The effects of abiotic stress on isoform composition in polyploid *Brassica napus*

by

Ryan Edward Bailey

B.Sc., The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Genome Science and Technology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

The effects of abiotic stress on isoform composition in polyploid *Brassica napus*

---

submitted by Ryan Bailey in partial fulfillment of the requirements for

the degree of Master of Science

in Genome Science and Technology

Examining Committee:

Dr. Keith Adams, Botany, UBC

Supervisor

Dr. Naomi Fast, Botany, UBC

Supervisory Committee Member

Dr. Inanc Birol, Medical Genetics, UBC

Supervisory Committee Member

Abstract

Polyploidy has played a major role throughout the evolution of plants and has long been considered a powerful driver of evolution across a broad range of plant lineages. Polyploidization events have occurred many times during the evolution of flowering plants (angiosperms). Following a polyploidization event, a set of duplicated genes is created which can diverge in function or new functions can evolve. *Brassica napus*, an allopolyploid derived from the hybridization of *B. rapa* and *B. oleracea*, more commonly known as canola, serves as an excellent model to study the complexities between duplicate gene pairs, also known as homeologous pairs. One of these complexities is the process of alternative splicing (AS) by which precursor mRNAs from multiexon genes are spliced to form mature mRNAs producing a vast repertoire of isoforms. The effects of abiotic stress conditions on isoform diversity and variability across homeologous pairs has received little attention.

We conducted a global isoform sequencing analysis of *Brassica napus* using long-read sequencing of plants subjected to heat and cold stress. Analysis of AS events reveal a heat-responsive increase in the number of AS events. Furthermore, we discovered that cold stress reduces the number of isoforms produced by a given gene, whereas heat stress increases the number of isoforms produced by a given gene. This heat-responsive increase in the number of isoforms produced was paired with the observation that heat-stress also induces a higher number of transcripts predicted to be likely targets of nonsense-mediated decay (NMD), a common mechanism by which AS exerts transcriptional regulation. Our analysis also revealed that across homeologous pairs, C_T homeologs are more likely to produce more isoforms relative to A_T homeologs, across all three conditions tested. These shifted isoform distributions do not lead to shifted distributions in the predicted likelihood of NMD-targeting. In all, our analysis reveals opposing shifts in isoform composition in response to cold and heat stress as well as skewed isoform distributions across subgenomes, in which C_T homeologs are more likely to produce more isoforms relative to A_T homeologs, across each of the conditions we tested.

Lay Summary

In this study we investigated how the genome of the canola plant responds to common environmental stressors. Canola is a hybrid plant formed from the merging of two different species of cabbage. This means the canola plant contains two genomes within it, each derived from one of the cabbage species. These are referred to as subgenomes and they each contain sets of genes that are highly similar. We investigated how genes from each subgenome express different versions, called isoforms, through a process known as alternative splicing. We discovered that cold stress triggers genes to produce fewer isoforms and heat stress triggers genes to produce more. Additionally, one of the subgenomes consistently produces more isoforms per gene than the other subgenome across different environmental conditions. These differences, across stress types and subgenomes, have important implications as to how the canola plant is able to respond to environmental stress.

Preface

Investigation into the various aspects of post-transcriptional regulation using long-read sequencing was conceived by Dr. Keith Adams. The design of the project, generation of plant material, RNA extraction, bioinformatic analysis, and data visualization were performed by myself, under the supervision of Dr. Keith Adams. Sequencing library preparation and isoform sequencing of the genetic material was performed at Génome Québec in Montreal, Québec.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A3	Alternative 3’ splice site
A5	Alternative 5’ splice site
ABA	Abscisic acid
AF	Alternative first exon
AL	Alternative last exon
ALTA	Alternative acceptor
ALTD	Alternative donor
ALTP	Alternative position
AS	Alternative splicing
AS-NMD	Alternative splicing-mediated nonsense-mediated decay
BP	Biological process
CC	Cellular compartment
CCS	Circular consensus sequence
COR	Cold regulated
ES	Exon skipping
FLNC	Full-length non-chimeric
FSM	Full splice match
GO	Gene ontology
GTF	Gene transfer format
IGV	Integrated Genome Viewer
IQR	Interquartile range
IR	Intron retention
ISM	Incomplete splice match
Iso-Seq	Isoform sequencing
MF	Molecular function
MX	Mutually exclusive exon
NIC	Novel in catalogue
NMD	Nonsense-mediated decay
NNC	Novel not in catalogue
ORF	Open reading frame
PTC	Premature termination codon
RI	Retained intron
SE	Skipped exon
SGS	Second-generation sequencing
siRNA	Short-interfering ribonucleic acid
SMRT	Single-molecule real-time
SQANTI2	Structural and quality annotation of novel transcript isoforms 2
SR	Serine/Arginine-rich
TE	Transposable element
ToFu	Transcripts isoforms: full-length and unassembled
WGD	Whole genome duplication

Acknowledgements

I offer my gratitude to Dr. Keith Adams for investing in myself as a trainee while simultaneously investing a substantial amount of resources into the project. I arrived in the Adams Lab with little bioinformatic expertise, yet I was given patience and support along the way as I built up my skills which culminated in this thesis. I am grateful to be given the opportunity to be the first person in the research group to pursue long-read sequencing analysis. I would also like to thank the members of the Adams Lab, past and present, for their participation and helpful critiques of the project as it developed. The high-level of scientific discourse in our lab discussions provided a fertile ground for continued scientific inquiry and high-calibre research.

Additionally, thank you to my committee members Dr. Naomi Fast and Dr. Inanc Birol for their insights. Their support and expertise in their respective fields has strengthened the analysis within thesis.

Furthermore, I extend a generous thanks to my parents for not once questioning my pursuit and passion for genome science. Despite their unfamiliarity with life sciences research and the opportunities it affords, I received unwavering support during this venture. I am proud to represent the second-generation of UBC graduate students in the family, completing my M. Sc. degree over 30 years after my father completed his M. A. Sc., in a building only steps away from the Biodiversity Research Centre.

Lastly, my time at UBC has provided me with numerous blessing, not the least of which are my friends. Their continued passion for adventure and spontaneity provided me endless support and balance, for that I am grateful.

To all those undeterred in their pursuit of science and fact.

Chapter 1: Introduction

1.1 Polyploidy in plants

Polyploidy, or whole genome duplication (WGD), has played a major role throughout the evolution of plants and has long been considered a major driver of evolution across a broad range of plant lineages.[1–3] Polyploidization events have occurred many times during the evolution of flowering plants (angiosperms). Polyploidy occurred at the base of angiosperms and thus all angiosperms have experienced at least one round of polyploidy in their evolutionary history.[4] Many angiosperm lineages have experienced additional polyploidy events during their evolutionary history. For example, analysis of the complete genome sequence of *Arabidopsis thaliana* revealed two recent WGDs (denoted α and β) which occurred within the Brassicaceae lineage and a triplication event (denoted as γ) which is likely to be shared by all eudicots.[5,6] Regarding the role polyploidy plays in the evolution of angiosperm lineages, it has been estimated that roughly 15% of angiosperm speciation events can be linked to polyploidy.[7] Additionally, polyploidization has contributed to the evolution of novel functions such as the evolution of floral structures, increased disease resistance, and has been implicated in the adaptation to various environmental stressors.[8]

Polyploidization leads to a sudden and drastic increase in genetic material and a doubling of the entire gene set. This sudden proliferation of genetic material can lead to a cascade of consequences for the organism, both short and long-term. Newly formed polyploids, termed neopolyploids, may undergo substantial genome reorganization, exchanges between genomes, gene loss, and significant alterations in gene expression.[2,9] This cascade of genomic changes in response to the increase in ploidy is collectively termed genomic "shock." [10,11]

Polyploidy results in a set of duplicated genes which are each contained within the respective parental genome of the polyploid, referred to as subgenomes. These duplicate pairs can undergo a multitude of fates, with some pairs being retained for millions of years.[9] It is these retained pairs which provide the evolutionary material for adaptation via divergence.[9] In allopolyploids, duplicate gene pairs are referred to as homeologous pairs, or homeologs. Several evolutionary fates for these duplicated genes exist, including i) pseudogenization, in which one copy loses function or expression; ii) retention, in which both copies retain the original function or expression pattern; iii) neofunctionalization, in which one copy gains a new function or expression pattern; and iv) subfunctionalization in which the original function of the expression pattern is subdivided between the duplicates.

Gene expression patterns in allopolyploids

Once established, the unique parental history of the progenitor species can lead allopolyploids to often display distinct patterns of gene regulation, gene expression, epigenetic differences, and preferential gene loss across the subgenomes.[12] One important aspect of genomic bias across subgenomes is dominance at the level of gene expression, in which genes from one subgenome show consistently higher levels of expression relative to the other subgenome, this phenomenon is referred to as subgenome dominance. Homeolog expression biases are of particular importance in the study of polyploid evolution, as some plant species have demonstrated widespread transcriptional differences between homeologous gene pairs.[13,14] The impact that these expression differences have on the stress response have been characterized in allopolyploid cotton.[15,16] As with other aspects of hybrid genomes, studies have shown that subgenome dominance can be established early in a allopolyploid's evolutionary history and its magnitude can increase over successive generations.[17] What causes the initial biased expression

patterning has been discussed in numerous studies and is postulated to be linked to higher transposable element (TE) density in a given subgenome, with the effect attributed to higher TE activity interrupting gene expression at a greater rate.[12] Other factors which may drive subgenome dominance are increased levels of DNA methylation in a given subgenome, with gene expression being inversely correlated to DNA methylation, and biases in the relative amounts of small interfering RNAs (siRNAs) present in a given subgenome.[18,19]

Recent transcriptomic studies have revealed patterns of subgenome dominance across several well-studied allopolyploids. For example, in *Gossypium hirsutum*, subgenome dominance has been observed among A-subgenome homeologs in natural allopolyploids and, interestingly, a reversed pattern of dominance among D-subgenome homeologs in resynthesized allopolyploids.[20] Edger et al. (2017) revealed subgenome dominance in *Mimulus peregrinus*, both in natural and resynthesized allopolyploids, as well as in *M. robertsii*, a resynthesized interspecies triploid hybrid, instantly following the hybridization event and demonstrated that these patterns significantly increase over generations.[19] In *Brassica napus*, the organism used in this investigation, an excess of homeologous pairs showing expression bias towards the *B. rapa* (A_T) subgenome has been observed in leaves, whereas the opposite relationship, a bias towards *B. oleracea* (C_T) subgenome, has been observed in roots.[21,22] The observed A_T subgenome bias is also accompanied with the observation that the A_T subgenome contains a greater number of retained genes.[21,23] Interestingly, separate expression analyses of resynthesized allopolyploid lines of *B. napus* have revealed both A_T and C_T subgenome expression biases.[12,24] Seemingly, there is some level of plasticity in the patterns of subgenome dominance in *B. napus* which may be cultivar-specific and tissue-dependent.

Patterns of subgenome dominance have been shown to be affected by abiotic stress possibly providing the mechanistic basis by which polyploids could adapt to environmental stressors. In a study of 30 duplicated gene pairs in *G. hirsutum*, researchers discovered that over 70% of the gene pairs exhibited relative expression changes across the homeologous pairs in response to abiotic stress.[16] Recent analysis of homeologous gene expression in *B. napus* revealed a consistent A_T subgenome bias in response to heat, cold, and drought stress.[22] The newly formed allopolyploid *Glycine dolichocarpa* displayed a higher capacity for non-photochemical quenching, a photoprotective mechanism, compared to its diploid progenitors providing functional evidence for the role of allopolyploidy in an enhanced abiotic stress response.[25]

1.3 Alternative splicing in plants

Alternative splicing (AS) is the process by which precursor messenger-RNAs (pre-mRNAs) from multiexon genes are spliced to form mature messenger RNAs (mRNAs) which produce a vast repertoire of mRNA isoforms. AS is thought to be a major contributor to the transcriptomic and proteomic complexity observed across wide variety of life forms.[26,27] In plants, transcriptome-wide analysis has revealed that AS occurs extensively and it is estimated that over 60% of intron-containing genes undergo AS.[28] Initial evidence for the significance of AS in transcriptional gene regulation in plants was first uncovered by Lopato et al. (1996) in their analysis of the differential expression of Serine/Arginine-rich (SR) protein factors in *A. thaliana*, key regulators of the splicing process, across different organs and during development.[29] Their work uncovered organ-specific regulation of AS in plants.[29] Moreover,

mutant screens have identified splicing factors as key regulators of functional proteins, providing evidence that certain pathways are regulated through differential splicing.[30–32]

As a fundamental aspect of transcriptional regulation, AS is able to alter the number and types of different isoforms, and thus influence the proteomic diversity of the organism.[28] AS events are categorized into five distinct types: i) intron retention (IR) in which an intron is not spliced out of the transcript; ii) exon skipping (ES) in which a given exon is skipped and not spliced into the transcript; iii) alternative donor (ALTD) in which an alternative 5' splice site is used; iv) alternative acceptor (ALTA) in which an alternative 3' splice site is used; and v) alternative position (ALTP) in which both alternative 5' and 3' spice junctions are used.[28] Analysis of AS in several plant species has revealed that IR is the most predominant type of event with ES being the least common, in metazoans the trend is reversed with ES representing the most abundant category and IR representing the least.[28,33]

The putative mechanism by which AS exerts transcriptional regulation in plants is thought to be a result of nonsense-mediated-decay (NMD).[28] NMD can modify the relative expression dosage of a target gene if a given isoform undergoes an AS event which introduces a premature termination codon (PTC).[34] This process targets both aberrant transcripts and a broad array of mRNA isoforms setting NMD apart as a regulator capable of fine and coarse adjustments of RNA levels within the cell by which NMD can remove improperly transcribed isoforms generated by RNA polymerase errors resulting in an in-frame PTC, or via AS in which specific transcripts are downregulated via NMD through the inclusion of a PTC.[35] Work in *A. thaliana* has revealed that NMD plays a role in regulating expression levels of various genes including important regulatory genes involved in plant development.[36]

In the context of polyploidy, AS patterns have been shown to be divergent across subgenomes. Studies of *B. napus* have revealed that many homeologous pairs exhibit divergent AS patterns.[22,37] In their global transcriptome analysis of 27,360 homeologous pairs in *B. napus*, Lee & Adams (2020) revealed a C_T-subgenome bias in the extent of AS under heat, cold, and drought stress.[22] In a recent long-read sequencing analysis of *G. barbadense* it was estimated that 51.4% of homeologous genes produce divergent isoforms.[38] Furthermore, a study of hexaploid wheat found that homeologous genes exhibited differential AS responses under drought and heat stress, with homeologs belonging to the B subgenome exhibiting higher AS activity relative to the A and D subgenomes.[39] Functional analysis of these biased homeologs showed enrichment for abiotic stress-responsive pathways further inferring the role that AS may play in the transcriptional response or adaption to abiotic stress.[39]

1.4 Isoform sequencing of polyploid transcriptomes

Single-molecule real-time sequencing (SMRT), originally developed by Pacific Biosciences (PacBio), in which sequences are read during the replication of the target DNA, or complementary DNA (cDNA), has ushered in a new generation of sequencing, otherwise referred to as long-read sequencing or third-generation sequencing.[40,41] These new technologies are capable of generating read lengths that far exceed those of second-generation sequencing (SGS) platforms such as Illumina.[41,42] Despite its significant contribution to our understanding of plant transcriptomes, its reliance on short reads (100-150 bp) makes isoform reconstruction and thus isoform identification, the process by which the short reads are reassembled back into their putative transcript models, quite difficult and error-prone.[43–45] Particularly, AS events can introduce considerable ambiguities during this process as it can generate both divergent and/or

partially redundant isoforms at a given gene locus.[46] This makes accurate detection of splicing isoforms quite difficult when using a SGS approach due to the inability of software packages to resolve these ambiguities in order to determine the actual combinations of splice-site usage. Long-read sequencing, however, removes these ambiguities by capturing entire transcripts in a single read and thus allowing researchers to obtain splicing isoforms directly without assembly.[47] This has been demonstrated in a study of maize in which researchers found that short-read assemblers such as Cufflinks and Trinity were only able to reconstruct small percentages (22% and 8%, respectively) of the isoforms that were discovered using PacBio long-read sequencing.[48]

The direct detection of isoforms using long-read sequencing is referred to as isoform sequencing (Iso-Seq). Iso-Seq has now enabled researchers to more accurately assess the role that AS plays in transcriptional regulation by offering a clear view of how individual AS events manifest as distinct isoforms. Iso-Seq has been used in numerous recent analyses to assess the role that AS plays in isoform diversity. In their analysis of maize (*Z. mays*) Wang et al. (2016) showed that over 50% of transcripts produced novel or tissue-specific isoforms.[48] In allopolyploid cotton (*G. barbadense*) researchers discovered that over 50% of homeologous gene pairs produce divergent isoforms in each subgenome, furthering our understanding and providing insights into the complexity of AS in polyploid species.[38] The application of Iso-Seq to uncover the complexities of AS in response to abiotic stress has not been extensively examined. In *P. trichocarpa*, Iso-Seq analysis revealed that abiotic stress can induce broad changes in isoform profiles including the regulation of certain IR events in a stress-specific manner.[49] However, more work needs to be done to investigate the role in which AS plays in stress-induced splicing and thus post-transcriptional regulation. Moreover, how these trends change in response to

polyploidy and the degree to which patterns of subgenome dominance affect stress-dependent splice regulation has been largely unexplored.

There are numerous possible ways, some of which have been previously discussed, in which AS can exert changes in isoform diversity in the context of polyploidy and abiotic stress. Firstly, the relative number of isoforms produced by a gene can vary in response to stress. Secondly, the relative number of isoforms that are produced by each member of a homeologous gene pair may be divergent in some pairs and the relative number of isoforms each homeolog contributes to the transcriptome may change in response to abiotic stress. Thirdly, isoform structure(s) may diverge in response to stress or relative to the other duplicate gene in the homeologous pair. These possibilities will be explored in this analysis.

