	N/A	Glycosyl hydrolase superfamily protein

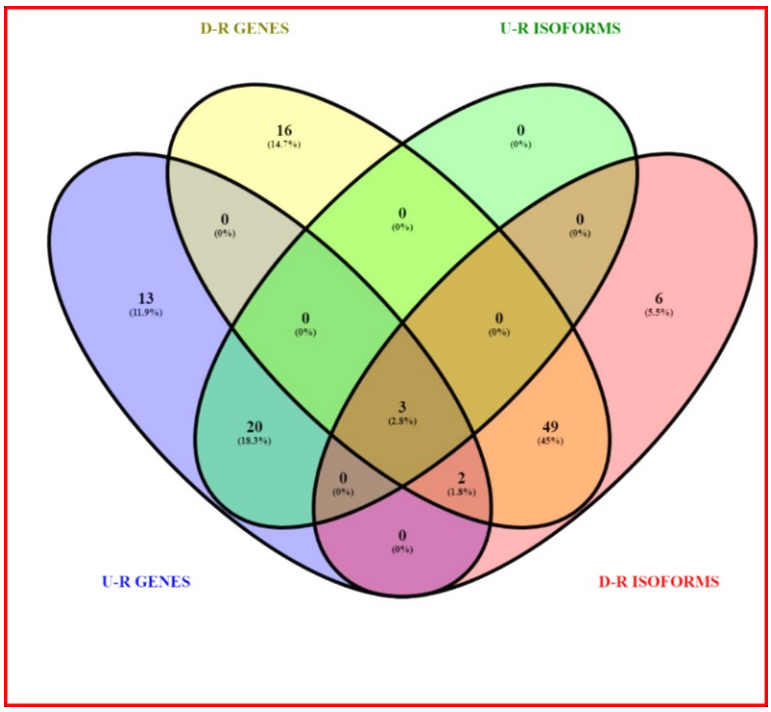


Figure 19. Identification of genes and isoforms that are frequently upregulated and down regulated in all the cultivars.

Figure 18 identifies gene functions and protein family that are frequently upregulated and down-regulated among differential expressed genes and isoforms in all the cultivars. Also, Peroxidase superfamily protein, jasmonic acid carboxyl methyltransferase were commonly up-regulated and down -regulated genes and isoforms

In down-regulated genes category, PYR1-like 6, Unknown protein Class, and Oxidoreductase family protein were common in the four cultivars, Uncharacterized protein family (UPF0497), related to ABI3/VP1 2, Plant invertase/pectin methyl esterase inhibitor superfamily protein, & N-terminal nucleophile amino hydrolases (Ntn hydrolases) superfamily protein) were common in

Desiree, Milva and Saturna. Only Homocysteine S-methyltransferase 3 is common to both Milva and Saturna cultivars.

Among up-regulated gene category, only 70B, 70T-2, & chaperones -like heat shock protein were common in the four cultivars. Whereas, in up -regulated isoforms, heat shock protein (20,70T-2, & 90.1) were common in Desire, Saturna and Milva cultivars with heat shock 70B common to both Desiree and Milva cultivars. Chloroplast heat shock protein 70-2, heat shock cognate protein 70-1, and mitochondrion-localized small heat shock protein 23.6 were the common isoforms in Milva and Saturna cultivars.

Table 10 Represents commonly upregulated genes in all the four cultivars and their protein families.

GENE ID	PEPTIDE NAME	PROTEIN FAMILY	GENE DESCRIPTION
PGSC0003DMG400014212	PGSC0003DMP400025001	PF00012	heat shock protein 70B
PGSC0003DMG400014956	PGSC0003DMP400026309	PF00011	HSP20-like chaperones superfamily protein
PGSC0003DMG400030089	PGSC0003DMP400052405	PF00012	heat-shock protein 70T-2
PGSC0003DMG400024887	PGSC0003DMP400043144	PF00012	chloroplast heat shock protein 70-2
PGSC0003DMG400027750	PGSC0003DMP400048253	PF00012	heat shock cognate protein 70-1
PGSC0003DMG400004808	PGSC0003DMP400008519	PF00011	mitochondrion-localized small heat shock protein 23.6
PGSC0003DMG400011197	PGSC0003DMP400019802	PF00012	heat shock protein 70B
PGSC0003DMG400010238	PGSC0003DMP400018101	PF10604	PYR1-like 6

