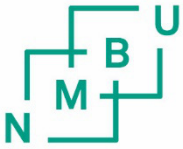


Norges miljø- og
biovitenskapelige
universitet

Master's Thesis 2016 60 ECTS
Department of Chemistry, Biotechnology and Food Science

Homeolog Regulation in Hexaploid Wheat

Ine Birgitta Hallsberg
Bioinformatics and applied statistics

Acknowledgements

This thesis is written as a part of my master's degree in bioinformatics at the Norwegian University of Life Sciences, Department of Chemistry, Biotechnology and Food Sciences. The work with this thesis was carried out from spring 2015 to spring 2016

I would like to thank my supervisors, Torgeir Rhoden Hvidsten and Simen Rød Sandve, for all their help during my work with this thesis. Especially I would like to thank Torgeir for introducing me to the topic of wheat gene expression. I would also like to thank Øystein for helping with the final improvements and for always being ready to provide LaTex support.

Ski, May 2016

Ine Birgitta Hallsberg

Abstract

Bread wheat (Triticum aestivum L.) is an allohexaploid hybrid composed of three closely related diploid subgenomes (AABBDD, 2n=6x=42). For most genes, bread wheat therefore has three gene copies, referred to as homeolog triplets. Normally there is no need for maintaining several copies of a gene performing the same function. This means that two of the copies are free to evolve, leading to diverged gene expression between homeologs. Previous studies have found that in some of the triplets there is differential expression between homeologs, but which subgenome that is preferentially expressed varies between tissues. However, no one has really focused on developmental regulation of the subgenomes, which is the main topic of this thesis.

In this thesis two RNA-seq datasets from different tissues and developmental stages were used to investigate subgenome divergence both in steady state expression levels, and in developmental regulation through tests for temporal changes in expression within tissues. The proportion of homeologous genes with differential expression between pairs of subgenomes varied from 25% to 50% depending on the tissue. In all tissues, the largest number of differentially expressed homeologs was between the A and B subgenomes. This suggests that these two subgenomes are least related, thereby confirming previous results regarding the evolutionary history of bread wheat. Interestingly, although the expression levels varied between homeologs on different subgenomes, the developmental regulation was very similar. For example, a triplet could have significantly higher leaf expression in the D subgenome compared to the A and B subgenomes, while still exhibiting significant up-regulation in *all* subgenomes from one time point to the next during leaf development.

The highly similar regulation of bread wheat subgenomes demonstrated in this thesis, coupled with the fact that subgenome expression levels often are highly variable, suggests that epigenetic regulation is a more important mechanism underlying subgenome specific differences in expression levels, compared to regulatory logics hard-wired in the DNA.

Sammendrag

Brødhvete (Triticum aestivum L.) er en allohexaploid hybrid bestående av tre nært beslektede diploide subgenomer (AABBDD, 2n=6x=42). Hvete har derfor tre kopier av de fleste gener. Disse blir omtalt som homeologe tripletter. Vanligvis er det ikke behov for å ivareta tre genkopier som utfører den samme funksjonen. Dette gir to av kopiene muligheten til å evolvere, som igjen kan føre til divergering i genuttrykk mellom homeologe gener. Tidligere studier har vist at det er differensielt uttrykk mellom homeologer i noen tripletter, men hvilket subgenom som er høyest uttrykt varierer mellom vev. Det vært lite fokus på hvordan subgenomene reguleres under utviklingen av hvete, som er hovedtemaet for denne oppgaven.

I denne oppaven ble det brukt to RNA-seq datasett med genuttrykk fra ulike vev og utviklingsstadier til å undersøke hvor like subgenomene er med hensyn til både hvor høyt genene er uttrykt og regulering av genene mellom ulike tidspunkter under hvetens utvikling. Andelen homeologe gener med differensielt uttrykk varierte mellom 25% og 50%, avhengig av vev. Antallet var høyest mellom A og B subgenomene i alle vevene, noe som tyder på at disse to subgenomene er minst i slekt og som dermed bekrefter resultater fra tidligere studier om hvetens evolusjonære historie. Et interessant resultat var at selv om uttrykksnivået varierte mellom homeologer på ulike subgenomer, så var reguleringen av subgenomene veldig lik. Som et eksempel kunne det være mye høyere uttrykk i D subgenomet i en triplett sammenliknet med A og B i et vev, selv om *alle* subgenomene var signifikant høyere uttrykt mellom to tidspunkter i utviklingen av vevet.

Det at subgenomene er veldig likt regulert, i kombinasjon med at de ofte har veldig ulike uttryksnivåer, tyder på at epigenetisk regulering er en viktigere mekanisme bak ulikehetene vi ser i uttrykk mellom subgenomene enn forskjeller i genenes regulatoriske elementer.

Contents

1 Introduction

1.1 The evolution of hexaploid bread wheat

Triticum aestivum L. (AABBDD, 2n=6x=42), or bread wheat, is an angiosperm and a member of the Poaceae family (the grasses). Wheat alone accounts for about 20% of the worlds food consumption (Pfeifer et al., 2014), and wheat together with rice and maize make up the most important cereal crops worldwide. As we face challenges like an ever increasing food demand and a changing climate, studying gene expression in wheat is important to gain knowledge of underlying molecular biology that can be used to further improve upon its agricultural characteristics.

Like many other crop plants, wheat is polyploid. A polyploid organism is defined as having more than two sets of homologous chromosomes. The chromosome sets may be from the same species (autopolyploids), or from closely related species (allopolyploids). Bread wheat is an allohexaploid composed of three closely related subgenomes, denoted the A-, B- and D-subgenomes, which were originally derived from three diploid (2n=2x=14) species within the tribe Triticeae (Petersen et al., 2006).

It is believed that bread wheat originated through a hybridization between T. turgidum (AABB) and Ae. tauschii (DD) with the beginning of agriculture about 10 000 years ago (Petersen et al., 2006; Salamini et al., 2002), and that the A and B genomes gave rise to the D genome (Ae. tauschii, DD) through homoploid species hybridization about 5.5 million years ago (Marcussen et al., 2014) (Figure 1).

1.2 Genetic effects of polyploidy

It is believed that a large part of the success of wheat as an agricultural species is due to it being an allohexaploid (The International Wheat Genome Sequencing Consortium (IWGSC), 2014). Polyploidization often reduces fertility and survival, but if this is overcome then polyploid species can be very successful (Van de Peer et al., 2009). This is seen in angiosperms in particular, as most angiosperm species have evolved from polyploid ancestors (Van de Peer et al., 2009).

One of the difficulties newly formed polyploids have to face is the tendency to produce aneuploid cells during mitosis and meisosis (Comai, 2005). This is an an error in cell division that results in cells with an abnormal number of chromosomes. Also, in order to secure fertility, it must be ensured that homologous chromosomes (from the same subgenome) rather than homeologous chromosomes (from different subgenomes) pair up during meiosis (Feldman

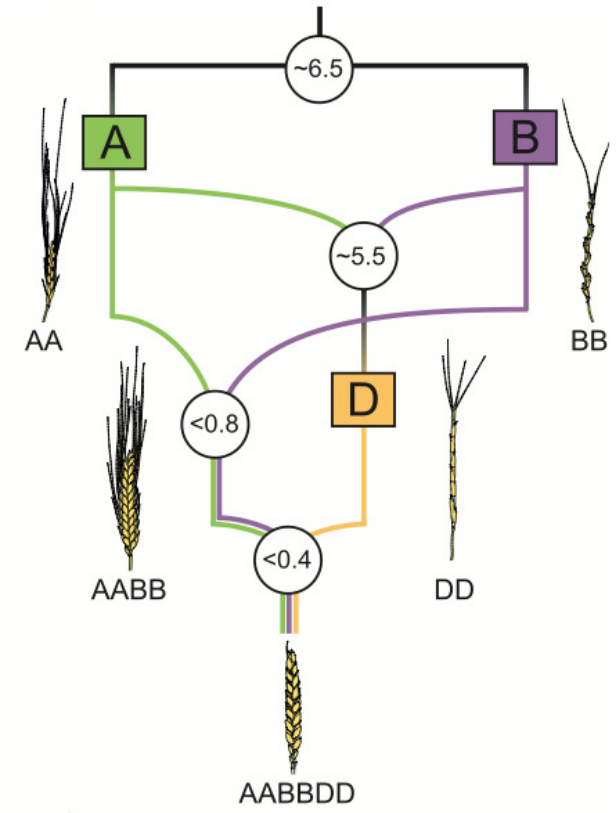


Figure 1: Model of the phylogenetic history of bread wheat. The numbers in the white circles give the approximate dates for divergence and hybridizations in units of million years ago. (Marcussen et al., 2014)

et al., 2012). In wheat homologous pairing during meiosis is achieved by Ph genes (Martinez-Perez et al., 2001).

Furthermore, the need to deal with redundant and conflicting patterns of gene expression leads to various genetic and epigenetic changes and changes in gene expression in new polyploids (Feldman et al., 2012). This may lead to e.g. loss of genome sequences, favored expression of one homeologous gene over the other(s) in different tissues and developmental stages, and global expression dominance of a subgenome (Pfeifer et al., 2014).

The changes that happen after polyploid hybridization can lead to advantages that give polyploids an increased ability to adapt to new and/or extreme environmental conditions, and giving them the opportunity to survive in habitats that are not accessible to their ancestors (Hegarty et al., 2008). There are several possible reasons for this. One reason is that polyploidy can increase heterozygosity, which again can lead to increased variation in

the regulation and expression of genes (Van de Peer et al., 2009). Also, the genetic and epigenetic changes that happen after hybridization can lead to transgressive gene expression, i.e. extreme gene expression that is outside the range of gene expression in either parent (Hegarty et al., 2008; Van de Peer et al., 2009). If this novel gene expression leads to a phenotype that is favourable, the polyploid offspring can have higher fitness than the parents.

Another reason for the success of polyploid species has to do with the genetic redundancy that follows polyploidization. Normally, there is no need for maintaining several copies of a gene performing the same function, which means that homeologous genes can face several possible fates. Diverged homeologs can be subject to subfunctionalization (partitioning of the ancestral functions among the homeologous genes), neofunctionalization (evolution of novel functions for one of the homeologs) (Chaudhary et al., 2009), or one of the homeologs can become completely silenced (non-functionalization) (Yoo et al., 2014). Homeolog expression divergence can vary between tissues and during development. The genetic redundancy of polyploids also increases the ability to tolerate damaging mutations (Pumphrey et al., 2009).

Lastly, an increase in the number of regulatory genes is of the essence when developing more complex biological systems (Van de Peer et al., 2009). A biological system consists of several interacting components and can be e.g. cells, tissues, the nervous system or whole organisms. Having more regulatory genes is advantageous also because the change in gene regulation can happen much faster than changes in the gene sequence, leading to faster adaptation. Furthermore, such an increase in the number of regulatory genes is much easier to accomplish through the duplication of whole genomes, as is the case with polyploidy, than with the duplication of single genes. For example, in the Arabidopsis lineage it was shown that more than 90% of the increase in regulatory genes for the past 150 million years comes from whole genome duplications (Van de Peer et al., 2009). Birchler et al. (2005) claims that the reason that an increase in regulatory genes preferentially happens through whole genome duplications, is that duplication of a single gene will lead to an imbalance in gene dosages between regulatory genes involved in the same regulatory pathway. Eventually though, mutations can alter the need to preserve this balance in dosages and the regulatory genes are free to evolve, leading to greater variation (Van de Peer et al., 2009; Birchler et al., 2005).

1.3 Gene expression in allopolyploid wheat

Many early studies have used synthetic polyploids (made in the lab) to find out what happens shortly after polyploidization. A big advantage of this

method is that the diploid parents will be available for comparison. Several of these studies found that gene silencing by cytosine methylation and gene loss were common immediate responses of the wheat genome after hybridization (Shaked et al., 2001; Kashkush et al., 2002; He et al., 2003). Furthermore, it has been demonstrated that sequence elimination affects a large fraction of the genome and that the results are reproducible (Shaked et al., 2001). To determine the extent of sequence elimination in polyploids, Ozkan et al. (2003) determined the DNA content of six newly synthesized wheat allopolyploids. They found that the allopolyploids had a genome size significantly smaller than the expected additive value of the two parents, and that the change was rapid.

While these studies have mainly looked at what happens to the genome as a whole, later studies have focused more on homeolog specific expression. Mochida et al. (2004) developed a method for SNP analysis in wheat that uses sequencing in combination with a nullisomic-tetrasomic series of hexaploid wheat to determine the expression profiles of homoeologous genes in ten different tissues. Nullisomic-tetrasomic wheat lacks both representatives of a pair of homologous chromosomes from one subgenome, but has two pairs of the chromosome from one of the other subgenomes. In this way the total number of chromosomes is not altered. They found that out of 90 sets of homeologous genes (triplets), 11 were silenced in one of the three subgenomes. No preference for silencing of particular subgenomes or chromosomes were observed. Of the triplets that were expressed in all three subgenomes, the number that showed significant preferential expression in a subgenome varied between the tissues and there was no overall preference for a certain subgenome.

Bottley et al. (2006) used single strand conformation polymorphism (SSCP) to investigate the extent of homeolog specific gene silencing in hexaploid wheat. SSCP analysis is based on the fact that small differences between homeologous sequences can cause differences in folding properties of single stranded DNA molecules. Such differences can then be measured using gel electrophoresis. SSCP was used to distinguish between both homoeologous wheat gDNA sequences and homeologous cDNA sequences. One of the homeologs was determined to be silenced if it was represented in the SSCP profile of gDNA, but not in the profile of the equivalent cDNA. The result showed that in leaf about 27% of the triplets had one silenced homeolog and in root about 26% had one (or rarely two) silenced homeolog. Reciprocal silencing, where one of the homeologs is transcribed in leaf but not root (and vice versa) was shown to be common. They did not find any evidence of a particular subgenome being preferentially silenced over the others. Homeolog specific silencing and unequal expression was also found in cotton using the same

method (Adams et al., 2003).

Bottley and Koebner (2008) also used SSCP to determine the extent to which homoeolog specific silencing can vary between genotypes within a species. 15 gene triplets in 16 varieties of Triticum aestivum were investigated. In 8 of the 15 triplets at least one of the three homoeologous genes varied in expression in either root or leaf, and only two varieties shared the same pattern of silencing.

Two studies have shown examples of methylation and histone modification being the cause of differences in homeolog expression in wheat. Shitsukawa et al. (2007) studied gene expression in Class E MADS box genes. They found that the B homeolog was expressed significantly lower than the A and D homeologs. They further showed that silencing of the B homeolog was not caused by alterations in cis-elements (transcription factor binding sites), but rather by cytosine methylation. While both the A and D homeologs are expressed, sequence changes in the A homeolog has caused it to not produce functional proteins, and of the three homeologs only the D homeolog is functional. Hu et al. (2013) found that regulation mechanisms involving cytosine methylation and histone modifications caused expression levels to be significantly different between the three TaEXPA1 homoeologs. The expression levels varied in different tissues and at different developmental stages. As with the the other study mentioned, no differences were found in the cis-elements in the promoter sequences for the three homoeologous genes.

Several microarray studies have investigated the occurence of nonadditive expression in polyploid wheat by measuring the expression levels in synthetic allopolyploid wheat and their parent species. Nonadditive gene expression means that the total expression level for all homeologous genes deviate from the average of the parental expression levels. There are at least three possibilities for nonadditive expression in polyploids. The first is expression level dominance, where the total level of gene expression for the homeologs is similar to the expression level in one of the parents (Yoo et al., 2014). The second is transgressive expression, where the total level of gene expression is lower or higher than in both parents (Yoo et al., 2014). And the third is homeolog expression bias, where the homeologs contribute differently to the total gene expression (Yoo et al., 2014). An extreme version of homeolog expression bias is when a homeolog is completely silenced. Both Pumphrey et al. (2009), Akhunova et al. (2010) and Chagué et al. (2010) compared gene expression in synthetic T. aestivum to the mid-parent value (average of the parents) found from parental T. turgidum and Ae. tauschii expression levels. Pumphrey et al. (2009) found that approximately 16% of genes displayed nonadditive expression, Akhunova et al. (2010) found that 19% of genes had nonadditive expression, while Chagué et al. (2010) found that

only about 7% of genes had nonadditive expression. Nonadditive expression in wheat has also been studied using cDNA-amplified fragment length polymorphism (AFLP). This was done by He et al. (2003) who found that 7.7% of genes had nonadditive expression (Chagué et al., 2010). For comparison, the occurence of nonadditive expression has been found in several polyploid species including about 5% of genes in Arabidopsis synthetic allotetraploids and 1–6.1% in cotton (Chagué et al., 2010).

1.4 RNA-seq studies of homeolog specific gene expression

Although microarrays are widely used, they are not ideal for studying gene expression. Microarray technology is based on hybridizing RNA from collected samples to probes on an array, and hence we need to have prior knowledge of the sequences we wish to study in order to construct the array. Microarrays are especially problematic to use for studying gene expression in polyploids, as the hybridization between the probes and the target sequences may not be sensitive enough to be able to distinguish between homeologs (Yoo et al., 2014). Furthermore, microarrays lack sensitivity for genes expressed at either very low or very high levels, and comparing expression levels across different experiments can often be difficult (Wang et al., 2009).

RNA-seq (methods chapter 2.2) on the other hand will report whatever is transcribed, even things we do not expect to find beforehand. In addition, RNA-seq does not have any problems with very low or high expression levels, and the results are more accurate and reproducible (Wang et al., 2009).

The recently released homoeologue-specific draft assembly of the bread wheat genome and annotation of protein coding genes by The International Wheat Genome Sequencing Consortium (IWGSC) (2014), and the rapid advances in next generation sequencing technologies has made it much easier to study homeolog specific expression in wheat using RNA-seq. The draft sequence was produced by sequencing isolated chromosome arms which made it possible to differentiate between homeologs.

IWGSC (2014) also studied gene expression using RNA-seq data from five organs (leaf, root, grain, spike, and stem) for sets of homeologous genes (triplets) with only one copy present in each of the subgenomes. Their results showed that gene expression cluster according to subgenomes using correlation distances, with the exception of root. They concluded that this indicates that the subgenomes are transcriptionally and regulatory autonomous. Moreover, pairwise tests of differential expression between homeologous genes revealed expression bias in 21% of the cases, but there was a similar number

of preferentially transcribed genes in each subgenome.

Pfeifer et al. (2014) studied gene expression using RNA-seq data from the three main cell types of the wheat endosperm (starchy endosperm (SE), the aleurone layer (AL) and transfer cells (TCs)) at three different developmental stages (10, 20, or 30 days post anthesis (DPA)). The IWGSC bread wheat genome survey sequence and annotations were used as reference for mapping transcripts to each the three subgenomes. They found that there was a low number of genes that were preferentially expressed in each cell type and developmental stage (genes that are up-regulated in that sample as compared with the other samples) and that the number varied between samples. The number of preferentially expressed genes in each cell type and stage was about equal for the subgenomes. Different subgenomes dominated gene expression in different tissues and developmental stages, but no subgenome dominated globally. Furthermore, the subgenomes contributed unequally to particular functions using GO enrichment. Of the genes expressed during endosperm development (all cell types and developmental stages), there was about the same number of genes from each of the subgenomes. The three subgenomes contributed almost equally to the number of expressed genes in the individual cell types and developmental stages. Hierarchical clustering based on gene expression grouped the samples according to subgenomes rather than tissues, which is the same as IWGSC (2014) observed for different RNA-seq data.

1.5 Study aims

In this thesis, homeolog specific expression and regulation will be studied. As we have seen in chapter 1.3 and chapter 1.4, previous studies of wheat gene expression show that in some triplets there is preferential expression of a subgenome. However, which subgenome that is preferentially expressed varies between tissues. Homeolog silencing is commonly observed, and one study found tissue specific reciprocal silencing. Although homeologs differ in expression levels, no overall expression dominance for any of the subgenomes has been found. Both IWGSC (2014) and Pfeifer et al. (2014) concluded that the subgenomes show regulatory and transcriptional autonomy, i.e. that genome specific gene expression dominates over tissue specific gene expression.

In earlier work on homeologous gene expression in wheat the focus has been on how steady state expression levels vary between subgenomes. In this theses the aim is to go beyond comparisons of expression levels and focus on the developmental regulation of subgenomes in bread wheat.

The thesis is divided into four topics:

1. Similarity between the subgenomes: do the subgenomes have similar regulation patterns, and which two of the subgenomes are most alike? Specifically it will be investigated if the subgenomes show regulatory autonomy, as it has been suggested previously.

2. Is regulation of the subgenomes similar between tissues?

3. How does regulation of the subgenomes relate to homeolog specific expression levels?

4. Do the subgenomes contribute differently to any molecular functions or biological processes?

2 Materials and methods

2.1 The data

In this thesis, two RNA-seq data sets with gene expression values (measured in fpkm) from wheat have been used. The first data set is from IWGSC (2014) and has expression values from five organs (leaf, root, grain, spike, and stem) at different developmental stages. The second data set is expression from the whole endosperm (W) and the starchy endosperm (SE), each at two different developmental stages, from Pfeifer et al. (2014) (Figure 2).

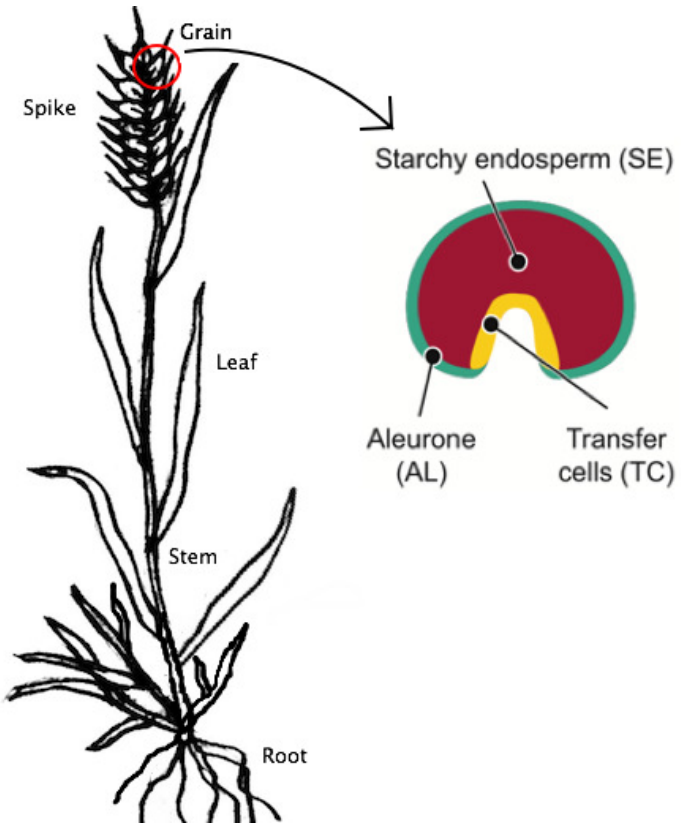


Figure 2: Tissues in wheat. The picture of the wheat endosperm is from Pfeifer et al. (2014).

The time points in the first data set are measured in zadoks scale, one of the most commonly used scales for measuring development in cereals. The zadoks scale uses a two-digit system of developmental stages. There are 10 primary stages, each of which is divided into 10 secondary stages (Table 1) (Herbek and Lee, 2009). The time points in the second data set is given in number of days post anthesis (DPA).