1.5 The model organism *Brassica napus*

B. napus (Brassicaceae) (AACC; oilseed rape) is a recently formed allopolyploid. The hybridization between its progenitors *B. oleracea* (CC: Mediterranean cabbage) and *B. rapa* (AA; Asian cabbage) occurred approximately 7,500 years ago.[21,50] It represents a suitable model to study aspects of AS and its effects on isoform diversity due to the availability of a high quality reference genome and annotation, which includes the identification and annotation of homeologous gene pairs.[21]

1.6 Research goals

The goal of this thesis is to investigate the impact that AS has on isoform diversity in the allopolyploid *B. napus,* through the use of recently developed long-read sequencing technology. Specifically, I aimed to determine the degree to which differences in AS patterns manifest

themselves as differences at the isoform level, in the context of polyploidy and in response abiotic stress. The overall goal of this thesis can be broken down into three distinct research goals:

1. To assess the total number of isoforms per gene across conditions and subgenomes.
2. To determine how isoform diversity responds to abiotic stress.
3. To assess the relative isoform diversity across homeologous gene pairs.

Chapter 2: Materials & Methods

2.1 Plant material and abiotic stress treatments

Plant material was prepared for the abiotic stress treatments and subsequent isoform sequencing following the protocol of Lee & Adams (2020).[22] 450 seedlings of *B. napus* (Sentry Summer Rape) were grown in growth chambers with 16/8 hour photoperiods at 22°C, relative humidity 50%, and a light intensity of 230-240 μmol m^{-2} sec^{-1}.[51] After 4 weeks of growth 15 healthy seedlings were selected for each condition (control, cold, and heat). Seedlings in the cold treatment group were subjected to 4°C for 24 hours, seedlings in the hot treatment were subjected to 35°C for 24 hours, and control seedlings remained at 22°C for 24 hours. All other conditions were held constant. For each biological replicate the first true leaves were pooled from three individuals and three biological replicates were prepared for each condition. The selected leaflets were flash frozen in liquid nitrogen and stored at -80°C until the RNA extraction. Total RNA was extracted using the Ambion RNAaqueous kit (AM1912) and possible DNA contamination was removed using the Ambion Turbo DNA-free kit (AM1907). The quantity and quality of the whole RNA was assessed using a NanoDrop ND-1000 UV-Vis Spectrophotometer and through the inspection of the 23s and 16s ribosomal RNA bands using agarose gel electrophoresis.[52] Preparation of cDNA (Clontech SMARTer PCR cDNA synthesis kit) and isoform sequencing via Pacific Biosciences Sequel platform were performed at Génome Québec (Québec, Canada).

2.2 Isoform structure analysis

Raw sequence data was processed into circular consensus sequences (CCSs) and full-length reads were detected and sorted using Pacific Biosciences' IsoSeq3 pipeline

(https://github.com/PacificBiosciences/IsoSeq). Full-length non-chimeric (FLNC) reads were then pooled from each of the biological replicates within each stress condition. FLNC reads were mapped to the *B. napus* reference genome using minimap2 with the parameters -ax splice -t 30 -uf --secondary=no -C5.[21,53] Identical mapped isoforms were then collapsed to obtain a final set of unique full-length isoforms using the python script collapse_isoforms_by_sam.py from the Cupcake-ToFu (Transcripts Isoforms: Full-length and Unassembled) python library.[54,55]

Filtering of the dataset was completed in order to remove genes which produced spuriously high numbers of isoforms per gene. Such genes were reported as having isoform counts that lie considerably outside of the known distribution of isoforms per gene based on the *B. napus* annotation and thus likely represent outliers.[21] Outliers were identified using a Z-score approach in which Z-scores were calculated for each gene, across all three conditions, and genes which reported a Z-score above 3, representing data points which are 3 standard deviations above the mean, were removed from the dataset.

Structural characterization, splice junction analysis, coding potential, and NMD prediction of full-length unique isoforms was performed using sqanti_qc.py from the SQANTI2 (Structural and Quality Annotation of Novel Transcript Isoforms) python library.[56] Downstream analysis, statistical analysis, and graphical representation of the resultant output was performed in R.[57]

2.3 Identification of stress responsive changes in isoform production

Log_2 ratios were calculated from isoform counts per gene in the abiotic stress condition relative to isoform counts per gene in the normal condition. Each gene was categorized into one of three discrete categories: i) decrease: log_2 ratio $\leq$ -1, ii) equal: log_2 ratio between -1 and 1, and

iii) increase: $\log_2$ ratio ≥ 1. Functional annotation of genes within each category was completed as described in section 2.5.

2.4 Alternative splicing analysis

AS event detection was performed using the generateEvents command from the SUPPA2 pipeline using the parameters -i $INPUT_GTF -o $LIBRARY -f ioe -e SE SS MX RI FL.[58] Events were generated using the GTF files generated with sqanti_qc.py. Downstream analysis and statistical tests of the resultant output was completed in R.[57]

2.5 Homeolog identification and comparison

Homeologous pairs were identified and sorted from the resultant SQANTI2 and SUPPA2 output for downstream analysis using the *B. napus* homeolog list provided by Chalhoub et al (2014).[21] Sorting and data manipulation was performed in R.[57] Isoform diversity across a given homeologous pair was calculated following the protocol developed by Wang et al. (2018) as the $\log_2$-fold change of isoforms produced by the A_T-subgenome homeolog relative to the C_T-subgenome homeolog ($\log_2$[number of isoforms in A_T homeolog/number of isoforms in C_T homeolog]).[38] Homeologous pairs were then classified into three groups: i) pairs with a $\log_2$ ratio ≤ -1; ii) pairs with a $\log_2$ ratio between -1 and 1; and iii) pairs with a $\log_2$ ratio ≥1. Downstream analysis and statistical tests of the resultant output was completed in R.[57]

2.6 Functional annotation

Gene ontology (GO) enrichment analysis was performed using the R package topGO.[59] Genes which showed significant changes in isoform diversity in response to stress, homeologous

pairs in each of the three groups listed in section 2.4, and homeologous pairs which changed categorization in response to stress were functionally categorized by their GO IDs using the reference genome and their *A. thaliana* orthologs.[21] Statistical tests of enrichment were done using Fisher's exact test in R.[57]

2.7 Analysis of known cold and heat responsive genes

Structural analysis of isoforms produced by homeologous pairs known to be involved in cold and heat responses, identified by Lee & Adams (2020), were investigated and visualized using the Integrated Genome Viewer (IGV).[22]

Chapter 3: Results

3.1 *Brassica napus* isoform sequencing and read alignment

Iso-Seq was done with a total of 9 samples: three biological replicates for each of the three conditions. This resulted in a total in a total of 3,015,152 full-length non-chimeric (FLNC) reads following the processing of the subreads and subsequent circular consensus sequences (CCS). Figure 1 describes the filtering process; read counts for each stage of filtering are reported in Table 1. FLNC reads from each condition were pooled and mapped to the *B. napus* reference genome.[21] Mapping accuracies of the uniquely mapped reads were 99.92%, 99.94%, and 99.91% for normal, cold, and hot, respectively (Table 1). Mapped transcripts were then collapsed into a set of non-redundant isoforms used for downstream analysis (Table 1).

3.2 Isoform length and count distributions across conditions and subgenomes

A total of 32,449 genes were captured across all three conditions with a set of 31,953 genes shared by all three conditions (Figure 2) representing 98.5% of the total genes captured. Only a small number of captured genes were unique to each condition (Figure 2). To analyze differences across the A_T and C_T subgenomes, we used a set of 8,744 homeologous pairs for which both the A_T and C_T subgenomes were represented across all three conditions. Mean isoform lengths were 1,619 base pairs (bp), 1,726 bp, and 1,655 bp, for the normal, cold, and hot conditions, respectively (Figure 3). The interquartile ranges (IQR), representing isoforms for the central 50% of the distribution, were tightly conserved across all three conditions with 1,116-2,003 bp for normal, 1,206-2,120 bp for cold, and 1,124-2,057 bp for hot.

Following the collapse step, a unique set of non-redundant isoforms for each condition was identified. A total 98,861 isoforms were detected from 32,268 genes in the normal condition,

resulting in an average of 3.06 isoforms per gene. For the cold condition, a total of 96,965 isoforms were detected from 32,212 genes, resulting in an average of 3.02 isoforms per gene. For the hot condition, 111,976 isoforms were detected from 32,233 genes, resulting in an average of 3.47 isoforms per gene, the highest of all three conditions (Table 2).

In order to investigate any differences across abiotic stress conditions relative to the control (normal) condition, the distribution of isoform counts per gene across the shared set of 31,953 genes was investigated. Globally across the shared set of 31,953 genes, each condition showed a similar distribution, with genes that produce a single isoform at the highest frequency. As the number of isoforms per gene increases, the frequencies drop steadily (Figure 4a). Similar trends are also seen when we inspect the distributions across subgenomes, using the shared set of 8,744 homeologous genes (Figure 4b). Subgenomic distributions are tightly conserved across all three conditions. Interestingly, in contrast to the other conditions and subgenomes, genes that produce two isoforms show the highest frequency in the hot condition subgenomic distribution (Figure 4b).

Table 3 summarizes the number of isoforms produced from the shared set of 8,744 homeologous genes across the subgenomes in all three conditions as well as the isoform per gene calculations. It is evident that in each condition there is a slight and consistent increase towards the C_T subgenome in the number of isoforms produced by each gene.

3.3 Isoform characterization reveals previously unidentified splice junctions

Using the non-redundant set of isoforms produced by the shared set of 31,953 genes from each condition, the resultant isoform models were compared to the *B. napus* annotation provided by Chalhoub et al (2014).[21] Isoform characterization was determined by comparing the structures

of each isoform to the known transcript models contained within the annotation. Each isoform was categorized into one of five categories: 1) full-splice match (FSM): isoform matches the reference annotation perfectly, 2) incomplete-splice match (ISM): isoform matches consecutive, but not all, splice junctions, 3) novel in catalogue (NIC): isoform uses a novel combination of previously annotated splice junctions, 4) novel not in catalogue (NNC): isoform uses at least one novel splice junction, and 5) genic genomic: isoform overlaps with introns and exons of another gene. The distributions of these categories are visualized in Figure 5. Total counts for all categories are reported in Table 4. The distributions of the categorizations across all three conditions remain conserved. The categories which represent isoforms containing previously annotated splice junctions and thus match the reference closely (FSM, ISM, and NIC) represent roughly 51%-52% of the isoforms detected across all three conditions, whereas isoforms containing at least one novel junction (NNC) represent 37%-39% of the isoforms detected across all three conditions (Table 4, Figure 5). The NNC category represents the largest percentage of isoforms across any individual category for each condition.

3.4 Stress responsive changes in isoform production result in skewed distributions in response to abiotic stress

To examine whether there was a stress responsive shift in isoform profiles in our abiotic stress conditions we examined the $\log_2$ ratios of isoforms produced in the abiotic stress condition relative to the normal condition across the shared set of 31,953 genes. Our analysis revealed that in response to both cold and heat stress the distribution of the $\log_2$ ratios were asymmetric (Figure 6a). Tests of symmetry were performed based on the method developed Miao et al. (2006).[60] The distribution of the cold-responsive $\log_2$ ratios showed a negatively skewed, or left-

skewed, distribution (Test of Symmetry, $P<2.2 \times 10^{-16}$). For the heat-responsive distribution, the opposite was observed in which $\log_2$ ratios were significantly skewed to the right, or positively-skewed (Test of Symmetry, $P<2.2 \times 10^{-16}$) (Figure 6a). These results indicate fewer isoforms are produced per gene in response to the cold treatment, thus shifting the $\log_2$ ratio distribution towards the negative direction. In contrast, there is an overall increase in the number of isoforms produced per gene in response to heat, as shown by the shift in the positive direction.

In order to further examine these distributions, genes were sorted into three discrete categories based upon their $\log_2$ ratio (Figure 6b, Table 5). Genes in the "decrease" category correspond to genes which produce a $\log_2$ ratio ≤ -1, genes in the "equal" category represent genes which produce a $\log_2$ ratio between -1 and 1, and genes in the "increases" category correspond to genes producing $\log_2$ ratios ≥ 1. Generally, the vast majority of genes in either stress response were placed in the "equal" category representing 70% and 77% of all genes in the cold response and heat response, respectively, suggesting that the majority of genes experience a relatively equally distributed isoform distribution in response to stress (Figure 6b). However, the negatively-skewed distribution in the cold response is also evident when we examine the proportions of genes in each of the other categories: 21% in the "decrease" category and 9% in the "increase" category (Figure 6b). Comparisons of these categories in the heat response do not show such major deviations (11.46% decrease and 11.25% increase). This is likely due to the fact that the deviations in the $\log_2$ ratios responsible for the positively-shifted $\log_2$ distribution occur within the bounds of -1 and 1, in other words, the shift occurs more centrally in the distribution rather than at the tails.

In order to begin to investigate the types of genes in each of the categories, we examined the degree of overlap among genes in the "decrease" or "increase" categories across both stress

responses (Figure 6c). A relatively small proportion of the genes were given the same categorization (increase or decrease) in both stress responses (17.4% for increase, 20.9% for decrease), thus indicating that the genes which undergo isoform profile shifts are largely unique to each stress response.

To further understand the genes in each of these categories, GO analysis was performed on the set of genes in each category across each stress type to predict possible enriched functions. The most significantly enriched function in the cold response: increase category was "response to abscisic acid" (GO:0009737), a known regulator of the abiotic stress response in plants (Table 6). This suggests that genes involved in the abscisic acid (ABA) pathway undergo an increase in the number of isoforms produced by each gene in response to cold. In the cold response: decrease category, several of the enriched terms pertain to chloroplast localization, including "chloroplast stroma" (GO:0009570), "chloroplast envelope" (GO:0009941), and "chloroplast thylakoid membrane" (GO:0009535) (Table 8). This suggests that genes whose gene products are localized to the chloroplast experience an overall decrease in their isoform repertoires when exposed to cold stress. In the cold response: equal category the most enriched term was "cytosol" (GO:0005829) (Table 8). In the heat response increase: category, "unfolded protein binding" (GO:0051082) was among the most enriched terms, suggesting that genes which undergo a proliferation in the number of isoforms produced in response to heat stress are enriched for functions which may respond to proteins that have been denatured by the heat stress (Table 7). In the heat response: decrease category "DNA-binding transcription factor activity" (GO:0003700) was shown to be among the most enriched terms, however the significance level was not comparatively high relative to the scores obtained in the cold response sets (Table 7). As with the cold response: equal category, "cytosol" (GO:0005829) was the most enriched term in

the hot response equal category, thus indicating that genes which maintain an equally distributed isoform distribution of isoforms in response to both abiotic stresses are commonly localized to the cytosol (Table 8).

3.5 Persistent increase in AS in response to heat and slight C_T increase across all three conditions

In order to further investigate the dynamics resulting in the skewed isoform distributions discussed in section 3.4 we sought to detect and classify AS events across each condition and between subgenomes. Since isoform repertoires are a result of AS, closer inspection of the AS profiles may shed more light on how skewed isoform distributions result from differences in AS, whether in response to stress, between subgenomes, or both.

In our analysis we detected a significant increase in AS events in response to heat stress (Figure 7a) (Wilcoxon signed rank test, P=.02). However, a significant result was not reported for cold stress despite a small increase relative to the normal condition (Figure 7a) (Wilcoxon signed rank test, P=.94). Across each of the conditions distributions of AS events remained relatively conserved (Table 9), with intron retention (RI) being the most abundant category at 46.2%, 45.7%, 48.6% for normal, cold, and hot, respectively (Table 9). The second most abundant category was alternative 3' splice site, representing 26.9%, 27.6%, and 26.0% for normal, cold and hot, respectively (Table 9). The least abundant category in each of the 3 categories was alternative last exon (AL), representing less than 1% of all total events in each condition (Table 1). Despite significant increases for the heat condition or the slight increase in the cold condition, relative to normal, the distributions of AS event types remain conserved, suggesting that although abiotic stress may result in a stress-responsive shift in total number of

events, the distribution of total events across each of the categories remains largely unaffected. The significant increase in total events in response to heat stress aligns with our previous observation that the $\log_2$ ratio of isoforms produced in the hot condition relative to normal condition distribution is positively skewed. This skew is likely due to the increased AS activity generating more isoforms per gene following exposure to heat.

We conducted the same analysis using the set of 8,744 homeologous pairs shared across all three conditions in order to examine these dynamics across the subgenomes. In each condition, the C_T subgenome exhibited higher levels of AS relative to the A_T subgenome, however these biases were not significant (Wilcoxon rank sum test, P=0.71 (normal), P=0.80 (cold), P=0.71(hot)) (Figure 7b, Table 10). Once again, if we compare the cumulative AS event counts for both the A_T and C_T subgenomes, there is a significant increase in response to heat (Wilcoxon signed rank test, P=0.04), yet not for cold (Wilcoxon signed rank test, P=0.57). The persistent significant signal for increased AS in response to heat across both sets of genes further supports the idea that AS is likely heat-responsive. AS event profiles of the various event types again show a high degree of conservation across stress types and subgenomes. The distributions also closely reflect the AS profiles reported in the global analysis above (Table 10), with RI being the most abundant category across conditions and between subgenomes and AL being the least abundant across conditions and between subgenomes (Table 10). Summaries of the counts of alternatively spliced genes and the average AS event counts per gene are reported in Tables 11 and 12.

3.6 Homeologous pairs exhibit transcriptome-wide increase in C_T isoforms

To further explore post-transcriptional dynamics across the A_T and C_T subgenomes, we categorized each of the 8,744 homeologous gene pairs based upon their isoform distribution across both subgenomes. Each homeologous pair was placed into one of three categories based on the following $\log_2$ transformation, adapted from Wang et al. (2018):[38]

$$\text{Log}_2\left(\frac{\text{Number of isoforms produced by the } A_T \text{ homeologous gene}}{\text{Number of isoforms produced by the } C_T \text{ homeologous gene}}\right)$$

The three categories were as follows: i) $C_T > A_T$: $\log_2$ ratio ≤ -1, ii) $A_T = C_T$: $\log_2$ ratio between -1 and 1, and iii)) $A_T > C_T$: $\log_2$ ratio ≥ 1. The distributions and density plots are shown in Figure 8a-b. Across the 8,744 homeologous pairs, a range of $\log_2$ values are observed for each condition indicating both isoform divergence and isoform conservation, in terms of isoform counts, depending on the particular homeologous pair. To investigate these distributions in finer detail we began to examine if any of the distributions, across conditions, exhibit skewness. Presence of skewness in the distributions will indicate whether homeologs from a certain subgenome consistently produce more isoforms relative to the homeolog from the opposite subgenome. In each of the distributions the mean is observed to be less than the median (median = 0 for all conditions, mean = -0.029 (normal), -0.017 (cold), -0.022 (hot)) (Figure 8b). In order to further examine possible asymmetry amongst these distributions we once again applied the test of symmetry developed by Miao et al (2006).[60] We find that, across all three conditions, the distributions show left-skewness, meaning the C_T homeologs are more likely to produce a greater number of isoforms relative to their A_T homeologs and this pattern remains conserved in response to both stresses (Test of Symmetry, left-skewness, P=0.001 (normal), P=0.032 (cold),

P=0.007 (hot)) (Figure 8b). These skewed distributions are also evident when we compare the distributions of counts associated with each of the three categories across stress types (Figure 8c). Here we can see a consistent, yet slight enrichment for homeologous pairs categorized as $C_T > A_T$ relative to those categorized as $A_T > C_T$ reflecting the left-skewed distributions identified above. $C_T > A_T$ homeologous pairs represent 21.7%-22.7% of the homeologous pairs we tested relative to 20.1%-21.6% being categorized as $A_T > C_T$ homeologs (Figure 8c). In each condition the majority of pairs (55.6%-58.2%) were categorized as $A_T=C_T$ (Figure 8c).

