Journal of Plant Sciences
2023; 11(3): 98-106
http://www.sciencepublishinggroup.com/j/jps
doi: 10.11648/j.jps.20231103.19
ISSN: 2331-0723 (Print); ISSN: 2331-0731 (Online)

Systematic Collection and Analysis of Alternative Splicing Events in Potato Plants

Atinuke Ogungbayi[1], Jessica Lee[1], Vishwa Vaghela[1], Feng Yu[2], Xiangjia Min[1, *]

[1]Department of Chemical and Biological Sciences, Youngstown State University, Youngstown, USA

[2]Department of Computer Science and Information Systems, Youngstown State University, Youngstown, USA

Email address:

aeogungbayi@student.ysu.edu (Atinuke Ogungbayi), jalee01@student.ysu.edu (Jessica Lee), vsvaghela@student.ysu.edu (Vishwa Vaghela), fyu@ysu.edu (Feng Yu), xmin@ysu.edu (Xiangjia Min)

[*]Corresponding author

To cite this article:

Atinuke Ogungbayi, Jessica Lee, Vishwa Vaghela, Feng Yu, Xiangjia Min. Systematic Collection and Analysis of Alternative Splicing Events in Potato Plants. *Journal of Plant Sciences*. Vol. 11, No. 3, 2023, pp. 98-106. doi: 10.11648/j.jps.20231103.19

Received: April 10, 2023; **Accepted**: June 19, 2023; **Published**: June 27, 2023

Abstract: Potato [*Solanum tuberosum* L.] is one of the major food plants with complete genome sequences available. Plant genes are subjected to alternative splicing (AS), a process increases both transcriptome and proteome diversities. The work reports a systematic genome-wide study on identification and analysis of AS events by integrating multiple sources of sequencing data in potato plants. A collection of 291,071 expressed sequence tags (ESTs) and mRNA sequences were cleaned and assembled into 150,435 unique transcripts and 87,992 of them were mapped to potato genome. In addition, a total of 5.8 billion out of 7.7 billion RNA-sequencing (RNA-seq) reads, which were collected from 227 samples deposited from 10 published projects, were mapped to potato genome. Combining all mapping data results in identification of a total of 226,769 AS events, which were further classified into basic events (49.0%) and complex events (51.0%), that were generated from 24,650 genes. The basic AS events include intron retention (9.5%), alternative acceptor sites (19.2%), alternative donor site (8.2%), and exon skipping (12.1%). The AS rate of annotated gene models was estimated to be ~45.8% in potato plants. Comparative analysis identified 2,929 alternatively splice genes conserved among potato, tomato, soybean and maize plants. The work provides an important resource for further functional characterization of these genes in potato biology.

Keywords: Alternative Splicing, Expressed Sequence Tags, mRNA, RNA Sequencing, Transcriptome, Potato

1. Introduction

Potato [*Solanum tuberosum* L] ranks as the world's third most important crop for human consumption, after rice and wheat, the most important non-cereal food, playing an important role in the food security of developing countries [1]. The first draft potato genome represented approximately 86% of the 844-megabase genome with 39,031 protein coding genes identified from a homozygous doubled-monoploid potato clone [2]. The genome sequence super-scaffolds were further assembled into pseudomolecules corresponding to 12 potato chromosomes, representing 674 Mb of the 723 Mb genome assembly and 37,482 protein coding genes [3].

Plant genes are subjected to alternative splicing (AS) during the transcription process. AS generates two or more types of RNA transcripts from a single gene, thus, increasing both transcriptome and proteome diversities. AS plays important roles in regulating growth and development as well as in coping various stresses in plant species [4, 5]. Recent genome-wide analysis revealed that ~70% of genes may be subjected to AS in plant species [6, 7]. A number of genes in potato were reported undergoing AS. These genes include an invertase gene [8], BRANCHED1a (BRC1a) gene [9], StPPCK2 (phosphoenolpyruvate carboxylase kinase, PPCK) gene [10], and four metacaspases (SotubMCs) genes [11]. However, there is a lack of systematic collection and analysis of genome-wide AS events in potato plants.

RNA sequencing (RNA-seq) is a technology using the next-generation sequencing (NGS) to sequence and quantify RNA transcripts of a transcriptome [12, 13]. A number of RNA-seq experiments have been reported in potato plants [14-17], however, none of these experiments have carried out AS analysis. Thus, the available genome sequences and

RNA-seq data make it feasible to carry out a comprehensive genome-wide identification and analysis of AS events in potato plants. In this work we combined all available mRNA and expressed sequence tags (ESTs) sequences as well as 227 RNA-seq data generated from 10 published projects, which were carried out with diverse biological treatments, to identify AS events in potato plants [14-23], with an aim to generate a catalog of genes subjecting to AS in potato. The collection of these genes with their respective transcript isoforms may serve as a foundation for potato researchers to further characterize their biological functions.