Table 1: The primary stages of the zadoks scale for wheat development.

Primary stage	Description
0	Germination
1	Seedling growth
2	Tillering
3	Stem Elongation
4	Booting
5	Inflorescence emergence
6	Anthesis
7	Milk development
8	Dough development
9	Ripening

As the three subgenomes in wheat are highly similar, there is a large number of homeologous gene triplets, which are genes in each of the three subgenomes that share a common ancestor gene (Figure 3). Only genes with a copy in each of the subgenomes have been analyzed in this thesis.

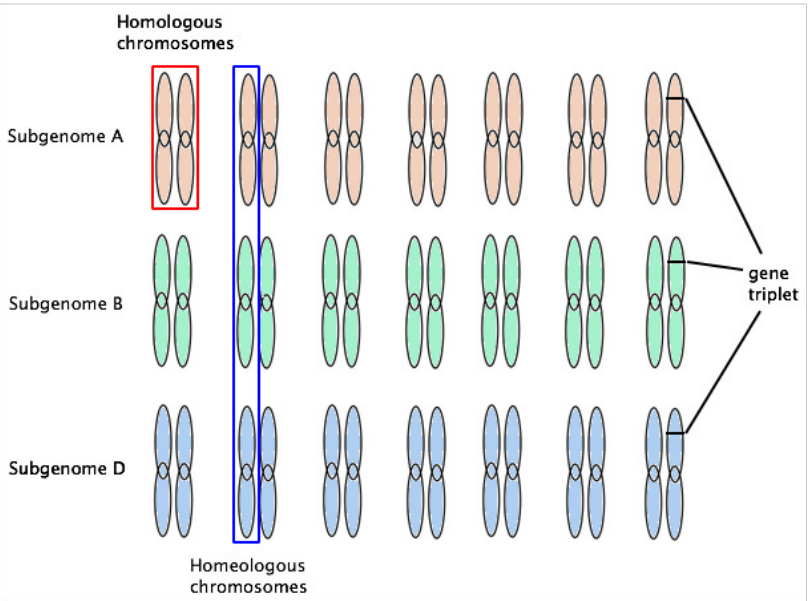


Figure 3: The bread wheat genome consists of three subgenomes, A, B and D, which each contains 7 pairs of homologous chromosomes.

2.2 RNA-seq

The transcriptome of a cell is the complete set of transcripts (i.e. RNA molecules) it contains. Transcriptomics, the study of the transcriptome, has many uses like studying changes in gene expression, alternative splicing and post-transcriptional modifications, or determining exon/intron boundaries and the start and stop sites of genes. All types of transcripts can be studied including messenger RNA (mRNA) and non-coding RNAs such as transfer RNA (tRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), micro RNA (miRNA), small interfering RNA (siRNA) and small nuclear RNA (snRNA) (Wang et al., 2009). Traditionally microarrays have commonly been used to study transcriptomes, but as previously mentioned RNA-seq has several advantages over microarrays.

RNA-seq (RNA sequencing) works by first isolating RNA from a cell, then fragmenting it and reverse transcribing it into cDNA. Each cDNA molecule is then sequenced from one end (single-end sequencing) or both ends (paired-end sequencing) using next generation sequencing, a group of sequencing technologies that parallelize the sequencing process and thereby reduces both the time and the cost used. The sequencing produces short sequences of normally a few hundred base pairs called reads. After sequencing, the reads need to be preprocessed by e.g. removing or trimming low quality reads and masking repeats to make assembly easier. Then lastly, the reads are either aligned to a reference genome or assembled de novo if no reference is available (Wang et al., 2009).

RNA-seq is often used to compare gene expression for example under different external conditions, at different times or between healthy and diseased individuals. Gene expression is quantified using the counts of each mRNA transcript, but in order to compare gene expression across samples the counts have to be normalized. In the data used in this thesis the gene expression is measured in fpkm (Fragments Per Kilobase of transcript per Million mapped reads), which is the number of fragments that have been mapped to the gene/transcript normalized by the length of the transcript in kb and the total number of mapped reads in the sample in million. For paired-end reads, each fragment corresponds to two reads. If one of the reads is of poor quality it might not be mappable, and we therefore count fragments rather than reads (Parnell et al., 2011). Rpkm (Reads Per Kilobase of transcript per Million mapped reads) is also used as a measure for gene expression and for single-end reads this is the same as fpkm (The farrago, 2014).

2.3 Differential expression analysis

Differential expression analysis was done using the R package limma (Ritchie et al., 2015). This is originally a package for analysis of microarray data, but the methods for single channel microarray data can also be applied to normalized and log transformed RNA-seq data. The analysis in limma is done by fitting a linear model to each gene. A method called Empirical Bayes (Berkeley, 2004) is used to adjust the variance of each gene towards an estimated pooled variance. This results in a more stable analysis when we have few samples, which is a common problem in biology due to e.g. high cost or practical issues. P-values from the differential expression tests are adjusted for multiple testing. Most commonly the Benjamini and Hochberg method is used (same as false discovery rate). If we reject the null hypothesis, that the gene is not differentially expressed, for adjusted p-values less than or equal to 0.05, we expect 5 % of the significant genes to be false positives. Without any adjustment of p-values, 5 % of all genes tested is expected to be false positives.

For the differential expression analyses in this thesis a table where each gene had between two and four replicate expression values for each combination of tissue and developmental time was used. As limma expects log expression values, the fpkm values were first transformed to log2(fpkm + 1). The number 1 was added to each expression value to avoid getting values of minus infinity or negative values.

2.3.1 Differential expression between different time points in a tissue

Tests were done to find which genes that were differentially expressed between two different developmental times in the same tissue. To do this a linear model with two coefficients was applied to each gene. The first coefficient was the intercept, which estimated the mean expression for the replicates at time 1. The second coefficient estimated the added effect we get from time 2. The genes that had a significant effect for the second coefficient (adjusted p-value ≤ 0.05) were considered to be differentially expressed.

Furthermore, when a gene has no change in expression this could mean either that the gene has a constant non zero expression, or it could mean that the gene has zero expression at both developmental times. To find the triplets where all three homeologous genes have zero expression, meaning that these triplets are not expressed at all in the tissue, a one sided one sample t-test using the replicate expression values for both developmental times was used for each of the genes. The null hypothesis was that the mean value is equal

to zero, and the alternative hypothesis was that the mean value is greater than zero. The p-values were adjusted using false discovery rate and genes with adjusted p-values smaller than or equal to 0.05 were considered to have a constant non zero expression.

The triplets where at least one of the homeologous genes had a non zero expression were then divided into regulation categories based on whether or not each of the three homeologs were differentially expressed between the two time points. There are four main categories, which tells us how many of the homeologs that had changed expression. Further subdivision of the categories gives information on which of the homeologs that had changed expression and if the changes were up or down (Table 2).

Table 2: Triplet regulation categories with increasing level of details.

0 changes	zero.zero.zero	zero.zero.zero
1 change	change.zero.zero	up.zero.zero
		down.zero.zero
	zero.change.zero	zero.up.zero
		zero.down.zero
	zero.zero.change	zero.zero.up
		zero.zero.down
2 changes	change.change.zero	up.up.zero
		down.down.zero
		up.down.zero
		down.up.zero
	change.zero.change	up.zero.up
		down.zero.down
		up.zero.down
		down.zero.up
	zero.change.change	zero.up.up
		zero.down.down
		zero.up.down
		zero.down.up
3 changes	change.change.change	up.up.up
		down.down.down
		up.up.down
		up.down.up
		down.up.up
		down.down.up
		down.up.down
		up.down.down

2.3.2 Differential expression between homeologs

Pairs of homeologous genes were tested for differential expression in each sample. A linear model like the one in section 2.3.1 was applied to each gene, but now the intercept estimated the mean expression for subgenome 1 and the second coefficient estimated the added effect from subgenome 2. The genes that had a significant effect for the second coefficient (adjusted p-value ≤ 0.05) were considered to be differentially expressed between the two subgenomes.

2.4 Test for significant overlap of triplets in the regulation categories between tissues

The triplets were divided into regulation categories based on how the expression of the three homeologous genes changes between two different developmental times in the same tissue (chapter 2.3.1). Fisher's exact test was used to look for significant overlaps of triplets between the tissues in these categories.

There are 9 combinations of tissues and developmental times. When testing for overlaps of triplets in one of the regulation categories, tests were performed between pairs of these combinations. In total 36 tests were done for each regulation category. Afterwards the p-values were adjusted using false discovery rate.

In order to do the Fisher's exact test, the genes were divided into two categories: genes that are classified as being in a particular category and genes that are not. The null hypothesis was that the genes are divided into these two categories in tissue 1 independently of which category they belong to for tissue two. The alternative hypothesis was that it is more likely for a gene to be in this category for tissue 1 if it is also in this category for tissue 2. The function `fisher.test` in R was used on a 2 x 2 contingency table (Table 3) to do a one sided fisher test. This gives the p-value, which is the probability of getting an overlap of size a or larger if the null hypothesis is true.

2.5 Finding the regulation similarity between subgenomes

Based on the t-values obtained from comparing the expression between two different developmental times (chapter 2.3.1), an analysis was done to investigate which two of the subgenomes that were most similar. This may vary between the tissues, so this was first done for each tissue separately. In order to find the distances between subgenomes, the t-values were stored

Table 3: Contingency table showing the number of genes found in the category tested (in this case zero.zero.zero) in both tissues (a), in just tissue 1 (c), in just tissue 2 (b) and in neither of the tissues (d). The total number of genes is N=a+b+c+d.

		Tissue1	
		0.0.0	not 0.0.0
Tissue 2	0.0.0	a	b
	not 0.0.0	c	d

in a matrix where each row corresponded to a subgenome and each column corresponded to a triplet of genes. Then the correlation distances between pairs of rows were calculated. The distances were normalized by the largest distance for each tissue to be able to compare distances between tissues.

To find the two subgenomes that are most similar overall, 95% confidence intervals for the mean value for each of the pairwise distances were calculated based on the distances for all the tissues (Table 5). Bootstrap confidence intervals were used since the distributions of these distances do not resemble normal distributions.

In statistics, bootstrapping is a useful method of approximating a sample distribution when you have a limited amount of data from an unknown distribution. For estimating the distribution of a sample statistic using nonparametric bootstrapping you do the following: create a bootstrap sample by sampling with replacement from the original sample, using the same sample size as the original sample. Then compute the statistic (in our case the mean) from the bootstrap sample. Do this many times to get a bootstrap distribution of the statistic. This distribution can then be used to obtain a confidence interval for the statistic.

In this thesis, the confidence intervals were calculated using the R package boot (Canty and Ripley, 2015) and three different methods for nonparametric bootstrap confidence intervals. The formulas for these confidence intervals can be found in chapter 5 of Davison and Hinkley (1997).

2.6 Bootstrap probabilities for assessing the uncertainty of hierarchical clustering

The R package pvclust (Suzuki and Shimodaira, 2006) was used to evaluate the uncertainty of the hierarchical clustering. For each cluster, pvclust gives two types of p-values: BP is the bootstrap probability and AU is the Approximately Unbiased p-value, calculated via multiscale bootstrap resampling (Shimodaira, 2002). It is claimed that the AU value has superiority in bias over the BP value (Suzuki and Shimodaira, 2006). For clusters with BP or AU values above 0.95 we reject the null hypothesis that the cluster does not exist at a 0.05 significance level.

The hierarchical clustering done in this thesis is based on correlation distances, using Pearson correlation, and Ward's method which minimizes the total within-cluster distance. Ward's method was chosen because it was observed that this method resulted in higher BP and AU values than using average/single/complete linkage.

2.7 Gene ontology enrichment

The Gene Ontology (GO) project is a collaborative effort to describe gene products across species (Ashburner et al., 2000; Gene Ontology Consortium and others, 2015). The GO terms that are used to describe genes are organized as nodes in a hierarchical graph and the edges describe the relationship between the terms. At the top of the graph there are three root nodes, cellular component, molecular function and biological process, one for each independent GO domain. All other terms can be traced back to these. Further down the graph the terms become more specialized.

Having a background set of genes (e.g. a whole genome) and a smaller set of genes from this background sharing some interesting properties, a GO enrichment analysis can be done to gain biological understanding. This is a test to find if any GO terms are overrepresented in this gene set. The p-value then gives the probability that we simply by chance get an overlap as big as we observe in our data or larger between the genes in the gene set and all genes in the background set annotated to a specific GO term (Figure 4).

The GO enrichment analyses in this thesis were done using GO annotations from IWGSC (2014). The goal was to see if the subgenomes contribute differently to any molecular functions or biological processes, by investigating if enriched GO terms differ between subgenomes. Gene sets for each of the three subgenomes were made both based on differences in regulation between the subgenomes (differential expression between time points, chapter 2.3.1), and differences in expression levels (differential expression be-

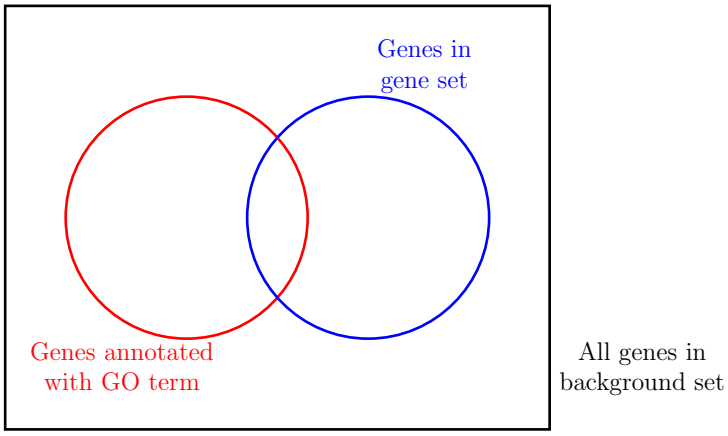


Figure 4: GO enrichment. The p-value gives the probability of getting an overlap of this size or larger by chance.

tween homeologs, chapter 2.3.2). For the first of these methods, a gene set consisted of the genes from a subgenome from all the regulation categories where this subgenome was upregulated, except from the category where all three homeologs were upregulated at the same time. This means that for the A subgenome for instance, the gene set consisted of all A homeologs from the categories up.zero.zero, up.zero.down, up.down.zero, up.zero.up, up.up.zero, up.up.down, up.down.up and up.down.down. For the second method, a gene set consisted of the genes from a subgenome that had an expression equal to or higher than both its homeologs and significantly higher than at least one of its homeologs. In both cases, the background set was all wheat genes that were used in the analyses in this thesis that are also annotated with GO terms. The analyses were done seperately for each of the combinations of tissues and time points. GO terms with $p \leq 0.025$ were considered to be enriched in the gene set.

Instead of only using a Fisher's exact test for the enrichment analyses, the R package topGO (Alexa and Rahnenfuhrer, 2010) was used, since this package has several different algorithms that also take the relationships between the GO terms into account. Specifically, the weight01 algorithm, which is the default algorithm in topGO, was used in combination with Fisher's exact test. Weight01 is a combination of the two methods elim and weight described by Alexa et al. (2006). Their simulations showed that compared to the classical approach (where each GO term is tested independently), the weigth algorithm reduced the rate of false positives and at the same time found more true enriched GO terms (higher sensitivity), while the elim method reduced the false positive rate even further, but also found fewer of

the true positives. Alexa and Rahnenfuhrer (2010) points out that since GO terms are not independent, the multiple testing theory does not apply. They further say that the p-values returned from the methods that account for the relationships between GO terms should be interpreted as already being corrected for multiple testing. The the p-values from the GO analyses in this thesis are therefore the ones given by the weight01 algorithm, without any further correction.

3 Results

3.1 Classification of triplets into regulation categories

When testing for differential expression between different time points (chapter 2.3.1), each of the three homeologous genes in a triplet can have a significantly higher or lower expression or no change in expression in the second time point. Based on this, the triplets were classified into regulation categories for each tissue. This was done to investigate if the subgenomes have similar regulation patterns.

The triplets were first classified into 5 categories (Table 4). The category "not expressed" means that all three homeologous genes have zero expression at both developmental times. For the other four categories at least one of them has a non-zero expression at one or both of the developmental times. In addition, either none of the three genes has changed expression ("0 changes"), one of them has changed expression ("1 change"), two of them have changed expression ("2 changes") or all three of the them have changed expression ("3 changes"). A total of 8605 triplets were tested.

Table 4: The counts and percentages of triplets in the different regulation categories that are explained in the text.

	not expressed	0 changes	1 change	2 changes	3 changes
W, 10DPA to 20DPA	1611 (19%)	6538 (76%)	380 (4%)	54 (0.6%)	22 (0.3%)
SE, 20DPA to 30DPA	1847 (21%)	6646 (77%)	100 (1%)	9 (0.1%)	3 (0.04%)
GRAIN, Z71 to Z75	445 (5%)	1730 (20%)	2213 (26%)	1989 (23%)	2228 (26%)
LEAF, Z10 to Z23	400 (5%)	4167 (48%)	1967 (23%)	1075 (12%)	996 (12%)
LEAF, Z23 to Z71	399 (5%)	1657 (19%)	1942 (23%)	1847 (21%)	2760 (32%)
ROOT, Z10 to Z13	464 (5%)	8141 (95%)	0	0	0
ROOT, Z13 to Z39	416 (5%)	8189 (95%)	0	0	0
SPIKE, Z32 to Z39	432 (5%)	4693 (55%)	1755 (20%)	969 (11%)	756 (9%)
SPIKE, Z39 to Z65	305 (4%)	3744 (44%)	1937 (23%)	1303 (15%)	1316 (15%)
STEM, Z30 to Z32	346 (4%)	2996 (35%)	2157 (25%)	1520 (18%)	1586 (18%)
STEM, Z32 to Z65	390 (5%)	2146 (25%)	2110 (25%)	1655 (19%)	2304 (27%)

In the barplots in figure 5 and figure 6 the regulation is further subdivided into which subgenome that has changed expression and if the change is up or down. The triplets that are not expressed at all in the tissues are left out. Root is also left out as there were no genes with significant changes in expression for this tissue.

Table 4 shows that there are some triplets that are not expressed at all in each tissue and that this number is largest for W and SE. Compared to the other tissues, W and SE also have a large percentage of triplets where none of the three homeologous genes changes expression. The exception is root,

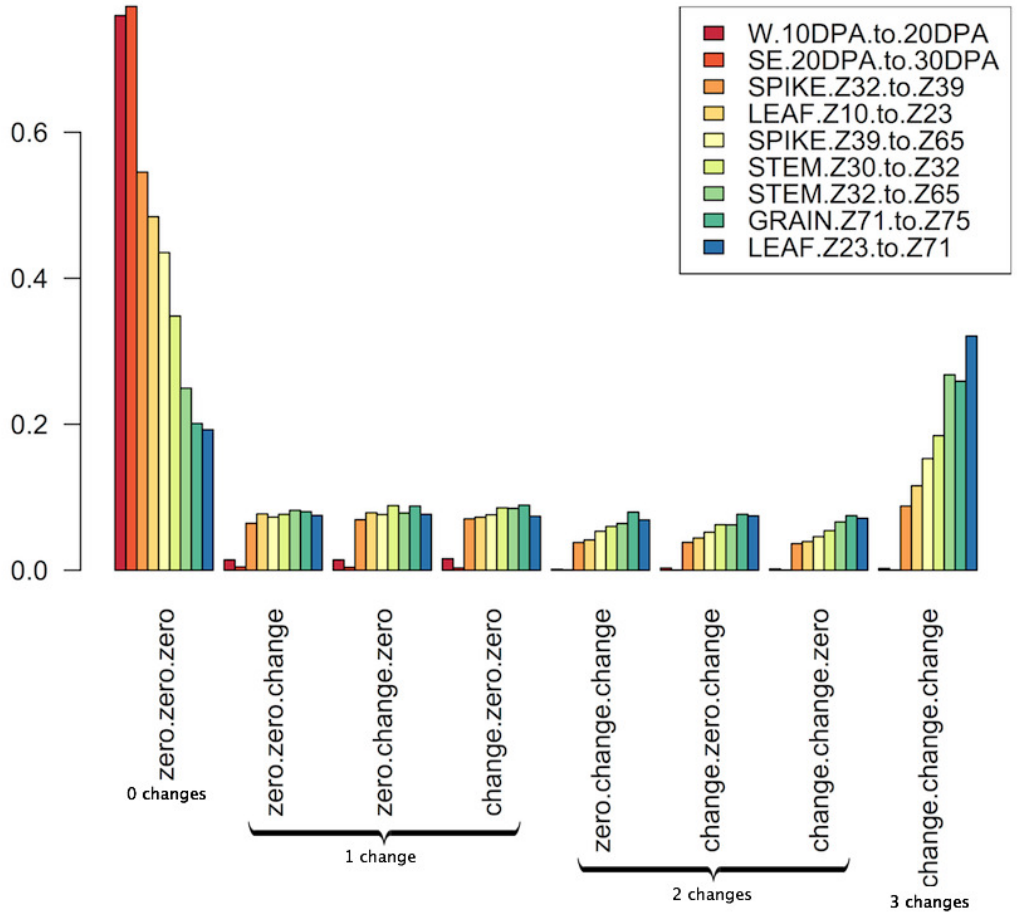


Figure 5: Fraction of triplets in the regulation categories for all combinations of change and zero change. The names of the categories show the triplet expression on the form A.B.D.

which has no significant changes for any of the triplets. Gene regulation is to a large degree conserved between subgenomes, as the largest categories for all tissues are either "0 changes" or "3 changes"

When the regulation categories get more detailed (Figure 5), we see that in the "1 change" category, the fraction of triplets is about the same for each of the subgenomes, indicating that none of the subgenomes are regulated more actively than the others. Also, the number of triplets in "1 change" seems to vary less between tissues than the number of triplets in the other categories. In the "2 changes" category, the different combinations of the A, B and D subgenomes seem to occur at about the same frequency.