In light of the results of the asymmetry analysis we wanted to investigate whether or not the $\log_2$ distributions from either abiotic stress condition were independent from the distribution observed in the normal condition. We uncovered a significant difference in the heat stress distribution at a significance threshold of $P<0.05$ (Two-Sample Kolmogorov-Smirnov Test, $P=0.013$). However, we did not detect a significant deviation from the normal distribution in response to cold stress (Two-Sample Kolmogorov-Smirnov Test, $P=0.227$). These results indicate that heat stress significantly perturbs the distribution of isoforms generated across homeologous pairs, but cold stress does not, and that the perturbation observed in the heat response, and lack thereof in cold response, does not affect the consistent $C_T > A_T$ trend, or left-skewness, of the distributions in each condition.

Finally, to further examine the homeologous pairs in each of the categories we conducted a GO enrichment analysis. Predicting the enriched functions in each of the categories will allow us to increase our understanding of the types of functions enriched among homeologous pairs which show an equally distributed isoform repertoire and ones that do not. *A. thaliana* orthologs were used to generate GO IDs for each of the 8,744 homeologous pairs in our dataset. These orthologs were previously annotated by Chalhoub et al (2014).[21] Across both the $C_T > A_T$ and A_T

> C_T categories few functions were consistently enriched across all three conditions. In other words, the set of enriched functions associated with the homeologous pairs in the $A_T > C_T$ category showed little overlap in the normal, cold, and hot conditions, likewise for the $A_T > C_T$ category (Table 13-15). However, in the $A_T=C_T$ category, "mRNA binding" (GO:0003729) was among the most highly enriched functions for the normal, cold, and hot conditions (Table 14). Consistent enrichment in the $A_T=C_T$ category possibly suggests that homeologous pairs involved in mRNA binding are more likely to show a conserved repertoire of isoforms across a homeologous pair rather than a divergent repertoire.

3.7 Homeologous isoform repertoire shifts show similar patterns in response to cold and heat stress

Once each of the 8,744 homeologous pairs had been categorized based on its isoform repertoire we sought to detect stress-responsive changes in these categorizations for each homeologous pair. We found that although a slim majority of pairs did not exhibit a category change in response to abiotic stress (representing 49.5% and 51.5% of homeologous pairs tested in the cold and heat response, respectively) a large proportion of homeologous pairs shift their category in response to stress (50.5% and 48.5% of homeologous pairs tested in the cold and heat response, respectively) (Figure 9, Table 16). Among homeologous pairs that exhibit a stress-responsive category shift, one of the most common shifts observed among homeologous pairs are pairs that move from the $C_T > A_T$ category (in the normal condition) to the $A_T=C_T$ category (in the stress condition) (11.6% for cold and 11.0% for heat stress) (Table 16). These represent homeologous pairs that, in response to stress, move towards a more equal distribution of isoforms across the homeologous pair when exposed to an abiotic stressor. The least common

category shifts were shifts between the $C_T > A_T$ and $A_T > C_T$ categories, and vice versa, as represented by the narrowest yellow alluvial flows in Figure 9a-b. The patterns observed in the various frequencies associated with each of the category shifts in response to cold and heat stress are tightly conserved (Wilcoxon signed rank test, $P>0.99$). The similarity in the alluvial flow patterning depicted in each panel of Figure 9a-b demonstrates this. Interestingly, a seemingly concomitant "switching" is observed between category shifts across both conditions. The proportions of homeologous genes which undergo a given switch (i.e. $C_T > A_T \rightarrow A_T = C_T$) are consistently similar to the opposite, or reversed, category shift (i.e. $A_T = C_T \rightarrow C_T > A_T$) (Table 16). This applies to shifts between the $A_T = C_T$ category and the $C_T > A_T$ or $A_T > C_T$ categories, across both conditions (Figure 9a and b). Overall, these results indicate that a large proportion of homeologous pairs display shifts in their isoform distributions across homeologs indicating the complex dynamics of the polyploid transcriptome. Additionally, neither of the abiotic stressors tested produce unique patterns of transcriptome-wide isoform repertoire changes, in other words cold and heat stress do not produce independent distributions of category shifts.

To further explore the connection between these category shifts and the underlying AS events responsible, we categorized the AS events in each subgenome across each of the major category shifts (as represented by the four widest alluvial flows in Figure 9a and 9b). The counts of AS events produced by each subgenome reflect the isoform repertoire defined in each category. Across both stress responses, if a homeologous pair is categorized as $C_T > A_T$, $A_T > C_T$, or $A_T=C_T$ they also reflect a similar distribution in AS events across homeologs (Table 17 and 18). These counts reflect that the category shifts are a result of shifting AS dynamics across both subgenomes.

Due to the large proportion of homeologous pairs that move towards or away from equally distributed isoform distributions in response abiotic stimuli, we were interested in investigating whether or not the genes identified in these category shifts were enriched for certain functions. Particularly, we were interested in enriched functions that arise in homeologous pairs that either move towards an equally distributed isoform distribution, away from an equally distributed isoform distribution, or ones that maintain such a distribution (Table 19-21). We see similarly enriched functions across the various shifts for both the cold and heat stresses. For the $A_T=C_T \rightarrow A_T=C_T$ shift, which includes homeologous pairs which are seemingly resistant to stress-induced shifts in isoform repertoires, we again see an enrichment for "mRNA binding" (GO:0003729) (Table 20). Similar enrichment for terms pertaining to the chloroplast are observed in each stress type for this category as well, including "chloroplast thylakoid membrane" (GO:0009535) and "chloroplast envelope" (GO:0009941). Interestingly, "apoplast" (GO:0048046) is also enriched across the cold and heat responses for the $A_T=C_T \rightarrow A_T=C_T$ shift (Table 20). Despite the highly similar patterning of category shifts across both stress responses shown in Figure 9a-b, there was little conservation in the enriched functions for the homeologous pairs which move away from an equally distributed isoform distribution, in response to stress (Table 19-21). For example, the top enriched term for the $A_T=C_T \rightarrow C_T > A_T$ category shift in the cold response was "microtubule bundle formation" (GO:0001578), however in the hot response it was "microtubule severing" (GO:0051013), a seemingly opposite function (Table 19).

3.8 Analysis of known cold and heat responsive genes

In order to further explore the post-transcriptional dynamics in regards to isoform repertoires we sought to characterize and visualize the isoform profiles of homeologous pairs

that were previously studied by Lee & Adams (2020) as known cold and heat responsive genes.[22] Due to our smaller subset of 8,744 homeologous pairs we were only able to identify a few homeologous pairs known to be involved in cold and heat stress that were shared across all three conditions. Thus, we are not able to comment broadly on the dynamics of isoform profiles across a larger subset of genes previously characterized to be involved in abiotic stress. This is likely a result of the decreased sensitivity of Iso-Seq when compared to short-read sequencers, such as Illumina.[61] However, we were able to collect and visualize a small number of pairs, two of which are shown in Figure 10. The homeologous pair BnaA03g38950 (A_T homeolog) and BnaC03g45990 (C_T homeolog), homologous to the cold response (COR) gene AT2G15970 in *A. thaliana*, show partially divergent isoform profiles in response to cold stress and across subgenomes (Figure 10a). Interestingly, in the normal condition BnaA03g38950 lacks the first exon and part of the second exon, each of which are fully captured within the ORF when exposed to cold stress. Additionally, these first two exons are intact in the majority of isoforms produced in the C_T homeolog (BnaC03g45990) with only one isoform (PB.4685.4) showing divergent intron/exon structures within the ORF. Notably, both the 5' and 3' UTRs vary across each homeolog and condition. The homeologous pair BnaA06g07260 (A_T homeolog) and BnCnng18070 (C_T homeolog), which is homologous to heat shock protein AT1G11660 in *A. thaliana*, is visualized in Figure 10b. Across the A_T homeolog, an overall increase in isoforms is noted in response to heat stress, with divergent structures produced by subtle changes in the 5' splice site of the first exon, as well as a premature termination codon in isoform PB.10401.2 in the 7th exon. In the C_T homeolog, an opposite change is seen in response to heat in which fewer isoforms are produced following exposure to the abiotic stressor (Figure 10b). Although this does not represent a thorough examination of known abiotic stress genes in the context of the

dynamics discussed in previous sections, for reasons explained above, these visualizations underscore the dynamic and complex nature of fluctuating isoform repertoires across the subgenomes in the context of abiotic stress.

3.9 Incidence of predicted NMD increases in response to heat stress

In order to examine whether or not the patterns discussed in earlier sections lead to nonsense-mediated decay (NMD), the process by which a premature termination codon (PTC) is introduced via alternative splicing leading to subsequent targeting of the isoform into a decay pathway or the production of a truncated protein, we categorized isoforms based off whether a PTC was detected within its ORF. We examined these likelihoods not only across conditions, but also across subgenomes in each of the conditions (Figure 11a). In each case, among the entire set of shared genes and across the set of homeologous pairs, the vast majority (82-84%) of isoforms are unlikely to be targets of NMD. However, slight differences in the relative frequencies of each NMD prediction are noted between the normal and heat conditions (Figure 11a). In fact, the frequencies of isoforms in each of the prediction categories varies significantly in response to heat (Fisher's Exact Test, $P = 4.6 \times 10^{-7}$), indicating a relationship between exposure to heat stress and an increase in the proportion of isoforms predicted to be likely targets of NMD. We did not detect a similar relationship in response to cold stress (Fisher's Exact Test, P=0.515).

Looking at these patterns across the subgenomes (Figure 11b), we can see that these frequencies are tightly conserved. Comparisons of the frequencies across the subgenomes in each condition reveal no significant differences, in other words the subgenomic origin of a given isoform and its NMD prediction are independent (Fisher's Exact Test, P=0.754 (normal), P=0.287 (cold), P=0.218 (hot)) (Figure 11b).

3.10 Figures & Tables

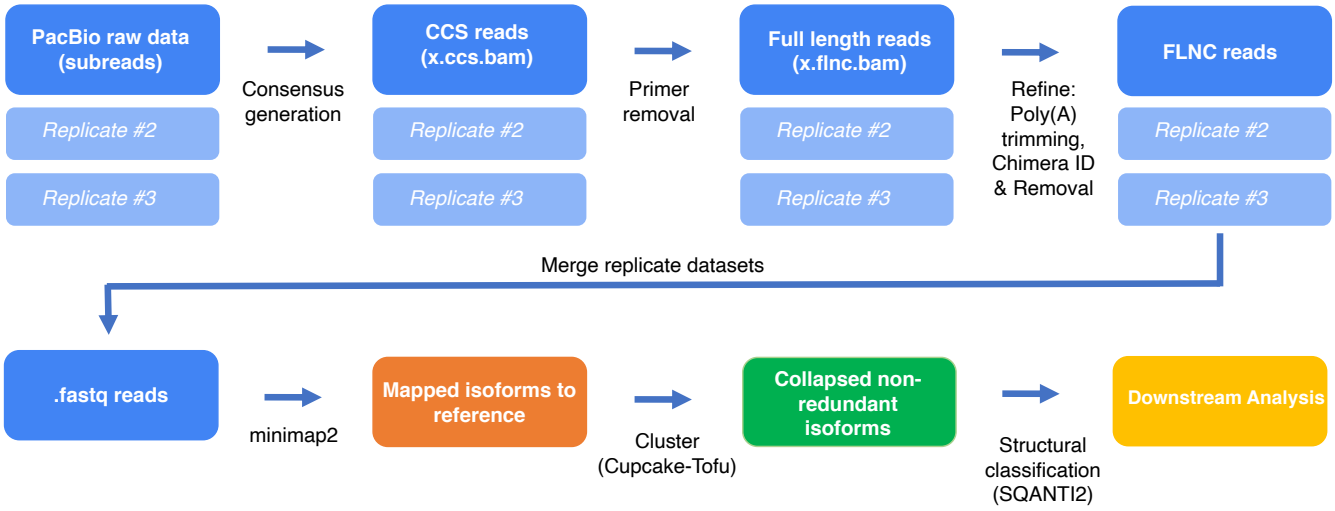

Figure 1. Flow chart describing the processing of raw isoform sequence data, alignment, and the collapsing of redundant isoforms.

Table 1. Summary of read counts at each filtering step with mapping accuracies

		Subreads	
Replicate	*Normal*	*Cold*	*Hot*
1	14,848,397	13,384,860	13,626,459
2	15,121,875	11,571,192	13,805,623
3	15,913,397	8,557,455	13,201,831
		CCS reads	
1	438,871	403,062	410,997
2	472,682	336,920	372,842
3	426,020	189,787	419,008
		FLNC reads	
1	375,265	356,445	358,491
2	412,514	296,433	329,857
3	361,750	169,571	354,826
		Mapping Accuracy (%)	
	99.92	99.94	99.91
		Collapsed Isoforms	
	131,749	125,266	156,029

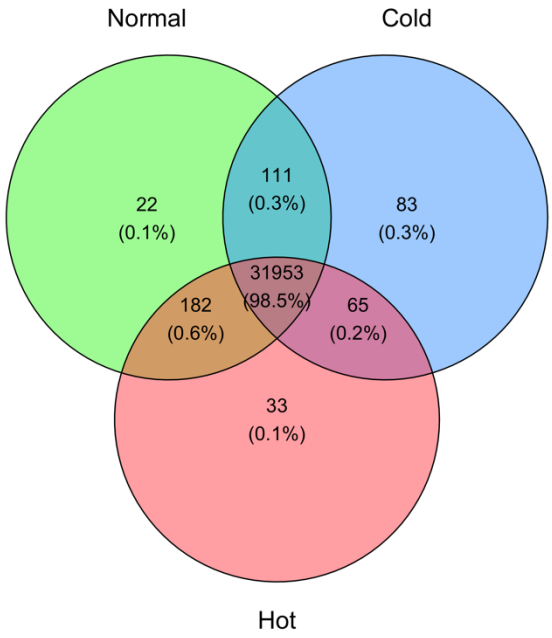

Figure 2. Genes captured in each condition, showing 32,449 total genes collected across all three conditions.

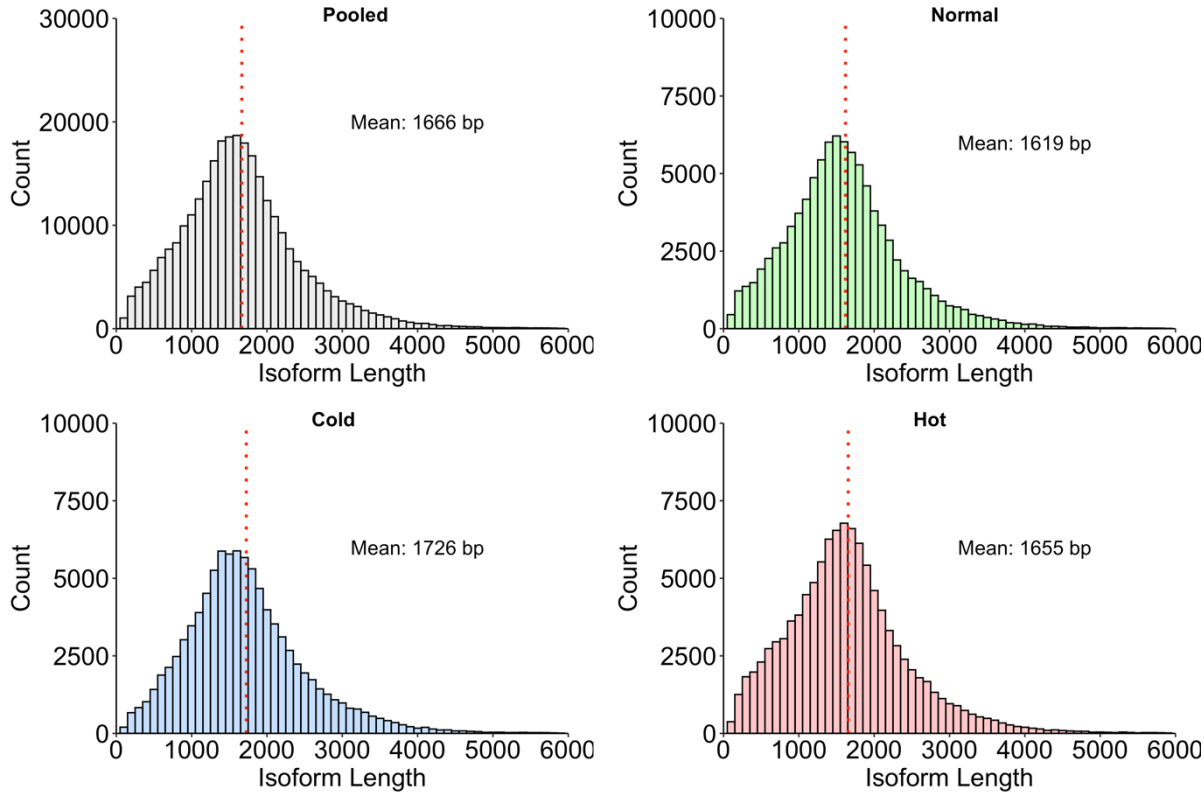

Figure 3. Distribution of isoform read lengths, pooled and separated by condition. Vertical dotted red line indicates the mean isoform length for each group.