Overall, different classes of heat shock proteins and PYR1-like proteins were identified in all the cultivars and constitute most commonly upregulated genes in all these four cultivars. PYR1-Like (Pyrabactin Resistance Like-1) positively regulates the abscisic acid signaling and improves adaptive responses to drought in plants (Yu et al.,2017). However, heat shock protein 20-like protein 1 is the only class of this protein that is unique to Milva cultivar and among the up-regulated genes in this particular cultivar. Heat shock protein 20- like protein 1, according to several studies, is different from HSP70 and HSP90 chaperone families in which sequence identity is significant throughout the protein (Scharf et al., 2001). These HSP 70 and other class of heat shock proteins

play a role in plant development and response to stress conditions like drought (Park and Seo, 2015; Usman et al., 2017; Ul Haq et al., 2019)

CHAPTER VI: CONCLUSION

RNA sequencing technology has been extensively used to conduct transcriptome analysis in plants. In this work, 18,225 and 14,929 alternatively spliced (AS) genes were generated in drought-stressed and nitrogen-grown potato plants, respectively. From these AS genes, a total of 258,786 and 21,600 AS events were identified and categorized in drought-stressed and nitrogen-grown potato plants, respectively.

The AS events analysis reveals no appreciable difference between low nitrogen and high nitrogen grown potatoes in nitrogen grown potato plants. Previous studies by Tiwari et al., 2020 also confirm nitrogen efficient potato cultivar to yield higher tuber under low nitrogen when compared with high nitrogen treatment. The significance of efficient use of nitrogen at a low concentration is to save the environment from hazards and, consequently, maintain healthy living.

Furthermore, the RNA sequencing analysis on the leaves of drought-stressed potato plants reveals levels of differentially expressed genes. The heat shock protein family and PYR1-Like (Pyrabactin Resistance Like-1) are the most common, significant, and upregulated gene in all the cultivars, which correlates with an investigation done by Sprenger et al., 2016. According to several studies, heat shock protein families are a group of proteins abundantly produced during abiotic stress conditions. It is a self-defense mechanism for withstanding adverse conditions in higher plants (Peng Zhao et al., 2018).

The future work would be to identify the characteristics of exons and introns of heat shock protein-coding genes to understand the evolution of these genes in potato plants. This will further address

genome complexities in potato plants and how this gene can be manipulated to harness the developmental process in potato plants in adverse conditions.

REFERENCE

1. Tiwari, J. K., Buckseth, T., Devi, S., Varshney, S., Sahu, S., Patil, V. U., Zinta, R., Ali, N., Moudgil, V., & Singh, R. K. (2020). Physiological and genome-wide RNA-sequencing analyses identify candidate genes in a nitrogen-use efficient potato cv. Kufri Gaurav. Plant Physiology and Biochemistry, 154, 171-183.

2. Devaux, A., Goffart, J. P., Kromann, P., Andrade-Piedra, J., Polar, V., & Hareau, G. (2021). The potato of the future: opportunities and challenges in sustainable agri-food systems. Potato Research, 64(4), 681-720.

3. Burlingame, B., Mouillé, B., & Charrondiere, R. (2009). Nutrients, bioactive non-nutrients and anti-nutrients in potatoes. Journal of food composition and analysis, 22(6), 494-502.

4. Lemke, P., Moerschbacher, B. M., & Singh, R. (2020). Transcriptome analysis of solanum tuberosum genotype rh89-039-16 in response to chitosan. Frontiers in plant science, 1193.

5. Sprenger, H., Kurowsky, C., Horn, R., Erban, A., Seddig, S., Rudack, K., ... & Kopka, J. (2016). The drought response of potato reference cultivars with contrasting tolerance. Plant, Cell & Environment, 39(11), 2370-2389.