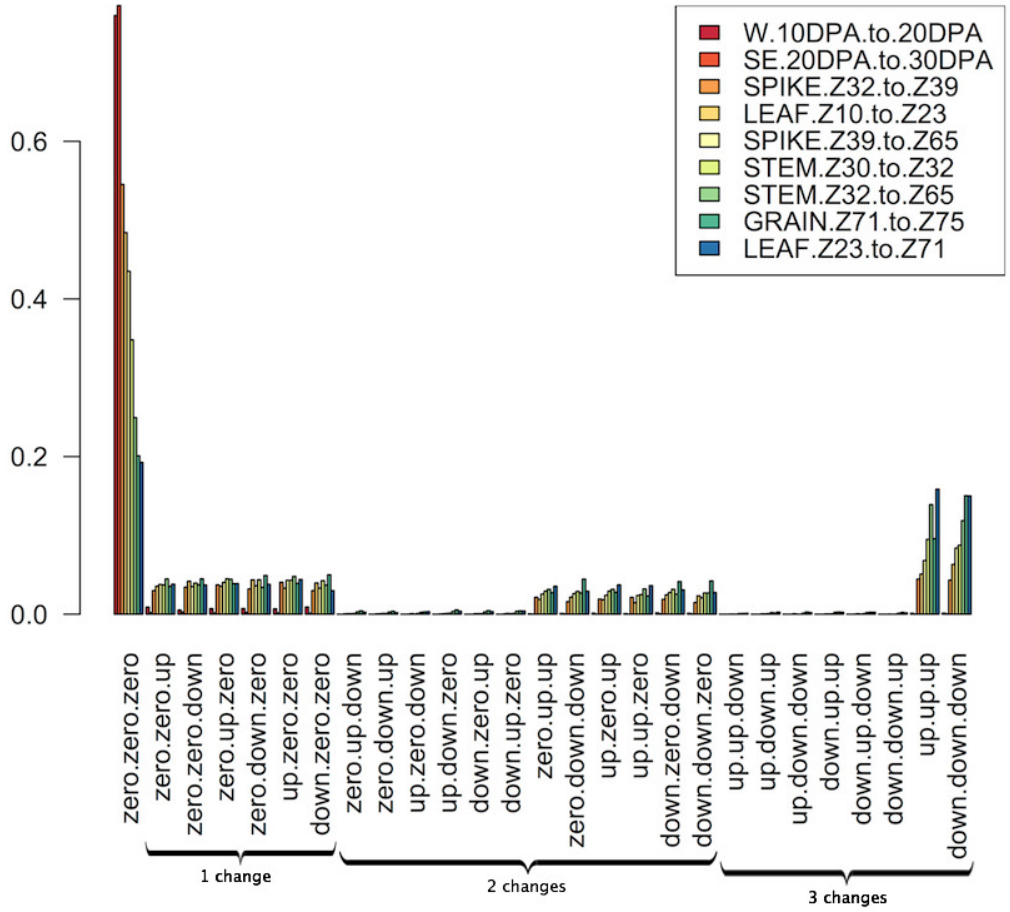


Figure 6: Fraction of triplets in the regulation categories for all combinations of up, down and zero change. The names of the categories show the triplet expression on the form A.B.D.

At the most detailed level (Figure 6) we see that when two or three of the homeologous genes have changed expression the change is almost always in the same direction, i.e. homeologs are rarely regulated in opposite directions.

3.2 Comparing the distribution of triplets in the regulaton categories between tissues

In order to see how similar the regulation of the subgenomes is between the tissues, the correlation between the tissues was calculated based on the fraction of triplets in each regulation category (the data plotted in figure 6). The type of correlation used was Spearman correlation, which is commonly used for measuring the relationship between rankings. Root has not been included, since root had no significant changes in expression. The results show that the ranking of the triplet categories is mostly very similar between tissues, but that W and SE behave a bit differently from the other tissues (Figure 7).

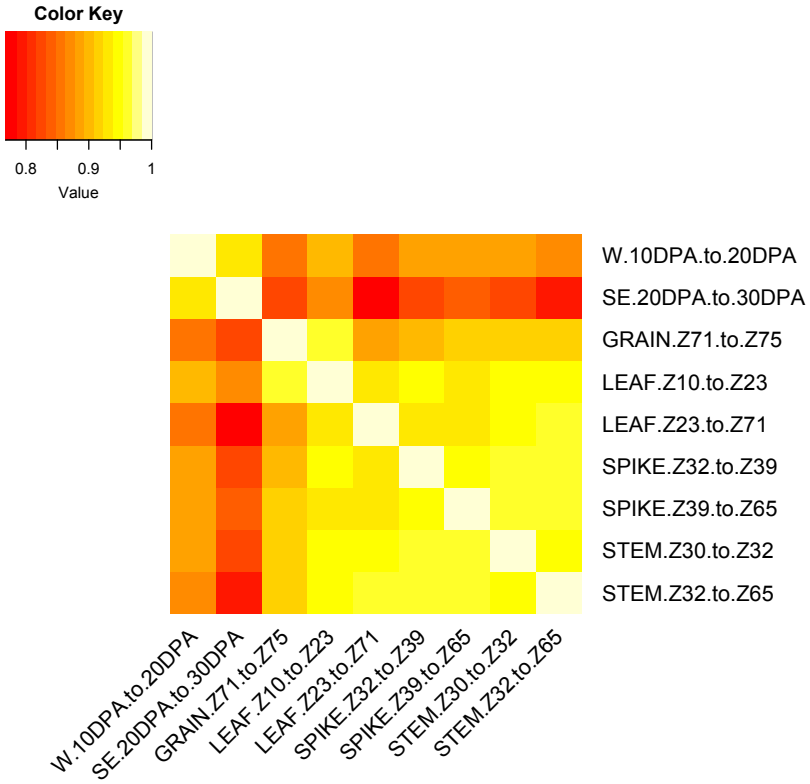


Figure 7: Heatmap of the Spearman correlation between tissues based on the fraction of triplets that are in each regulation category (all 27 combinations of up, down and zero change).

3.3 Do the triplets belong to the same regulation categories in different tissues?

In chapter 3.2 we saw that the number of triplets in the different regulation categories is similar between tissues. The results in this section show if these numbers harbor the same triplets in different tissues, or not

As explained in chapter 2.4, 36 pairwise tests between the tissues are done when testing for overlaps of triplets in a regulation category. Figure 8 shows the fraction of these tests that are significant ($p \leq 0.05$) for each of the regulation categories. More details for the categories "0 changes", "1 change", "2 changes" and "3 changes" can be found in appendix A.

We see that for the category "0 changes" we mostly have the same triplets in all tissues, as all but one of these tests are significant. For the rest of the categories we see that when all three homeologs behave in the same way there is a large tendency of these being the same triplets in several tissues. But for the triplets where one or two of them changes expression, it varies more which triplets these are in the different tissues. The fewest significant overlaps are seen for the categories where two or three of the homeologous genes changes expression in different directions.

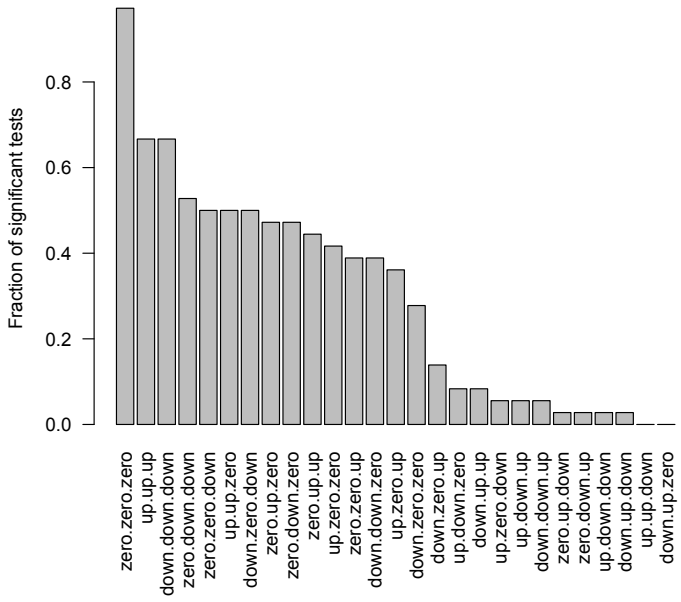


Figure 8: The fraction of significant pairwise Fisher's exact tests for overlap of triplets between tissues for all the regulation categories.

3.4 Regulation similarity between subgenomes

The results from calculating the distances between the subgenomes for each tissue, according to the method explained in chapter 2.5, can be seen in table 5. We see that which two subgenomes that are regulated most similarly varies between the tissues. A-B has the shortest distance two times, A-D has the shortest distance five times and B-D has the shortest distance four times.

Table 5: Distances between the subgenomes based on t-values from comparing gene expression between to developmental times in the tissues. The shortest distance in each row is in bold font.

	A - B	A - D	B - D
W 10DPA to 20DPA	**0.8814**	0.9403	1
SE 20DPA to 30DPA	0.9709	**0.9607**	1
GRAIN Z71 to Z75	1	**0.8183**	0.9268
LEAF Z10 to Z23	1	**0.8966**	0.9132
LEAF Z23 to Z71	0.9938	**0.9646**	1
ROOT Z10 to Z13	0.9929	1	**0.9128**
ROOT Z13 to Z39	0.9953	1	**0.9887**
SPIKE Z32 to Z39	**0.9785**	0.980	1
SPIKE Z39 to Z65	1	0.9799	**0.8946**
STEM Z30 to Z32	1	0.9577	**0.9514**
STEM Z32 to Z65	1	**0.9834**	0.9961

95% confidence intervals for the mean pairwise distances between subgenomes based on the distances in table 5 are shown in table 6. Even though the limits vary a little for each of the three methods used, the intervals for all three distances overlap for each of them. We therefore can not claim that there is a difference between the mean values of the three pairwise distances, meaning that none of the subgenomes are more similar to each other than they are to the third subgenome based on this analysis.

Table 6: Confidence intervals for the mean value for each of the distances A-B, A-D and B-D.

	Basic	Percentile	BCa
A - B	(0.9687, 1.0055)	(0.9604, 0.9973)	(0.9423, 0.9953)
A - D	(0.9272, 0.9859)	(0.9198, 0.9786)	(0.9083, 0.9746)
B - D	(0.9393, 0.9875)	(0.9368, 0.9850)	(0.9358, 0.9841)

3.5 Hierarchical clustering of expression levels and t-values

Figure 9 shows a heatmap of t-values from comparing gene expression at two different developmental times in the same tissue (chapter 2.3.1). The rows are triplets and the columns are all combinations of tissues and times that were tested for each of the subgenomes. The hierarchical clustering is based on Ward's method and correlation distance using Pearson correlation. The heatmap and clustering in figure 10 is done in the same way as in figure 9, only with expression values (log2(fpkm+1) for each sample) rather than t-values. Bootstrap probabilities for assessing the uncertainty for both of the clusters can be found in appendix B.

We see that expression regulation (i.e. t-values, figure 9) cluster according to tissues, but that expression values (Figure 10) mostly cluster according to subgenomes. This big difference in clustering indicates that there can be differential expression between subgenomes in the samples even though the subgenomes are mostly regulated in the same way.

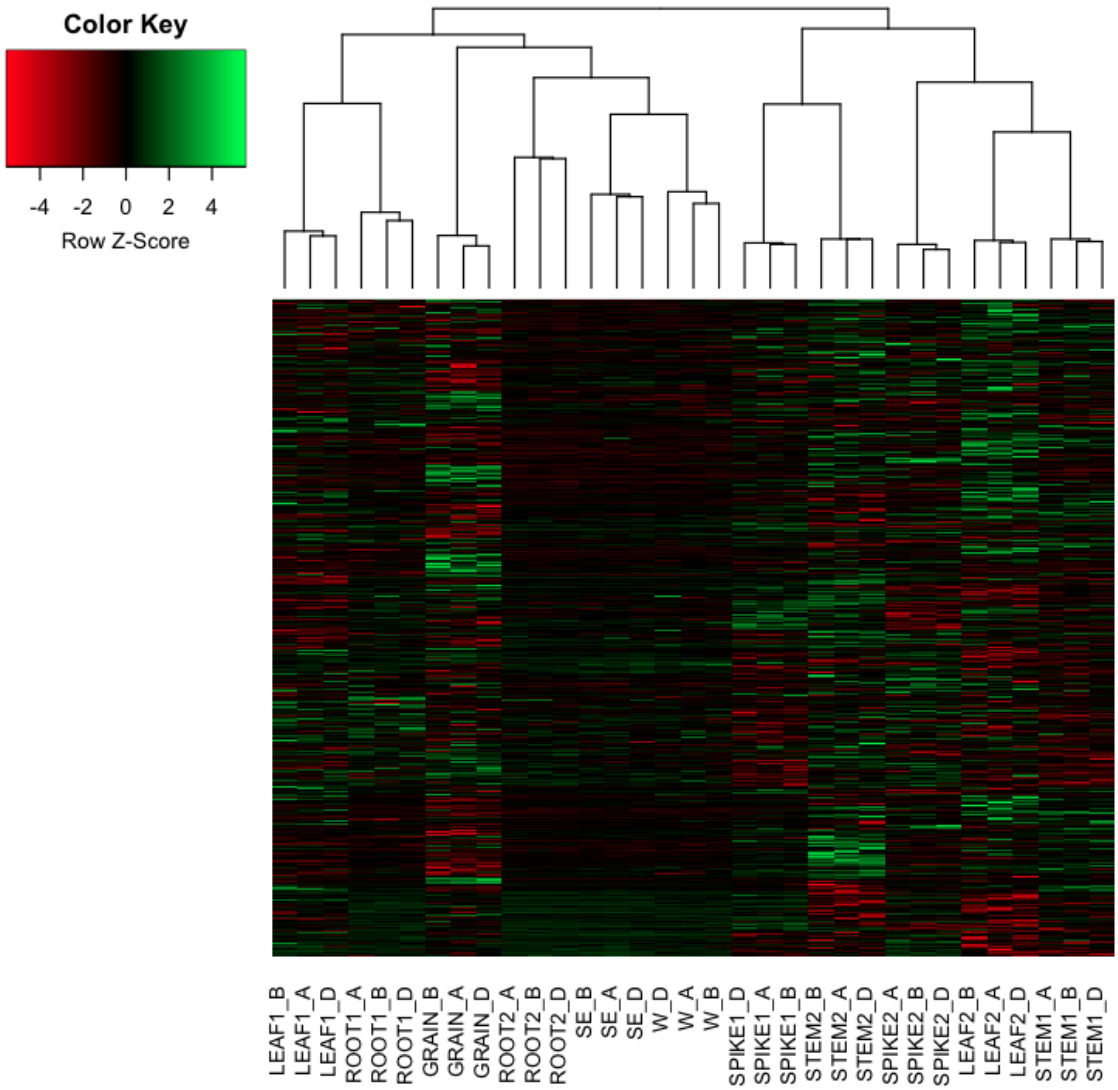


Figure 9: Heatmap of t-values and hierarchical clustering based on Ward's method and correlation distances. For practical reasons when plotting, the column names are in short form, but for example W_A is really short for W 10DPA to 20DPA for subgenome A. The rows are scaled to better show the differences in t-values.

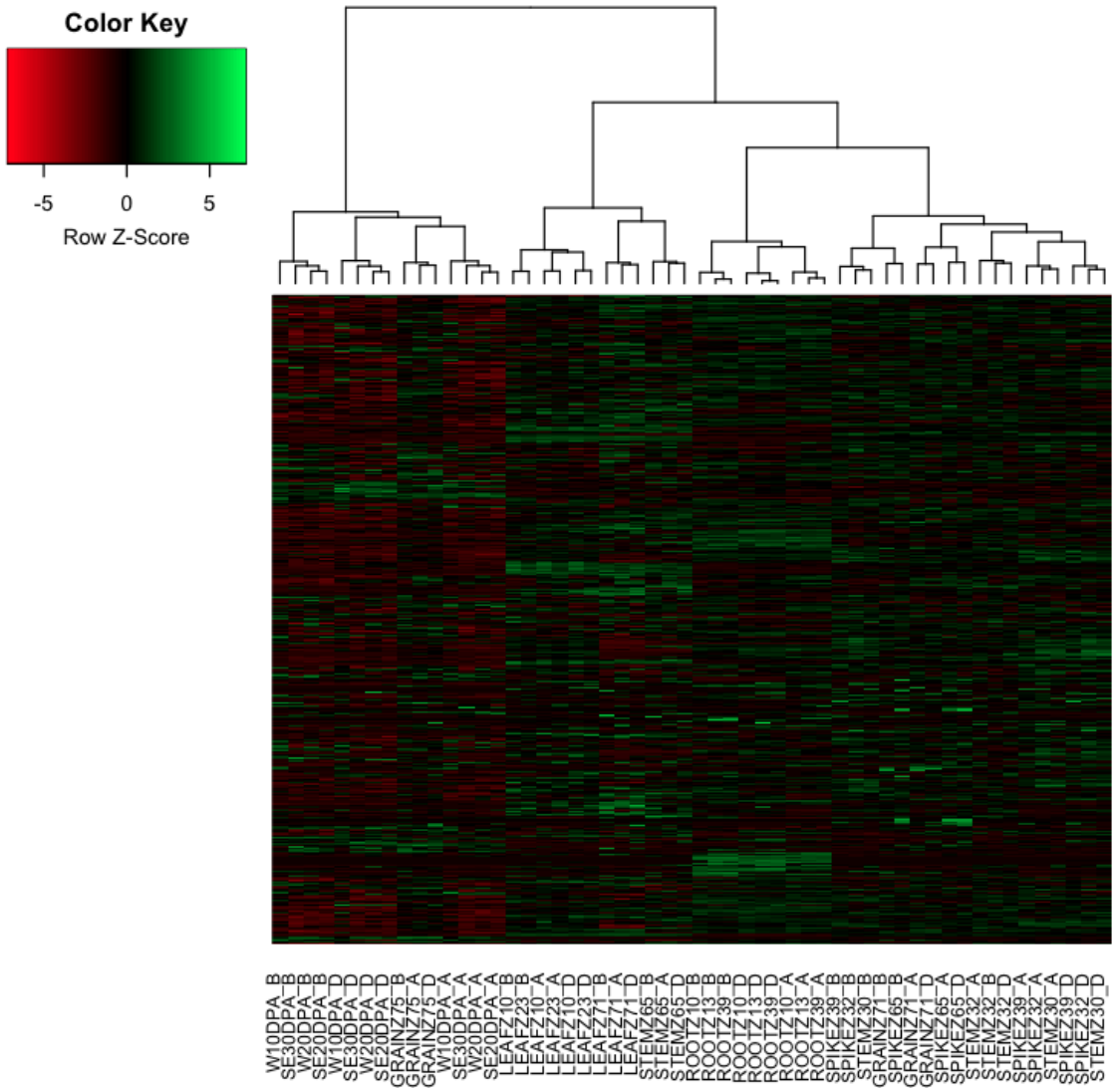


Figure 10: Heatmap of log2(fpkm+1) and hierarchical clustering based on Ward's method and correlation distances. The rows are scaled to better show the differences in expression values.

3.6 Comparing expression levels to expression regulation in the subgenomes

As the results in chapter 3.5 required some more investigations, analyses were done to compare homeolog specific expression values in the samples to how the subgenomes are regulated.

The initial hypothesis was that the triplets in the "0 changes" category might cause the difference we see in clustering in figure 9 and figure 10. In this category none of the three homeologs changes expression, but this does not necessarily mean that their expression levels are similar. If the expression levels differ greatly between subgenomes, even though they are regulated in the same way, this could possibly cause the subgenome specific clustering we see in figure 10.

The approach for testing this hypothesis was to remove triplets from the "0 changes" category with large differences in expression levels between homeologs, before doing the clustering again. If the hypothesis was correct we would now expect to see more clustering of tissues. This was done by first doing pairwise tests for differential expression between subgenomes (chapter 2.3.2) in the triplets that are in the "0 changes" category in any of the tissues. As we saw in chapter 3.3, these triplets have large overlaps between the tissues. The triplets with differential expression between any of the subgenomes in any of the samples, where one of the homeologs also had expression over a threshold value, were removed. Then the clustering of expression values was done without these triplets. Root was not included in this analysis because so many of the triplets are in "0 changes" in this tissues that this would result in the removal of many triplets that are not in "0 changes" in the other tissues. The clustering was done for fpkm thresholds of 2000 (24 triplets removed), 1000 (48 triplets removed), 500 (120 triplets removed) and 250 (258 triplets removed). They all gave the clustering in figure 11, which is almost the same as in figure 10.

As removing triplets in the "0 changes" category did not have any effect on the clustering of expression values, pairwise tests for differential expression between the subgenomes were done for each sample for all the triplets. The goal was to be able to see how triplet regulation relates to differential expression between homeologs in all the regulation categories. Figure 12 shows a couple of examples of the gene expression we could expect to see if there is a connection between triplet regulation and expression levels. In figure 12A, the three homeologous genes are all regulated in the same way and they also have similar expression levels. In this case, we would expect the number of differentially expressed genes to be similar between all pairs of subgenomes and at both time points. In figure 12B, only the homeolog

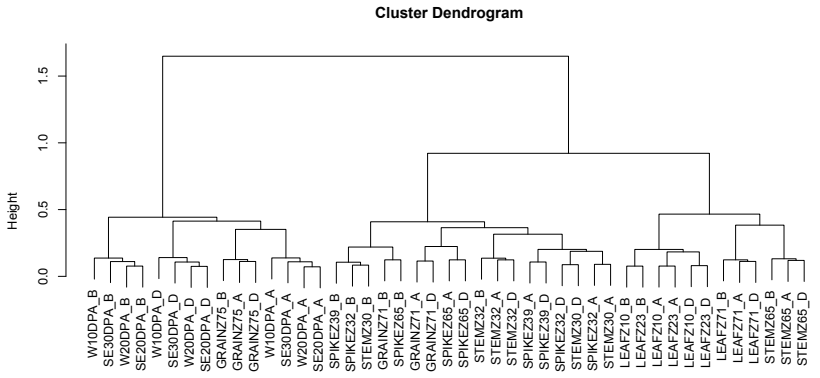


Figure 11: Hierarchical clustering of log2(fpkm+1) values where triplets in the zero.zero.zero category with expression over a threshold value have been removed. The clustering is based on Ward's method and correlation distances.

from the D subgenome has changed expression, which leads to differential expression between D and the other two subgenomes in the second time point. Now we would expect to see an increase in the number of genes that are differentially expressed between D and the other two subgenomes in the second time point compared to he first.

However, this clear connection between regulation and expression is not what we see in our data. Figures 13-15 show the fraction of triplets in each regulation category that is differentially expressed between subgenomes in the first and second time point for each tissue (the same information is shown as tables in appendix D). What we see is that in some of the tissues the changes in number of differentially expressed genes resemble what we could expect to find, at least for some of the expression categories. Some examples are the categories zero.change.zero, zero.zero.change and change.change.change in LEAF Z23 to Z71 (Figure 14). For zero.change.zero the increase in differentially expressed genes in the second time point is largest for AB and BD, which is what we would expect since only the B homeolog changes expression. In the same way, the number of differentially expressed genes increases most for AD and BD for zero.zero.change. For change.change.change all three combinations increase equally much. However, for most tissues the number of triplets that are differentially expressed between subgenomes does not seem to vary in accordance with the subgenomes that changes expression in the regulation categories.

In addition to this, we also note that for all the tissues the total number of differentially expressed homeologs is larger between the A and B subgenomes than between A and D or B and D (Table 7).

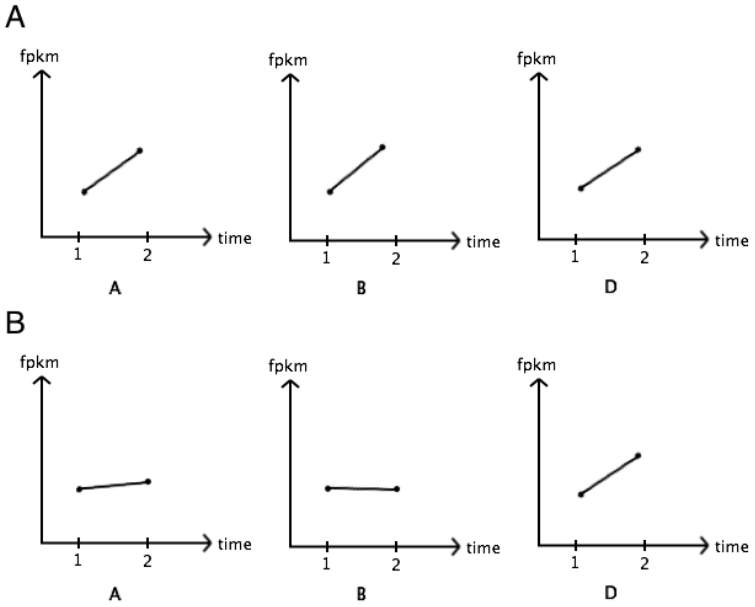


Figure 12: The gene expression we could expect to see if there is a connection between triplet regulation and expression. (A) Expression for a triplet in the regulation category change.change.change. (B) Expression for a triplet in the regulation category zero.zero.change.