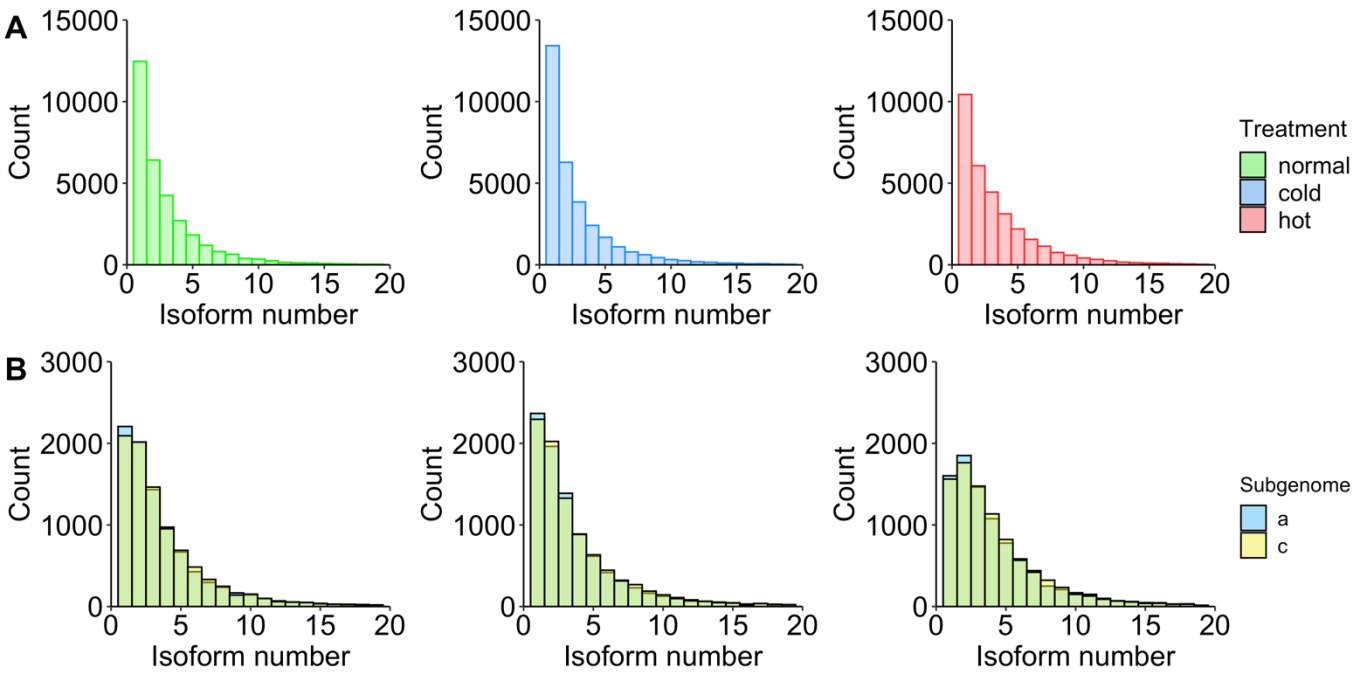

Figure 4. Distributions of isoforms per gene counts. a) Across abiotic conditions n=31,953, b) Across subgenomes. Blue represents the A_T subgenome, yellow represents the C_T subgenome, green represents where the distributions overlap.

Table 2. Genes and isoforms obtained across the three abiotic conditions.

	Normal	Cold	Hot
Unique Genes	32,268	32,212	32,233
Unique Isoforms	98,861	96,965	111,976
Isoforms per gene	3.06	3.02	3.47

Table 3. Genes and isoforms obtained across the three abiotic conditions, per subgenome.

	Normal		Cold		Hot	
Subgenome	A_T	C_T	A_T	C_T	A_T	C_T
Genes	8,744	8,744	8,744	8,744	8,744	8,744
Isoforms	30,623	31,013	30,702	30,939	35,058	35,388
Isoforms/Gene	3.50	3.55	3.51	3.54	4.01	4.05

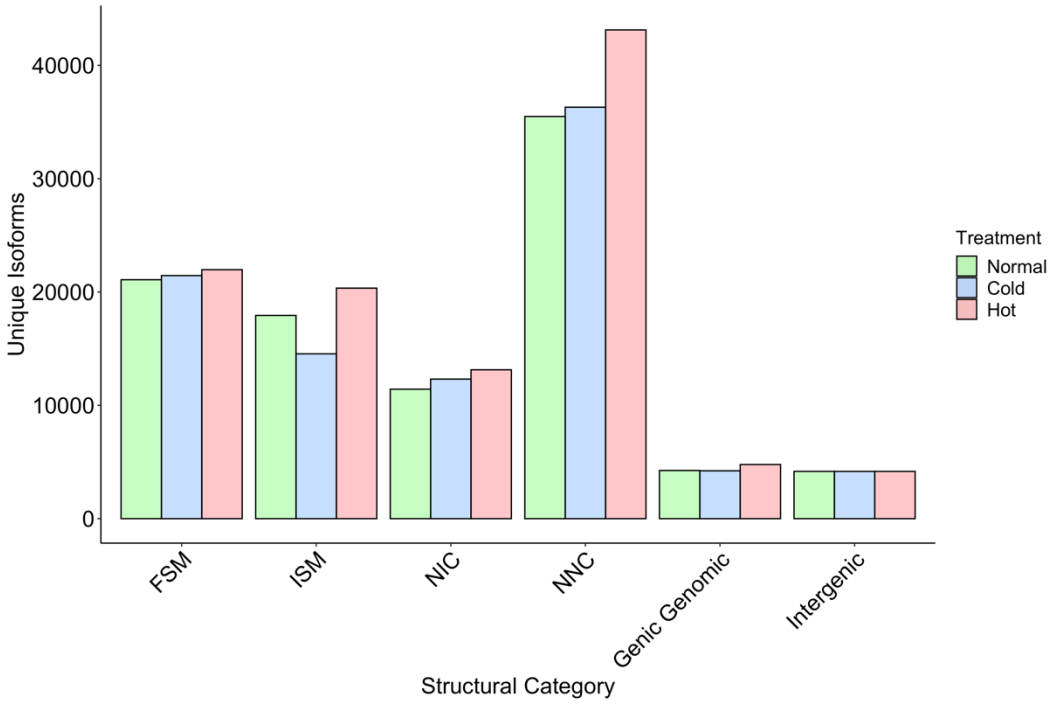

Figure 5. Distribution of isoforms across structural categories in each condition. FSM: full splice match, ISM: incomplete splice match, NIC: novel in catalogue, NNC: novel not in catalogue.

Table 4. Summary of isoform classifications per category across all three conditions.

	Normal	Cold	Hot
FSM	21,083 (22.1%)	21,450 (22.8%)	21,970 (20.2%)
ISM	17,933 (18.8%)	14,545 (15.5%)	20,342 (18.7%)
NIC	11,424 (12.0%)	12,318 (13.1%)	13,138 (12.1%)
NNC	35,487 (37.3%)	36,308 (38.7%)	43,124 (39.7%)
Genic Genomic	4,254 (4.5%)	4,224 (4.5%)	4,781 (4.4%)
Other	5,107 (5.4%)	5,089 (5.4%)	5,215 (4.8%)
Total	95,227	93,934	108,570

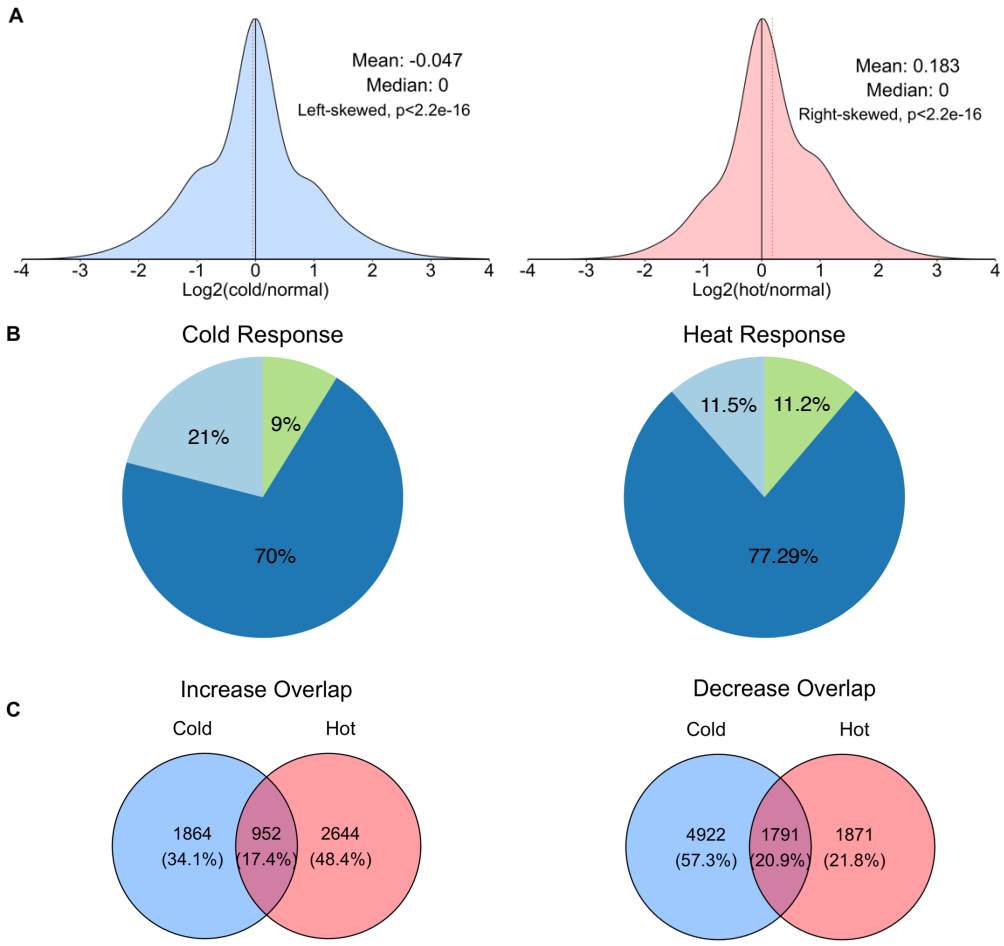

Figure 6. a) Density plot of the $\log_2$ ratios of isoforms produced in the abiotic stress condition relative the normal condition. Cold response distribution is left-skewed (Test of Symmetry, left-skewness, $P < 2.2 \times 10^{-16}$). Heat response distribution if right-skewed (Test of Symmetry, right-skewness, $P < 2.2 \times 10^{-16}$). Vertical dotted red line represents mean. b) Distributions of isoforms across 3 categories (decrease: $\log_2$ ratio <-1, equal: $\log_2$ratio between -1 and 1, increase: $\log_2$ ratio >1. c) Overlap of the genes which show both an increase or decrease in isoforms produced in response to each abiotic condition

Table 5. Counts of genes which undergo large shifts in the number of isoforms produced in response to cold and heat treatments.

	Cold	**Hot**
Decrease	6,713	3,662
Equal	22,424	24,695
Increase	2,816	3,596

Table 6. Summary of the top 10 enriched GO terms for the biological process (BP) domain, for the cold and heat responses.

Cold								
Decrease			Equal			Increase		
GO ID	Term	p-value	GO ID	Term	p-value	GO ID	Term	p-value
GO:0006412	translation	4.20E-05	GO:0046907	intracellular transport	0.00035	GO:0009737	response to abscisic acid	9.00E-07
GO:0002181	cytoplasmic translation	0.0021	GO:0006996	organelle organization	0.00049	GO:0009611	response to wounding	1.50E-06
GO:0051260	protein homooligomerization	0.0026	GO:0006457	protein folding	0.00064	GO:0035556	intracellular signal transduction	2.30E-05
GO:0030154	cell differentiation	0.0032	GO:0006412	translation	0.00097	GO:0046686	response to cadmium ion	3.60E-05
GO:0090391	granum assembly	0.0033	GO:0009735	response to cytokinin	0.00107	GO:0042742	defense response to bacterium	0.00012
GO:0048236	plant-type sporogenesis	0.0041	GO:0048366	leaf development	0.00197	GO:0050832	defense response to fungus	0.00012
GO:0046148	pigment biosynthetic process	0.0043	GO:0048193	Golgi vesicle transport	0.00263	GO:0002221	pattern recognition receptor signaling p...	0.00017
GO:0010152	pollen maturation	0.0045	GO:0006397	mRNA processing	0.00276	GO:0080142	regulation of salicylic acid biosyntheti...	0.00049
GO:0097502	mannosylation	0.0048	GO:0009723	response to ethylene	0.00431	GO:0009751	response to salicylic acid	0.00053
GO:0055114	oxidation-reduction process	0.0053	GO:0070646	protein modification by small protein re...	0.00445	GO:0009414	response to water deprivation	0.00066
Hot								
Decrease			Equal			Increase		
GO ID	Term	p-value	GO ID	Term	p-value	GO ID	Term	p-value
GO:0018107	peptidyl-threonine phosphorylation	0.001	GO:0007165	signal transduction	0.00066	GO:0046686	response to cadmium ion	0.0001
GO:0080167	response to karrikin	0.0017	GO:0006413	translational initiation	0.00083	GO:0042026	protein refolding	0.00014
GO:0043161	proteasome-mediated ubiquitin-dependent ...	0.0018	GO:0006612	protein targeting to membrane	0.00098	GO:0034976	response to endoplasmic reticulum stress	0.00028
GO:0062033	positive regulation of mitotic sister ch...	0.0019	GO:0009737	response to abscisic acid	0.00109	GO:0006511	ubiquitin-dependent protein catabolic pr...	0.00036
GO:0009805	coumarin biosynthetic process	0.0032	GO:0045893	positive regulation of transcription, DN...	0.00143	GO:0051301	cell division	0.00052
GO:0071922	regulation of cohesin loading	0.0036	GO:0010118	stomatal movement	0.00218	GO:1901566	organonitrogen compound biosynthetic pro.	0.00054
GO:0018105	peptidyl-serine phosphorylation	0.0045	GO:1901701	cellular response to oxygen-containing c...	0.00239	GO:0046907	intracellular transport	0.00061
GO:0010499	proteasomal ubiquitin-independent protei...	0.0087	GO:0010016	shoot system morphogenesis	0.00348	GO:0042254	ribosome biogenesis	0.00084
GO:0001678	cellular glucose homeostasis	0.0087	GO:0048366	leaf development	0.00367	GO:0006487	protein N-linked glycosylation	0.00107
GO:1990414	replication-born double-strand break rep...	0.0102	GO:0009409	response to cold	0.00377	GO:0009987	cellular process	0.00141

Table 7. Summary of the top 10 enriched GO terms for the molecular function (MF) domain, for the cold and heat responses.

Cold								
Decrease			Equal			Increase		
GO ID	Term	p-value	GO ID	Term	p-value	GO ID	Term	p-value
GO:0003735	structural constituent of ribosome	9.90E-08	GO:0003735	structural constituent of ribosome	0.00012	GO:0005524	ATP binding	3.90E-05
GO:0003729	mRNA binding	0.0002	GO:0043621	protein self-association	0.00035	GO:0005509	calcium ion binding	0.00052
GO:0005507	copper ion binding	0.00021	GO:0003729	mRNA binding	0.00151	GO:0004674	protein serine/threonine kinase activity	0.00081
GO:0010011	auxin binding	0.00048	GO:0051082	unfolded protein binding	0.00411	GO:0000978	RNA polymerase II cis-regulatory region ...	0.00104
GO:0016814	hydrolase activity, acting on carbon-nit...	0.00082	GO:0005515	protein binding	0.00551	GO:0005516	calmodulin binding	0.00134
GO:0015250	water channel activity	0.0013	GO:0016417	S-acyltransferase activity	0.00772	GO:0043424	protein histidine kinase binding	0.00255
GO:0004675	transmembrane receptor protein serine/th...	0.00207	GO:0005198	structural molecule activity	0.00984	GO:0030276	clathrin binding	0.00297
GO:0004312	fatty acid synthase activity	0.00273	GO:0046982	protein heterodimerization activity	0.01119	GO:0005546	phosphatidylinositol-4,5-bisphosphate bi...	0.00426
GO:0000030	mannosyltransferase activity	0.0043	GO:0019899	enzyme binding	0.01278	GO:0004630	phospholipase D activity	0.00502
GO:0016491	oxidoreductase activity	0.00564	GO:0019787	ubiquitin-like protein transferase activ...	0.01637	GO:0140327	flippase activity	0.00502
Hot								
Decrease			Equal			Increase		
GO ID	Term	p-value	GO ID	Term	p-value	GO ID	Term	p-value
GO:0003700	DNA-binding transcription factor activit...	0.001	GO:0003729	mRNA binding	0.00023	GO:0051082	unfolded protein binding	2.70E-06
GO:0080044	quercetin 7-O-glucosyltransferase activi...	0.0016	GO:0005515	protein binding	0.00137	GO:0008017	microtubule binding	1.50E-05
GO:0034595	phosphatidylinositol phosphate 5-phospha...	0.0072	GO:0003735	structural constituent of ribosome	0.00174	GO:0003756	protein disulfide isomerase activity	0.00019
GO:0004674	protein serine/threonine kinase activity	0.0079	GO:0003743	translation initiation factor activity	0.00178	GO:0003924	GTPase activity	0.00042
GO:0019901	protein kinase binding	0.0083	GO:0004175	endopeptidase activity	0.00866	GO:0005524	ATP binding	0.00073
GO:0004842	ubiquitin-protein transferase activity	0.0092	GO:0042802	identical protein binding	0.01247	GO:0004579	dolichyl-diphosphooligosaccharide-protei...	0.00079
GO:0004448	isocitrate dehydrogenase activity	0.0127	GO:0016765	transferase activity, transferring alkyl...	0.0137	GO:0003682	chromatin binding	0.00097
GO:0043495	protein-membrane adaptor activity	0.0127	GO:0016301	kinase activity	0.01652	GO:0004553	hydrolase activity, hydrolyzing O-glycos...	0.00142
GO:0015172	acidic amino acid transmembrane transpor...	0.0127	GO:0008320	protein transmembrane transporter activi...	0.02555	GO:0016887	ATPase activity	0.00219
GO:0008889	glycerophosphodiester phosphodiesterase ...	0.0127	GO:0022884	macromolecule transmembrane transporter ...	0.02555	GO:0031072	heat shock protein binding	0.00263

Table 8. Summary of the top 10 enriched GO terms for the cellular compartment (CC) domain, for the cold and heat responses.