6. Sablok, G., Powell, B., Braessler, J., Yu, F., & Min, X. J. (2017). Comparative landscape of alternative splicing in fruit plants. Current Plant Biology, 9, 29-36.

7. Pennisi, E. (2008). Getting to the root of drought responses. Science, 320(5873), 173-173.

8. Tiwari, J. K., Buckseth, T., Devi, S., Varshney, S., Sahu, S., Patil, V. U., ... & Rai, A. (2020). Physiological and genome-wide RNA-sequencing analyses identify candidate genes in a nitrogen-use efficient potato cv. Kufri Gaurav. Plant Physiology and Biochemistry, 154, 171-183.

9. Reddy, A. S., Marquez, Y., Kalyna, M., Barta, A. (2013). Complexity of the alternative splicing landscape in plants. Plant Cell 25, 3657-3683. doi: 10.1105/tpc.113.117523.

10. Iñiguez, L. P., & Hernández, G. (2017). The evolutionary relationship between alternative splicing and gene duplication. Frontiers in genetics, 8, 14.

11. Li, Q., Qin, Y., Hu, X., Li, G., Ding, H., Xiong, X., & Wang, W. (2020). Transcriptome analysis uncovers the gene expression profile of salt-stressed potato (Solanum tuberosum L.). Scientific reports, 10(1), 1-19.

12. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., ... & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology, 28(5), 511-515.

13. Zhang, C., Yang, H., & Yang, H. (2015). Evolutionary character of alternative splicing in plants. Bioinformatics And Biology Insights, 9, BBI. S33716.

14. McGuire, A. M., Pearson, M. D., Neafsey, D. E., & Galagan, J. E. (2008). Cross kingdom patterns of alternative splicing and splice recognition. Genome Biology, 9(3), 1-19.

15. Mei, W., Boatwright, L., Feng, G., Schnable, J. C., & Barbazuk, W. B. (2017). Evolutionarily conserved alternative splicing across monocots. Genetics, 207(2), 465-480.

16. Foissac, S., Sammeth, M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res. 2007; 35:W297–299. doi: 10.1093/nar/gkm311

17. Reddy, A. S., Marquez, Y., Kalyna, M., Barta, A. (2013). Complexity of the alternative splicing landscape in plants. Plant Cell 25, 3657-3683. doi: 10.1105/tpc.113.117523

18. Melo, J. P., Kalyna, M., & Duque, P. (2020, 2020-March-24). Current Challenges in Studying Alternative Splicing in Plants: The Case of Physcomitrella patens SR Proteins [Perspective]. Frontiers In Plant Science, 11(286). https://doi.org/10.3389/fpls.2020.00286

19. Min, X. (2017). Comprehensive cataloging and analysis of alternative splicing in maize. Computational Molecular Biology, 7.

20. Min, X. J. (2013). ASFinder: a tool for genome-wide identification of alternatively splicing transcripts from EST-derived sequences. International Journal Of Bioinformatics Research And Applications, 9(3), 221-226.

21. Min, X. J. (2018). A survey of alternative splicing in allotetraploid cotton (Gossypium hirsutum L.). Computational Molecular Biology, 8.

22. Min, X. J., Butler, G., Storms, R., & Tsang, A. (2005a). OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Research, 33(suppl_2), W677-W680.

23. Min, X. J., Butler, G., Storms, R., & Tsang, A. (2005b). TargetIdentifier: a webserver for identifying full-length cDNAs from EST sequences. Nucleic Acids Research, 33(suppl_2), W669-W672.

24. Min, X. J., Powell, B., Braessler, J., Meinken, J., Yu, F., & Sablok, G. (2015). Genome-wide cataloging and analysis of alternatively spliced genes in cereal crops. BMC Genomics, 16(1), 1-13.

25. Staiger, D., Korneli, C., Lummer, M., & Navarro, L. (2013). Emerging role for RNA-based regulation in plant immunity. New Phytologist, 197(2), 394-404.

26. Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T., & Soreq, H. (2005). Function of alternative splicing. Gene, 344, 1-20.

27. 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.