Table 7: The total number of differentially expressed homeologous genes between pairs of subgenomes for each sample.

Sample	AB	AD	BD
W 10 DPA	967	865	829
W 20 DPA	542	456	446
SE 20 DPA	749	723	729
SE 30 DPA	745	637	659
GRAIN Z71	4190	3991	3922
GRAIN Z75	3450	3114	3197
LEAF Z10	2318	2148	2316
LEAF Z23	3816	3446	3501
LEAF Z71	4468	4252	4209
ROOT Z10	4921	4569	4575
ROOT Z13	204	152	144
ROOT Z39	247	60	71
SPIKE Z32	2465	1992	2191
SPIKE Z39	2656	2232	2334
SPIKE Z65	3823	3432	3438
STEM Z30	3144	2666	2836
STEM Z32	4022	3734	3800
STEM Z65	4260	4165	4141

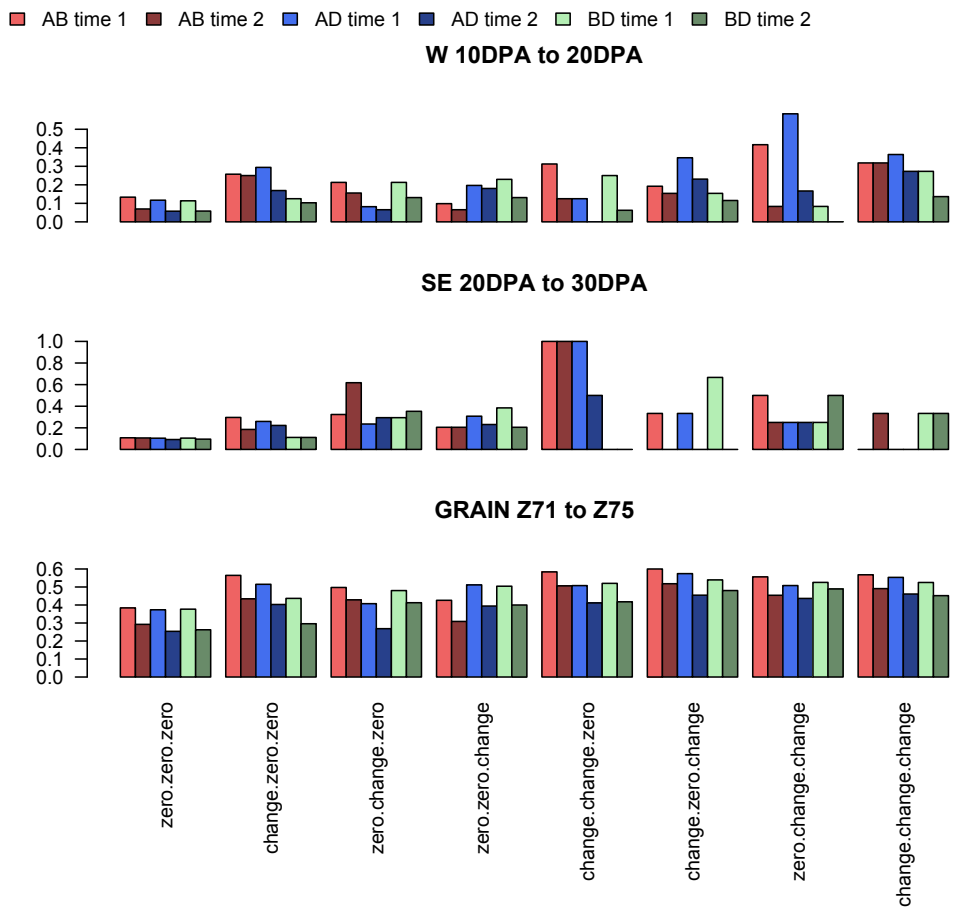


Figure 13: The fraction of triplets in each regulation category that is differentially expressed between pairs of subgenomes for time 1 and time 2. For W 10DPA to 20DPA, SE 20DPA to 30DPA and GRAIN Z71 to Z75.

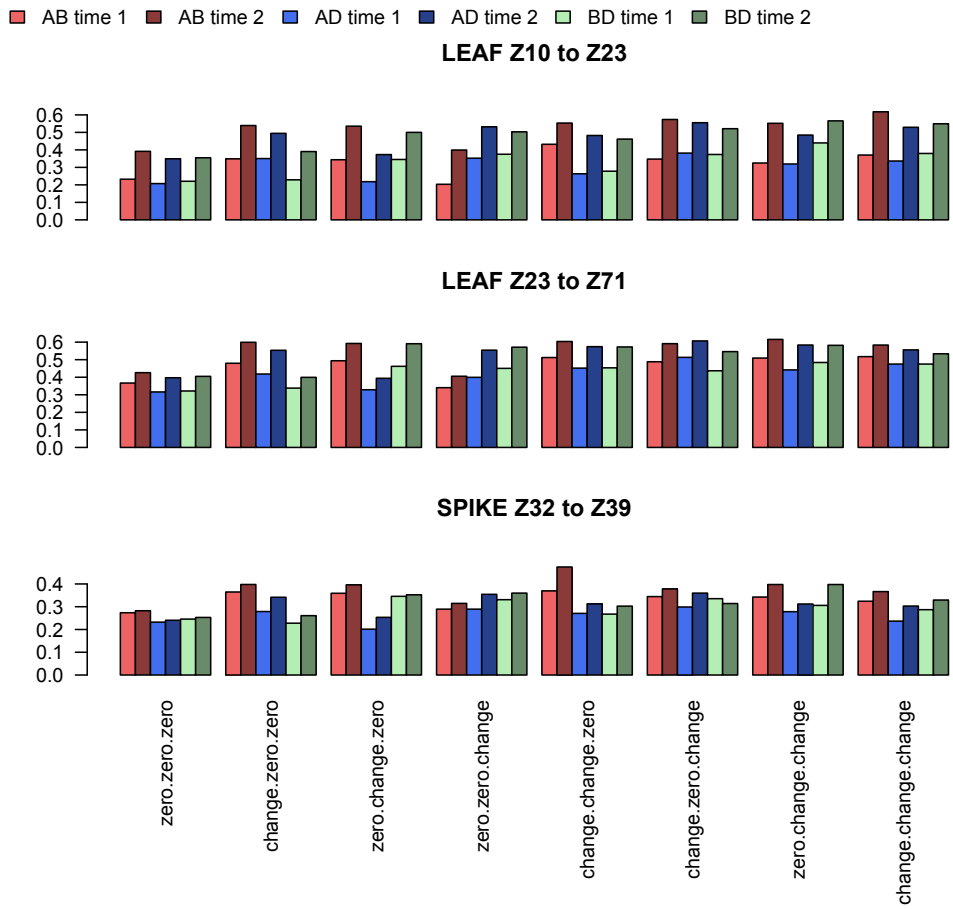


Figure 14: The fraction of triplets in each regulation category that is differentially expressed between pairs of subgenomes for time 1 and time 2. For LEAF Z10 to Z23, LEAF Z23 to Z71 and SPIKE Z32 to Z39.

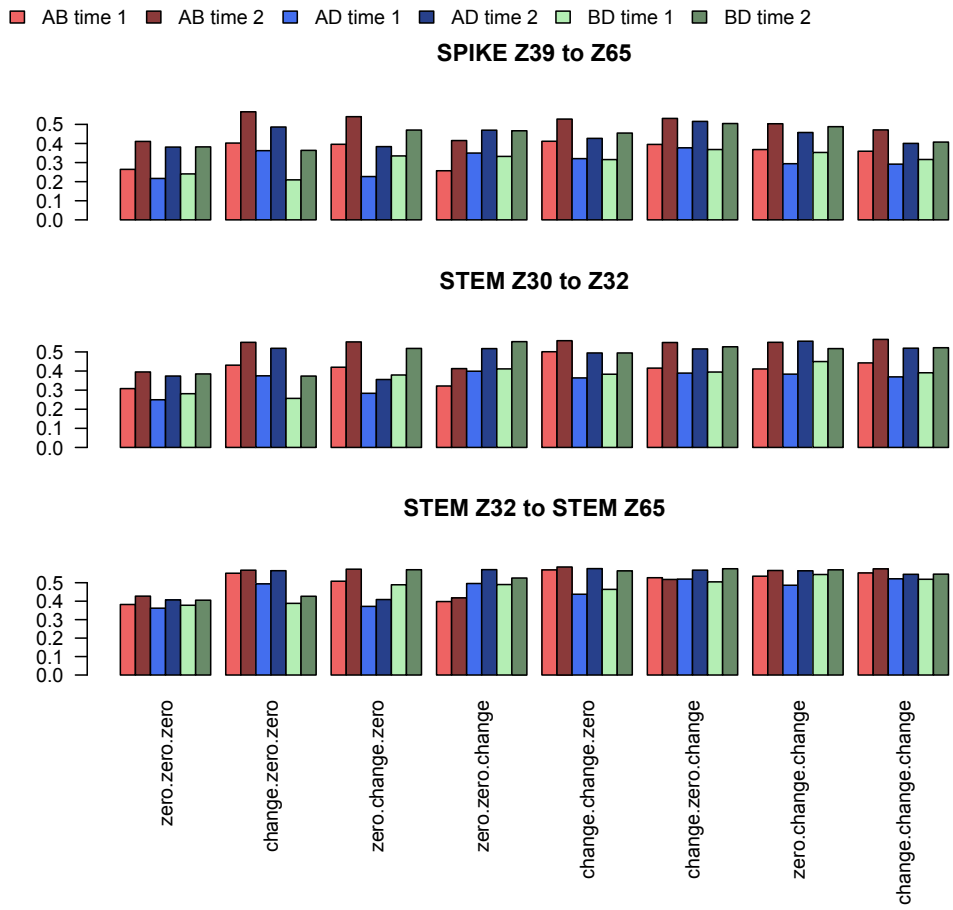


Figure 15: The fraction of triplets in each regulation category that is differentially expressed between pairs of subgenomes for time 1 and time 2. For SPIKE Z39 to Z65, STEM Z30 to Z32 and STEM Z32 to STEM Z65.

3.7 GO enrichment

GO enrichment analyses were done based both on how the subgenomes are regulated (differential expression between time points) and their levels of expression (differential expression between homeologs) (see methods chapter 2.7). The goal in both cases was to investigate if enriched GO terms differ between subgenomes, but also to compare the results between the two different approaches.

Tables with the results for GO enrichment based on both regulation and expression for all three subgenomes and for both the molecular function and biological process ontologies can be found in appendix D. GO terms with $p \leq 0.025$ are included.

For both regulation and expression there are some GO terms that are enriched in more than one of the subgenomes (Tables 8-11). The number GO terms that are shared between subgenomes is larger for regulation than it is for expression levels. However, most of the enriched GO terms are specific for one of the subgenomes for both methods, indicating that subgenomes contribute differently to certain molecular functions and biological processes.

Table 8: GO terms that are enriched in more than one subgenome for GO enrichment based on expression using molecular function GO terms.

GO ID	Term	Sample in A	Sample in B	Sample in D
GO:0008270	zinc ion binding	SPIKE Z65	SE 20 DPA	LEAF Z71
GO:0003984	acetolactate synthase activity		W 10 DPA	STEM Z65
GO:0003993	acid phosphatase activity		LEAF Z10	SPIKE Z39
GO:0016787	hydrolase activity		LEAF Z71	STEM Z32
				STEM Z65

Table 9: GO terms that are enriched in more than one subgenome for GO enrichment based on expression using biological process GO terms.

GO ID	Term	Sample in A	Sample in B	Sample in D
GO:0006457	protein folding	W 10 DPA	SE 20 DPA	
GO:0006096	glycolytic process	W 20 DPA	W 10 DPA	
			STEM Z30	
GO:0005978	glycogen biosynthetic process	W 20 DPA		W 20 DPA
				W 10 DPA

Table 10: GO terms that are enriched in more than one subgenome for GO enrichment based on regulation using molecular function GO terms.

GO ID	Term	Samples in A	Samples in B	Samples in D
GO:0005215	transporter activity	W 10 DPA to 20 DPA STEM Z30 to Z32	STEM Z32 to Z65 STEM Z30 to Z32	LEAF Z23 to Z71 SPIKE Z32 to Z39
GO:0004416	hydroxyacylglutathione hydrolase activit...	W 10 DPA to 20 DPA	W 10 DPA to 20 DPA	STEM Z30 to Z32
GO:0008270	zinc ion binding	GRAIN Z71 to Z75	GRAIN Z71 to Z75	LEAF Z23 to Z71 STEM Z32 to Z65
GO:0003824	catalytic activity	LEAF Z23 to Z71 SPIKE Z39 to Z65 STEM Z30 to Z32	STEM Z30 to Z32	STEM Z30 to Z32 SPIKE Z39 to Z65
GO:0004806	triglyceride lipase activity	SPIKE Z39 to Z65	SE 20 DPA to 30 DPA STEM Z32 to Z65	
GO:0022891	substrate-specific transmembrane transpo...	SPIKE Z39 to Z65 STEM Z30 to Z32	SPIKE Z32 to Z39	
GO:0043531	ADP binding	STEM Z30 to Z32	LEAF Z10 to Z23	LEAF Z10 to Z23
GO:0004674	protein serine/threonine kinase activity	LEAF Z10 to Z23		LEAF Z23 to Z71
GO:0004185	serine-type carboxypeptidase activity	SPIKE Z32 to Z39		GRAIN Z71 to Z75
GO:0042578	phosphoric ester hydrolase activity	SPIKE Z32 to Z39		SPIKE Z32 to Z39
GO:0004089	carbonate dehydratase activity	SPIKE Z39 to Z65		GRAIN Z71 to Z75 SPIKE Z39 to Z65 STEM Z30 to Z32
GO:0015095	magnesium ion transmembrane transporter ...	STEM Z30 to Z32		LEAF Z23 to Z71
GO:0005524	ATP binding		LEAF Z10 to Z23	LEAF Z10 to Z23
GO:0016787	hydrolase activity		STEM Z32 to Z65 SPIKE Z32 to Z39 LEAF Z23 to Z71	STEM Z30 to Z32
GO:0017057	6-phosphogluconolactonase activity		STEM Z30 to Z32	STEM Z30 to Z32 LEAF Z23 to Z71
GO:0003942	N-acetyl-gamma-glutamyl-phosphate reduct...		STEM Z32 to Z65	SPIKE Z39 to Z65

Table 11: GO terms that are enriched in more than one subgenome for GO enrichment based on regulation using biological process GO terms.

GO ID	Term	Samples in A	Samples in B	Samples in D
GO:0030154	cell differentiation	GRAIN Z71 to Z75	GRAIN Z71 to Z75	
GO:0055114	oxidation-reduction process	SPIKE Z39 to Z65 SPIKE Z32 to Z39 GRAIN Z71 to Z75	SPIKE Z39 to Z65	SPIKE Z32 to Z39
GO:0055085	transmembrane transport	SPIKE Z39 to Z65 LEAF Z10 to Z23	SPIKE Z32 to Z39	
GO:0005985	sucrose metabolic process	SPIKE Z32 to Z39	STEM Z32 to Z65	
GO:0006952	defense response	STEM Z30 to Z32	LEAF Z10 to Z23	LEAF Z10 to Z23
GO:0006810	transport	LEAF Z23 to Z71 STEM Z30 to Z32		LEAF Z23 to Z71
GO:0006807	nitrogen compound metabolic process	SPIKE Z39 to Z65		SPIKE Z32 to Z39
GO:0015976	carbon utilization	SPIKE Z39 to Z65		STEM Z30 to Z32
GO:0009408	response to heat	SPIKE Z39 to Z65		LEAF Z10 to Z23
GO:0015693	magnesium ion transport	STEM Z30 to Z32		LEAF Z23 to Z71
GO:0010380	regulation of chlorophyll biosynthetic p...	STEM Z32 to Z65		STEM Z32 to Z65
GO:0006108	malate metabolic process		SE 20 DPA to 30 DPA	W 10 DPA to 20 DPA
GO:0008152	metabolic process		SPIKE Z39 to Z65	STEM Z30 to Z32
GO:0016559	peroxisome fission		SPIKE Z39 to Z65	SPIKE Z39 to Z65

4 Discussion

4.1 Similarity between the subgenomes

Previous studies of homeolog specific gene expression in wheat have mainly compared expression levels between homeologous genes in a triplet in the same sample. Different studies have found varying numbers of triplets with preferential expression of a subgenome (Mochida et al., 2004; Pumphrey et al., 2009; Akhunova et al., 2010; Chagué et al., 2010; The International Wheat Genome Sequencing Consortium (IWGSC), 2014; Pfeifer et al., 2014). However, which subgenome that is preferentially expressed varies between tissues and no overall expression dominance for any of the subgenomes has been found.

In this thesis an approach has been used that investigates the similarity in regulation rather than expression levels, through tests for temporal changes in expression (i.e. regulation) within tissues. The goal was to see if we could come to some general conclusions about the regulation of subgenomes, and especially to test the hypotheis that the subgenomes are regulatory autonomous. However, it was also tested for differential expression between subgenomes in the tissues at each of the time points sampled, which is more directly comparable to previous studies.

The results show that all three subgenomes have very similar regulation patterns. First of all, gene regulation is most often conserved between subgenomes, as the largest regulation categories are either "3 changes" or "0 changes" (Figures 5-6). In addition, when two of the three genes in a triplet has changed expression, the different combinations of the A, B and D homeologs seem to occur at about the same frequency (Figures 5-6), indicating that the regulation is equally similar for all three combinations. We also see that homeologs are almost never regulated in opposite directions (Figure 6).

Although no significant difference in regulation similarity was found between subgenomes overall (Table 6), the A and B subgenomes were regulated most differently in many tissues (Table 5). These results are also supported by analysis of differential expression between subgenomes, as the total number of differentially expressed homeologs is larger between the A and B subgenomes than between A and D or B and D for all tissues (Table 7). This is consistent with the evolutionary history of bread wheat based on gene contents and sequence evolution (Marcussen et al., 2014; The International Wheat Genome Sequencing Consortium (IWGSC), 2014).

4.2 Regulation similarity between tissues

The number of triplets in the different regulation categories is highly correlated between tissues (Figure 7). However, which triplets these are varies (Figure 8). This variation is smallest when the three homeologous genes are regulated in the same way and larger when only one or two of them changes expression (Figure 8). In other words, triplets that are conserved within tissues are also conserved between tissues (Figure 16).

It is interesting to note that triplets where all three homeologs have constant expression ("zero changes") are largely the same in all tissues (Figure 8). This suggests that they are housekeeping genes, as housekeeping genes are required to maintain basic cellular functions and are expressed in all tissues under normal conditions.

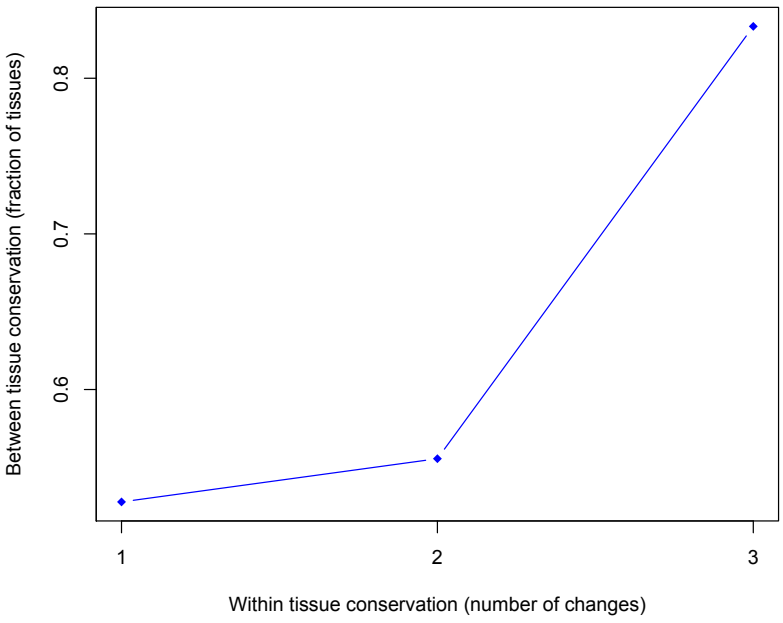


Figure 16: The fraction of tissues with a significant overlap of triplets plotted against the number of homeologs in the triplets that changes expression.

4.3 Differences in results between the two data sets

We have seen that there are few expressed genes and also few genes with significant changes in expression in the W (whole endosperm) and SE (starchy endosperm) samples compared with the other samples (Table 4). Especially we would expect W and SE samples to show similar results to the grain samples (Figure 2).

One theory that could explain what we see, is that there are some very highly expressed genes in these samples that account for a large proportion of the reads. However, looking at the the gene expression for the most highly expressed genes in these and some other samples, we see that this is not the case (Figure 17).

Another theory is that there probably are some differences in the methods used for generating fpkm values, as these samples are from a different data set than the other samples.

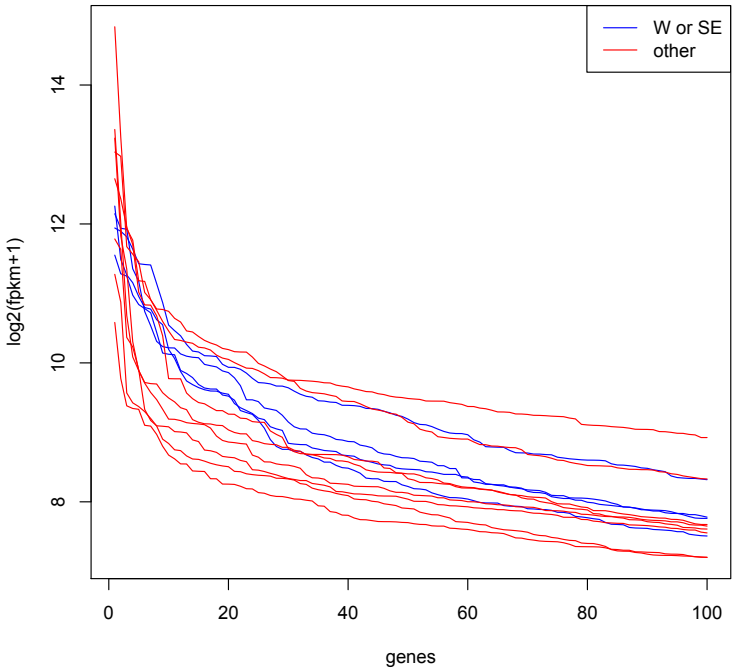


Figure 17: Expression values measured in log2(fpkm+1) for the 100 genes with the highest expression for W- and SE-samples and some of the other samples.