Cold								
Decrease			Equal			Increase		
GO ID	Term	p-value	GO ID	Term	p-value	GO ID	Term	p-value
GO:0009570	chloroplast stroma	1.80E-08	GO:0005829	cytosol	1.70E-08	GO:0005886	plasma membrane	1.30E-06
GO:0009941	chloroplast envelope	3.90E-08	GO:0005634	nucleus	3.40E-05	GO:0005634	nucleus	1.30E-06
GO:0009507	chloroplast	1.30E-07	GO:0005794	Golgi apparatus	8.20E-05	GO:0009705	plant-type vacuole membrane	0.00022
GO:0009535	chloroplast thylakoid membrane	2.80E-07	GO:0098588	bounding membrane of organelle	8.40E-05	GO:0009506	plasmodesma	0.00044
GO:0042788	polysomal ribosome	3.10E-07	GO:0098805	whole membrane	0.00019	GO:0005774	vacuolar membrane	0.00276
GO:0022625	cytosolic large ribosomal subunit	4.40E-07	GO:0022625	cytosolic large ribosomal subunit	0.00028	GO:0071007	U2-type catalytic step 2 spliceosome	0.00486
GO:0005829	cytosol	8.00E-06	GO:0005783	endoplasmic reticulum	0.00092	GO:0010008	endosome membrane	0.00571
GO:0022627	cytosolic small ribosomal subunit	3.70E-05	GO:0098796	membrane protein complex	0.00102	GO:0005794	Golgi apparatus	0.00596
GO:0010287	plastoglobule	0.00044	GO:1990234	transferase complex	0.00158	GO:0071012	catalytic step 1 spliceosome	0.01246
GO:0009522	photosystem I	0.0019	GO:0042788	polysomal ribosome	0.00341	GO:0071006	U2-type catalytic step 1 spliceosome	0.01246
Hot								
Decrease			Equal			Increase		
GO ID	Term	p-value	GO ID	Term	p-value	GO ID	Term	p-value
GO:0032116	SMC loading complex	0.0034	GO:0005829	cytosol	1.30E-05	GO:0005788	endoplasmic reticulum lumen	3.70E-09
GO:0005839	proteasome core complex	0.0079	GO:0044391	ribosomal subunit	0.0019	GO:0005829	cytosol	1.10E-08
GO:0005634	nucleus	0.0104	GO:0009535	chloroplast thylakoid membrane	0.0036	GO:0005618	cell wall	2.00E-08
GO:0000152	nuclear ubiquitin ligase complex	0.0115	GO:0005844	polysome	0.0067	GO:0005768	endosome	1.30E-05
GO:0000164	protein phosphatase type 1 complex	0.0122	GO:0009506	plasmodesma	0.0068	GO:0005802	trans-Golgi network	2.00E-05
GO:0090694	Scc2-Scc4 cohesin loading complex	0.0122	GO:1990204	oxidoreductase complex	0.0073	GO:0032991	protein-containing complex	2.90E-05
GO:0005886	plasma membrane	0.0133	GO:0022626	cytosolic ribosome	0.0088	GO:0005730	nucleolus	3.40E-05
GO:0019774	proteasome core complex, beta-subunit co...	0.0206	GO:0009536	plastid	0.0093	GO:0005774	vacuolar membrane	0.00024
GO:0009840	chloroplastic endopeptidase Clp complex	0.0206	GO:0042788	polysomal ribosome	0.0129	GO:0005832	chaperonin-containing T-complex	0.00026
GO:0031597	cytosolic proteasome complex	0.0206	GO:0098552	side of membrane	0.0161	GO:0005874	microtubule	0.00036

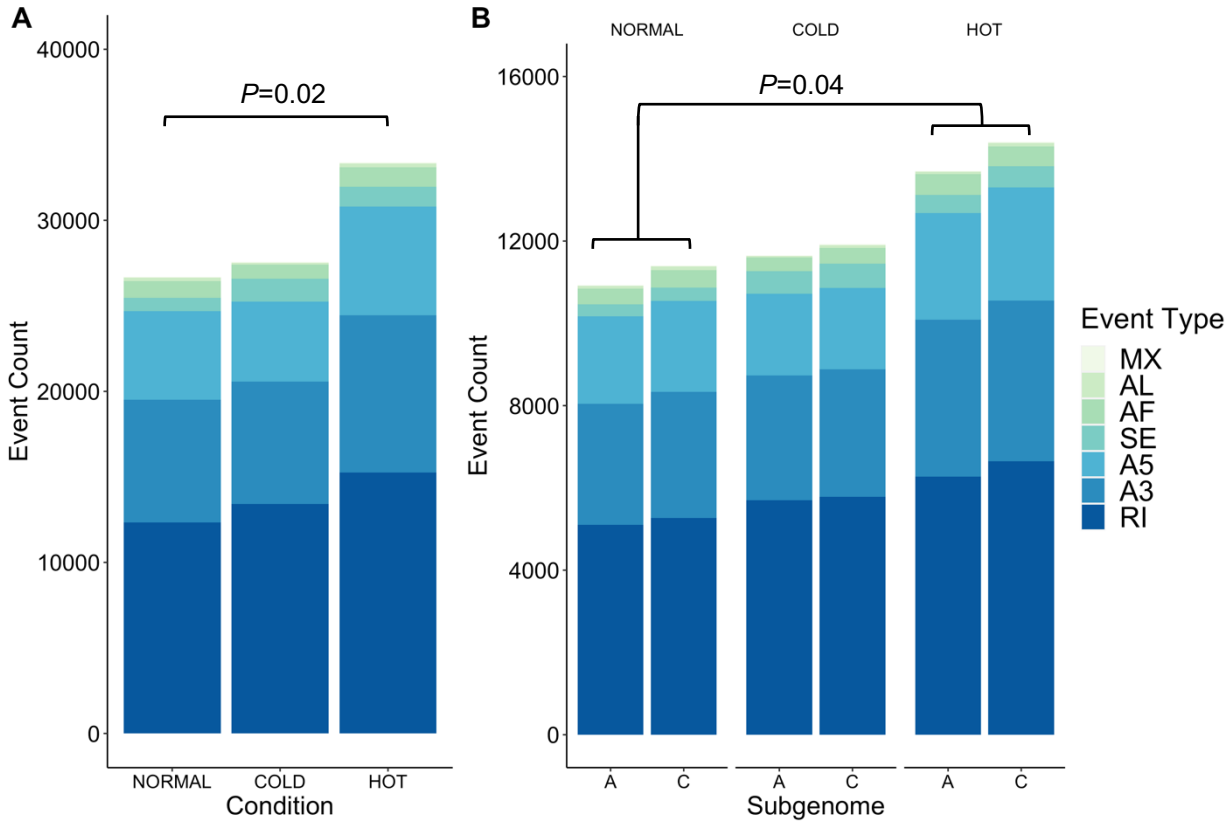

Figure 7. a) AS events detected across 31,953 genes shared by each condition. A significant difference between the number of alternative splicing events in the normal condition and the hot condition was detected (Wilcoxon signed rank test, n = 7 categories of alternative splicing events, *P*=0.02). b) AS events detected in the 8,744 homeologous pairs, separated by subgenome. Significant difference in cumulative AS events across both subgenomes in the response to heat (Wilcoxon signed rank test, n = 7 categories of alternative splicing events; $P > 0.05$). No significant differences were detected across subgenomes.

Table 9. AS event counts and percentages across abiotic conditions, n=31,953 genes.

	Normal	**Cold**	**Hot**
Exon skipping (SE)	781 (2.9%)	1,341 (4.9%)	1,153 (3.5%)
Mutually exclusive exon (MX)	35 (0.1%)	48 (0.2%)	52 (0.16%)
Alternative 5' splice site (A5)	5,191 (19.5%)	4,668 (16.9%)	6,359 (19.0%)
Alternative 3' splice site (A3)	7,170 (26.9%)	7,169 (26.0%)	9,200 (27.6%)
Intron retention (RI)	12,332 (46.2%)	13,407 (48.6%)	15,251 (45.7%)
Alternative first exon (AF)	972 (3.6%)	813 (3.0%)	1139 (3.4%)
Alternative last exon (AL)	202 (0.8%)	113 (0.4%)	227 (0.7%)
Total	26,683	27,579	33,381

Table 10. AS event counts and percentages across abiotic conditions per subgenome, n=8,744 homeologous gene pairs.

	Normal		Cold		Hot	
	A_T	C_T	A_T	C_T	A_T	C_T
Exon skipping (SE)	297 (2.7%)	327 (2.9%)	552 (4.73%)	585 (4.9%)	444 (3.2%)	516 (3.6%)
Mutually exclusive exon (MX)	12 (0.10%)	17 (0.1%)	15 (0.1%)	23 (0.2%)	21 (0.2%)	22 (0.2%)
Alternative 5' exon (A5)	2,216 (19.5%)	2,209 (19.3%)	1,983 (17.0%)	1,982 (16.6%)	2,589 (18.9%)	2,748 (19.1%)
Alternative 3' exon (A3)	2,934 (26.9%)	3,068 (26.9%)	3,032 (26.0%)	3,097 (26.0%)	3,816 (27.8%)	3,909 (27.1%)
Intron retention (RI)	5,099 (46.7%)	5,264 (46.2%)	5,703 (49.0%)	5,783 (48.5%)	6,272 (45.8%)	6,646 (46.12%)
Alternative first exon (AF)	375 (3.4%)	427 (3.7%)	333 (2.9%)	385 (3.2%)	510 (3.7%)	482 (3.3%)
Alternative last exon (AL)	73 (0.7%)	90 (0.8%)	30 (0.3%)	70 (0.6%)	52 (0.8%)	85 (0.6%)
Total	10,925	11,402	11,648	11,925	13,704	14,408

Table 11. AS event counts and percentages across abiotic conditions per subgenome, n=8,744 homeologous gene pairs.

	Normal	Cold	Hot
Total Genes	31,953	31,953	31,953
AS Events	26,683	27,579	33,381
AS Genes	10,384	10,294	12,187
AS events/gene	2.57	2.68	2.74

Table 12. Summary of the counts of homeologous genes which undergo AS, by condition and subgenome

	Normal (A_T)	Normal (C_T)	Cold (A_T)	Cold (C_T)	Hot (A_T)	Hot (C_T)
Total Genes	8,744	8,744	8,744	8,744	8,744	8,744
AS Events	10,925	11,402	11,648	11,925	13,704	14,408
AS Genes	4,344	4,476	4,395	4,472	5,132	5,292
AS events/gene	2.51	2.55	2.65	2.67	2.67	2.72

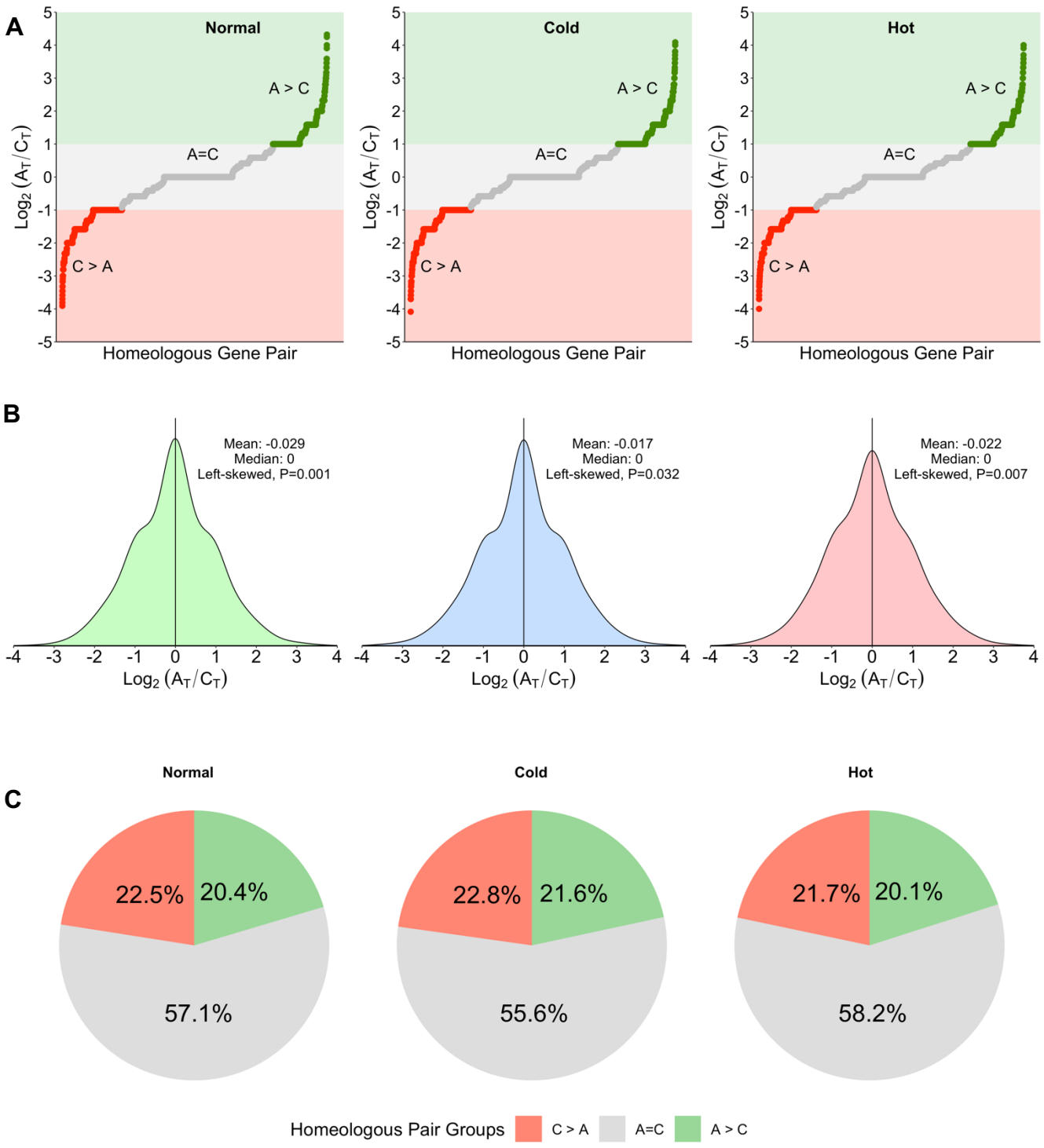

Figure 8. a) Log_2 ratio of isoform numbers of homeologous genes in A_T and C_T subgenomes, separated by condition. All genes pairs are categorized into three difference groups: C>A (log_2 ratio ≤ -1), A=C (log_2 ratio >-1, and log_2 ratio <1), and A>C (log_2 ratio ≥1). b) Density plots of the log_2 ratios for each condition (green = normal, blue = cold, red = hot), mean, medians, and the results of the symmetry analysis are listed on the graphs. c) Percentages of homeologs belonging to each of the homeologous pair groups, separated by condition.

Table 13. Summary of the top 5 enriched GO terms for the biological process (BP) domain across homeologous pair categories.

Normal								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0051592	response to calcium ion	0.0005	GO:0009768	photosynthesis, light harvesting in phot...	0.0034	GO:0006468	protein phosphorylation	0.00066
GO:0042255	ribosome assembly	0.0019	GO:0009628	response to abiotic stimulus	0.0035	GO:0006376	mRNA splice site selection	0.001
GO:0045815	positive regulation of gene expression, ...	0.0045	GO:0048316	seed development	0.0035	GO:1903829	positive regulation of cellular protein ...	0.0011
GO:0080148	negative regulation of response to water...	0.0046	GO:0019684	photosynthesis, light reaction	0.005	GO:0033206	meiotic cytokinesis	0.00182
GO:0051865	protein autoubiquitination	0.0052	GO:0048827	phyllome development	0.0053	GO:0009063	cellular amino acid catabolic process	0.00263
Cold								
C bias			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0009089	lysine biosynthetic process via diaminop...	0.0017	GO:0009414	response to water deprivation	0.0017	GO:0006536	glutamate metabolic process	0.0011
GO:0000723	telomere maintenance	0.0052	GO:0071310	cellular response to organic substance	0.0021	GO:0071281	cellular response to iron ion	0.0018
GO:0072503	cellular divalent inorganic cation homeo...	0.0073	GO:0009787	regulation of abscisic acid-activated si...	0.0033	GO:0009157	deoxyribonucleoside monophosphate biosyn...	0.0024
GO:0015749	monosaccharide transmembrane transport	0.0088	GO:0009725	response to hormone	0.0037	GO:0046835	carbohydrate phosphorylation	0.0025
GO:0060969	negative regulation of gene silencing	0.0132	GO:0009644	response to high light intensity	0.0042	GO:0032786	positive regulation of DNA-templated tra...	0.0028
Hot								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0042255	ribosome assembly	0.0011	GO:0046686	response to cadmium ion	3.40E-05	GO:0006281	DNA repair	0.0011
GO:0006412	translation	0.0023	GO:0009411	response to UV	0.004	GO:0000209	protein polyubiquitination	0.0014
GO:0006561	proline biosynthetic process	0.0039	GO:0009817	defense response to fungus, incompatible...	0.0043	GO:0009157	deoxyribonucleoside monophosphate biosyn...	0.0023
GO:1902288	regulation of defense response to oomyce...	0.0039	GO:0072594	establishment of protein localization to...	0.0046	GO:0043254	regulation of protein-containing complex...	0.0033
GO:0042273	ribosomal large subunit biogenesis	0.0045	GO:0006623	protein targeting to vacuole	0.007	GO:0016102	diterpenoid biosynthetic process	0.0037

Table 14. Summary of the top 5 enriched GO terms for the molecular function (MF) domain across homeologous pair categories.