28. Zhang, Y., Zhao, G., Li, Y., Zhang, J., Shi, M., Muhammad, T., et al. (2017). Transcriptome profiling of tomato uncovers an involvement of cytochrome P450s and peroxidases in stigma color formation. Front. Plant Sci. 8, 897. doi: 10.3389/fpls.2017.00897

29. [7] Chaudhary S., Khokhar W., Jabre I., et al., 2019. Alternative splicing and protein diversity: plants versus animals, Frontiers in plant science, 10, p.708.

30. [8] Clark S., Yu F., Gu L., et al., 2019, Expanding alternative splicing identification by integrating multiple sources of transcription data in tomato, Frontiers in Plant Sciences. 10:689. doi:10.3389/fpls.2019.00689.

31. Riolo, G., Cantara, S., & Ricci, C. (2021). What's Wrong in a Jump? Prediction and Validation of Splice Site Variants. Methods and protocols, 4(3), 62.

32. Wanniarachchi, V. R., Dametto, L., Sweetman, C., Shavrukov, Y., Day, D. A., Jenkins, C. L., & Soole, K. L. (2018). Alternative respiratory pathway component genes (AOX and ND) in rice and barley and their response to stress. International Journal of Molecular Sciences, 19(3), 915.

33. Zhu, B., Yang, Y., Li, R., Fu, D., Wen, L., Luo, Y., et al. (2016). RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. J. Exp. Bot. 66, 4483-4495. doi: 10.1093/jxb/erv203

34. Carvalho, R. F., Feijão, C. V., & Duque, P. (2013). On the physiological significance of alternative splicing events in higher plants. Protoplasma, 250(3), 639-650.

35. Zhao, P., Wang, D., Wang, R., Kong, N., Zhang, C., Yang, C., ... & Chen, Q. (2018). Genome-wide analysis of the potato Hsp20 gene family: identification, genomic organization and expression profiles in response to heat stress. BMC genomics, 19(1), 1-13.

36. Chen, C., Chen, H., He, Y., & Xia, R. (2018). TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. BioRxiv, 289660(10.1101), 289660.

37. Tress, M. L., Abascal, F., & Valencia, A. (2017). Alternative splicing may not be the key to proteome complexity. Trends in biochemical sciences, 42(2), 98-110.

38. Shang, X., Cao, Y., & Ma, L. (2017). Alternative splicing in plant genes: a means of regulating the environmental fitness of plants. International journal of molecular sciences, 18(2), 432.

39. Laloum, T., Martín, G., & Duque, P. (2018). Alternative splicing control of abiotic stress responses. Trends in plant science, 23(2), 140-150.

40. Sprenger, H., Kurowsky, C., Horn, R., Erban, A., Seddig, S., Rudack, K., ... & Kopka, J. (2016). The drought response of potato reference cultivars with contrasting tolerance. Plant, Cell & Environment, 39(11), 2370-2389.

41. Qu, Z., Qi, X., Shi, R., Zhao, Y., Hu, Z., Chen, Q., & Li, C. (2020). Reduced n fertilizer application with optimal blend of controlled-release urea and urea improves tomato yield and quality in greenhouse production system. Journal of Soil Science and Plant Nutrition, 20(4), 1741-1750.

42. Gálvez, J. H., Tai, H. H., Lagüe, M., Zebarth, B. J., & Strömvik, M. V. (2016). The nitrogen responsive transcriptome in potato (Solanum tuberosum L.) reveals significant gene regulatory motifs. Scientific reports, 6(1), 1-15.

43. Sun, X., Xu, L., Wang, Y., Yu, R., Zhu, X., Luo, X., ... & Liu, L. (2015). Identification of novel and salt-responsive miRNAs to explore miRNA-mediated regulatory network of salt stress response in radish (Raphanus sativus L.). BMC genomics, 16(1), 1-16.

44. Beatty, P. H., Klein, M. S., Fischer, J. J., Lewis, I. A., Muench, D. G., & Good, A. G. (2016). Understanding plant nitrogen metabolism through metabolomics and computational approaches. Plants, 5(4), 39.