4.4 Limitations of the differential expression analysis

As all the analyses in this thesis are initially based on differential expression with limma, it is worth mentioning that the fact that we have very few replicates might influence the results. One consequence of this is that the t-test does not find a significant difference between highly expressed genes even though the fold change is large. This is due to the large variance we get when we have few replicates even with the Empirical Bayes variance adjustment. I therefore tried doing the clustering in figure 9 based on fold change instead of t-values and this gave the same result regarding clustering of tissues rather than subgenomes.

4.5 GO enrichment

The goal of the GO enrichment analyses was to see if the subgenomes contribute differently to any molecular functions or biological processes, by testing if any GO terms are enriched for only one of the subgenomes. This was tested based both on how the subgenomes are regulated and their expression levels.

The results show that there are some GO terms that are enriched in more than one subgenome for both the molecular function and biological process ontologies and both methods tested, but most often this is in different tissues (Tables 8-11). However, most of the enriched GO terms are subgenome specific. That some of the GO terms are shared between two or three of the subgenomes also suggests that there are different selection pressures on different functions.

Pfeifer et al. (2014) also found that subgenomes contributed differently to particular functions. They mostly found different enriched GO terms than what was found here, but apart from W and SE they also analysed different tissues. They reported no enriched terms in W and one of the two enriched terms in SE was also enriched in this analysis. This was the term regulation of autophagy which they found to be enriched in the D subgenome. In this thesis the term was also enriched in the D subgenome in the analysis based on regulation, but in grain instead of SE.

The results also show that regulation has more shared GO terms between subgenomes than expression. This is not surprising as we have seen that the subgenomes are very similar in regulation even though the expression levels are different.

4.6 Comparing expression levels to expression regulation in the subgenomes

4.6.1 The subgenomes are regulated similarly, but differ in expression levels

Both IWGSC (2014) and Pfeifer et al. (2014) found that the bread wheat subgenomes show regulatory and transcriptional autonomy, i.e. that genome-specific gene expression dominates over tissue-specific gene expression. These general results regarding expression levels were reproduced and comfirmed in this thesis using similar approaches (Figure 10). In this thesis we did however observe some clustering of tissues, but this difference between studies is likely an effect of clustering all time points individually rather than the mean for each tissue.

However, in this thesis it was also shown that clustering based on expression regulation resulted in clustering of tissues (Figure 9). This difference in clustering between expression levels and expression regulation indicates that there is differential expression between subgenomes in the samples, even though the subgenomes are regulated very similarly. This is also what we see in the results, as there are many triplets with differential expression between pairs of subgenomes (Table 7). The exact number varies between tissues, but is usually between 25% and 50%. The exceptions are W, SE and root. Furthermore, figures 13-15 show that for most tissues there is no clear connection between the number of triplets that are differentially expressed between pairs of subgenomes and which regulation category the triplets are in. For comparison, IWGSC (2014) found that only about 21% of the triplets showed an expression bias in one of the pairwise comparisons between subgenomes based on the same data. This difference is possibly because they used an average across tissues. Or it can be the result of using a different method for differential expression analysis than what was used in this thesis.

4.6.2 Mechanisms underlying gene expression divergence of subgenomes

We have seen that there are large differences in homeolog specific expression in the samples, even though the subgenomes are very similar in regulation. What possible mechanisms can explain this observation?

One hypothesis is divergence of regulatory elements. Hexaploid wheat is quite young and therefore it is not likely that mutations that has happened after polyploidization is the main reason for the differences we see in expression. If on the other hand the variations we see are caused by differences that already existed in the diploid ancestors before polyploidization, this would

have enabled mutations to happen over a much longer time. However, the fact that we also observe very similar regulation of the subgenomes imply that the homeologous genes are still turned on by the binding of the same transcription factors, but that the expression levels are different.

The alternative hypothesis is that homeologous gene expression is regulated through epigenetics, which is changes in gene expression that is not due to differences in DNA sequence. This is more likely, as epigentic changes happen much faster. Two epigenetic mechanisms that function on the transcriptional level are histone modifications and DNA methylation.

Chromatin is a complex of DNA and proteins that makes up eukaryotic chromosomes. The basic structural unit of chromatin is the nucleosome, which is double stranded DNA wrapped around octamers of histone proteins. The primary function of chromatin is to package the DNA into a smaller volume so that it can fit into the cell. Both the degree of chromatin packaging in the region surrounding a gene and the exact positioning of nucleosomes determine the accessibility of the DNA strand to transcription factors and RNA polymerase II, and thereby the activity of a gene. Histone modifications are covalent modifications of amino terminal tails of histone proteins, e.g. acetylation, methylation and phosphorylation. These modifications can influence the interactions both between the DNA and histone proteins, and between adjacent nucleosomes causing alterations in chromatine structure. Histone modifications also provide binding sites that are recognized by other proteins. Strahl and Allis (2000) proposed the hypothesis of the histone code, where histone modifications act in combination to form a pattern or code that is recognized by different proteins. Some proteins may cause the chromatin to be less accessible and thereby silence gene transcription, and other proteins may promote transcription.

DNA methylation silences gene expression by the attachment of methyl groups to cytosine bases in the DNA. Many genes contain sequences called CpG islands near their promoters, which are clusters of cytosines and guanines connected by a phosphodiester linkage. Unmethylated CpG islands are usually correlated with active genes and methylated CpG islands with repressed genes. Silencing of genes by methylation can happen in two ways: methylation can physically block transcription factors from binding to the DNA, or methylated DNA can be bound by methyl-CpG-binding proteins which recruit additional proteins that can modify histones to increase the degree of chromatin packaging.

Several studies support the hypothesis that homeolog expression divergence in wheat may be the result of epigentic regulation. The studies by Shitsukawa et al. (2007) and Hu et al. (2013), mentioned in chapter 1.3, both found examples of homeologous genes that were differentially expressed

due to epigenetic differences rather than alterations in transcription factor binding sites. Based on the knowledge that epigenetic mechanisms often work on neighbouring genes, Pfeifer et al. (2014) investigated the relationship between gene expression and chromosomal locations of the genes in wheat. They found that along all chromosomes, gene expression changed according to chromosome domains and that this expression was similar between subgenomes. But they also found some chromosomal domains where the gene expression was not similar between homeologs. Based on this they suggest that epigenetic regulatory mechanisms act differently on certain corresponding domains of homeologous chromosomes.

Epigentic regulation of gene expression has also been studied in other species. One example is the study by Ha et al. (2011) where it was found that histone acetylation and deacetylation together with histone methylation is connected with variation in gene expression between homeologous genes in Arabidopsis. Another example is Karlić et al. (2010) who found that the level of gene expression is highly correlated with the level of histone modification in humans. Furthermore, they were able to predict gene expression levels of one cell type using a model trained on another cell type, which shows that the relationship between gene expression and level of histone modification is general.

Epigenetic gene regulation could possibly also be linked to the activation of transposable elements adjacent to the genes. Transposable elements (TEs) are segments of DNA that can move from one position to another in the genome. They were first discovered by Barbara McClintock in the 1940s while studying maize (McClintock, 1956). Already then it was suggested that the presence of these element at or near the locus of a gene may affect gene expression and that the changes are heritable. A more recent study in maize by Makarevitch et al. (2015) shows strong correlation between stress responsive upregulation of gene expression and the upstream insertions of TEs.

When TEs insert into the genome, this will have consequences that are most often harmful to the organism. TEs can for example cause chromosome breakage and genome rearrangements and if TEs insert into the regulatory or coding region of a gene this can inactivate the gene or alter gene expression (Muñoz-López and García-Pérez, 2010; Slotkin and Martienssen, 2007).

Because of the potentially damaging effects of active TEs, most TEs are controlled through mechanisms that recognize and silence them using epigenetic mechanisms including modifications of histone tails and DNA methylation (Slotkin and Martienssen, 2007; Lisch, 2009). In many cases this epigenetic silencing of TEs also affect the regulation of nearby genes (Lisch, 2009; Slotkin and Martienssen, 2007).

A variety of conditions, like stress in response to the environment or hybridization by allopolyploidization, has been shown to increase the number of active transposons (Lisch, 2009; Fedoroff, 2012). As transposons are known to contribute to the regulation of nearby genes, changes in transposon activity as a result of polyploidization is therefore likely to have an effect on epigenetic regulation of homeologous genes in wheat. One example of this is the study by Kashkush et al. (2003). They showed that the transcriptional activity of Wis 2-1A retrotransposons is much higher in newly synthesized wheat amphiploids (having at least one diploid set of chromosomes from each parent species), and that this is associated with silencing or activation of adjacent genes.

To sum up, the results in this thesis suggest that epigenetics is the primary mechanism underlying homeolog expression divergence in wheat. The main argument in support of this hypothesis is that the developmental regulation, i.e. the turning on and off of genes, largely seems to be conserved between subgenomes, while the expression levels are highly different. This will be expected if there are systematic chromatin packaging differences in syntenic genomic regions of the subgenomes. If for instance packaging in a region is higher in the B subgenome, then the transcription factors, that most likely bind to all subgenomes, will turn on the genes in B to a lesser degree compared to A and D.

4.7 Future work

The results in this thesis provide a contribution towards a better understanding of how the expression of homeologous genes in wheat is regulated. However, it is not possible to know for certain the exact mechanisms that cause the variations we see in homeolog expression from the data that this thesis is based on. The hypothesis, that epigenetic regulation is the main cause of homeolog expression divergence, is therefore something that should be investigated further. This could be done by testing for correlation between subgenome expression levels and chromatin states at each homeolog loci, where we would expect the correlation to be high if the hypothesis is true. It could also be tested for correlation between expression divergence and divergence in regulatory elements. In this case we would predict a low correlation. It would also be interesting to investigate if the results we observe regarding homeolog regulation in wheat apply to allopolyploid species in general.

4.8 Conclusion

Differential expression between homeologous genes is common, and the largest differences in expression levels are between the A and B subgenomes, suggesting that these two subgenomes are evolutionary least related. But although the expression levels vary between homeologs, the developmental regulation of the subgenomes is very similar. These results suggest that homeologous genes are turned on and off by the same transcription factors, but that differences in chromatin packaging in syntenic genomic regions of the subgenomes result in varying levels of expression. In this thesis it is therefore proposed that homeolog expression divergence in wheat is in large part caused by epigenetic regulation, perhaps in connection with an increase in the activity of transposable elements as a result of polyploidization. This is something that should be investigated further.

Bibliography

Adams, K. L., Cronn, R., Percifield, R., and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of sciences*, 100(8):4649–4654.

Akhunova, A. R., Matniyazov, R. T., Liang, H., and Akhunov, E. D. (2010). Homoeolog-specific transcriptional bias in allopolyploid wheat. *BMC genomics*, 11(1):1.

Alexa, A. and Rahnenfuhrer, J. (2010). *topGO: topGO: Enrichment analysis for Gene Ontology*. R package version 2.20.0.

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating go graph structure. *Bioinformatics*, 22(13):1600–1607.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29.

Berkeley, C. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *E-book available at http://www. bepress. com/sagmb/vol3/iss1/art3.[PubMed]*.

Birchler, J. A., Riddle, N. C., Auger, D. L., and Veitia, R. A. (2005). Dosage balance in gene regulation: biological implications. *Trends in Genetics*, 21(4):219–226.

Bottley, A. and Koebner, R. (2008). Variation for homoeologous gene silencing in hexaploid wheat. *The Plant Journal*, 56(2):297–302.

Bottley, A., Xia, G., and Koebner, R. (2006). Homoeologous gene silencing in hexaploid wheat. *The Plant Journal*, 47(6):897–906.

Canty, A. and Ripley, B. D. (2015). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-17.

Chagué, V., Just, J., Mestiri, I., Balzergue, S., Tanguy, A.-M., Huneau, C., Huteau, V., Belcram, H., Coriton, O., Jahier, J., et al. (2010). Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytologist*, 187(4):1181–1194.

Chaudhary, B., Flagel, L., Stupar, R. M., Udall, J. A., Verma, N., Springer, N. M., and Wendel, J. F. (2009). Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). *Genetics*, 182(2):503–517.

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature reviews genetics*, 6(11):836–846.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2.

Fedoroff, N. V. (2012). Transposable elements, epigenetics, and genome evolution. *Science*, 338(6108):758–767.

Feldman, M., Levy, A. A., Fahima, T., and Korol, A. (2012). Genomic asymmetry in allopolyploid plants: wheat as a model. *Journal of experimental botany*.

Gene Ontology Consortium and others (2015). Gene ontology consortium: going forward. *Nucleic acids research*, 43(D1):D1049–D1056.

Ha, M., Ng, D. W., Li, W.-H., and Chen, Z. J. (2011). Coordinated histone modifications are associated with gene expression variation within and between species. *Genome research*, 21(4):590–598.

He, P., Friebe, B. R., Gill, B. S., and Zhou, J.-M. (2003). Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant molecular biology*, 52(2):401–414.

Hegarty, M. J., Barker, G. L., Brennan, A. C., Edwards, K. J., Abbott, R. J., and Hiscock, S. J. (2008). Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in senecio. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1506):3055–3069.

Herbek, J. and Lee, C. (2009). Section 2: Growth and development. *A Comprehensive Guide to Wheat Management in Kentucky*, pages 6–13.

Hu, Z., Han, Z., Song, N., Chai, L., Yao, Y., Peng, H., Ni, Z., and Sun, Q. (2013). Epigenetic modification contributes to the expression divergence of three taexpa1 homoeologs in hexaploid wheat (triticum aestivum). *New Phytologist*, 197(4):1344–1352.

Karlić, R., Chung, H.-R., Lasserre, J., Vlahoviček, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences*, 107(7):2926–2931.

Kashkush, K., Feldman, M., and Levy, A. A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*, 160(4):1651–1659.

Kashkush, K., Feldman, M., and Levy, A. A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature genetics*, 33(1):102–106.

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. *Annual review of plant biology*, 60:43–66.

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., and Springer, N. M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet*, 11(1):e1004915.

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K. S., Wulff, B. B., Steuernagel, B., Mayer, K. F., Olsen, O.-A., et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, 345(6194):1250092.

Martinez-Perez, E., Shaw, P., and Moore, G. (2001). The ph1 locus is needed to ensure specific somatic and meiotic centromere association. *Nature*, 411(6834):204–207.

McClintock, B. (1956). Controlling elements and the gene. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 21, pages 197–216. Cold Spring Harbor Laboratory Press.

Mochida, K., Yamazaki, Y., and Ogihara, Y. (2004). Discrimination of homoeologous gene expression in hexaploid wheat by snp analysis of contigs grouped from a large number of expressed sequence tags. *Molecular Genetics and Genomics*, 270(5):371–377.

Muñoz-López, M. and García-Pérez, J. L. (2010). Dna transposons: nature and applications in genomics. *Current genomics*, 11(2):115–128.

Ozkan, H., Tuna, M., and Arumuganathan, K. (2003). Nonadditive changes in genome size during allopolyploidization in the wheat (aegilops-triticum) group. *Journal of Heredity*, 94(3):260–264.

Parnell, L. D., Lindenbaum, P., Shameer, K., Dall'Olio, G. M., Swan, D. C., Jensen, L. J., Cockell, S. J., Pedersen, B. S., Mangan, M. E., Miller, C. A., et al. (2011). Biostar: an online question & answer resource for the bioinformatics community. *PLoS Comput Biol*, 7(10):e1002216.

Petersen, G., Seberg, O., Yde, M., and Berthelsen, K. (2006). Phylogenetic relationships of triticum and aegilops and evidence for the origin of the a, b, and d genomes of common wheat (triticum aestivum). *Molecular phylogenetics and evolution*, 39(1):70–82.

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., Mayer, K. F., Olsen, O.-A., et al. (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science*, 345(6194):1250091.

Pumphrey, M., Bai, J., Laudencia-Chingcuanco, D., Anderson, O., and Gill, B. S. (2009). Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. *Genetics*, 181(3):1147–1157.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47.

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*, 3(6):429–441.

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *The Plant Cell*, 13(8):1749–1759.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic biology*, 51(3):492–508.

Shitsukawa, N., Tahira, C., Kassai, K.-i., Hirabayashi, C., Shimizu, T., Takumi, S., Mochida, K., Kawaura, K., Ogihara, Y., and Murai, K. (2007). Genetic and epigenetic alteration among three homoeologous genes of a class e mads box gene in hexaploid wheat. *The Plant Cell*, 19(6):1723–1737.

Slotkin, R. K. and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, 8(4):272–285.

Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765):41–45.

Suzuki, R. and Shimodaira, H. (2006). Pvclust: an r package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12):1540–1542.

The farrago (2014). What the fpkm? a review of rna-seq expression units `https://haroldpimentel.wordpress.com/2014/05/08/what-the-fpkm-a-review-rna-seq-expression-units/`.

The International Wheat Genome Sequencing Consortium (IWGSC) (2014). A chromosome-based draft sequence of the hexaploid bread wheat (triticum aestivum) genome. *Science*, 345(6194):1251788.

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, 10(10):725–732.

Wang, Z., Gerstein, M., and Snyder, M. (2009). Rna-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.

Yoo, M.-J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annual review of genetics*, 48:485–517.

Appendix A: Tables for the Fisher's exact tests for the regulation categories zero changes, one change, two changes and three changes.

Table 12: P-values from Fisher's exact tests for the triplets with zero change. There are 25815 genes in total.

Pairs of tissues tested	Adjusted p-values	Overlap	Genes in tissue 1	Genes in tissue 2
W 10DPA to 20DPA and SE 20DPA to 30DPA	0	18186	19614	19938
STEM Z30 to Z32 and STEM Z32 to Z65	7.8e-240	3357	8988	6438
LEAF Z10 to Z23 and STEM Z32 to Z65	6.9e-198	4158	12501	6438
LEAF Z23 to Z71 and STEM Z30 to Z32	1.1e-178	2610	4971	8988
SPIKE Z32 to Z39 and SPIKE Z39 to Z65	3.9e-162	7200	14079	11232
LEAF Z23 to Z71 and STEM Z32 to Z65	4.1e-159	2007	4971	6438
LEAF Z10 to Z23 and LEAF Z23 to Z71	2.5e-152	3237	12501	4971
GRAIN Z71 to Z75 and STEM Z32 to Z65	3.1e-144	2034	5190	6438
SPIKE Z32 to Z39 and STEM Z30 to Z32	3.1e-129	5820	14079	8988
SPIKE Z39 to Z65 and STEM Z32 to Z65	3.5e-104	3552	11232	6438
W 10DPA to 20DPA and LEAF Z10 to Z23	1.8e-90	10188	19614	12501
LEAF Z10 to Z23 and SPIKE Z39 to Z65	6.4e-90	6240	12501	11232
LEAF Z10 to Z23 and STEM Z30 to Z32	5.7e-87	5109	12501	8988
GRAIN Z71 to Z75 and STEM Z30 to Z32	6.8e-84	2412	5190	8988
GRAIN Z71 to Z75 and LEAF Z10 to Z23	1.7e-82	3132	5190	12501
GRAIN Z71 to Z75 and LEAF Z23 to Z71	7.2e-68	1458	5190	4971
LEAF Z23 to Z71 and SPIKE Z39 to Z65	4.7e-67	2709	4971	11232
SE 20DPA to 30DPA and LEAF Z10 to Z23	4e-63	10218	19938	12501
SE 20DPA to 30DPA and SPIKE Z32 to Z39	1.2e-57	11412	19938	14079
GRAIN Z71 to Z75 and SPIKE Z39 to Z65	1.4e-54	2757	5190	11232
W 10DPA to 20DPA and SPIKE Z32 to Z39	2.1e-54	11229	19614	14079
SPIKE Z39 to Z65 and STEM Z30 to Z32	1.1e-53	4497	11232	8988
W 10DPA to 20DPA and SPIKE Z39 to Z65	1.9e-53	9054	19614	11232
SPIKE Z32 to Z39 and STEM Z32 to Z65	5.2e-45	3996	14079	6438
LEAF Z23 to Z71 and SPIKE Z32 to Z39	7.7e-43	3141	4971	14079
SE 20DPA to 30DPA and SPIKE Z39 to Z65	1.7e-39	9111	19938	11232
GRAIN Z71 to Z75 and SPIKE Z32 to Z39	2.4e-38	3243	5190	14079
W 10DPA to 20DPA and STEM Z32 to Z65	2.9e-37	5262	19614	6438
SE 20DPA to 30DPA and STEM Z32 to Z65	3.7e-24	5262	19938	6438
W 10DPA to 20DPA and LEAF Z23 to Z71	4.8e-17	3999	19614	4971
W 10DPA to 20DPA and STEM Z30 to Z32	3e-14	7074	19614	8988
SE 20DPA to 30DPA and LEAF Z23 to Z71	1.1e-13	4032	19938	4971
LEAF Z10 to Z23 and SPIKE Z32 to Z39	2.3e-10	7068	12501	14079
SE 20DPA to 30DPA and STEM Z30 to Z32	3.2e-09	7128	19938	8988
W 10DPA to 20DPA and GRAIN Z71 to Z75	0.0099	4008	19614	5190
SE 20DPA to 30DPA and GRAIN Z71 to Z75	0.79	3987	19938	5190

Table 13: P-values from Fisher's exact tests for triplets with one change. There are 25815 genes in total.

Pairs of tissues tested	Adjusted p-values	Overlap	Genes in tissue 1	Genes in tissue 2
STEM Z30 to Z32 and STEM Z32 to Z65	3.4e-40	1998	6471	6330
GRAIN Z71 to Z75 and STEM Z32 to Z65	3.4e-16	1884	6639	6330
LEAF Z23 to Z71 and STEM Z32 to Z65	3.4e-16	1674	5826	6330
LEAF Z23 to Z71 and STEM Z30 to Z32	9.2e-13	1677	5826	6471
SPIKE Z32 to Z39 and SPIKE Z39 to Z65	4.2e-10	1362	5265	5811
SPIKE Z39 to Z65 and STEM Z30 to Z32	2.1e-09	1638	5811	6471
LEAF Z23 to Z71 and SPIKE Z39 to Z65	6.3e-07	1458	5826	5811
GRAIN Z71 to Z75 and LEAF Z23 to Z71	2e-06	1644	6639	5826
GRAIN Z71 to Z75 and SPIKE Z39 to Z65	0.00011	1614	6639	5811
GRAIN Z71 to Z75 and SPIKE Z32 to Z39	0.00014	1467	6639	5265
LEAF Z10 to Z23 and LEAF Z23 to Z71	0.00016	1443	5901	5826
LEAF Z10 to Z23 and STEM Z30 to Z32	0.00039	1587	5901	6471
LEAF Z10 to Z23 and STEM Z32 to Z65	0.0016	1542	5901	6330
LEAF Z10 to Z23 and SPIKE Z39 to Z65	0.0019	1419	5901	5811
SPIKE Z32 to Z39 and STEM Z30 to Z32	0.005	1401	5265	6471
W 10DPA to 20DPA and LEAF Z10 to Z23	0.012	297	1140	5901
GRAIN Z71 to Z75 and STEM Z30 to Z32	0.014	1740	6639	6471
SPIKE Z39 to Z65 and STEM Z32 to Z65	0.018	1494	5811	6330
SE 20DPA to 30DPA and SPIKE Z32 to Z39	0.021	78	300	5265
GRAIN Z71 to Z75 and LEAF Z10 to Z23	0.076	1569	6639	5901
SE 20DPA to 30DPA and SPIKE Z39 to Z65	0.14	78	300	5811
W 10DPA to 20DPA and SE 20DPA to 30DPA	0.19	18	1140	300
LEAF Z23 to Z71 and SPIKE Z32 to Z39	0.22	1218	5826	5265
SE 20DPA to 30DPA and LEAF Z10 to Z23	0.31	75	300	5901
LEAF Z10 to Z23 and SPIKE Z32 to Z39	0.33	1224	5901	5265
W 10DPA to 20DPA and SPIKE Z39 to Z65	0.43	264	1140	5811
SE 20DPA to 30DPA and GRAIN Z71 to Z75	0.64	78	300	6639
SE 20DPA to 30DPA and STEM Z30 to Z32	0.69	75	300	6471
SPIKE Z32 to Z39 and STEM Z32 to Z65	0.8	1281	5265	6330
W 10DPA to 20DPA and SPIKE Z32 to Z39	0.87	225	1140	5265
SE 20DPA to 30DPA and LEAF Z23 to Z71	0.89	63	300	5826
W 10DPA to 20DPA and GRAIN Z71 to Z75	1	219	1140	6639
W 10DPA to 20DPA and LEAF Z23 to Z71	1	225	1140	5826
W 10DPA to 20DPA and STEM Z30 to Z32	1	228	1140	6471
W 10DPA to 20DPA and STEM Z32 to Z65	1	249	1140	6330
SE 20DPA to 30DPA and STEM Z32 to Z65	1	57	300	6330

Table 14: P-values from Fisher's exact tests for triplets with two changes. There are 25815 genes in total.