Normal								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0003735	structural constituent of ribosome	3.10E-05	GO:0003729	mRNA binding	1.80E-06	GO:0004672	protein kinase activity	0.0012
GO:0003906	DNA-(apurinic or apyrimidinic site) endo...	0.0049	GO:0016168	chlorophyll binding	0.0016	GO:0016639	oxidoreductase activity, acting on the C...	0.0019
GO:0015078	proton transmembrane transporter activit...	0.0067	GO:0019904	protein domain specific binding	0.0016	GO:0070568	guanylyltransferase activity	0.0019
GO:0016835	carbon-oxygen lyase activity	0.0091	GO:0043022	ribosome binding	0.0076	GO:0036094	small molecule binding	0.0054
GO:0038023	signaling receptor activity	0.0124	GO:0003779	actin binding	0.0084	GO:0071617	lysophospholipid acyltransferase activit...	0.0061
Cold								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0015078	proton transmembrane transporter activit...	0.00069	GO:0003735	structural constituent of ribosome	1.80E-05	GO:0004252	serine-type endopeptidase activity	0.0019
GO:0015145	monosaccharide transmembrane transporter...	0.00605	GO:0003729	mRNA binding	4.50E-05	GO:0030170	pyridoxal phosphate binding	0.002
GO:0004185	serine-type carboxypeptidase activity	0.01026	GO:0003700	DNA-binding transcription factor activit...	0.00043	GO:0003984	acetolactate synthase activity	0.0022
GO:0004180	carboxypeptidase activity	0.01071	GO:0005515	protein binding	0.00419	GO:0047334	diphosphate-fructose-6-phosphate 1-phosp...	0.0022
GO:0005355	glucose transmembrane transporter activi...	0.01303	GO:0005048	signal sequence binding	0.00512	GO:0051117	ATPase binding	0.0037
Hot								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0003735	structural constituent of ribosome	2.30E-05	GO:0003729	mRNA binding	3.80E-05	GO:0061631	ubiquitin conjugating enzyme activity	0.0042
GO:0080043	quercetin 3-O-glucosyltransferase activi...	0.003	GO:0052689	carboxylic ester hydrolase activity	0.0022	GO:0004124	cysteine synthase activity	0.0081
GO:0022821	potassium ion antiporter activity	0.004	GO:0016651	oxidoreductase activity, acting on NAD(P...	0.0111	GO:0043813	phosphatidylinositol-3,5-bisphosphate 5-...	0.0087
GO:0080044	quercetin 7-O-glucosyltransferase activi...	0.0041	GO:0044390	ubiquitin-like protein conjugating enzym...	0.0113	GO:0000036	acyl carrier activity	0.0087
GO:0051139	metal ion:proton antiporter activity	0.0048	GO:0042803	protein homodimerization activity	0.0117	GO:0052744	phosphatidylinositol monophosphate phosp...	0.0087

Table 15. Summary of the top 5 enriched GO terms for the cellular compartment (CC) domain across homeologous pair categories.

Normal								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0042788	polysomal ribosome	6.20E-05	GO:0009535	chloroplast thylakoid membrane	2.40E-06	GO:0030660	Golgi-associated vesicle membrane	0.0014
GO:0022627	cytosolic small ribosomal subunit	0.0019	GO:0009522	photosystem I	0.00015	GO:0030662	coated vesicle membrane	0.0019
GO:0022625	cytosolic large ribosomal subunit	0.0019	GO:1990904	ribonucleoprotein complex	0.002	GO:0030658	transport vesicle membrane	0.0083
GO:0005797	Golgi medial cisterna	0.0078	GO:0048046	apoplast	0.0026	GO:0030117	membrane coat	0.0122
GO:0033177	proton-transporting two-sector ATPase co...	0.0085	GO:0015629	actin cytoskeleton	0.00413	GO:0048475	coated membrane	0.0122
Cold								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0009705	plant-type vacuole membrane	0.001	GO:0005747	mitochondrial respiratory chain complex ...	0.0012	GO:0005948	acetolactate synthase complex	0.0024
GO:0000784	nuclear chromosome, telomeric region	0.0012	GO:0005634	nucleus	0.0026	GO:0031410	cytoplasmic vesicle	0.0089
GO:0005881	cytoplasmic microtubule	0.0074	GO:0009535	chloroplast thylakoid membrane	0.0033	GO:0005852	eukaryotic translation initiation factor...	0.0096
GO:0042788	polysomal ribosome	0.0153	GO:0022625	cytosolic large ribosomal subunit	0.0041	GO:0008287	protein serine/threonine phosphatase com...	0.0155
GO:1904949	ATPase complex	0.0163	GO:0031967	organelle envelope	0.0063	GO:1903293	phosphatase complex	0.0155
Hot								
C > A			A=C			A > C		
GO ID	Term	*P*	GO ID	Term	*P*	GO ID	Term	*P*
GO:0022625	cytosolic large ribosomal subunit	7.20E-05	GO:0016607	nuclear speck	0.00045	GO:0061631	ubiquitin conjugating enzyme activity	0.0045
GO:0042788	polysomal ribosome	0.00024	GO:0048046	apoplast	0.00059	GO:0004124	cysteine synthase activity	0.0046
GO:0005753	mitochondrial proton-transporting ATP sy...	0.00323	GO:0031967	organelle envelope	0.00342	GO:0043813	phosphatidylinositol-3,5-bisphosphate 5-...	0.0048
GO:0005851	eukaryotic translation initiation factor...	0.00371	GO:0016020	membrane	0.00361	GO:0000036	acyl carrier activity	0.0079
GO:0034045	phagophore assembly site membrane	0.01216	GO:0005730	nucleolus	0.00829	GO:0052744	phosphatidylinositol monophosphate phosp...	0.0116

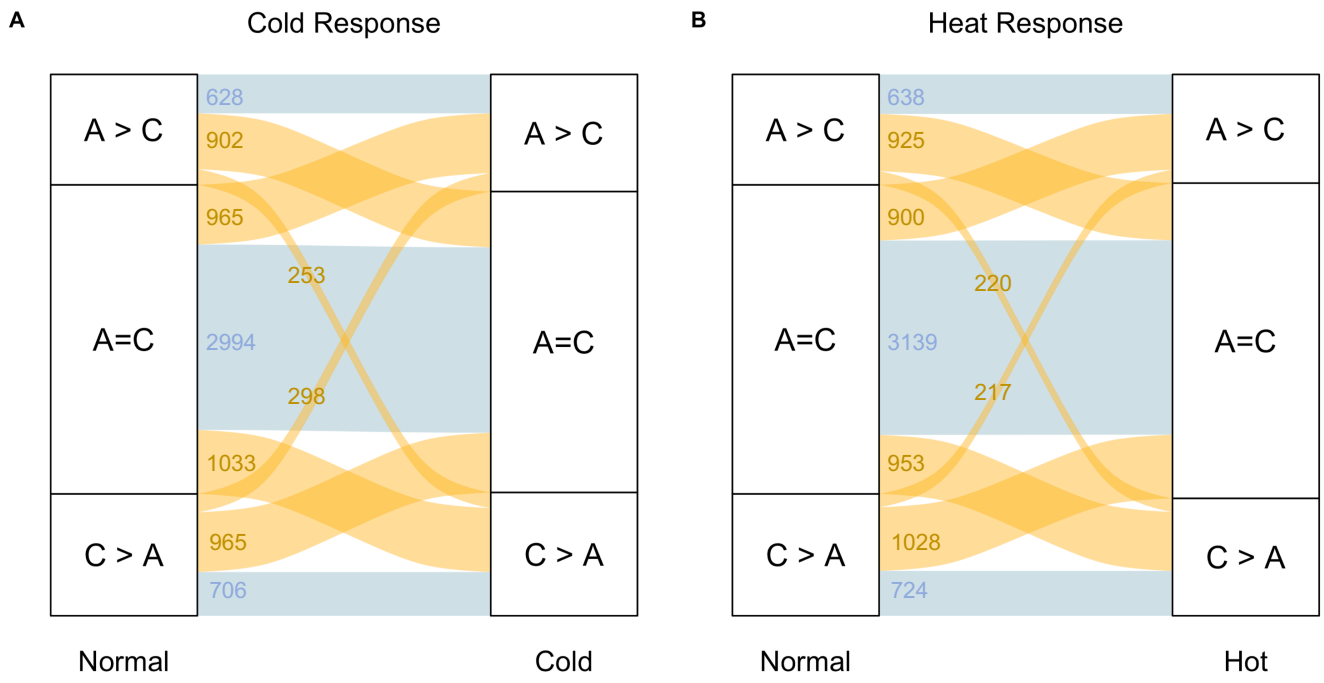

Figure 9. Changes in A_T and C_T isoform distribution categories in response to stress. Yellow alluvial flow represents homeologs that change categorization in response to stress, blue alluvial flow represents homeologs that do not change categorization in response to stress across the 8,744 homeologous pairs. a) Cold response. b) Heat response.

Table 16. Counts of genes homeologous pairs for each type of stress responsive isoform ratio category shift.

Category Shift (Normal → Stress)	**Pairs (Cold Response)**	**Pairs (Heat Response)**
$C_T > A_T \rightarrow C_T > A_T$	706 (8.1%)	724 (8.3%)
$C_T > A_T \rightarrow A_T = C_T$	965 (11.0%)	1028 (11.6%)
$C_T > A_T \rightarrow A_T > C_T$	298 (3.4%)	217 (2.5%)
$A_T = C_T \rightarrow C_T > A_T$	1033 (11.8%)	953 (10.9%)
$A_T = C_T \rightarrow A_T = C_T$	2994 (34.2%)	3139 (35.9%)
$A_T = C_T \rightarrow A_T > C_T$	965 (11.0%)	900 (10.3%)
$A_T > C_T \rightarrow C_T > A_T$	253 (2.9%)	220 (2.5%)
$A_T > C_T \rightarrow A_T = C_T$	902 (10.3%)	925 (10.6%)
$A_T > C_T \rightarrow A_T > C_T$	628 (7.2%)	638 (7.3%)

Table 17. AS event counts per subgenome among homeologous pairs which demonstrate a cold responsive shift in isoform ratio category.

		AS Events			
Category Shift (Normal → Cold)	**Genes**	**Normal (A_T)**	**Normal (C_T)**	**Cold (A_T)**	**Cold (C_T)**
$C_T > A_T \rightarrow A_T = C_T$	1,152	270	1,627	1,018	1,085
$A_T = C_T \rightarrow C_T > A_T$	1,185	886	1,176	852	1,999
$A_T = C_T \rightarrow A_T > C_T$	631	915	633	1,468	229
$A_T = C_T \rightarrow A_T = C_T$	1,160	1,160	173	703	547

Table 18. AS event counts per subgenome among homeologous pairs which demonstrate a heat responsive shift in isoform ratio category.

		AS Events			
Category Shift (Normal → Hot)	**Genes**	**Normal (A_T)**	**Normal (C_T)**	**Hot (A_T)**	**Hot (C_T)**
$C_T > A_T \rightarrow A_T = C_T$	1,153	270	1,627	931	1,534
$A_T = C_T \rightarrow C_T > A_T$	1,100	886	1,176	368	1,568
$A_T = C_T \rightarrow A_T > C_T$	629	915	633	1,209	835
$A_T = C_T \rightarrow A_T = C_T$	555	1,160	173	927	560

Table 19. Summary of the top 5 enriched GO terms for the biological process (BP) domain across isoform repertoire shift categories.

Cold								
A=C → C > A			A=C → A=C			A=C → A > C		
GO ID	Term	P	GO ID	Term	P	GO ID	Term	P
GO:0001578	microtubule bundle formation	0.0039	GO:0009767	photosynthetic electron transport chain	0.00072	GO:0006536	glutamate metabolic process	0.00057
GO:0045324	late endosome to vacuole transport	0.0048	GO:0009644	response to high light intensity	0.00113	GO:0009099	valine biosynthetic process	0.00139
GO:0008202	steroid metabolic process	0.0059	GO:0006402	mRNA catabolic process	0.00134	GO:0046835	carbohydrate phosphorylation	0.00196
GO:0048468	cell development	0.0083	GO:0010608	posttranscriptional regulation of gene expression	0.00159	GO:0048354	mucilage biosynthetic process involved i...	0.00261
GO:0061024	membrane organization	0.0092	GO:0015979	photosynthesis	0.00355	GO:0071281	cellular response to iron ion	0.00602
Hot								
A=C → C > A			A=C → A=C			A=C → A > C		
GO ID	Term	P	GO ID	Term	P	GO ID	Term	P
GO:0051013	microtubule severing	0.0034	GO:0046686	response to cadmium ion	1.90E-05	GO:0006281	DNA repair	0.00027
GO:0055129	L-proline biosynthetic process	0.0034	GO:0009644	response to high light intensity	0.0015	GO:0018258	protein O-linked glycosylation via hydro...	0.0053
GO:0002239	response to oomycetes	0.0043	GO:0006094	gluconeogenesis	0.0015	GO:0036092	phosphatidylinositol-3-phosphate biosynt...	0.0053
GO:0006412	translation	0.0056	GO:0009826	unidimensional cell growth	0.0024	GO:0007010	cytoskeleton organization	0.00603
GO:0030048	actin filament-based movement	0.0058	GO:0009765	photosynthesis, light harvesting	0.0025	GO:0016024	CDP-diacylglycerol biosynthetic process	0.00706

Table 20. Summary of the top 5 enriched GO terms for the molecular function (MF) domain across isoform repertoire shift categories.

Cold								
A=C → C > A			A=C → A=C			A=C → A > C		
GO ID	Term	P	GO ID	Term	P	GO ID	Term	P
GO:0004185	serine-type carboxypeptidase activity	0.0033	GO:0003729	mRNA binding	8.20E-10	GO:0003984	acetolactate synthase activity	0.00056
GO:0016646	oxidoreductase activity, acting on the C...	0.0068	GO:0003735	structural constituent of ribosome	0.00078	GO:0047334	diphosphate-fructose-6-phosphate 1-phosp...	0.00056
GO:0051119	sugar transmembrane transporter activity	0.0193	GO:0044390	ubiquitin-like protein conjugating enzyme	0.00216	GO:0102786	stearoyl-[acp] desaturase activity	0.00212
GO:0004860	protein kinase inhibitor activity	0.0245	GO:0008143	poly(A) binding	0.00394	GO:0045300	acyl-[acyl-carrier-protein] desaturase a...	0.00212
GO:0004416	hydroxyacylglutathione hydrolase activit...	0.0245	GO:0005244	voltage-gated ion channel activity	0.00477	GO:1990610	acetolactate synthase regulator activity	0.00685
Hot								
A=C → C > A			A=C → A=C			A=C → A > C		
GO ID	Term	P	GO ID	Term	P	GO ID	Term	P
GO:0005516	calmodulin binding	0.0064	GO:0003729	mRNA binding	1.60E-07	GO:0043813	phosphatidylinositol-3,5-bisphosphate 5-...	0.0023
GO:0004724	magnesium-dependent protein serine/threo...	0.0073	GO:0019904	protein domain specific binding	0.00057	GO:1990714	hydroxyproline O-galactosyltransferase a...	0.0053
GO:0003735	structural constituent of ribosome	0.0103	GO:0008266	poly(U) RNA binding	0.00141	GO:0001091	RNA polymerase II general transcription ...	0.0072
GO:0000156	phosphorelay response regulator activity	0.0118	GO:0044390	ubiquitin-like protein conjugating enzyme	0.00254	GO:1990883	rRNA cytidine N-acetyltransferase activi...	0.0072
GO:0043175	RNA polymerase core enzyme binding	0.0163	GO:0005381	iron ion transmembrane transporter activity	0.00521	GO:0004124	cysteine synthase activity	0.0164

Table 21. Summary of the top 5 enriched GO terms for the cellular compartment (CC) domain across isoform repertoire shift categories.

Cold								
A=C → C > A			A=C → A=C			A=C → A > C		
GO ID	Term		GO ID	Term		GO ID	Term	P
GO:0042788	polysomal ribosome	0.0087	GO:0009535	chloroplast thylakoid membrane	7.50E-07	GO:0005948	acetolactate synthase complex	0.00064
GO:0016282	eukaryotic 43S preinitiation complex	0.0124	GO:0022626	cytosolic ribosome	0.00038	GO:0061700	GATOR2 complex	0.02106
GO:0001401	SAM complex	0.014	GO:0048046	apoplast	0.0005	GO:0035859	Seh1-associated complex	0.02106
GO:0071944	cell periphery	0.0141	GO:0009941	chloroplast envelope	0.0005	GO:1903293	phosphatase complex	0.03002
GO:0033290	eukaryotic 48S preinitiation complex	0.0198	GO:1990904	ribonucleoprotein complex	0.00057	GO:0008287	protein serine/threonine phosphatase com...	0.03002
Hot								
A=C → C > A			A=C → A=C			A=C → A > C		
GO ID	Term		GO ID	Term		GO ID	Term	P
GO:0022625	cytosolic large ribosomal subunit	0.00091	GO:0048046	apoplast	2.70E-06	GO:0043813	phosphatidylinositol-3,5-bisphosphate 5-...	0.013
GO:0042788	polysomal ribosome	0.00171	GO:0009941	chloroplast envelope	7.60E-05	GO:1990714	hydroxyproline O-galactosyltransferase a...	0.014
GO:0005753	mitochondrial proton-transporting ATP sy...	0.00932	GO:0009535	chloroplast thylakoid membrane	0.00021	GO:0001091	RNA polymerase II general transcription ...	0.021
GO:0045261	proton-transporting ATP synthase complex...	0.01154	GO:0016607	nuclear speck	0.00033	GO:1990883	rRNA cytidine N-acetyltransferase activi...	0.03
GO:0005762	mitochondrial large ribosomal subunit	0.01154	GO:0009522	photosystem I	0.00044	GO:0004124	cysteine synthase activity	0.037

Figure 10. Isoform models of known cold and heat responsive homeologous pairs. a) COR (cold response) homeologous pair BnaA03g38950 and BnaC03g45990 across the normal and cold conditions. b) Heat shock protein homeologous pair BnaA06g07260 and BnaCnng18070 across the normal and hot conditions.