2. Materials and Methods

2.1. Genome, EST and mRNA Sequences, and RNA-seq Datasets

Potato genome sequences, predicted transcripts and annotation GFF3 files (version 4.03) were downloaded from the Phytozome database (https://phytozome-next.jgi.doe.gov/) [3]. A total of 291,071 transcript sequences including 250,140 EST and 40,931 mRNA transcripts were downloaded from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) nucleotide database by setting limit to RNA within "Solanum tuberosum" organism. RNA-seq data were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra/docs/sradownload/) using SRA Toolkit. We selected 10 recently published RNA-seq projects (Table 1). These projects include samples from tuber and leaf transcriptomes in response to the late blight pathogen *Phytophthora infestans* [14, 15], effects of different nitrogen treatments [21, 22], potato virus X infection [23], virus Y infection [18], *Pectobacterium carotovorum* infection [17], drought stress in different cultivars [19, 20], and different light exposure [16].

Table 1. *Potato RNA-seq data sources.*

Project ID	Dataset	Treatments	References
PRJNA203403	36	Late blight pathogen *Phytophthora infestans*	Gao et al. (2013) [14]
PRJNA283568	19	Virus Y infection	Goyer et al. (2015) [18]
PRJNA301698	30	*Pectobacterium carotovorum* ssp. brasiliense infection	Kwenda et al. (2016) [17]
PRJNA311702	48	Drought response of different cultivars	Sprenger et al. (2016) [19]
PRJNA318049	36	Pathogen *Phytophthora infestans* foliar samples	Gao and Bradeen (2016) [15]
PRJNA476484	10	Drought stress of different cultivars	Moon et al. (2018) [20]
PRJNA565618	6	Nitrogen treatments	Tiwari et al. (2020) [21]
PRJNA529319	6	Nitrogen treatments	Tiwari et al. (2020) [22]
PRJNA544876	24	Biotic stress response induced by light exposure	Zhang et al. (2019) [16]
PRJNA679820	12	Potato virus X infection	Herath and Verchot (2021) [23]

2.2. EST and mRNA Sequence Mapping, RNA-seq Reads Mapping, and AS Identification

The procedure for cleaning EST and mRNA sequences and further assembling these sequences into a non-redundant set of unique transcripts were described in our previous work [7, 24, 25]. The final cleaned transcripts consisting of 290,504 sequences were further assembled into a non-redundant set of 150,435 sequences. The assembled nucleotide sequences were mapped to potato genome sequences using cutoff values of a minimum 95% identity and >75% length coverage using ASFinder and Sim4 programs [26, 27].

The RNA-seq reads were mapped to potato genome sequences using TopHat (v2.2.6) with default parameters [28]. Then the transcript alignment file together with annotation GFF3 were used as input for Cufflinks (v2.2.1) (http://cole-trapnell-lab.github.io/cufflinks/) [29]. The GTF (Gene Transfer Format) files generated from each RNA-seq dataset after Cufflinks were merged using cuffmerge script within the Cufflinks package [29]. The GTF file generated from merged RNA-seq GTF files then was further merged, using Cuffcompare script, with the GTF file of mapping assembled EST/mRNA sequences to the genome that was generated by ASFinder to generate a final GTF file for AS analysis. AStalavista was used for AS event classification [30]. AS events are generally classified as exon skipping (ExonS), alternative donor site (AltD), alternative accepter site (AltA), intron retention (IntronR) and complex event. Complex events are formed by two or more basic events in comparison of a pair of isoforms.

2.3. Functional Annotation of Potato Transcripts

The transcript sequences were retrieved using gtf_to_fasta tool in the Tophat package [28], based on the GTF file generated by Cuffcompare program after merging all mapping GTF files. These transcripts were functionally annotated, including open reading frame (ORF) prediction, protein family (Pfam), and comparison with predicted gene models.

2.4. Identification of Conserved Alternatively Spliced Genes in Four Plant Species

Previously we carried out genome-wide identification of genes undergoing AS in tomato, soybean, and maize plants [7, 25, 36]. To identify conserved alternatively spliced orthologous genes in these four species we retrieved protein sequences of 24,643 from potato, 34,415 from tomato, 30,394 from soybean, and 20,860 from maize plants. The longest ORFs among the isoforms generated by alternatively spliced genes were used for performing reciprocal BLASTP with a cutoff E-value of 1e-10 ($1x10^{-10}$).

2.5. Data Availability

The transcripts and AS events identified in the work are available through Plant Alternative Splicing Database (http://proteomics.ysu.edu/altsplice). This database holds the data from our published work including *Brachypodium distachyon*, pineapple, sacred lotus (*Nelumbo nucifera*), rice, maize, sorghum, tomato, cotton and soybean plants [24, 25, 31-36]. The BLAST server is provided for searching the transcripts generated in our published work. The data used for database construction and the supplementary files mentioned in the work are publicly available at: http://proteomics.ysu.edu/publication/data/Potato/.

3. Result & Analysis

3.1. Mapping RNA-seq Data to the Genome

We mapped 227 RNA-seq datasets, which were collected from 10 published projects, including 85 datasets with single reads and 142 datasets with paired reads, to potato reference genome (Table 2). The accession numbers and detailed mapping information of these RNA-seq data can be found in supplementary Table 1. A total of 7.7 billion reads were collected with 5.8 billion reads (~75.0%) mapped to the genome. Among the mapped reads, ~4.4% reads (0.34 billion) were mapped to two or more genomic loci (Table 2).

AS events were identified from 10 published projects (Table 3). Since each project had different numbers of samples and associated transcriptomes, thus the AS events varied greatly among them. However, among the basic AS events, AltA was the predominant AS type in all the RNA-seq projects, this is different from findings in tomato and other plant species, which is often reported that IntronR is the predominant basic AS type [7, 36-38].