Pairs of tissues tested	Adjusted p-values	Overlap	Genes in tissue 1	Genes in tissue 2
STEM Z30 to Z32 and STEM Z32 to Z65	6e-25	1140	4560	4965
GRAIN Z71 to Z75 and LEAF Z10 to Z23	3.5e-17	945	5967	3225
LEAF Z10 to Z23 and STEM Z32 to Z65	2.9e-14	789	3225	4965
GRAIN Z71 to Z75 and STEM Z30 to Z32	1.3e-13	1254	5967	4560
LEAF Z23 to Z71 and SPIKE Z39 to Z65	9.4e-12	1008	5541	3909
SPIKE Z32 to Z39 and SPIKE Z39 to Z65	1.7e-11	570	2907	3909
LEAF Z10 to Z23 and STEM Z30 to Z32	3e-11	711	3225	4560
GRAIN Z71 to Z75 and SPIKE Z32 to Z39	6.4e-10	810	5967	2907
GRAIN Z71 to Z75 and SPIKE Z39 to Z65	1.2e-09	1056	5967	3909
LEAF Z23 to Z71 and STEM Z30 to Z32	1.2e-09	1137	5541	4560
SPIKE Z32 to Z39 and STEM Z30 to Z32	1.3e-09	636	2907	4560
LEAF Z23 to Z71 and STEM Z32 to Z65	3.1e-09	1224	5541	4965
LEAF Z10 to Z23 and LEAF Z23 to Z71	3.5e-09	825	3225	5541
GRAIN Z71 to Z75 and STEM Z32 to Z65	1.5e-08	1302	5967	4965
LEAF Z10 to Z23 and SPIKE Z39 to Z65	8e-08	594	3225	3909
SPIKE Z39 to Z65 and STEM Z30 to Z32	2.2e-05	786	3909	4560
SPIKE Z39 to Z65 and STEM Z32 to Z65	4.8e-05	846	3909	4965
GRAIN Z71 to Z75 and LEAF Z23 to Z71	0.00062	1377	5967	5541
SE 20DPA to 30DPA and SPIKE Z39 to Z65	0.028	9	27	3909
W 10DPA to 20DPA and LEAF Z10 to Z23	0.031	30	162	3225
SPIKE Z32 to Z39 and STEM Z32 to Z65	0.075	594	2907	4965
W 10DPA to 20DPA and STEM Z30 to Z32	0.34	33	162	4560
LEAF Z23 to Z71 and SPIKE Z32 to Z39	0.45	636	5541	2907
SE 20DPA to 30DPA and STEM Z32 to Z65	0.63	6	27	4965
W 10DPA to 20DPA and LEAF Z23 to Z71	0.93	33	162	5541
SE 20DPA to 30DPA and LEAF Z10 to Z23	0.93	3	27	3225
W 10DPA to 20DPA and SE 20DPA to 30DPA	1	0	162	27
W 10DPA to 20DPA and GRAIN Z71 to Z75	1	30	162	5967
W 10DPA to 20DPA and SPIKE Z32 to Z39	1	9	162	2907
W 10DPA to 20DPA and SPIKE Z39 to Z65	1	21	162	3909
W 10DPA to 20DPA and STEM Z32 to Z65	1	24	162	4965
SE 20DPA to 30DPA and GRAIN Z71 to Z75	1	3	27	5967
SE 20DPA to 30DPA and LEAF Z23 to Z71	1	0	27	5541
SE 20DPA to 30DPA and SPIKE Z32 to Z39	1	0	27	2907
SE 20DPA to 30DPA and STEM Z30 to Z32	1	3	27	4560
LEAF Z10 to Z23 and SPIKE Z32 to Z39	1	351	3225	2907

Table 15: P-values from Fisher's exact tests for triplets with three changes. There are 25815 genes in total.

Pairs of tissues tested	Adjusted p-values	Overlap	Genes in tissue 1	Genes in tissue 2
LEAF Z23 to Z71 and STEM Z30 to Z32	0	2862	8280	4758
LEAF Z23 to Z71 and STEM Z32 to Z65	0	3600	8280	6912
STEM Z30 to Z32 and STEM Z32 to Z65	0	2427	4758	6912
LEAF Z10 to Z23 and STEM Z32 to Z65	3e-274	1656	2988	6912
GRAIN Z71 to Z75 and STEM Z32 to Z65	7.5e-220	2805	6684	6912
GRAIN Z71 to Z75 and LEAF Z10 to Z23	4.4e-156	1407	6684	2988
LEAF Z10 to Z23 and LEAF Z23 to Z71	2.9e-153	1614	2988	8280
SPIKE Z32 to Z39 and STEM Z32 to Z65	1.2e-134	1134	2268	6912
LEAF Z23 to Z71 and SPIKE Z39 to Z65	4.8e-128	1935	8280	3948
LEAF Z23 to Z71 and SPIKE Z32 to Z39	3.1e-123	1248	8280	2268
SPIKE Z32 to Z39 and STEM Z30 to Z32	5.1e-123	873	2268	4758
GRAIN Z71 to Z75 and STEM Z30 to Z32	1.5e-105	1848	6684	4758
LEAF Z10 to Z23 and STEM Z30 to Z32	6.5e-104	1014	2988	4758
GRAIN Z71 to Z75 and LEAF Z23 to Z71	1.9e-87	2805	6684	8280
SPIKE Z32 to Z39 and SPIKE Z39 to Z65	1.1e-85	699	2268	3948
SPIKE Z39 to Z65 and STEM Z32 to Z65	2.1e-72	1533	3948	6912
GRAIN Z71 to Z75 and SPIKE Z39 to Z65	9.7e-71	1488	6684	3948
SPIKE Z39 to Z65 and STEM Z30 to Z32	5.7e-39	1032	3948	4758
LEAF Z10 to Z23 and SPIKE Z39 to Z65	2.4e-24	654	2988	3948
GRAIN Z71 to Z75 and SPIKE Z32 to Z39	2e-23	792	6684	2268
W 10DPA to 20DPA and GRAIN Z71 to Z75	6e-18	51	66	6684
LEAF Z10 to Z23 and SPIKE Z32 to Z39	4.4e-08	345	2988	2268
W 10DPA to 20DPA and LEAF Z23 to Z71	2.1e-07	42	66	8280
W 10DPA to 20DPA and SE 20DPA to 30DPA	2e-06	3	66	9
W 10DPA to 20DPA and STEM Z32 to Z65	6.9e-05	33	66	6912
SE 20DPA to 30DPA and SPIKE Z39 to Z65	0.00098	6	9	3948
W 10DPA to 20DPA and LEAF Z10 to Z23	0.01	15	66	2988
SE 20DPA to 30DPA and GRAIN Z71 to Z75	0.015	6	9	6684
SE 20DPA to 30DPA and STEM Z32 to Z65	0.018	6	9	6912
SE 20DPA to 30DPA and LEAF Z23 to Z71	0.042	6	9	8280
SE 20DPA to 30DPA and LEAF Z10 to Z23	0.089	3	9	2988
SE 20DPA to 30DPA and STEM Z30 to Z32	0.25	3	9	4758
W 10DPA to 20DPA and SPIKE Z39 to Z65	0.33	12	66	3948
W 10DPA to 20DPA and SPIKE Z32 to Z39	0.56	6	66	2268
W 10DPA to 20DPA and STEM Z30 to Z32	0.91	9	66	4758
SE 20DPA to 30DPA and SPIKE Z32 to Z39	1	0	9	2268

Appendix B: Bootstrap probabilities for hierarchical clustering

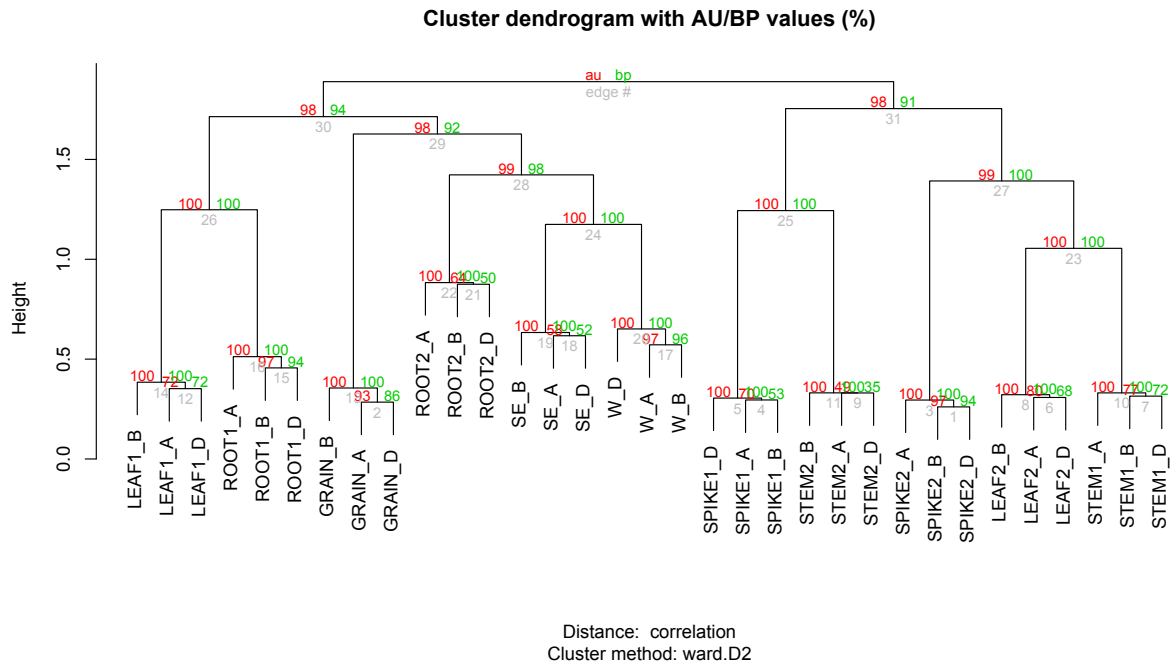


Figure 18: Hierarchical clustering of t-values based on Ward's method and correlation distances with bootstrap probabilities.

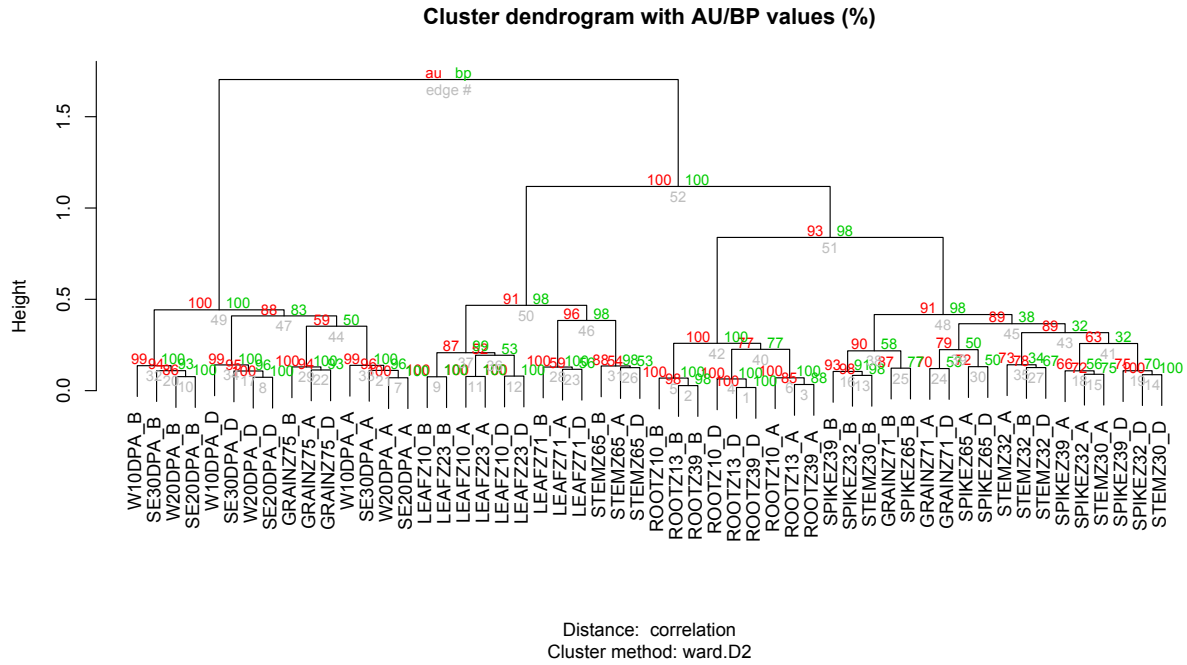


Figure 19: Hierarchical clustering of log2(fpkm+1) based on Ward's method and correlation distances with bootstrap probabilities.

Appendix C: Tables for comparing expression levels to expression regulation

Table 16: W 10DPA to 20DPA: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	1611	0	0	0	12	12	12
zero.zero.zero	6538	872	765	743	455	377	381
zero.zero.change	122	12	24	28	8	22	16
zero.change.zero	122	26	10	26	19	8	16
zero.change.change	12	5	7	1	1	2	0
change.zero.zero	136	35	40	17	34	23	14
change.zero.change	26	5	9	4	4	6	3
change.change.zero	16	5	2	4	2	0	1
change.change.change	22	7	8	6	7	6	3
total	8605	967	865	829	542	456	446

Table 17: SE 20DPA to 30DPA: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	1847	0	0	0	0	0	0
zero.zero.zero	6646	717	692	697	707	610	633
zero.zero.change	39	8	12	15	8	9	8
zero.change.zero	34	11	8	10	21	10	12
zero.change.change	4	2	1	1	1	1	2
change.zero.zero	27	8	7	3	5	6	3
change.zero.change	3	1	1	2	0	0	0
change.change.zero	2	2	2	0	2	1	0
change.change.change	3	0	0	1	1	0	1
total	8605	749	723	729	745	637	659

Table 18: GRAIN Z71 to Z75: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	445	5	2	3	1	0	0
zero.zero.zero	1730	664	646	652	506	439	454
zero.zero.change	690	294	353	348	213	272	276
zero.change.zero	756	376	308	363	324	203	312
zero.change.change	685	381	348	360	311	299	335
change.zero.zero	767	433	395	335	333	309	227
change.zero.change	660	396	379	356	342	300	317
change.change.zero	644	376	327	335	326	265	269
change.change.change	2228	1265	1233	1170	1094	1027	1007
total	8605	4190	3991	3922	3450	3114	3197

Table 19: LEAF Z10 to Z23: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	400	1	0	0	2	2	2
zero.zero.zero	4167	968	864	919	1632	1455	1479
zero.zero.change	664	135	234	249	265	353	334
zero.change.zero	678	233	148	234	363	253	339
zero.change.change	357	116	114	157	197	173	202
change.zero.zero	625	218	219	143	337	309	244
change.zero.change	380	132	145	142	218	211	198
change.change.zero	338	146	89	94	187	163	156
change.change.change	996	369	335	378	615	527	547
total	8605	2318	2148	2316	3816	3446	3501

Table 20: LEAF Z23 to Z71: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	399	0	2	1	4	3	6
zero.zero.zero	1657	608	524	533	706	658	671
zero.zero.change	646	220	258	291	262	358	369
zero.change.zero	660	326	217	305	391	260	390
zero.change.change	593	302	262	287	365	346	345
change.zero.zero	636	305	266	215	381	352	254
change.zero.change	641	313	329	280	379	389	350
change.change.zero	613	314	277	278	370	352	351
change.change.change	2760	1428	1311	1311	1610	1534	1473
total	8605	3816	3446	3501	4468	4252	4209

Table 21: ROOT Z10 to Z13: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	464	20	19	26	0	0	0
zero.zero.zero	8141	4901	4550	4549	204	152	144
zero.zero.change	0	0	0	0	0	0	0
zero.change.zero	0	0	0	0	0	0	0
zero.change.change	0	0	0	0	0	0	0
change.zero.zero	0	0	0	0	0	0	0
change.zero.change	0	0	0	0	0	0	0
change.change.zero	0	0	0	0	0	0	0
change.change.change	0	0	0	0	0	0	0
total	8605	4921	4569	4575	204	152	144

Table 22: ROOT Z13 to Z39: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	416	0	0	0	0	0	0
zero.zero.zero	8189	204	152	144	247	60	71
zero.zero.change	0	0	0	0	0	0	0
zero.change.zero	0	0	0	0	0	0	0
zero.change.change	0	0	0	0	0	0	0
change.zero.zero	0	0	0	0	0	0	0
change.zero.change	0	0	0	0	0	0	0
change.change.zero	0	0	0	0	0	0	0
change.change.change	0	0	0	0	0	0	0
total	8605	204	152	144	247	60	71

Table 23: SPIKE Z32 to Z39: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	432	0	0	0	0	2	3
zero.zero.zero	4693	1284	1090	1153	1325	1129	1187
zero.zero.change	553	160	160	183	174	196	199
zero.change.zero	596	214	120	206	236	151	210
zero.change.change	327	112	91	100	130	102	130
change.zero.zero	606	221	169	138	241	207	158
change.zero.change	328	113	98	110	124	118	103
change.change.zero	314	116	85	84	149	98	95
change.change.change	756	245	179	217	277	229	249
total	8605	2465	1992	2191	2656	2232	2334

Table 24: SPIKE Z39 to Z65: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	305	0	0	0	2	4	2
zero.zero.zero	3744	990	812	901	1538	1427	1431
zero.zero.change	626	161	219	208	260	294	292
zero.change.zero	657	260	149	220	355	252	309
zero.change.change	459	169	135	162	231	210	224
change.zero.zero	654	263	237	137	370	318	238
change.zero.change	448	177	169	165	238	231	226
change.change.zero	396	163	127	125	209	169	180
change.change.change	1316	473	384	416	620	527	536
total	8605	2656	2232	2334	3823	3432	3438

Table 25: STEM Z30 to Z32: The total number of triplets in each regulation category and the number these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	346	1	0	1	3	1	3
zero.zero.zero	2996	923	748	843	1184	1120	1153
zero.zero.change	659	212	263	271	272	341	365
zero.change.zero	762	320	216	289	421	271	395
zero.change.change	516	212	198	232	284	287	267
change.zero.zero	736	317	276	189	405	382	275
change.zero.change	537	223	209	212	295	277	283
change.change.zero	467	234	170	179	261	231	231
change.change.change	1586	702	586	620	897	824	828
total	8605	3144	2666	2836	4022	3734	3800

Table 26: STEM Z32 to Z65: The total number of triplets in each regulation category and the number and percentage of these that are differentially expressed between pairs of subgenomes in the first and second time point.

	tot. triplets	AB time1	AD time1	BD time1	AB time2	AD time2	BD time2
not expressed	390	1	0	1	0	1	2
zero.zero.zero	2146	820	777	811	917	873	870
zero.zero.change	706	281	350	346	295	403	371
zero.change.zero	675	343	251	330	387	276	385
zero.change.change	551	295	268	300	312	311	314
change.zero.zero	729	402	360	283	414	412	311
change.zero.change	535	282	278	270	277	304	308
change.change.zero	569	324	249	264	333	328	321
change.change.change	2304	1274	1201	1195	1325	1257	1259
total	8605	4022	3734	3800	4260	4165	4141

Appendix D: Tables for GO enrichment analysis

GO enrichment based on regulation

Table 27: GO terms that are enriched in the A subgenome for GO enrichment based on regulation using molecular function GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W	GO:0046524	sucrose-phosphate synthase activity	2	1	0.01	0.0066
2	W	GO:0003844	1,4-alpha-glucan branching enzyme activi...	5	1	0.02	0.0163
3	W	GO:0005215	transporter activity	990	8	3.25	0.0175
4	W	GO:0004416	hydroxyacylglutathione hydrolase activit...	6	1	0.02	0.0196
5	W	GO:0003876	AMP deaminase activity	6	1	0.02	0.0196
6	SE	GO:0016811	hydrolase activity, acting on carbon-nit...	25	1	0.02	0.018
7	GRAIN	GO:0031369	translation initiation factor binding	12	3	0.41	0.0068
8	GRAIN	GO:0004523	RNA-DNA hybrid ribonuclease activity	5	2	0.17	0.0108
9	GRAIN	GO:0008270	zinc ion binding	1133	53	38.51	0.0112
10	LEAF1	GO:0004674	protein serine/threonine kinase activity	1132	45	25.43	0.00018
11	LEAF2	GO:0009881	photoreceptor activity	5	2	0.22	0.018
12	LEAF2	GO:0003824	catalytic activity	9156	408	401.46	0.019
13	LEAF2	GO:0008138	protein tyrosine/serine/threonine phosph...	37	5	1.62	0.022
14	LEAF2	GO:0004725	protein tyrosine phosphatase activity	38	5	1.67	0.024
15	SPIKE1	GO:0004185	serine-type carboxypeptidase activity	54	6	1.54	0.0042
16	SPIKE1	GO:0003978	UDP-glucose 4-epimerase activity	15	3	0.43	0.0081
17	SPIKE1	GO:0003872	6-phosphofructokinase activity	30	4	0.85	0.0098
18	SPIKE1	GO:0004719	protein-L-isoaspartate (D-aspartate) O-m...	6	2	0.17	0.0112
19	SPIKE1	GO:0016157	sucrose synthase activity	18	3	0.51	0.0136
20	SPIKE1	GO:0030060	L-malate dehydrogenase activity	18	3	0.51	0.0136
21	SPIKE1	GO:0042578	phosphoric ester hydrolase activity	268	10	7.62	0.0208
22	SPIKE1	GO:0015079	potassium ion transmembrane transporter ...	81	7	2.30	0.0210
23	SPIKE1	GO:0004806	triglyceride lipase activity	39	4	1.11	0.0242
24	SPIKE1	GO:0022891	substrate-specific transmembrane transpo...	568	28	16.15	0.0244
25	SPIKE2	GO:0010333	terpene synthase activity	12	3	0.37	0.0050
26	SPIKE2	GO:0016149	translation release factor activity, cod...	14	3	0.43	0.0080
27	SPIKE2	GO:0003824	catalytic activity	9156	286	278.95	0.0096
28	SPIKE2	GO:0004482	mRNA (guanine-N7-)-methyltransferase act...	6	2	0.18	0.0128
29	SPIKE2	GO:0004089	carbonate dehydratase activity	7	2	0.21	0.0176
30	SPIKE2	GO:0050660	flavin adenine dinucleotide binding	153	10	4.66	0.0189
31	STEM1	GO:0003824	catalytic activity	9156	323	305.75	0.0010
32	STEM1	GO:0005215	transporter activity	990	54	33.06	0.0018
33	STEM1	GO:0030170	pyridoxal phosphate binding	140	12	4.68	0.0026
34	STEM1	GO:0016491	oxidoreductase activity	1401	69	46.78	0.0041
35	STEM1	GO:0022891	substrate-specific transmembrane transpo...	568	32	18.97	0.0086
36	STEM1	GO:0004351	glutamate decarboxylase activity	15	3	0.50	0.0125
37	STEM1	GO:0015095	magnesium ion transmembrane transporter ...	43	5	1.44	0.0138
38	STEM1	GO:0050662	coenzyme binding	423	29	14.13	0.0149
39	STEM1	GO:0043531	ADP binding	143	10	4.78	0.0218
40	STEM2	GO:0016972	thiol oxidase activity	6	2	0.25	0.023