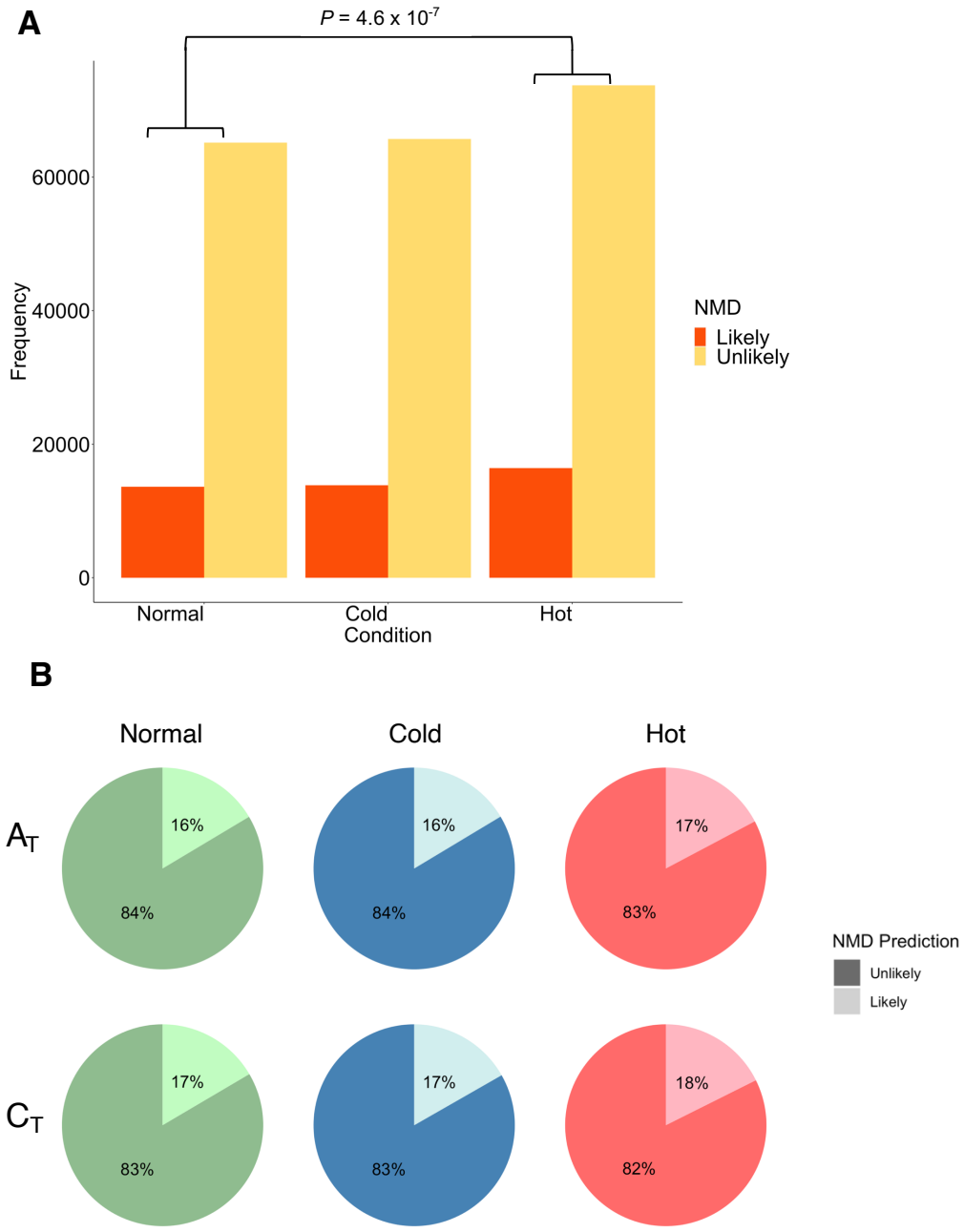

Figure 11. a) Proportion of isoforms predicted to undergo NMD across conditions, n=31,953 genes. A "likely" prediction denotes the case in which a termination codon was detected within the ORF, "unlikely" represents the absence of a premature termination codon detected within the ORF. Frequencies of isoforms in each of the prediction categories varies significantly in response to heat (Fisher's Exact Test, P=4.6 x 10^{-7}). b) Across conditions and subgenomes, n = 8,744 homeologous pairs shared across all 3 conditions. Lighter colour represents "likely", darker colour represents "unlikely." In each condition the subgenomic origin of a given isoform and its prediction of NMD are independent (Fisher's Exact Test, P=0.754 (normal), P=0.287 (cold), P=0.218 (hot)).

Chapter 4: Discussion

4.1 Isoform distributions

Through our use of third generation sequencing, or long-read sequencing, we have generated a comprehensive isoform level transcriptome of *B. napus* leaves from seedlings subjected to heat and cold stress treatments as well as untreated. This analysis facilitates our understanding of the complexities of this allopolyploid transcriptome, including how isoform distributions respond to abiotic stimuli. In regard to isoform length, we found the average length of isoforms within our dataset to be 1,666 bp in length, pooled across each of the three conditions (Figure 3). Our analysis reveals that the largest proportion of genes captured are genes which produce a single isoform; however, a substantial number of genes produce more than one isoform. These distributions result in a range of average estimates of isoforms per gene, from 3.06 in the normal condition, 3.02 in the cold condition, and 3.47 in the heat condition (Table 3). A substantial proportion of the isoforms captured in our dataset represent newly discovered isoforms arising from previously unannotated, or novel, splice junctions. In our structural characterization of our isoform set we discovered that a large proportion of isoforms, across each condition, were characterized as “novel in catalogue” (NIC) or “novel not in catalogue” (NNC), representing isoforms that contain a novel combination of known splice sites or include previously unannotated splice sites (Figure 5). Thus, our Iso-Seq approach has facilitated the discovery of previously uncharacterized isoforms. While this study was in progress, another report of isoforms derived from long-read sequencing in *Brassica napus* was published.[62] In their Iso-Seq analysis of root, leaf, bud, silique, and callus tissue from the *B. napus* cultivar ZS11, the researchers found an average of 3.81 isoforms per gene which resulted from their discovery of

128,967 novel isoforms.[62] Their use of a broad range of tissue types is likely responsible for the increased estimate relative to our own, as some isoforms may be unique to certain tissues.

Interestingly, across other plant species, Iso-Seq data has revealed broad ranges in estimated isoform per gene counts. In their study of allopolyploid cotton, researchers detected 3.93 isoforms per gene which was paired with the discovery of large proportion of genes which produced at least five isoforms.[38] As with our own dataset, their observation is paired with the detection of a substantial amount of novel transcripts.[38] In maize, Iso-Seq also facilitated the detection of a large proportion of novel isoforms leading to an estimated 6.56 isoforms per gene.[48] In their Iso-Seq analysis of red clover (*Trifolium pratense* L.) researchers detected 2.61 isoforms per gene within a pooled sample of leaf, stem, root, and flower tissue.[63] Across several different species, Iso-Seq reveals an expanded range of isoform distributions which likely manifest as a result of its ability to capture novel transcripts.

4.2 Abiotic stress influences isoform distributions

This study revealed that both cold and heat stress can significantly shift isoform composition. When exposed to an environmentally relevant cold stress, genes are more likely to produce fewer isoforms, as detected by a significant shift in the $\log_2$ ratio of isoforms produced in the cold condition relative to the normal condition (Figure 6a). Conversely, when exposed to heat stress, isoform composition shifts in the opposite direction, as genes are more likely to produce a greater number of isoforms when exposed to heat relative to the normal condition (Figure 6a). The observed isoform composition shift detected in the heat stress mirrors what was detected in our AS event analysis. We found that exposure to heat significantly increases the number of AS events (Figure 7b). Despite our detection of a shifted isoform distribution in

response to cold stress, no significant differences were detected in the number of AS event counts in responses to cold stress (Figure 7b). Interestingly, despite the observed heat-responsive increases in the number of AS events, the relative frequencies of AS event types are remarkably conserved, in other words no particular event type seems to be enriched in response to heat stress (Table 9).

As for the molecular mechanism by which AS could modulate an abiotic stress response, and possibly contribute to the shifts described above, emerging evidence has pointed to the role that splicing factors may play in modulating the stress response through their targeting of components of the ABA pathway, a central component of the abiotic stress response.[64] Several splicing regulators have been reported to affect ABA sensitivity providing further evidence that AS is an important mediator in the post-transcriptional regulation of plant stress responses.[65–71] Furthermore, studies have shown that plants treated with splicing inhibitors display responses that mimic stress signals in plants leading to the activation of ABA-inducible promoters and stomatal closure.[72,73] Seemingly, ABA signalling, and thus the abiotic stress response, is fine-tuned through splicing activity among the genes involved within the ABA pathways. Within our dataset, genes which showed a large increase in the number isoforms in response to cold stress were significantly enriched for functions that relate to ABA signalling (Table 6). Thus, genes involved in ABA signalling during the cold stress response undergo increased amounts of alternative splicing leading to a diversified set of isoforms in response to an abiotic stress, providing further evidence for the adaptive role AS may have in fine-tuning the ABA pathways in order to adapt to environmental stress.

Aspects of transcriptional regulation in *B. napus* have previously been shown to be stress-dependent. In their global transcriptomic analysis Lee & Adams (2020) demonstrated overall

gene expression decreases across cold, heat, and drought stresses, with cold stress having the greatest impact.[22] Interestingly, whereas the cold stress showed overall gene expression decreases, AS increased in response to the cold stress, thus representing an inverse relationship between gene expression and AS.[22] This inverse relationship across these fundamental aspects of transcriptional regulation was also observed across the homeologous pairs.[22] Although we did not detect a similar cold-responsive increase in AS events in our own dataset, the observation of AS event counts being inversely correlated to gene expression begs the question whether the characteristic stress-induced isoform composition shifts we observed in this study are paired with relative increases or decreases in overall expression of these isoforms. Other studies have made similar suggestions that AS responses to stress are independent relative to gene expression, showing that genes which are differentially spliced in response to stress are unlikely to be identified as differentially expressed genes.[74,75]

Across other species, AS has also been shown to change in a stress-dependent manner, however the effects of abiotic stress on the actual isoform composition has remained relatively understudied. In their analysis of *P. trichocarpa* researchers found that several abiotic stresses (heat, cold, drought, and salinity) profoundly perturbed isoform profiles.[49] Particularly, they found that stress treatments were often shown to increase or decrease the retention of specific introns, a phenomenon they describe as differential intron retention (DIR).[49] In *Arabidopsis* researchers discovered that approximately 50% of all intron-containing genes were subject to AS under high salinity stress and that AS event counts increased significantly in response to salt stress.[74] Furthermore, in grapevine (*Vitis vinifera*) heat stress was shown to substantially increase the incidence of intron retention events, exon skipping events, and alternative donor/acceptor site events.[76] These studies, in conjunction with our own, provide emerging evidence for the role that

AS can play in modulating the transcriptome in a stress-specific manner and that stress-mediated AS regulation may be distinct from gene expression levels.

4.3 Homeologous gene pair analysis reveals C_T-subgenome skewed isoform distributions

In our analysis of 8,744 homeologous gene pairs across the A_T and C_T subgenomes we found that for any given homeologous pair the C_T homeolog was more likely to produce a greater number of isoforms (Figure 8b). This pattern of uneven isoform distributions across the homeologous pairs was detected in all three conditions. These skewed isoform distributions across the subgenomes might be the result of the slightly increased levels of AS event counts detected in the C_T subgenome (Figure 7b). The majority of homeologous pairs tested fall into what we refer to as the "A_T=C_T" category, in which the differences in isoform counts between the two homeologs in a given pair are within the bounds of two-fold increase or decrease, or in other words, a $\log_2$ ratio between -1 and 1 (Figure 8c). This indicates that despite the negatively skewed distribution towards producing more isoforms from the C_T homeologs, the differences across a majority of pairs were not extreme.

As part of our analysis of homeologous gene pairs we were interested to see whether the isoform count distributions across the pairs changed in a stress-responsive manner. To do this we compared a given isoform count categorization in the normal condition to the abiotic stress condition (Figure 9). We found that a large proportion of homeologous pairs do not retain their original categorization when subjected to either cold or heat stress. Interestingly, the subgenomic origin of the AS events responsible for a category shift are highly dynamic, and a given stress-responsive category shift could be the result of opposing AS events across both subgenomes. In other words, if a homeologous pair moves to having a larger set of C_T isoforms compared to A_T

isoforms (i.e., a $A_T = C_T \rightarrow C_T > A_T$) this is a result of increased number of AS events in the C_T subgenome and reduced number of AS events in the A_T subgenome (Table 17). This underlines the highly complex nature of AS dynamics at work in the polyploid.

To further understand the homeologous pairs that show similar or divergent distributions of isoforms across duplicates, we conducted a GO analysis to examine the predicted functions enriched in each of the groups. Although there was little consistency among the top enriched terms among the homeologous pairs that show divergent isoform distributions, "mRNA binding" (GO:0003729) was among the most enriched terms in each of the three conditions among homeologous pairs which show relatively balanced isoform distributions (Table 14). Additionally, when we examined the enriched functions of homeologous pairs that did or did not change their categorization in response to abiotic stress, genes which remained in $A_T = C_T$ categories despite stress exposure also were enriched for "mRNA binding" (Table 20). Seemingly, this fundamental cellular process is enriched among homeologous gene pairs which show no obvious or extreme differences in the number of isoforms generated across each subgenome. Despite mRNA binding encompassing a broad range of potential functional relevance, it was not enriched in any other category, possibly indicating that the splicing of homeologous gene pairs responsible for such functions remains tightly regulated in order to preserve such functions. The importance of RNA binding, and thus the regulation of RNA metabolism, has previously been shown to play important roles in plant development, growth, and in response to a variety of environmental stimuli.[77,78] Previous research has shown that RNA binding proteins, such as the SR proteins involved in AS, are precisely regulated and are involved in the abiotic stress response in rice.[79] For example, *sr* mutants have revealed that SR splicing factors play a regulatory role in maintaining nutrient homeostasis in rice.[79] While the

critical role that RNA binding proteins play in the cell has been well characterized, our discovery of such functions being enriched among homeologous pairs which do not show divergent isoform counts across the pair is novel and may underline a potential selective pressure for tighter regulation of the various splice isoforms produced by the duplicate pair. The less divergence in splice isoforms that exists between the pair would likely result in a narrower range of functions encoded by the subsequent proteins which may be a result of the precise regulation of RNA binding proteins in a polyploid. More work needs to be done to elucidate the possible evolutionary and functional significance of relatively equal isoform counts produced by a homeologous pair.

The detectable C_T subgenome skew uncovered in our analysis is somewhat analogous to the C_T subgenome bias in the extent of AS under three abiotic stress treatments detected by Lee & Adams (2020).[22] Where they were able to discover increased levels of AS in the C_T subgenome, we showed that a more numerous set of isoforms is generated from the C_T subgenome. Cases in which a certain subgenome exhibits more AS relative to another is not a phenomenon unique to *B. napus*. In their analysis of hexaploid wheat, researchers found that the genes belonging to the B subgenome displayed higher counts of stress-induced AS events relative to the A and D subgenomes.[39]

Although there are many other aspects of transcriptional regulation not assessed in this study, our findings indicate that the C_T sub-transcriptome generated by the C_T homeologs contains more isoforms and thus may be more complex. Similar work in allopolyploid cotton has assessed the relative isoform distributions across subgenomes and has found that slightly more pairs produce more isoforms from the A_T homeolog than the D_T homeolog, relative to pairs showing the opposite distribution.[38] However, they found that the majority of pairs tested did not

show extreme differences in the number of isoforms generated by the two subgenomes, similar to our findings.[38] Furthermore, our work adds to the emerging evidence that homeologous genes in allopolyploids show divergent counts of splicing isoforms. It is these divergent isoform counts across the subgenomes which may lead to subfunctionalization or neofunctionalization, wherein new splice isoforms may provide the opportunity for new functions to evolve between the duplicates or possibly the original function of the homeologous pair to be subdivided between the pair.

4.4 Stress-responsive NMD

We uncovered a significant shift in the proportion of isoforms predicted to be likely targets of NMD in response to heat stress. Our detection predictions were based off whether the resultant isoform had a premature termination codon (PTC) within the ORF. If that criterion was met, the isoform was categorized as a likely target of NMD, and if not, it was designated as an unlikely target (Figure 11a). In the case of heat stress, we are able to piece together our analysis of a stress-responsive increase in AS to a positively-shifted isoform distribution, wherein plants under heat stress produced a higher number of isoforms per gene, to the heat responsive increase in likely NMD targets. Despite this connection, we are unable to trace a similar line through the cold response, or across the subgenomes, where we can detect differences in isoform distributions across conditions or between subgenomes, but unlike the heat stress, our skewed stress-responsive and subgenomic isoform distributions do not manifest skewed, or shifted, proportions of likely or unlikely NMD targets (Figure 11b). In other words, within our dataset, NMD does not act in an independent manner across subgenomes or in response to cold.

The detection of a stress-responsive signal in the increase of NMD candidates provides further evidence for the role that AS-mediated NMD (AS-NMD) may play in an adaptive stress response. It is estimated that one-third of all AS events produce mRNA isoforms containing an in-frame PTC, thus providing evidence that AS plays a prominent role in generating putative NMD substrates.[35] Many have speculated that NMD may be a critical mechanism that plants use to adapt to abiotic stress, through fine-tuned post-transcriptional adjustments of a given set of isoforms which can alter the dosage level of any given transcript.[28,34,36,80–82] Cases of stress-responsive NMD have been discovered in plants. For example, in their analysis of the NMD factor UPF3 in *Arabidopsis*, involved in the proper maintenance of NMD homeostasis, researchers revealed that *upf3* mutants have increased sensitivity to salt stress, emphasizing the important role that UPF3 and thus proper NMD functioning play in adapting to an abiotic stressor.[83] Similar findings have shown relative increases in NMD-targeted splicing variants following salt stress further underlining the ability of NMD to react in a stress-responsive manner and raising the possibility that NMD can act as adaptive mechanism.[82] In their analysis of circadian clock genes in *Arabidopsis,* researchers were able to pinpoint changes in AS leading to NMD-targeted degradation of a select group of these genes following plant exposure to heat and cold stress.[84] Filichkin et al. (2018) showed evidence of differentially retained introns (DIRs) in response to cold, heat, and drought stress and among the genes which harboured the DIRs, many resulted in mRNAs with unusually long 3' untranslated regions (3' UTR) thus making them likely targets of NMD.[49] This evidence, along with our own, adds to an emerging body of literature that implicates the role of NMD as a possible mediator of the abiotic stress response.

Our detection method represents a primary, or top-level, assessment of putative NMD targets. Although our full-length reads allow us to directly detect a PTC within an ORF, which

strengthens the relevance of our approach, this does not guarantee an isoform is targeted to an NMD pathway or results in the subsequent production of a truncated non-functional protein.[85] Isoforms with in-frame PTCs may be rescued through the reversible sequestration of splicing intermediates, by which IR-containing isoforms escape cytoplasmic NMD via sequestration into the nucleus where the removal of the retained introns and subsequent release of mature mRNAs occurs.[85,86] Conversely, studies of *Arabidopsis* NMD mutants have revealed that a small proportion NMD-sensitive transcripts do not contain the characteristic NMD-inducing in-frame PTCs indicating that NMD-eliciting features likely encompass more than just in-frame PTCs.[36] Although our data presents further evidence for an adaptive AS-NMD response to environmental stressors, more work needs to be done to investigate these processes in finer detail. Detailed expression analysis of NMD factors in *B. napus*, such as UP1, UP2, and UP3 homologs, will offer further insights into how NMD is precisely regulated. Furthermore, similar analyses of splicing factors in *B. napus* could help uncover precisely how NMD-targeted degradation could be regulated via AS and thus help determine exactly how AS-NMD operates in the abiotic stress response within this important crop species.