3.2. Mapping EST and mRNA Assembled Transcripts to the Genome

After going through the cleaning procedure including trimming poly (A/T) ends and removing contaminants and repetitive sequences starting with a total of 291,071 EST and mRNA sequences, we obtained 290,541 sequences that were further assembled into a non-redundant set of 150,435 transcripts for mapping to potato genome. A total of 87,992 (58.5%) assembled transcripts were mapped to 11,760 genomic loci and among them 1,976 transcripts (1.3%) were mapped to two or more genomic loci. Mapping assembled EST/mRNA transcripts to the genome identified a total of 6,306 AS events including 1,653 ExonS, 832 AltD, 1,617 AltA, 930 IntronR, and 1,274 complex events (Table 3).

3.3. Merging All Mapping Information for AS Identification

Since our main purpose of this work was to identify all AS events from all data we collected, thus we merged all RNA-seq data with the EST/mRNA mapping data. From the merged dataset a total of 226,769 AS events were obtained and categorized into four types of basic events (111,162, 49.0%) and complex events (115,607, 51.0%). The basic AS events include intron retention (9.5%), alternative acceptor sites (19.2%), alternative donor site (8.2%), and exon skipping (12.1%) (Table 3).

Combining all the mapping data generated a non-redundant set of 295,952 transcripts from 70,772 genomic loci, i. e. genes, including 46,122 genes with only one transcript and 24,650 genes with two or more transcripts (Table 3). These genes have two or more transcripts undergoing AS with an average of 10 transcripts. Thus, the estimated AS rate was 34.8% (Table 3). However, if we only used the annotated gene models downloaded from Phytozome, among a total of 42,483 gene models, 19,450 (45.8%) of them were alternatively spliced. We noticed 84 alternatively spliced genes each generated 100 or more isoforms, for example, 294 (the highest number) AS isoforms were generated by PGSC0003DMG400028454. It should be noted that these isoforms were assembled by Cufflinks, whether these isoforms are real or not need to be validated by cloning the full-length of mRNA transcripts. In the annotated GFF3 file (version 4.03), a total of 39,028 genes with 56,215 transcripts were predicted in the reference genome. Among the predicted genes, there are 29,153 genes have only one transcript each gene and 9,875 genes have two to 18 isoforms, that is, these genes are subject to AS, thus, the AS rate is estimated 25.3%. In addition, a total of 29,446 new genomic loci with a total of 94,866 transcripts were identified, they were likely real expressed genes as they were supported by mapped RNA-seq reads and EST/mRNA sequences in this work.

3.4. Functional Annotation of Transcripts

A total of 295,952 RNA transcript sequences were retrieved and further annotated, including ORF prediction, coding region completeness assessment, functional annotation based on BLASTX and protein family (Pfam) prediction. These basic features of these transcripts were summarized in Table 4. The transcripts have an average length of 1,565 bp and 70% have a BLASTX match. A total of 256,945 (86.8%) ORFs with an average length of 332 amino acids were predicted. In addition, using BLASTN search with a cutoff of 95% identity and a minimum length of 80 bp in alignment 207,694 (70.2%) transcripts were matched with predicted transcripts of gene models.

The predicted proteins from retrieved mRNA transcripts were annotated to Pfam, that facilitates database search and examination of the functional domains in proteins encoded by different transcript isoforms. Since the isoforms in alternatively spliced genes may encode a truncated protein, thus with a domain loss, due to a pre-mature stop codon or may not be able to translate to a protein due to a translation frame shift. A total of 154,753 (60.2%) out of 256,945 predicted ORFs matched to Pfam (Table 5). We compared the Pfam of protein sequences encoded by the isoform pairs of all identified AS events. Among a total of 226,769 pairs, 70,842 (31.2%) pairs had no Pfam hit, 110,510 (48.7%) pairs

had identical Pfam, 34,044 (15.0%) pairs had one isoform with Pfam but the other isoform not having a Pfam hit, suggesting the transcript could not be translated or the domain was lost, and 11,373 (5.0%) pairs had different Pfam, suggesting the functional domains were changed in the isoforms (Supplementary Table 2).

To compare the AS rates of genes encoding different protein families we extracted only one Pfam annotation for genes having multiple isoforms. Among a total of 24,800 genes which were matched to protein families, 16,798 (67.7%) genes were alternatively spliced, though different gene families had variable AS rates (Table 5; Supplementary Table 3). These gene families were well conserved in tomato, pineapple, soybean and cereal plants [7, 24, 25, 31]. The varying AS rates among different gene families are expected as different genes had different exon and intron numbers and structures. However, the observed much higher AS rates in these genes having Pfam annotation indicate AS in these genes play important roles in regulation of various types of cellular processes.