Table 28: GO terms that are enriched in the B subgenome for GO enrichment based on regulation using molecular function GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W	GO:0004750	ribulose-phosphate 3-epimerase activity	3	1	0.01	0.0077
2	W	GO:0004416	hydroxyacylglutathione hydrolase activit...	6	1	0.02	0.0153
3	W	GO:0004356	glutamate-ammonia ligase activity	9	1	0.02	0.0229
4	SE	GO:0004356	glutamate-ammonia ligase activity	9	1	0.01	0.0054
5	SE	GO:0004470	malic enzyme activity	14	1	0.01	0.0083
6	SE	GO:0004806	triglyceride lipase activity	39	1	0.02	0.0231
7	GRAIN	GO:0008270	zinc ion binding	1133	59	38.92	0.00089
8	GRAIN	GO:0004560	alpha-L-fucosidase activity	5	2	0.17	0.01099
9	GRAIN	GO:0004499	N,N-dimethylaniline monooxygenase activi...	15	3	0.52	0.01348
10	GRAIN	GO:0032549	ribonucleoside binding	3355	106	115.24	0.01441
11	GRAIN	GO:0017176	phosphatidylinositol N-acetylglucosaminy...	6	2	0.21	0.01612
12	GRAIN	GO:0004707	MAP kinase activity	18	3	0.62	0.02240
21	LEAF1	GO:0005509	calcium ion binding	303	18	7.51	0.00059
31	LEAF1	GO:0043531	ADP binding	143	11	3.55	0.00088
41	LEAF1	GO:0004672	protein kinase activity	1486	69	36.84	0.00091
51	LEAF1	GO:0005524	ATP binding	2923	93	72.46	0.00533
61	LEAF1	GO:0008883	glutamyl-tRNA reductase activity	9	2	0.22	0.01967
18	LEAF2	GO:0016787	hydrolase activity	3077	130	126.65	0.018
19	LEAF2	GO:0004360	glutamine-fructose-6-phosphate transamin...	6	2	0.25	0.023
20	SPIKE1	GO:0022891	substrate-specific transmembrane transpo...	568	29	15.51	0.00014
21	SPIKE1	GO:0016787	hydrolase activity	3077	95	84.00	0.00411
22	SPIKE1	GO:0050897	cobalt ion binding	5	2	0.14	0.00704
23	SPIKE1	GO:0008017	microtubule binding	153	10	4.18	0.00935
24	SPIKE1	GO:0003989	acetyl-CoA carboxylase activity	6	2	0.16	0.01037
25	SPIKE1	GO:0008686	3,4-dihydroxy-2-butanone-4-phosphate syn...	6	2	0.16	0.01037
26	SPIKE1	GO:0003935	GTP cyclohydrolase II activity	6	2	0.16	0.01037
27	SPIKE1	GO:0045430	chalcone isomerase activity	6	2	0.16	0.01037
28	SPIKE1	GO:0008378	galactosyltransferase activity	36	4	0.98	0.01619
29	SPIKE1	GO:0004075	biotin carboxylase activity	8	2	0.22	0.01867
30	SPIKE1	GO:0004571	mannosyl-oligosaccharide 1,2-alpha-manno...	9	2	0.25	0.02358
31	SPIKE2	GO:0010181	FMN binding	43	6	1.37	0.0023
32	SPIKE2	GO:0004733	pyridoxamine-phosphate oxidase activity	4	2	0.13	0.0058
33	STEM1	GO:0005215	transporter activity	990	53	34.66	0.00037
34	STEM1	GO:0003707	steroid hormone receptor activity	3	2	0.11	0.00358
35	STEM1	GO:0003824	catalytic activity	9156	367	320.51	0.00736
36	STEM1	GO:0017057	6-phosphogluconolactonase activity	5	2	0.18	0.01140
37	STEM1	GO:0046422	violaxanthin de-epoxidase activity	5	2	0.18	0.01140
38	STEM1	GO:0010181	FMN binding	43	5	1.51	0.01661
39	STEM1	GO:0004140	dephospho-CoA kinase activity	6	2	0.21	0.01671
40	STEM2	GO:0005215	transporter activity	990	45	38.68	0.0037
41	STEM2	GO:0004326	tetrahydrofolylpolyglutamate synthase ac...	9	3	0.35	0.0042
42	STEM2	GO:0008171	O-methyltransferase activity	42	6	1.64	0.0086
43	STEM2	GO:0043682	copper-transporting ATPase activity	12	3	0.47	0.0100
44	STEM2	GO:0004806	triglyceride lipase activity	39	5	1.52	0.0172
45	STEM2	GO:0003942	N-acetyl-gamma-glutamyl-phosphate reduct...	6	2	0.23	0.0206
46	STEM2	GO:0016787	hydrolase activity	3077	124	120.21	0.0238

Table 29: GO terms that are enriched in the D subgenome for GO enrichment based on regulation using molecular function GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W	GO:0008131	primary amine oxidase activity	3	1	0.01	0.013
2	W	GO:0003949	1-(5-phosphoribosyl)-5-[(5-phosphoribosy...	3	1	0.01	0.013
3	W	GO:0003904	deoxyribodipyrimidine photo-lyase activi...	3	1	0.01	0.013
4	W	GO:0008915	lipid-A-disaccharide synthase activity	3	1	0.01	0.013
5	SE	GO:0004042	acetyl-CoA:L-glutamate N-acetyltransfera...	5	1	0.00	0.0036
6	SE	GO:0003849	3-deoxy-7-phosphoheptulonate synthase ac...	6	1	0.00	0.0043
7	GRAIN	GO:0005515	protein binding	4435	181	156.05	0.0067
8	GRAIN	GO:0004185	serine-type carboxypeptidase activity	54	6	1.90	0.0114
9	GRAIN	GO:0003676	nucleic acid binding	2606	104	91.69	0.0115
10	GRAIN	GO:0000166	nucleotide binding	4145	139	145.84	0.0226
11	GRAIN	GO:0004089	carbonate dehydratase activity	7	2	0.25	0.0231
12	GRAIN	GO:0031625	ubiquitin protein ligase binding	18	3	0.63	0.0239
13	LEAF1	GO:0005524	ATP binding	2923	103	74.91	0.00031
14	LEAF1	GO:0008987	quinolinate synthetase A activity	4	2	0.10	0.00380
15	LEAF1	GO:0043531	ADP binding	143	9	3.66	0.01146
16	LEAF1	GO:0003968	RNA-directed RNA polymerase activity	9	2	0.23	0.02094
17	LEAF2	GO:0030247	polysaccharide binding	31	6	1.25	0.0013
18	LEAF2	GO:0005215	transporter activity	990	54	39.80	0.0043
19	LEAF2	GO:0015095	magnesium ion transmembrane transporter ...	43	6	1.73	0.0071
20	LEAF2	GO:0017057	6-phosphogluconolactonase activity	5	2	0.20	0.0149
21	LEAF2	GO:0008270	zinc ion binding	1133	59	45.55	0.0243
22	LEAF2	GO:0004674	protein serine/threonine kinase activity	1132	60	45.51	0.0247
23	SPIKE1	GO:0042578	phosphoric ester hydrolase activity	268	12	6.50	0.0014
24	SPIKE1	GO:0004367	glycerol-3-phosphate dehydrogenase [NAD+...	15	3	0.36	0.0052
25	SPIKE1	GO:0003855	3-dehydroquinate dehydratase activity	5	2	0.12	0.0056
26	SPIKE1	GO:0050661	NADP binding	70	6	1.70	0.0070
27	SPIKE1	GO:0004764	shikimate 3-dehydrogenase (NADP+) activi...	6	2	0.15	0.0083
28	SPIKE1	GO:0005215	transporter activity	990	37	24.01	0.0092
29	SPIKE1	GO:0003924	GTPase activity	181	10	4.39	0.0130
30	SPIKE1	GO:0005525	GTP binding	271	13	6.57	0.0153
31	SPIKE1	GO:0004563	beta-N-acetylhexosaminidase activity	9	2	0.22	0.0189
32	SPIKE1	GO:0003913	DNA photolyase activity	9	2	0.22	0.0189
33	SPIKE1	GO:0046872	metal ion binding	2333	60	56.58	0.0227
34	SPIKE1	GO:0003838	sterol 24-C-methyltransferase activity	1	1	0.02	0.0243
35	SPIKE2	GO:0033897	ribonuclease T2 activity	15	4	0.45	0.00086
36	SPIKE2	GO:0003824	catalytic activity	9156	300	276.21	0.00596
37	SPIKE2	GO:0005506	iron ion binding	262	16	7.90	0.00612
38	SPIKE2	GO:0003755	peptidyl-prolyl cis-trans isomerase acti...	40	5	1.21	0.00675
39	SPIKE2	GO:0004516	nicotinate phosphoribosyltransferase act...	6	2	0.18	0.01257
40	SPIKE2	GO:0003942	N-acetyl-gamma-glutamyl-phosphate reduct...	6	2	0.18	0.01257
41	SPIKE2	GO:0004619	phosphoglycerate mutase activity	6	2	0.18	0.01257
42	SPIKE2	GO:0004089	carbonate dehydratase activity	7	2	0.21	0.01725
43	SPIKE2	GO:0004616	phosphogluconate dehydrogenase (decarbox...	7	2	0.21	0.01725
44	STEM1	GO:0003824	catalytic activity	9156	353	307.39	0.00062
45	STEM1	GO:0016787	hydrolase activity	3077	119	103.30	0.00088
46	STEM1	GO:0004089	carbonate dehydratase activity	7	3	0.24	0.00119
47	STEM1	GO:0016887	ATPase activity	564	18	18.93	0.00340
48	STEM1	GO:0017057	6-phosphogluconolactonase activity	5	2	0.17	0.01052
49	STEM1	GO:0008531	riboflavin kinase activity	6	2	0.20	0.01543
50	STEM1	GO:0004416	hydroxyacylglutathione hydrolase activit...	6	2	0.20	0.01543
51	STEM2	GO:0003995	acyl-CoA dehydrogenase activity	9	3	0.36	0.0043
52	STEM2	GO:0008270	zinc ion binding	1133	60	44.87	0.0129
53	STEM2	GO:0008889	glycerophosphodiester phosphodiesterase ...	24	4	0.95	0.0138
54	STEM2	GO:0004556	alpha-amylase activity	6	2	0.24	0.0211
55	STEM2	GO:0008479	queuine tRNA-ribosyltransferase activity	6	2	0.24	0.0211

Table 30: GO terms that are enriched in the A subgenome for GO enrichment based on regulation using biological process GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W	GO:0006750	glutathione biosynthetic process	18	2	0.06	0.0018
2	W	GO:0051716	cellular response to stimulus	594	3	2.12	0.0107
3	W	GO:0008299	isoprenoid biosynthetic process	61	2	0.22	0.0200
4	W	GO:0006471	protein ADP-ribosylation	6	1	0.02	0.0213
5	SE	GO:0000917	barrier septum assembly	5	1	0.00	0.0035
6	SE	GO:0006672	ceramide metabolic process	8	1	0.01	0.0056
7	SE	GO:0001510	RNA methylation	24	1	0.02	0.0166
8	GRAIN	GO:0030154	cell differentiation	6	2	0.19	0.014
9	GRAIN	GO:0006413	translational initiation	48	5	1.51	0.017
10	GRAIN	GO:0055114	oxidation-reduction process	1262	51	39.62	0.020
11	LEAF1	GO:0055085	transmembrane transport	821	34	18.69	0.0041
12	LEAF2	GO:0017006	protein-tetrapyrrole linkage	5	2	0.21	0.017
13	LEAF2	GO:0009585	red, far-red light phototransduction	5	2	0.21	0.017
14	LEAF2	GO:0006904	vesicle docking involved in exocytosis	24	4	1.03	0.018
15	LEAF2	GO:0006810	transport	1851	95	79.59	0.022
16	LEAF2	GO:0009228	thiamine biosynthetic process	6	2	0.26	0.025
17	SPIKE1	GO:0006096	glycolytic process	112	10	3.38	0.0020
18	SPIKE1	GO:0055114	oxidation-reduction process	1262	54	38.08	0.0065
19	SPIKE1	GO:0006886	intracellular protein transport	326	16	9.84	0.0178
20	SPIKE1	GO:0044262	cellular carbohydrate metabolic process	235	12	7.09	0.0220
21	SPIKE1	GO:0005985	sucrose metabolic process	27	4	0.81	0.0241
22	SPIKE1	GO:0006012	galactose metabolic process	21	3	0.63	0.0242
23	SPIKE2	GO:0055114	oxidation-reduction process	1262	57	38.85	0.0020
24	SPIKE2	GO:0006807	nitrogen compound metabolic process	2464	82	75.86	0.0028
25	SPIKE2	GO:0015976	carbon utilization	6	2	0.18	0.0131
26	SPIKE2	GO:0006364	rRNA processing	49	5	1.51	0.0169
27	SPIKE2	GO:0045454	cell redox homeostasis	111	8	3.42	0.0211
28	SPIKE2	GO:0055085	transmembrane transport	821	35	25.28	0.0233
29	SPIKE2	GO:0009408	response to heat	36	4	1.11	0.0239
30	STEM1	GO:0009765	photosynthesis, light harvesting	21	4	0.75	0.0059
31	STEM1	GO:0015693	magnesium ion transport	43	5	1.54	0.0180
32	STEM1	GO:0006952	defense response	154	11	5.51	0.0223
33	STEM1	GO:0006810	transport	1851	77	66.19	0.0223
34	STEM1	GO:0006536	glutamate metabolic process	18	3	0.64	0.0248
35	STEM2	GO:0010380	regulation of chlorophyll biosynthetic p...	6	2	0.23	0.020

Table 31: GO terms that are enriched in the B subgenome for GO enrichment based on regulation using biological process GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W	GO:0006542	glutamine biosynthetic process	9	1	0.02	0.023
2	SE	GO:0006542	glutamine biosynthetic process	9	1	0.01	0.0070
3	SE	GO:0006777	Mo-molybdopterin cofactor biosynthetic p...	11	1	0.01	0.0086
4	SE	GO:0006108	malate metabolic process	32	1	0.03	0.0248
5	GRAIN	GO:0006351	transcription, DNA-templated	824	42	28.17	0.0019
6	GRAIN	GO:0030154	cell differentiation	6	2	0.21	0.0160
7	GRAIN	GO:0034968	histone lysine methylation	31	4	1.06	0.0204
8	LEAF1	GO:0006952	defense response	154	11	4.04	0.0025
9	LEAF2	GO:0045010	actin nucleation	25	4	1.02	0.014
10	LEAF2	GO:0051225	spindle assembly	6	2	0.24	0.022
11	SPIKE1	GO:0006633	fatty acid biosynthetic process	106	9	3.14	0.0066
12	SPIKE1	GO:0009813	flavonoid biosynthetic process	6	2	0.18	0.0122
13	SPIKE1	GO:0055085	transmembrane transport	821	34	24.35	0.0201
14	SPIKE2	GO:0008152	metabolic process	8774	302	283.90	0.00071
15	SPIKE2	GO:0008615	pyridoxine biosynthetic process	4	2	0.13	0.00600
16	SPIKE2	GO:0006099	tricarboxylic acid cycle	39	5	1.26	0.00804
17	SPIKE2	GO:0016559	peroxisome fission	6	2	0.19	0.01437
18	SPIKE2	GO:0055114	oxidation-reduction process	1262	57	40.83	0.02440
19	STEM1	GO:0006909	phagocytosis	12	3	0.46	0.0093
20	STEM1	GO:0015937	coenzyme A biosynthetic process	12	3	0.46	0.0093
21	STEM1	GO:0051225	spindle assembly	6	2	0.23	0.0195
22	STEM1	GO:0005975	carbohydrate metabolic process	970	45	36.88	0.0233
23	STEM2	GO:0006662	glycerol ether metabolic process	37	6	1.47	0.0032
24	STEM2	GO:0006825	copper ion transport	12	3	0.48	0.0106
25	STEM2	GO:0005985	sucrose metabolic process	27	4	1.08	0.0211

Table 32: GO terms that are enriched in the D subgenome for GO enrichment based on regulation using biological process GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W	GO:0006308	DNA catabolic process	18	2	0.09	0.0034
2	W	GO:0006108	malate metabolic process	32	2	0.16	0.0106
3	SE	GO:0006526	arginine biosynthetic process	20	1	0.02	0.016
4	GRAIN	GO:0010508	positive regulation of autophagy	6	2	0.20	0.015
5	LEAF1	GO:0009435	NAD biosynthetic process	19	4	0.47	0.0010
6	LEAF1	GO:0006952	defense response	154	10	3.77	0.0092
7	LEAF1	GO:0009408	response to heat	36	4	0.88	0.0112
8	LEAF1	GO:0009607	response to biotic stimulus	18	3	0.44	0.0191
9	LEAF2	GO:0006810	transport	1851	93	78.13	0.0043
10	LEAF2	GO:0015693	magnesium ion transport	43	6	1.82	0.0089
11	LEAF2	GO:0006744	ubiquinone biosynthetic process	14	3	0.59	0.0192
12	SPIKE1	GO:0007264	small GTPase mediated signal transductio...	127	10	3.43	0.0023
12	SPIKE1	GO:0019632	shikimate metabolic process	4	2	0.11	0.0042
14	SPIKE1	GO:0055114	oxidation-reduction process	1262	49	34.12	0.0054
15	SPIKE1	GO:0006807	nitrogen compound metabolic process	2464	46	66.62	0.0201
16	SPIKE2	GO:0006098	pentose-phosphate shunt	18	4	0.62	0.0028
17	SPIKE2	GO:0019358	nicotinate nucleotide salvage	6	2	0.21	0.0160
18	SPIKE2	GO:0016559	peroxisome fission	6	2	0.21	0.0160
19	STEM1	GO:0008152	metabolic process	8774	297	302.26	0.0015
20	STEM1	GO:0009058	biosynthetic process	2566	76	88.40	0.0079
21	STEM1	GO:0006470	protein dephosphorylation	110	9	3.79	0.0136
22	STEM1	GO:0015976	carbon utilization	6	2	0.21	0.0162
23	STEM1	GO:0006811	ion transport	507	16	17.47	0.0167
24	STEM2	GO:0010380	regulation of chlorophyll biosynthetic p...	6	2	0.24	0.022
25	STEM2	GO:0008616	queuosine biosynthetic process	6	2	0.24	0.022
26	STEM2	GO:0006071	glycerol metabolic process	27	4	1.09	0.022

GO enrichment based on expression

Table 33: GO terms that are enriched in the A subgenome for GO enrichment based on expression using molecular function GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W 10 DPA	GO:0003980	UDP-glucose:glycoprotein glucosyltransfe...	6	2	0.14	0.0074
2	W 10 DPA	GO:0003735	structural constituent of ribosome	204	11	4.67	0.0075
3	W 10 DPA	GO:0016851	magnesium chelatase activity	8	2	0.18	0.0133
4	W 10 DPA	GO:0004356	glutamate-ammonia ligase activity	9	2	0.21	0.0169
5	SE 20 DPA	GO:0004362	glutathione-disulfide reductase activity	3	2	0.06	0.0011
6	SE 20 DPA	GO:0004089	carbonate dehydratase activity	7	2	0.13	0.0072
7	SE 20 DPA	GO:0051082	unfolded protein binding	92	6	1.76	0.0085
8	SE 20 DPA	GO:0016851	magnesium chelatase activity	8	2	0.15	0.0095
9	SE 20 DPA	GO:0003735	structural constituent of ribosome	204	9	3.91	0.0171
10	SE 20 DPA	GO:0031369	translation initiation factor binding	12	2	0.23	0.0213
11	W 20 DPA	GO:0051082	unfolded protein binding	92	5	1.18	0.0065
12	W 20 DPA	GO:0003735	structural constituent of ribosome	204	7	2.61	0.0161
13	SE 30 DPA	GO:0008565	protein transporter activity	47	5	0.79	0.0061
14	SE 30 DPA	GO:0003735	structural constituent of ribosome	204	9	3.41	0.0075
15	SE 30 DPA	GO:0005509	calcium ion binding	303	11	5.07	0.0133
16	GRAIN Z71	GO:0004362	glutathione-disulfide reductase activity	3	2	0.26	0.020
17	GRAIN Z75	GO:0004252	serine-type endopeptidase activity	119	17	8.29	0.0036
18	GRAIN Z75	GO:0003735	structural constituent of ribosome	204	23	14.21	0.0152
19	GRAIN Z75	GO:0043169	cation binding	2356	182	164.10	0.0180
20	GRAIN Z75	GO:0003743	translation initiation factor activity	68	10	4.74	0.0187
21	GRAIN Z75	GO:0003995	acyl-CoA dehydrogenase activity	9	3	0.63	0.0206
22	LEAF Z10	GO:0030170	pyridoxal phosphate binding	140	15	7.33	0.0066
23	LEAF Z10	GO:0004362	glutathione-disulfide reductase activity	3	2	0.16	0.0079
24	LEAF Z10	GO:0005525	GTP binding	271	23	14.18	0.0156
25	LEAF Z10	GO:0009881	photoreceptor activity	5	2	0.26	0.0246
26	LEAF Z23	GO:0004362	glutathione-disulfide reductase activity	3	2	0.22	0.016
27	LEAF Z23	GO:0004644	phosphoribosylglycinamide formyltransfer...	3	2	0.22	0.016
28	LEAF Z23	GO:0016851	magnesium chelatase activity	8	3	0.59	0.017
29	LEAF Z23	GO:0005525	GTP binding	271	30	20.01	0.017
30	LEAF Z23	GO:0004066	asparagine synthase (glutamine-hydrolyzi...	9	3	0.66	0.024
31	LEAF Z71	GO:0004362	glutathione-disulfide reductase activity	3	2	0.25	0.020
32	LEAF Z71	GO:0004644	phosphoribosylglycinamide formyltransfer...	3	2	0.25	0.020
51	SPIKE Z32	GO:0004362	glutathione-disulfide reductase activity	3	2	0.15	0.0076
33	SPIKE Z39	GO:0003746	translation elongation factor activity	21	5	1.19	0.0056
34	SPIKE Z39	GO:0003872	6-phosphofructokinase activity	30	6	1.71	0.0061
35	SPIKE Z39	GO:0004362	glutathione-disulfide reductase activity	3	2	0.17	0.0093
36	SPIKE Z65	GO:0003872	6-phosphofructokinase activity	30	7	2.29	0.0063
37	SPIKE Z65	GO:0004362	glutathione-disulfide reductase activity	3	2	0.23	0.0166
38	SPIKE Z65	GO:0008565	protein transporter activity	47	9	3.59	0.0168
39	SPIKE Z65	GO:0008270	zinc ion binding	1133	105	86.43	0.0201
40	STEM Z30	GO:0003746	translation elongation factor activity	21	5	1.35	0.0093
41	STEM Z30	GO:0016851	magnesium chelatase activity	8	3	0.51	0.0116
42	STEM Z32	GO:0008237	metallopeptidase activity	79	8	6.40	0.015
43	STEM Z32	GO:0004362	glutathione-disulfide reductase activity	3	2	0.24	0.019
44	STEM Z32	GO:0004644	phosphoribosylglycinamide formyltransfer...	3	2	0.24	0.019
45	STEM Z32	GO:0016851	magnesium chelatase activity	8	3	0.65	0.022
46	STEM Z65	GO:0008565	protein transporter activity	47	10	3.84	0.0081
47	STEM Z65	GO:0005234	extracellular-glutamate-gated ion channe...	24	6	1.96	0.0109
48	STEM Z65	GO:0004970	ionotropic glutamate receptor activity	24	6	1.96	0.0109
49	STEM Z65	GO:0004965	G-protein coupled GABA receptor activity	7	3	0.57	0.0148
50	STEM Z65	GO:0008237	metallopeptidase activity	79	8	6.45	0.0148
51	STEM Z65	GO:0004362	glutathione-disulfide reductase activity	3	2	0.24	0.0189
52	STEM Z65	GO:0004644	phosphoribosylglycinamide formyltransfer...	3	2	0.24	0.0189
53	STEM Z65	GO:0016851	magnesium chelatase activity	8	3	0.65	0.0222