4.5 Conclusion

Overall, our study contributes to a fuller understanding of the complex transcriptional dynamics at play in the allopolyploid *B. napus* in the context of common environmental stressors. Our use of third-generation sequencing, or long-read sequencing, has enabled us to accurately assess the dynamic nature of AS and its overall effects on the transcriptome. Through the capture and sequencing of full-length mRNA isoforms we were able to precisely determine how AS activity manifests a complex transcriptome. We documented differential levels of AS in

response to heat and cold stress, and across subgenomes, including shifted isoform distributions that may facilitate NMD, an important aspect of post-transcriptional regulation. Iso-Seq has allowed us to investigate these dynamics with a level of certainty previously unavailable using a short-read RNA-seq approach, as direct isoform detection removes ambiguities introduced when short-reads are reconstructed to resolved isoforms.[43–45] Our study represents the first to use Iso-Seq to assess post-transcriptional dynamics in response to environmentally relevant heat and cold stresses in an allopolyploid plant species.

Our findings reveal a fuller picture of the effects that AS has on the transcriptome, including generating a more complex transcriptome in response to heat stress, and a reduced set of isoforms in response to cold stress. This adds to the complex body of literature that describes how transcriptional dynamics occur in a stress-responsive manner and thus facilitate a fuller understanding for how aspects of AS and divergent isoform patterns may play a role in the abiotic stress response.

We have provided evidence that uneven distributions of AS and gene expression across the subgenomes previously described by other researchers are also unequally distributed at the level of resolved isoforms, further underlining the highly complex nature of allopolyploid transcriptional regulation. Divergence across subgenomes has been well characterized among aspects of gene expression and AS, not only in *B. napus* but other allopolyploids as well and our analysis provides further evidence that this divergence persists to the level of the isoform.[22,38,39] Despite these differences across the subgenomes our approach did not lead us to detect subsequent differences in the predicted level of NMD across homeologous pairs, suggesting that NMD does not respond in a differential manner across homeologous pairs.

Although this study reveals the complexities of the allopolyploid transcriptome at the isoform level and how it responds to environmental stimuli, more work is needed to fully understand these dynamics. Due to our novel Iso-Seq approach, our sensitivity was decreased as we captured a reduced subset of the known gene models in *B. napus*. Furthermore, when assessing dynamics across homeologous pairs we are restricted to a reduced set of genes as our analytical approach required the presence of a complete homeologous pair captured in each of the three conditions. Future work will need to be done to investigate the molecular and mechanistic processes in order to ascribe functional significance to the isoform distribution shifts we observed. Our study provides a useful and novel glimpse into how environmentally relevant abiotic stimuli can produce skewed isoform distributions as a result of AS in this agronomically important polyploid species.

References

1. Soltis DE, Visger CJ, Soltis PS. The polyploidy revolution then...and now: Stebbins revisited. *Am J Bot*. 2014;101(7):1057-1078. doi:10.3732/ajb.1400178
2. Renny-Byfield S, Wendel JF. Doubling down on genomes: Polyploidy and crop plants. *Am J Bot*. 2014;101(10):1711-1725. doi:10.3732/ajb.1400119
3. Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS (Pat). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann Bot*. 2017;120(2):183-194. doi:10.1093/aob/mcx079
4. Jiao Y, Wickett NJ, Ayyampalayam S, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature*. 2011;473(7345):97-100. doi:10.1038/nature09916
5. Blanc G, Hokamp K, Wolfe KH. A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Res*. 2003;13(2):137-144. doi:10.1101/gr.751803
6. Jiao Y, Leebens-Mack J, Ayyampalayam S, et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol*. 2012;13(1):R3. doi:10.1186/gb-2012-13-1-r3
7. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A*. 2009;106(33):13875-13879. doi:10.1073/pnas.0811575106
8. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. *Plant Physiol*. 2016;171(4):2294-2316. doi:10.1104/pp.16.00523
9. Adams KL, Wendel JF. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol*. 2005;8(2):135-141. doi:10.1016/j.pbi.2005.01.001

10. Comai L. Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol Biol*. 2000;43(2-3):387-399. doi:10.1023/a:1006480722854

11. Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. The "Polyploid Hop": Shifting Challenges and Opportunities Over the Evolutionary Lifespan of Genome Duplications . *Front Ecol Evol* . 2018;6:117. https://www.frontiersin.org/article/10.3389/fevo.2018.00117

12. Bird KA, Niederhuth CE, Ou S, et al. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. *New Phytol*. Published online 2020. doi:10.1111/nph.17137

13. Flagel L, Udall J, Nettleton D, Wendel J. Duplicate gene expression in allopolyploid Gossypiumreveals two temporally distinct phases of expression evolution. *BMC Biol*. 2008;6(1):16. doi:10.1186/1741-7007-6-16

14. Buggs RJA, Zhang L, Miles N, et al. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr Biol*. 2011;21(7):551-556. doi:10.1016/j.cub.2011.02.016

15. Liu Z, Adams KL. Expression Partitioning between Genes Duplicated by Polyploidy under Abiotic Stress and during Organ Development. *Curr Biol*. 2007;17(19):1669-1674. doi:10.1016/j.cub.2007.08.030

16. Dong S, Adams KL. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytol*. 2011;190(4):1045-1057. doi:10.1111/j.1469-8137.2011.03650.x

17. Schnable JC, Springer NM, Freeling M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci U S A*.

2011;108(10):4069-4074. doi:10.1073/pnas.1101368108

18. Li A, Liu D, Wu J, et al. mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell*. 2014;26(5):1878-1900. doi:10.1105/tpc.114.124388

19. Edger PP, Smith R, McKain MR, et al. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell*. 2017;29(9):2150-2167. doi:10.1105/tpc.17.00010

20. Yoo MJ, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb)*. 2013;110(2):171-180. doi:10.1038/hdy.2012.94

21. Chalhoub B, Denoeud F, Liu S, et al. Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. *Science (80-)*. 2014;345(6199):950-953. doi:10.1126/science.1253435

22. Lee JS, Adams KL. Global insights into duplicated gene expression and alternative splicing in polyploid Brassica napus under heat, cold, and drought stress. *Plant Genome*. 2020;13(3). doi:10.1002/tpg2.20057

23. Lee H, Chawla HS, Obermeier C, Dreyer F, Abbadi A, Snowdon R. Chromosome-Scale Assembly of Winter Oilseed Rape Brassica napus . *Front Plant Sci* . 2020;11:496. https://www.frontiersin.org/article/10.3389/fpls.2020.00496

24. Wu J, Lin L, Xu M, et al. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid Brassica napus. *BMC Genomics*. 2018;19(1):586. doi:10.1186/s12864-018-4966-5

25. Coate JE, Powell AF, Owens TG, Doyle JJ. Transgressive physiological and

transcriptomic responses to light stress in allopolyploid Glycine dolichocarpa (Leguminosae). *Heredity (Edinb)*. 2013;110(2):160-170. doi:10.1038/hdy.2012.77

26. Barash Y, Calarco JA, Gao W, et al. Deciphering the splicing code. *Nature*. 2010;465(7294):53-59. doi:10.1038/nature09000

27. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010;463(7280):457-463. doi:10.1038/nature08909

28. Reddy ASN, Marquez Y, Kalyna M, Barta A. Complexity of the alternative splicing landscape in plants. *Plant Cell*. 2013;25(10):3657-3683. doi:10.1105/tpc.113.117523

29. Lopato S, Waigmann E, Barta A. Characterization of a Novel Arginine/Serine-Rich Splicing Factor in Arabidopsis. *Plant Cell*. 1996;8(12):2255-2264. doi:10.1105/tpc.8.12.2255

30. Lee BH, Kapoor A, Zhu J, Zhu JK. Stabilized1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in Arabidopsis. *Plant Cell*. 2006;18(7):1736-1749. doi:10.1105/tpc.106.042184

31. Monaghan J, Xu F, Gao M, et al. Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathog*. 2009;5(7). doi:10.1371/journal.ppat.1000526

32. Sugliani M, Brambilla V, Clerkx EJM, Koornneef M, Soppe WJJ. The conserved splicing factor SUA controls alternative splicing of the developmental regulator ABI3 in arabidopsis. *Plant Cell*. 2010;22(6):1936-1946. doi:10.1105/tpc.110.074674

33. Wang BB, Brendel V. Genomewide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci U S A*. 2006;103(18):7175-7180. doi:10.1073/pnas.0602039103

34. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative

splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A*. 2003;100(1):189-192. doi:10.1073/pnas.0136770100

35. Smith JE, Baker KE. Nonsense-mediated RNA decay - a switch and dial for regulating gene expression. *BioEssays*. 2015;37(6):612-623. doi:10.1002/bies.201500007

36. Kalyna M, Simpson CG, Syed NH, et al. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res*. 2012;40(6):2454-2469. doi:10.1093/nar/gkr932

37. Zhou R, Moshgabadi N, Adams KL. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc Natl Acad Sci U S A*. 2011;108(38):16122-16127. doi:10.1073/pnas.1109551108

38. Wang M, Wang P, Liang F, et al. A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol*. 2018;217(1):163-178. doi:10.1111/nph.14762

39. Liu Z, Qin J, Tian X, et al. Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (Triticum aestivum L.). *Plant Biotechnol J*. 2018;16(3):714-726. doi:10.1111/pbi.12822

40. Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science (80-)*. 2009;323(5910):133-138. doi:10.1126/science.1162986

41. An D, Cao HX, Li C, Humbeck K, Wang W. Isoform Sequencing and State-of-Art Applications for Unravelling Complexity of Plant Transcriptomes. *Genes (Basel)*. 2018;9(1). doi:10.3390/genes9010043

42. Zhao L, Zhang H, Kohnen M V., Prasad KVSK, Gu L, Reddy ASN. Analysis of transcriptome and epitranscriptome in plants using pacbio iso-seq and nanopore-based

direct RNA sequencing. *Front Genet*. 2019;10(MAR):1-14. doi:10.3389/fgene.2019.00253

43. Hardwick SA, Joglekar A, Flicek P, Frankish A, Tilgner HU. Getting the entire message: Progress in isoform sequencing. *Front Genet*. 2019;10(JUL):1-10. doi:10.3389/fgene.2019.00709

44. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511-515. doi:10.1038/nbt.1621

45. Wang X, You X, Langer JD, et al. Full-length transcriptome reconstruction reveals a large diversity of RNA and protein isoforms in rat hippocampus. *Nat Commun*. 2019;10(1):5009. doi:10.1038/s41467-019-13037-0

46. Steijger T, Abril JF, Engström PG, et al. Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods*. 2013;10(12):1177-1184. doi:10.1038/nmeth.2714

47. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol*. 2013;31(11):1009-1014. doi:10.1038/nbt.2705

48. Wang B, Tseng E, Regulski M, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun*. 2016;7:1-13. doi:10.1038/ncomms11708

49. Filichkin SA, Hamilton M, Dharmawardhana PD, et al. Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching. *Front Plant Sci*. 2018;9(February):1-17. doi:10.3389/fpls.2018.00005

50. Nagaharu U. Genome Analysis in Brassica with Special Reference to the Experimental

Formation of B. Napus and Peculiar Mode of Fertilization.e. *Japanese J Bot*. 1935;7:389-452.

51. Rimmer SR, Scarth R, McVetty PBE. Sentry summer rape. *Can J Plant Sci*. 1998;78(4):615-616. doi:10.4141/P97-128

52. Bhagwat AA, Ying ZI, Karns J, Smith A. Determining RNA quality for NextGen sequencing: some exceptions to the gold standard rule of 23S to 16S rRNA ratio§. *Microbiol Discov*. 2013;1(1):10. doi:10.7243/2052-6180-1-10

53. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191

54. Tseng E. Cupcake ToFU. Published online 2019.

55. Gordon SP, Tseng E, Salamov A, et al. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One*. 2015;10(7):1-15. doi:10.1371/journal.pone.0132628

56. Tardaguila M, de la Fuente L, Marti C, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res*. 2018;28(3):396-411. doi:10.1101/gr.222976.117

57. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Published online 2013.

58. Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyras E. Leveraging transcript quantification for fast computation of alternative splicing profiles. *Rna*. 2015;21(9):1521-1531. doi:10.1261/rna.051557.115

59. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. Published online 2020.

60. Miao W, Gel R, Gastwirth JL. A new test of symmetry about an unkown median. *Random Walk, Sequential Analysis and Related Topics*. WORLD SCIENTIFIC; 2006:199-214. doi:doi:10.1142/9789812772558_0013

61. Sahraeian SME, Mohiyuddin M, Sebra R, et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat Commun*. 2017;8(1):59. doi:10.1038/s41467-017-00050-4

62. Yao S, Liang F, Gill RA, et al. A global survey of the transcriptome of allopolyploid Brassica napus based on single-molecule long-read isoform sequencing and Illumina-based RNA sequencing data. *Plant J*. 2020;103(2):843-857. doi:10.1111/tpj.14754

63. Chao Y, Yuan J, Li S, Jia S, Han L, Xu L. Analysis of transcripts and splice isoforms in red clover (Trifolium pratense L.) by single-molecule long-read sequencing. *BMC Plant Biol*. 2018;18(1):300. doi:10.1186/s12870-018-1534-8

64. Laloum T, Martín G, Duque P. Alternative Splicing Control of Abiotic Stress Responses. *Trends Plant Sci*. 2018;23(2):140-150. doi:10.1016/j.tplants.2017.09.019

65. Kim Y-O, Pan S, Jung C-H, Kang H. A zinc finger-containing glycine-rich RNA-binding protein, atRZ-1a, has a negative impact on seed germination and seedling growth of Arabidopsis thaliana under salt or drought stress conditions. *Plant Cell Physiol*. 2007;48(8):1170-1181. doi:10.1093/pcp/pcm087

66. Hugouvieux V, Kwak JM, Schroeder JI. An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. *Cell*. 2001;106(4):477-487. doi:10.1016/s0092-8674(01)00460-3

67. Papp I, Mur LA, Dalmadi A, Dulai S, Koncz C. A mutation in the Cap Binding Protein 20 gene confers drought tolerance to Arabidopsis. *Plant Mol Biol*. 2004;55(5):679-686.

doi:10.1007/s11103-004-1680-2

68. Daszkowska-Golec A, Skubacz A, Marzec M, et al. Mutation in HvCBP20 (Cap Binding Protein 20) Adapts Barley to Drought Stress at Phenotypic and Transcriptomic Levels. *Front Plant Sci*. 2017;8:942. doi:10.3389/fpls.2017.00942

69. Zhang Z, Zhang S, Zhang Y, et al. Arabidopsis floral initiator SKB1 confers high salt tolerance by regulating transcription and pre-mRNA splicing through altering histone H4R3 and small nuclear ribonucleoprotein LSM4 methylation. *Plant Cell*. 2011;23(1):396-411. doi:10.1105/tpc.110.081356

70. Lee B, Kapoor A, Zhu J, Zhu J-K. STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in Arabidopsis. *Plant Cell*. 2006;18(7):1736-1749. doi:10.1105/tpc.106.042184

71. Huang C-F, Miki D, Tang K, et al. A Pre-mRNA-splicing factor is required for RNA-directed DNA methylation in Arabidopsis. *PLoS Genet*. 2013;9(9):e1003779. doi:10.1371/journal.pgen.1003779

72. Ling Y, Alshareef S, Butt H, et al. Pre-mRNA splicing repression triggers abiotic stress signaling in plants. *Plant J*. 2017;89(2):291-309. doi:10.1111/tpj.13383

73. AlShareef S, Ling Y, Butt H, Mariappan KG, Benhamed M, Mahfouz MM. Herboxidiene triggers splicing repression and abiotic stress responses in plants. *BMC Genomics*. 2017;18(1):260. doi:10.1186/s12864-017-3656-z

74. Ding F, Cui P, Wang Z, Zhang S, Ali S, Xiong L. Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. *BMC Genomics*. 2014;15(1):431. doi:10.1186/1471-2164-15-431

75. Dong C, He F, Berkowitz O, et al. Alternative Splicing Plays a Critical Role in

Maintaining Mineral Nutrient Homeostasis in Rice (Oryza sativa). *Plant Cell*. 2018;30(10):2267-2285. doi:10.1105/tpc.18.00051

76. Jiang J, Liu X, Liu C, Liu G, Li S, Wang L. Integrating Omics and Alternative Splicing Reveals Insights into Grape Response to High Temperature. *Plant Physiol*. 2017;173(2):1502 LP - 1518. doi:10.1104/pp.16.01305

77. Floris M, Mahgoub H, Lanet E, Robaglia C, Menand B. Post-transcriptional regulation of gene expression in plants during abiotic stress. *Int J Mol Sci*. 2009;10(7):3168-3185. doi:10.3390/ijms10073168

78. Lorković ZJ. Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci*. 2009;14(4):229-236. doi:10.1016/j.tplants.2009.01.007

79. Ganie SA, Reddy ASN. Stress-Induced Changes in Alternative Splicing Landscape in Rice: Functional Significance of Splice Isoforms in Stress Tolerance. *Biology (Basel)*. 2021;10(4). doi:10.3390/biology10040309

80. Filichkin S, Priest HD, Megraw M, Mockler TC. Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr Opin Plant Biol*. 2015;24:125-135. doi:https://doi.org/10.1016/j.pbi.2015.02.008

81. Chaudhary S, Khokhar W, Jabre I, et al. Alternative Splicing and Protein Diversity: Plants Versus Animals. *Front Plant Sci*. 2019;10:708. doi:10.3389/fpls.2019.00708

82. Drechsel G, Kahles A, Kesarwani AK, et al. Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. *Plant Cell*. 2013;25(10):3726-3742. doi:10.1105/tpc.113.115485

83. Vexler K, Cymerman MA, Berezin I, et al. The Arabidopsis NMD Factor UPF3 Is

Feedback-Regulated at Multiple Levels and Plays a Role in Plant Response to Salt Stress. *Front Plant Sci*. 2016;7:1376. doi:10.3389/fpls.2016.01376

84. Kwon Y-J, Park M-J, Kim S-G, Baldwin IT, Park C-M. Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. *BMC Plant Biol*. 2014;14(1):136. doi:10.1186/1471-2229-14-136

85. Boothby TC, Zipper RS, van der Weele CM, Wolniak SM. Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. *Dev Cell*. 2013;24(5):517-529. doi:10.1016/j.devcel.2013.01.015

86. Boothby TC, Wolniak SM. Masked mRNA is stored with aggregated nuclear speckles and its asymmetric redistribution requires a homolog of mago nashi. *BMC Cell Biol*. 2011;12(1):45. doi:10.1186/1471-2121-12-45