Table 2. *Summary of mapping rates of RNA-seq data generated in different projects.*

		Total reads	Total mapped	M. A.	Mapping rate (%)	M. A. (%)
PRJNA203403	paired	921149968	777587186	62810721	84.4	6.8
PRJNA283568	single	563198665	492413277	29519550	87.4	5.2
PRJNA301698	paired	1375032248	1098344225	41892005	79.9	3.0
PRJNA311702	single	687593470	623243410	41622626	90.6	6.1
PRJNA318049	paired	1030833894	547774342	39040236	53.1	3.8
PRJNA476484	paired	612616770	485996341	23287500	79.3	3.8
PRJNA529319	single	178768280	84186669	13199193	47.1	7.4
PRJNA544876	paired	975556406	668834427	29379255	68.6	3.0
PRJNA565618	paired	470382212	402720148	30622971	85.6	6.5
PRJNA679820	single	906778800	607062704	27683452	66.9	3.1
Total		7721910713	5788162729	339057509	75.0	4.4

M. A.: reads mapped to two or more loci.

Table 3. *Categories of alternative splicing events identified from different RNA-seq datasets and EST/mRNA data.*

Project	ExonS	AltD	AltA	IntronR	Complex	Total
PRJNA203403	1869	1578	3547	1292	4022	12308
PRJNA283568	6066	4918	11500	5693	12610	40787
PRJNA301698	3133	3040	7227	2944	8456	24800
PRJNA311702	6359	4606	11348	4098	13171	39582
PRJNA318049	1563	1368	3091	884	3340	10246
PRJNA476484	4903	3276	8045	2399	6459	25082
PRJNA529319	2092	1626	3957	1129	2779	11583
PRJNA544876	3783	3797	10167	3067	9574	30388
PRJNA679820	11153	6521	17987	5710	16872	58243
PRJNA565618	1959	1955	4138	973	2368	11393
RNA-seq merged	26500	18042	42656	20637	110584	218419
EST/mRNA data	1653	832	1617	930	1274	6306
All data merged	27435	18658	43573	21496	115607	226769

Table 4. *Basic features of assembled RNA transcripts and functional annotation in potato plants.*

Total genomic loci	70772
Non-AS loci	46122 (65.2%)
AS-loci	24650 (34.8%)
New genomic loci	29446
Total transcripts	295952
Transcripts match to gene models	207694 (70.2%)
Average length of transcripts (bp)	1565
BLASTX match against Swiss-Prot database	207302 (70.0%)
Total predicted ORFs	256945 (86.8%)
Average length of ORFs (aa)	332
Total Pfam matches of ORFs	154753 (60.2%)

3.5. Comparison of Alternatively Spliced Genes in Potato, Tomato, Soybean and Maize Plants

A total of 14,110 orthologous AS genes were conserved between potato and tomato plants in the current analysis and 6,620 AS genes were conserved among potato, tomato and soybean plants, which are all dicot plants (Figure 1). We also identified a total of 2,929 AS genes conserved among above three dicot plants with maize, a monocot plant (Figure 1; Supplementary Table 4). The higher ratio of conserved AS genes in potato and tomato is expected as these two species belongs to the same genus. Conserved AS events across evolutionarily related species may indicate these events play important functional roles [38]. These AS genes conserved among these four species may warrant further detailed examination of their biological functions as monocot-dicot species diverged ~200 million years ago [39].

3.6. Dynamic Nature of AS: Comparative Analysis of AS in Different Cultivars of Potato Plants in Responses to Drought Stresses

In this work we collected RNA-seq data from 10 published projects (Table 1). Since AS plays an important role in defining the diversity of transcriptomes, it is valuable to

perform more detailed comparative analysis of AS events in each project. In this project we selected the datasets generated by Gao and Bradeen (2016) [15] and Sprenger et al. (2016) [19] for further analysis to demonstrate the transcriptome complexities of different treatments. The results from the dataset of Gao and Bradeen (2016) were reported recently [40]. Here we include some results obtained by Ogungbayi (2022) [41] using the dataset of Springer et al. (2016) [19] to illustrate the dynamic nature of AS events identified in four potato cultivars in response to drought stresses. The RNA-seq data were generated from four cultivars including Milva, Alegria, Desiree and Saturna with an ascending order in drought tolerance [19]. Plants were drought stressed in field trials and in greenhouse experiments [19]. Ogungbayi (2022) [41] combined the genome mapping information of RNA-seq data obtained from all controls or drought stressed treatments for AS identification. It is clearly demonstrated that the profile shift of AS events from controls to drought treatments in all four cultivars (Table 6). The further combined data of both controls and drought treatments illustrate the conserved AS events and cultivar specific AS events of the four cultivars (Figure 2). These cultivar specific AS events likely play some important roles in determining the different drought resistances in different cultivars. As the original article already investigated the differentially expressed genes and carried out functional enrichment analysis of drought responsive genes in different cultivars [19], it is not needed for this work to carry out this type of analysis. However, for example, among those 102 "robustly drought-responsive genes" identified by Sprenger et al. (2016) [19], we found 75 of them were alternatively spliced, thus these AS isoforms may play different regulatory roles (Supplementary Table 5).

4. Discussion

We combined mapping information of a large amount of RNA-seq data with ESTs and mRNAs mapping information and performed a comprehensive identification and functional analysis of genome-wide AS events in potato. To our knowledge this is the first large scale genome-wide analysis of AS events in potato using RNA-seq data collected from multiple projects. The AS rate was estimated ~45.8% among all annotated genes, which was increased from 25.3% in current genome annotation. However, the AS rates have been reported ~70% in Arabidopsis, 55% in maize, and 65.0% in tomato previously [7, 36, 37]. Thus, we expect that more AS events and a higher AS rate can be obtained when more RNA-seq or transcripts data are used for genome mapping in potato. As we have made our data publicly available, future work can be expanded based upon our work as well as recently published RNA-seq data, particularly, long reads generated by PacBio sequencing technology [42, 43].

Plant transcriptomes vary greatly depending on the environmental factors and developmental stages. Our analysis demonstrates that there are dynamic changes of AS events between different cultivars within a species as well as in responses to different biotic and abiotic stresses (Table 6; Figure 2) [40]. However, how the AS information as well as RNA-seq data can be applied to plant molecular breeding remains to be explored. For the RNA-seq data we collected from these published projects, all original analysis focused on the differential gene expression. Since plant genes are subjected to AS with a high rate, we suggest AS analysis be an essential part of transcriptome analysis in addition to differential gene analysis. It should be noted that these isoforms were assembled by Cufflinks, validation by RT-PCR or cloning the full-length of mRNA transcripts are needed for further detailed functional analysis. Computational identification of genes undergoing AS and annotation of their associated transcript isoform sequences are useful for researchers to design more specific experiments to examine the functions of genes of interests. The data generated in this work are publicly available on our server. The current work provides an important resource for investigating alternatively spliced genes and their associated functional regulations in potato plants.

Table 5. *Genes encoding different protein families having variable alternative splicing rates in potato plants.*

Identifier	Total	AS gene	AS (%)	Pfam	Pfam description
pfam00069	693	513	74.0	Pkinase	Protein kinase domain
pfam00067	422	235	55.7	p450	Cytochrome P450
pfam07727	388	77	19.8	RVT_2	Reverse transcriptase (RNA-dependent DNA)
pfam07714	381	310	81.4	Pkinase_Tyr	Protein tyrosine kinase
pfam13041	356	253	71.1	PPR_2	PPR repeat family
pfam00931	308	218	70.8	NB-ARC	NB-ARC domain
pfam13639	259	144	55.6	zf-RING_2	Ring finger domain
pfam00078	237	45	19.0	RVT_1	Reverse transcriptase (RNA-dependent DNA
pfam14223	236	38	16.1	Retrotran_gag_2	gag-polypeptide of LTR copia-type
pfam13966	235	51	21.7	zf-RVT	zinc-binding in reverse transcriptase
pfam00201	206	120	58.3	UDPGT	UDP-glucoronosyl and UDP-glucosyl transferase
pfam00665	202	38	18.8	rve	Integrase core domain
pfam00076	185	165	89.2	RRM_1	RNA recognition motif
pfam00249	183	117	63.9	Myb_DNA-binding	Myb-like DNA-binding domain
pfam00847	162	73	45.1	AP2	AP2 domain
pfam10551	157	65	41.4	MULE	MULE transposase domain
pfam13456	157	40	25.5	RVT_3	Reverse transcriptase-like
pfam00319	147	36	24.5	SRF-TF	SRF-type transcription factor (DNA-binding)

Identifier	Total	AS gene	AS (%)	Pfam	Pfam description
pfam02458	139	42	30.2	Transferase	Transferase family
pfam14111	137	35	25.5	DUF4283	Domain of unknown function (DUF4283)
pfam14244	136	30	22.1	Retrotran_gag_3	gag-polypeptide of LTR copia-type
pfam02519	133	38	28.6	Auxin_inducible	Auxin responsive protein
pfam03171	126	89	70.6	2OG-FeII_Oxy	2OG-Fe(II) oxygenase superfamily
pfam00141	124	74	59.7	peroxidase	Peroxidase
pfam03106	95	64	67.4	WRKY	WRKY DNA -binding domain
pfam00010	89	70	78.7	HLH	Helix-loop-helix DNA-binding domain
pfam00083	88	51	58.0	Sugar_tr	Sugar (and other) transporter
pfam02992	87	27	31.0	Transposase_21	Transposase family tnp2
pfam14432	87	56	64.4	DYW_deaminase	DYW family of nucleic acid deaminases
pfam00481	84	70	83.3	PP2C	Protein phosphatase 2C
pfam03004	84	66	78.6	Transposase_24	Plant transposase (Ptta/En/Spm family)
pfam14291	83	39	47.0	DUF4371	Domain of unknown function (DUF4371)
pfam00561	80	67	83.8	Abhydrolase_1	alpha/beta hydrolase fold
Total	24750	16753	67.7		

This is only a partial list of Pfam. The complete list is provided as a Supplementary file.

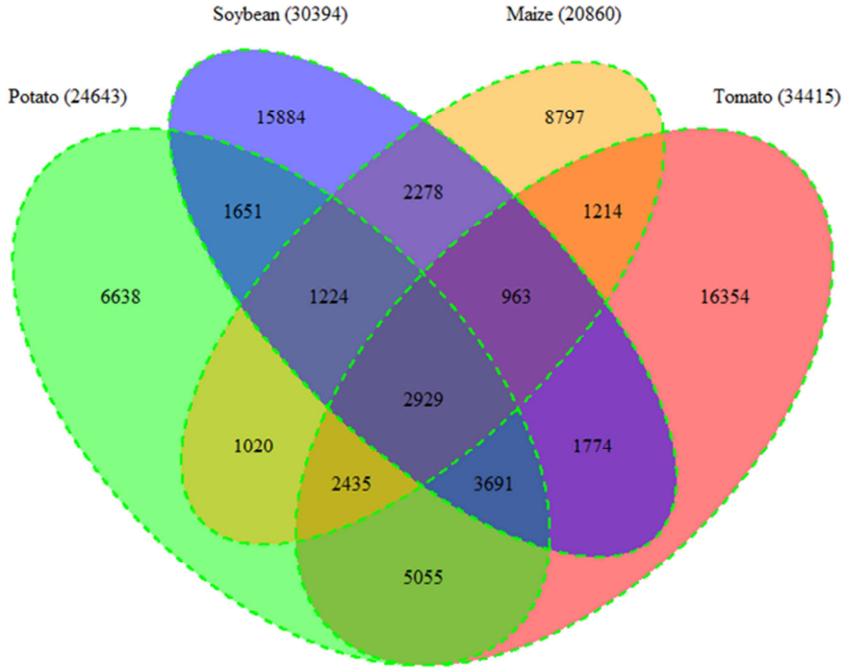

Figure 1. *Conserved alternatively spliced genes among potato, tomato, soybean and maize plants. The sequence data of alternatively spliced genes of these four plants can be found in the Plant Alternative Splicing Database.*

Table 6. *Alternative splicing events in four cultivars of potato plants under control and drought treatments.*

		ExonS	AltD	AltA	IntronR	Complex	Total
Cultivar: Alegria							
	Control	1579	1525	3497	1754	2326	10681
	Drought	1704	1576	3794	1878	2369	11321
	Conserved	879	783	1801	921	1166	5550
Cultivar: Desiree							
	Control	1579	1494	3547	1632	2259	10511
	Drought	1706	1593	3665	1847	2388	11199
	Conserved	897	761	1782	897	1175	5512
Cultivar: Milva							
	Control	1633	1501	3411	1670	2238	10453
	Drought	1699	1513	3706	1835	2448	11201
	Conserved	908	737	1776	873	1203	5497
Cultivar: Saturna							
	Control	1599	1531	3543	1695	2278	10646
	Drought	1718	1648	3813	1957	2403	11539
	Conserved	868	756	1801	914	1152	5491

Note: The raw RNA-seq data were generated by Springer et al. (2016) [19].

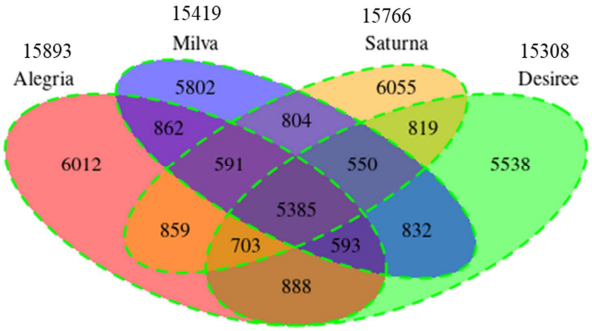

Figure 2. *Alternative splicing events identified in four cultivars of potato plants from combined transcriptome data of control and drought treatments.*

5. Conclusion

The work represents the first systematic collection of alternatively spliced genes and isoforms in potato plants. Combining RNA-seq mapping data generated from multiple projects and all mRNA and EST sequences of potato plants in NCBI database allowed us to identify a total of 226,769 AS events, which were produced from a total of 24,650 genes. Comparative analysis identified 2,929 alternatively splice genes conserved among potato, tomato, soybean and maize plants. The work provides an important resource for further functional characterization of these genes in potato biology. Thus, we recommend that researchers in potato biology use the resource and incorporate the AS information and newly identified genomic loci in their work in potato breeding and genomic applications.

Author Contributions

Conceptualization, X. M; Methodology, X. M and F. Y.; Software, X. M. and F. Y.; Formal Analysis, X. M., A. O., and F. Y.; Investigation, A. O., J. L., V. V., and X. M.; Resources, X. M. and F. Y.; Data Curation, X. M. and F. Y.; Writing – original draft, X. M.; Writing – review and editing, X. M. and F. Y.; all authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgements

The work was supported by Yongstown State University (YSU) Research Professorship award and reassigned time for scholarship from College of Science, Technology, Engineering, and Mathematics at YSU to XM. The Ohio Supercomputer Center provided computational resources for data processing.

References

[1] Wijesinha-Bettoni, R.; Mouillé, B. (2019). The contribution of potatoes to global food security, nutrition and healthy diets. Am. J. Potato Res. 2019, 96, 139-149.

[2] Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. Nature. 2011, 475, 189-195. https://doi:10.1038/nature10158

[3] Sharma, S. K.; Bolser, D.; de Boer, J.; Sønderkær, M.; Amoros, W.; Carboni, M. F.; D'Ambrosio, J. M.; de la Cruz, G.; Di Genova, A.; Douches, D. S.; Eguiluz, M. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 (Bethesda, Md.), 2013, 3 (11): 2031-2047. https://doi.org/10.1534/g3.113.007153

[4] Staiger, D.; Brown, J. W. Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell 2013, 25, 3640-3656. https://doi.org/10.1105/tpc.113.113803

[5] Reddy, A. S.; Marquez, Y.; Kalyna, M.; Bartab, A. Complexity of the alternative splicing landscape in plants, Plant Cell 2013, 25, 3657-3683. https://doi.org/10.1105/tpc.113.117523

[6] Chaudhary, S.; Khokhar, W.; Jabre, I.; Reddy, A. S.; Byrne, L. J.; Wilson, C. M.; Syed, N. H. Alternative splicing and protein diversity: plants versus animals. Front. Plant Sci. 2019, 10, 708. https://doi.org/10.3389/fpls.2019.00708

[7] Clark, S.; Yu F.; Gu, L.; Min, X. Expanding alternative splicing identification by integrating multiple sources of transcription data in tomato. Front. Plant Sci. 2019, 10, 689. https://doi:10.3389/fpls.2019.00689

[8] Bournay, A.-S.; Hedley, P. E.; Maddison, A.; Waugh, R.; Machray, G. C. Exon skipping induced by cold stress in a potato invertase gene transcript. Nucleic Acids Res. 1996, 24, 2347–2351. https://doi: 10.1093/nar/24.12.2347

[9] Nicolas, M.; Rodríguez-Buey, M. L.; Franco-Zorrilla, J. M.; Cubas, P. A recently evolved alternative splice site in the BRANCHED1a gene controls potato plant architecture. Curr. Biol. 2015, 25 (14), 1799-809. https://doi.org/10.1016/j.cub.2015.05.053

[10] Marsh, J. T.; Sullivan, S.; Hartwell, J.; Nimmo, H. G. Structure and expression of phospho enol pyruvate carboxylase kinase genes in Solanaceae. A novel gene exhibits alternative splicing. Plant Physiol. 2003, 133, 2021-2028. https://doi.org/10.1104/pp.103.030775

[11] Dubey, N.; Trivedi, M.; Varsani, S.; Vyas, V.; Farsodia, M.; Singh, S. K. Genome-wide characterization, molecular evolution and expression profiling of the metacaspases in potato (*Solanum tuberosum* L.). Heliyon. 2019, 5, e01162. https://doi.org/10.1016/j.heliyon.2019.e01162

[12] Wang, Z.; Gerstein, M.; Snyder, M. RNA-seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 2009, 10, 57-63. https://doi.org/10.1038/nrg2484

[13] Chu, Y.; Corey, D. R. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther. 2012, 22, 271-274. https://doi.org/10.1089/nat.2012.0367

[14] Gao, L.; Tu, Z. J.; Millett, B. P.; Bradeen, J. M. Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-Phytophthora infestans interactions. BMC Genomics. 2013, 14, 340. https://doi.org/10.1186/1471-2164-14-340

[15] Gao, L.; Bradeen, J. M. Contrasting potato foliage and tuber defense mechanisms against the late blight pathogen *Phytophthora infestans*. PLoS One. 2016, 11, e0159969. https://doi.org/10.1371/journal.pone.0159969

[16] Zhang, W.; Zuo, C.; Chen, Z.; Kang, Y.; Qin, S. RNA sequencing reveals that both abiotic and biotic stress-responsive genes are induced during expression of steroidal glycoalkaloid in potato tuber subjected to light exposure. Genes. 2019, 10, 920. http://dx.doi.org/10.3390/genes10110920

[17] Kwenda, S.; Motlolometsi, T. V.; Birch, P. R.; Moleleki, L. N. RNA-seq profiling reveals defense responses in a tolerant potato cultivar to stem infection by *Pectobacterium carotovorum* ssp. brasiliense. Front. Plant Sci. 2016, 7, 1905. https://doi.org/10.3389/fpls.2016.01905

[18] Goyer, A.; Hamlin, L.; Crosslin, J. M.; Buchanan, A.; Chang, J. H. RNA-Seq analysis of resistant and susceptible potato varieties during the early stages of potato virus Y infection. BMC Genomics. 2015, 16, 1-3. https://doi.org/10.1186/s12864-015-1666-2

[19] Sprenger, H.; Kurowsky, C.; Horn, R.; Erban, A.; Seddig, S.; Rudack, K.; Fischer, A.; Walther, D.; Zuther, E.; Köhl, K.; Hincha, D. K.; J. Kopka. The drought response of potato reference cultivars with contrasting tolerance. Plant Cell Environ. 2016, 39, 2370-2389. https://doi.org/10.1111/pce.12780

[20] Moon, K. B.; Ahn, D. J.; Park, J. S.; Jung, W. Y.; Cho, H. S.; Kim, H. R.; Jeon, J. H.; Park, Y. I.; Kim, H. S. Transcriptome profiling and characterization of drought-tolerant potato plant (*Solanum tuberosum* L.). Mol. Cells. 2018, 41, 979. http://dx.doi.org/10.14348/molcells.2018.0312

[21] Tiwari J. K.; Buckseth, T.; Devi, S.; Varshney, S.; Sahu, S.; Patil, V. U.; Zinta, R.; Ali, N.; Moudgil, V.; Singh, R. K.; Rawat, S. Physiological and genome-wide RNA-sequencing analyses identify candidate genes in a nitrogen-use efficient potato cv. Kufri Gaurav. Plant Physiol. Biochem. 2020, 154, 171-183. https://doi.org/10.1016/j.plaphy.2020.05.041

[22] Tiwari, J. K.; Buckseth, T.; Zinta, R.; Saraswati, A.; Singh, R. K.; Rawat, S.; Dua, V. K.; Chakrabarti, S. K. Transcriptome analysis of potato shoots, roots and stolons under nitrogen stress. Sci. Rep. 2020, 10, 1-8. https://doi.org/10.1038/s41598-020-58167-4

[23] Herath, V.; Verchot, J. Transcriptional regulatory networks associate with early stages of potato virus X infection of *Solanum tuberosum*. Int. J. Mol. Sci. 2021, 22, 2837. https://doi.org/10.3390/ijms22062837

[24] Wai, C. M.; Powell, B.; Ming, R.; Min, X. J. Analysis of alternative splicing landscape in pineapple (*Ananas comosus*). Tropical Plant Biol. 2016, 9, 150-160. https://doi.org/10.1007/s12042-016-9168-1

[25] Min, X.; Kasamias, T.; Wagner, M.; Ogungbayi, A.; Yu, F. Identification and analysis of alternative splicing in soybean plants. Proceedings of 14th Int. Conf. on Bioinformatics and Computational Biol. 2022, 83, 1-9. https://doi.org/10.29007/2qv6

[26] Min, X. J. ASFinder: a tool for genome-wide identification of alternatively spliced transcripts from EST-derived sequences. Int. J. Bioinform. Res. Appl. 2013, 9, 221-226. https://doi.org/10.1504/IJBRA.2013.053603

[27] Florea, L.; Hartzell, G.; Zhang, Z.; Rubin, G. M.; Miller, W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Res. 1998, 8, 967-974. https://doi.org/10.1101/gr.8.9.967

[28] Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013, 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36

[29] Trapnell, C.; Williams, B. A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M. J.; Salzberg, S. L.; Wold, B. J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 2010, 28, 511-515. https://doi.org/10.1038/nbt.1621

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

[31] Min, X. J.; Powell B.; Braessler J.; Meinken, J.; Yu, F.; Sablok, G. Genome-wide cataloging and analysis of alternatively spliced genes in cereal crops. BMC Genomics. 2015, 16, 721. https://doi.org/10.1186/s12864-015-1914-5

[32] Walters, B., Lum, G., Sablok, G., Min, X. J. Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. DNA Res. 2013, 20, 163-171. https://doi.org/10.1093/dnares/dss041

[33] VanBuren, R.; Walters, B.; Ming, R.; Min, X. J. Analysis of expressed sequence tags and alternative splicing genes in sacred lotus (*Nelumbo nucifera* Gaertn.). Plant Omics J. 2013, 6, 311-317.

[34] Sablok, G.; Powell, B.; Braessler, J.; Yu, F.; Min, X. J. Comparative landscape of alternative splicing in fruit plants. Current Plant Biol. 2017, 9, 29-36. https://doi.org/10.1016/j.cpb.2017.06.001.

[35] Min, X. J. A survey of alternative splicing in allotetraploid cotton (*Gossypium hirsutum* L.). Comp. Mol. Biol. 2018, 8, 1-13. https://doi.org/10.5376/cmb.2018.08.0001

[36] Min, X. J. Comprehensive cataloging and analysis of alternative splicing in maize. Comp. Mol. Biol. 2017, 7, 1-11. https://doi.org/10.5376/cmb.2017.07.0001

[37] Mei, W.; Boatwright, L.; Feng, G.; Schnable, J. C.; Barbazuk, W. B. Evolutionarily conserved alternative splicing across monocots. Genetics. 2017, 207, 465-480. https://doi.org/10.1534/genetics.117.300189

[38] Marquez, Y.; Brown, J. W.; Simpson, C.; Barta, A.; Kalyna, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res. 2012, 22, 1184-1195. https://doi.org/10.1101/gr.134106.111

[39] Wolfe, K. H.; Gouy, M.; Yang, Y. W.; Sharp, P. M.; Li, W. H. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. U. S. A. 1989, 86, 6201-6205.

[40] Lee, J.; Min, X. (2023). Comparative analysis of alternative splicing events in foliar transcriptomes of potato plants inoculated with *Phytophthora infestans*. Comput. Mol. Biol. 2023, 13, 1-8. https://doi.org/10.5376/cmb.2023.13.0001

[41] Ogungbayi, A. Analysis of alternative splicing events in the transcriptome of potato plants. Youngstown State University, Master's thesis. 2023, OhioLINK Electronic Theses and Dissertations Center, http://rave.ohiolink.edu/etdc/view?acc_num=ysu1651817160842415.

[42] Yan, C.; Zhang, N.; Wang, Q;, Fu, Y.; Zhao, H.; Wang, J.; Wu, G.; Wang, F.; Li, X; Liao, H. Full-length transcriptome sequencing reveals the molecular mechanism of potato seedlings responding to low-temperature. BMC Plant Biology. 2022, 22, 1-20. https://doi.org/10.1186/s12870-022-03461-8

[43] Petek, M.; Zagorščak, M.; Ramšak, Ž.; Sanders, S.; Tomaž, Š.; Tseng, E.; Zouine, M.; Coll, A.; Gruden, K. Cultivar-specific transcriptome and pan-transcriptome reconstruction of tetraploid potato. Scientific Data. 2020, 7, 249. https://doi.org/10.1038/s41597-020-00581-4