Table 34: GO terms that are enriched in the B subgenome for GO enrichment based on expression using molecular function GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W 10DPA	GO:0000287	magnesium ion binding	137	12	3.41	0.00016
2	W 10DPA	GO:0031072	heat shock protein binding	26	4	0.65	0.00367
3	W 10DPA	GO:0004634	phosphopyruvate hydratase activity	6	2	0.15	0.00869
4	W 10DPA	GO:0004609	phosphatidylserine decarboxylase activit...	6	2	0.15	0.00869
5	W 10DPA	GO:0003984	acetolactate synthase activity	9	2	0.22	0.01985
6	W 10DPA	GO:0004106	chorismate mutase activity	9	2	0.22	0.01985
7	W 10DPA	GO:0016491	oxidoreductase activity	1401	41	34.90	0.02278
8	SE 20DPA	GO:0004619	phosphoglycerate mutase activity	6	2	0.11	0.0046
9	SE 20DPA	GO:0051287	NAD binding	68	5	1.23	0.0076
10	SE 20DPA	GO:0008270	zinc ion binding	1133	31	20.44	0.0135
11	SE 20DPA	GO:0003774	motor activity	135	6	2.44	0.0149
12	SE 20DPA	GO:0043682	copper-transporting ATPase activity	12	2	0.22	0.0190
13	SE 20DPA	GO:0004652	polynucleotide adenylyltransferase activ...	12	2	0.22	0.0190
14	W 20DPA	GO:0008935	1,4-dihydroxy-2-naphthoyl-CoA synthase a...	1	1	0.01	0.014
15	W 20DPA	GO:0016491	oxidoreductase activity	1401	30	19.33	0.018
16	W 20DPA	GO:0010181	FMN binding	43	3	0.59	0.021
17	W 20DPA	GO:0016157	sucrose synthase activity	18	2	0.25	0.025
18	SE 30DPA	GO:0004749	ribose phosphate diphosphokinase activit...	5	2	0.09	0.0031
19	SE 30DPA	GO:0016308	1-phosphatidylinositol-4-phosphate 5-kin...	9	2	0.16	0.0107
20	SE 30DPA	GO:0000287	magnesium ion binding	137	7	2.47	0.0122
21	GRAIN Z71	GO:0046961	proton-transporting ATPase activity, rot...	21	5	1.65	0.021
22	GRAIN Z75	GO:0030955	potassium ion binding	27	6	1.81	0.0079
23	GRAIN Z75	GO:0004743	pyruvate kinase activity	27	6	1.81	0.0079
24	GRAIN Z75	GO:0005488	binding	11650	810	782.23	0.0156
25	GRAIN Z75	GO:0048038	quinone binding	9	3	0.60	0.0187
26	GRAIN Z75	GO:0000287	magnesium ion binding	137	16	9.20	0.0214
27	GRAIN Z75	GO:0010181	FMN binding	43	7	2.89	0.0233
28	LEAF Z10	GO:0046961	proton-transporting ATPase activity, rot...	21	5	1.09	0.0038
29	LEAF Z10	GO:0050661	NADP binding	70	9	3.64	0.0100
30	LEAF Z10	GO:0003993	acid phosphatase activity	39	6	2.03	0.0146
31	LEAF Z10	GO:0004733	pyridoxamine-phosphate oxidase activity	4	2	0.21	0.0151
32	LEAF Z10	GO:0003824	catalytic activity	9156	506	475.85	0.0229
33	LEAF Z10	GO:0004749	ribose phosphate diphosphokinase activit...	5	2	0.26	0.0243
34	LEAF Z10	GO:0050897	cobalt ion binding	5	2	0.26	0.0243
35	LEAF Z23	GO:0000287	magnesium ion binding	137	20	10.49	0.0039
36	LEAF Z23	GO:0051287	NAD binding	68	12	5.21	0.0051
37	LEAF Z23	GO:0003824	catalytic activity	9156	738	701.19	0.0145
38	LEAF Z23	GO:0004197	cysteine-type endopeptidase activity	17	5	1.30	0.0184
39	LEAF Z23	GO:0046961	proton-transporting ATPase activity, rot...	21	5	1.61	0.0189
40	LEAF Z23	GO:0016788	hydrolase activity, acting on ester bond...	713	71	54.60	0.0239
41	LEAF Z71	GO:0051287	NAD binding	68	13	5.63	0.0074
42	LEAF Z71	GO:0016779	nucleotidyltransferase activity	192	20	15.91	0.0174
43	LEAF Z71	GO:0000287	magnesium ion binding	137	19	11.35	0.0181
44	LEAF Z71	GO:0016787	hydrolase activity	3077	276	254.95	0.0214
45	SPIKE Z32	GO:0016597	amino acid binding	42	6	2.04	0.015
46	SPIKE Z39	GO:0050661	NADP binding	70	10	4.01	0.0063
47	SPIKE Z39	GO:0010181	FMN binding	43	7	2.46	0.0105
48	SPIKE Z39	GO:0048038	quinone binding	9	3	0.52	0.0121
49	SPIKE Z65	GO:0050661	NADP binding	70	14	5.25	0.0006
50	SPIKE Z65	GO:0051287	NAD binding	68	13	5.10	0.0014
51	SPIKE Z65	GO:0016779	nucleotidyltransferase activity	192	16	14.41	0.0111
52	SPIKE Z65	GO:0000287	magnesium ion binding	137	18	10.28	0.0140
53	SPIKE Z65	GO:0046961	proton-transporting ATPase activity, rot...	21	5	1.58	0.0174
54	SPIKE Z65	GO:0004367	glycerol-3-phosphate dehydrogenase [NAD+...	15	4	1.13	0.0221
55	SPIKE Z65	GO:0004345	glucose-6-phosphate dehydrogenase activi...	15	4	1.13	0.0221
56	STEM Z30	GO:0000287	magnesium ion binding	137	20	8.71	0.00041
57	STEM Z30	GO:0015171	amino acid transmembrane transporter act...	24	5	1.53	0.01586
58	STEM Z30	GO:0008883	glutamyl-tRNA reductase activity	9	3	0.57	0.01610
59	STEM Z30	GO:0031072	heat shock protein binding	26	5	1.65	0.02211
60	STEM Z30	GO:0004733	pyridoxamine-phosphate oxidase activity	4	2	0.25	0.02222
61	STEM Z32	GO:0046961	proton-transporting ATPase activity, rot...	21	6	1.64	0.0044
62	STEM Z32	GO:0010181	FMN binding	43	9	3.36	0.0053
63	STEM Z32	GO:0051287	NAD binding	68	12	5.32	0.0060
64	STEM Z65	GO:0051287	NAD binding	68	14	5.44	0.0019
65	STEM Z65	GO:0046961	proton-transporting ATPase activity, rot...	21	6	1.68	0.0049
66	STEM Z65	GO:0050661	NADP binding	70	13	5.60	0.0064
67	STEM Z65	GO:0016779	nucleotidyltransferase activity	192	17	15.37	0.0150
68	STEM Z65	GO:0010181	FMN binding	43	8	3.44	0.0193

Table 35: GO terms that are enriched in the D subgenome for GO enrichment based on expression using molecular function GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W 10DPA	GO:0004826	phenylalanine-tRNA ligase activity	7	2	0.15	0.009
2	W 10DPA	GO:0016818	hydrolase activity, acting on acid anhyd...	1197	23	25.67	0.015
3	SE 20DPA	GO:0004826	phenylalanine-tRNA ligase activity	7	2	0.12	0.0056
4	SE 20DPA	GO:0016818	hydrolase activity, acting on acid anhyd...	1197	20	20.16	0.0096
5	SE 20DPA	GO:0004519	endonuclease activity	97	5	1.63	0.0145
6	SE 20DPA	GO:0005484	SNAP receptor activity	12	2	0.20	0.0167
7	SE 20DPA	GO:0003684	damaged DNA binding	37	3	0.62	0.0241
8	W 20DPA	GO:0008878	glucose-1-phosphate adenylyltransferase ...	12	2	0.14	0.0086
9	W 20DPA	GO:0004553	hydrolase activity, hydrolyzing O-glycos...	401	15	4.77	0.0103
10	W 20DPA	GO:0004559	alpha-mannosidase activity	15	2	0.18	0.0133
11	SE 30DPA	GO:0004806	triglyceride lipase activity	39	4	0.58	0.0026
12	SE 30DPA	GO:0016818	hydrolase activity, acting on acid anhyd...	1197	14	17.88	0.0078
13	SE 30DPA	GO:0004298	threonine-type endopeptidase activity	33	3	0.49	0.0129
14	SE 30DPA	GO:0045300	acyl-[acyl-carrier-protein] desaturase a...	15	2	0.22	0.0205
15	GRAIN Z71	GO:0004713	protein tyrosine kinase activity	48	8	3.39	0.018
16	GRAIN Z71	GO:0008536	Ran GTPase binding	24	5	1.69	0.024
17	GRAIN Z75	GO:0004826	phenylalanine-tRNA ligase activity	7	3	0.43	0.0068
18	GRAIN Z75	GO:0000049	tRNA binding	15	4	0.92	0.0113
19	LEAF Z10	GO:0004124	cysteine synthase activity	11	3	0.53	0.014
20	LEAF Z10	GO:0005515	protein binding	4435	233	213.27	0.023
21	LEAF Z23	GO:0008987	quinolinate synthetase A activity	4	2	0.26	0.023
22	LEAF Z71	GO:0008270	zinc ion binding	1133	103	83.79	0.016
23	LEAF Z71	GO:0031683	G-protein beta/gamma-subunit complex bin...	9	3	0.67	0.024
24	LEAF Z71	GO:0010277	chlorophyllide a oxygenase [overall] act...	9	3	0.67	0.024
25	SPIKE Z32	GO:0031683	G-protein beta/gamma-subunit complex bin...	9	3	0.38	0.0051
26	SPIKE Z32	GO:0051536	iron-sulfur cluster binding	121	10	5.07	0.0085
27	SPIKE Z32	GO:0008987	quinolinate synthetase A activity	4	2	0.17	0.0099
28	SPIKE Z32	GO:0004871	signal transducer activity	147	9	6.16	0.0204
29	SPIKE Z32	GO:0000049	tRNA binding	15	3	0.63	0.0228
30	SPIKE Z32	GO:0005471	ATP:ADP antiporter activity	6	2	0.25	0.0235
31	SPIKE Z32	GO:0019001	guanyl nucleotide binding	277	12	11.60	0.0236
32	SPIKE Z39	GO:0003993	acid phosphatase activity	39	6	1.79	0.0082
33	SPIKE Z39	GO:0008987	quinolinate synthetase A activity	4	2	0.18	0.0119
34	SPIKE Z65	GO:0004806	triglyceride lipase activity	39	8	2.46	0.0026
35	SPIKE Z65	GO:0042802	identical protein binding	39	6	2.46	0.0106
36	SPIKE Z65	GO:0008987	quinolinate synthetase A activity	4	2	0.25	0.0219
37	STEM Z30	GO:0008987	quinolinate synthetase A activity	4	2	0.22	0.017
38	STEM Z32	GO:0016787	hydrolase activity	3077	219	211.75	0.010
39	STEM Z32	GO:0003774	motor activity	135	12	9.29	0.013
40	STEM Z65	GO:0016787	hydrolase activity	3077	240	229.21	0.015
41	STEM Z65	GO:0003984	acetolactate synthase activity	9	3	0.67	0.025

Table 36: GO terms that are enriched in the A subgenome for GO enrichment based on expression using biological process GO terms.

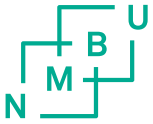
	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W 10DPA	GO:0006457	protein folding	188	11	4.54	0.006
2	W 10DPA	GO:0006950	response to stress	543	17	13.12	0.015
3	W 10DPA	GO:0006542	glutamine biosynthetic process	9	2	0.22	0.019
4	W 10DPA	GO:0015986	ATP synthesis coupled proton transport	9	2	0.22	0.019
5	W 10DPA	GO:0006412	translation	403	23	9.74	0.019
6	SE 20DPA	GO:0006749	glutathione metabolic process	24	3	0.45	0.0049
7	SE 20DPA	GO:0015976	carbon utilization	6	2	0.11	0.0049
8	SE 20DPA	GO:0006412	translation	403	18	7.49	0.0090
9	SE 20DPA	GO:0045454	cell redox homeostasis	111	6	2.06	0.0173
10	SE 20DPA	GO:0048280	vesicle fusion with Golgi apparatus	1	1	0.02	0.0186
11	W 20DPA	GO:0015986	ATP synthesis coupled proton transport	9	2	0.12	0.0065
12	SE 20DPA	GO:0006096	glycolytic process	112	5	1.55	0.0197
13	SE 20DPA	GO:0005978	glycogen biosynthetic process	17	2	0.24	0.0227
14	SE 30DPA	GO:0006886	intracellular protein transport	326	18	6.06	0.00046
15	SE 30DPA	GO:0015986	ATP synthesis coupled proton transport	9	2	0.17	0.01135
16	SE 30DPA	GO:0006606	protein import into nucleus	11	2	0.20	0.01692
17	SE 30DPA	GO:0048280	vesicle fusion with Golgi apparatus	1	1	0.02	0.01858
18	SE 30DPA	GO:0000902	cell morphogenesis	12	2	0.22	0.02005
19	GRAIN Z71	GO:0006486	protein glycosylation	75	15	6.42	0.015
20	GRAIN Z75	GO:0006413	translational initiation	48	9	3.33	0.011
21	GRAIN Z75	GO:0006412	translation	403	48	27.94	0.017
22	GRAIN Z75	GO:0009584	detection of visible light	9	3	0.62	0.020
23	GRAIN Z75	GO:0015986	ATP synthesis coupled proton transport	9	3	0.62	0.020
24	GRAIN Z75	GO:0018298	protein-chromophore linkage	9	3	0.62	0.020
25	LEAF Z10	GO:0006520	cellular amino acid metabolic process	442	34	22.78	0.0090
26	LEAF Z10	GO:0009058	biosynthetic process	2566	149	132.26	0.0097
27	LEAF Z10	GO:0017006	protein-tetrapyrrole linkage	5	2	0.26	0.0239
28	LEAF Z10	GO:0009585	red, far-red light phototransduction	5	2	0.26	0.0239
29	LEAF Z23	GO:0006561	proline biosynthetic process	8	3	0.61	0.019
30	LEAF Z71	GO:0016070	RNA metabolic process	1270	118	104.12	0.0072
31	SPIKE Z32	GO:0008612	peptidyl-lysine modification to peptidyl...	4	2	0.20	0.014
32	SPIKE Z39	GO:0006002	fructose 6-phosphate metabolic process	25	6	1.43	0.0024
33	SPIKE Z39	GO:0006414	translational elongation	27	6	1.54	0.0103
34	SPIKE Z39	GO:0008612	peptidyl-lysine modification to peptidyl...	4	2	0.23	0.0182
35	SPIKE Z39	GO:0009058	biosynthetic process	2566	175	146.81	0.0197
36	SPIKE Z65	GO:0006002	fructose 6-phosphate metabolic process	25	7	1.92	0.0022
37	STEM Z30	GO:0006520	cellular amino acid metabolic process	442	31	28.37	0.0072
38	STEM Z30	GO:0008612	peptidyl-lysine modification to peptidyl...	4	2	0.26	0.0226
39	STEM Z30	GO:0019538	protein metabolic process	3145	207	201.88	0.0246
40	STEM Z32	GO:0006857	oligopeptide transport	10	4	0.81	0.0061
41	STEM Z32	GO:0005985	sucrose metabolic process	27	6	2.19	0.0237

Table 37: GO terms that are enriched in the B subgenome for GO enrichment based on expression using biological process GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W 10DPA	GO:0006096	glycolytic process	112	8	2.71	0.0057
2	W 10DPA	GO:0046417	chorismate metabolic process	9	2	0.22	0.0187
3	W 10DPA	GO:0015991	ATP hydrolysis coupled proton transport	24	3	0.58	0.0194
4	SE 20DPA	GO:0006108	malate metabolic process	32	4	0.59	0.0027
5	SE 20DPA	GO:0015991	ATP hydrolysis coupled proton transport	24	3	0.44	0.0094
6	SE 20DPA	GO:0006825	copper ion transport	12	2	0.22	0.0197
7	SE 20DPA	GO:0006457	protein folding	188	8	3.46	0.0230
8	W 20DPA	GO:0009234	menaquinone biosynthetic process	1	1	0.01	0.014
9	W 20DPA	GO:0055114	oxidation-reduction process	1262	26	17.28	0.022
10	SE 30DPA	GO:0046488	phosphatidylinositol metabolic process	122	6	2.17	0.013
11	SE 30DPA	GO:0006777	Mo-molybdopterin cofactor biosynthetic p...	11	2	0.20	0.016
12	SE 30DPA	GO:0055114	oxidation-reduction process	1262	33	22.45	0.017
13	SE 30DPA	GO:0016572	histone phosphorylation	1	1	0.02	0.018
14	SE 30DPA	GO:0010212	response to ionizing radiation	1	1	0.02	0.018
15	SE 30DPA	GO:0090399	replicative senescence	1	1	0.02	0.018
16	SE 30DPA	GO:0006108	malate metabolic process	32	3	0.57	0.019
17	GRAIN Z71	GO:0006012	galactose metabolic process	21	5	1.69	0.023
18	GRAIN Z75	GO:0042254	ribosome biogenesis	93	13	6.50	0.0041
19	LEAF Z10	GO:0000910	cytokinesis	35	6	1.91	0.011
20	LEAF Z10	GO:0008615	pyridoxine biosynthetic process	4	2	0.22	0.017
21	LEAF Z10	GO:0000226	microtubule cytoskeleton organization	54	8	2.95	0.025
22	SPIKE Z32	GO:0042254	ribosome biogenesis	93	12	4.47	0.0019
23	SPIKE Z32	GO:0006479	protein methylation	57	8	2.74	0.0222
24	SPIKE Z65	GO:0046168	glycerol-3-phosphate catabolic process	9	4	0.69	0.0031
25	SPIKE Z65	GO:0006006	glucose metabolic process	35	8	2.67	0.0118
26	SPIKE Z65	GO:0042254	ribosome biogenesis	93	13	7.09	0.0227
27	STEM Z30	GO:0003333	amino acid transmembrane transport	30	6	1.95	0.011
28	STEM Z30	GO:0006096	glycolytic process	112	14	7.28	0.014
29	STEM Z30	GO:0008615	pyridoxine biosynthetic process	4	2	0.26	0.023
30	STEM Z32	GO:0042254	ribosome biogenesis	93	15	7.47	0.0092
31	STEM Z32	GO:0006098	pentose-phosphate shunt	18	5	1.45	0.0117
32	STEM Z65	GO:0000226	microtubule cytoskeleton organization	54	12	4.45	0.0017
33	STEM Z65	GO:0015991	ATP hydrolysis coupled proton transport	24	6	1.98	0.0114
34	STEM Z65	GO:0000910	cytokinesis	35	7	2.88	0.0219

Table 38: GO terms that are enriched in the D subgenome for GO enrichment based on expression using biological process GO terms.

	Tissue	GO.ID	Term	Annotated	Significant	Expected	p-value
1	W 10DPA	GO:0006357	regulation of transcription from RNA pol...	31	4	0.69	0.0047
2	W 10DPA	GO:0005978	glycogen biosynthetic process	17	3	0.38	0.0059
3	W 10DPA	GO:0006432	phenylalanyl-tRNA aminoacylation	7	2	0.16	0.0096
4	W 10DPA	GO:0006814	sodium ion transport	21	3	0.47	0.0108
5	W 10DPA	GO:0006952	defense response	154	9	3.42	0.0154
6	W 10DPA	GO:0000162	tryptophan biosynthetic process	11	2	0.24	0.0237
7	SE 20DPA	GO:0006432	phenylalanyl-tRNA aminoacylation	7	2	0.12	0.0059
8	W 20DPA	GO:0005978	glycogen biosynthetic process	17	3	0.23	0.0014
9	W 20DPA	GO:0006013	mannose metabolic process	9	2	0.12	0.0061
10	W 20DPA	GO:0006357	regulation of transcription from RNA pol...	31	3	0.42	0.0083
11	W 20DPA	GO:0005975	carbohydrate metabolic process	970	26	13.11	0.0097
12	W 20DPA	GO:0000105	histidine biosynthetic process	14	2	0.19	0.0149
13	W 20DPA	GO:0032784	regulation of DNA-templated transcriptio...	15	2	0.20	0.0170
14	W 20DPA	GO:0009082	branched-chain amino acid biosynthetic p...	18	2	0.24	0.0241
15	SE 30DPA	GO:0006357	regulation of transcription from RNA pol...	31	4	0.49	0.0014
16	SE 30DPA	GO:0051603	proteolysis involved in cellular protein...	161	6	2.54	0.0103
17	SE 30DPA	GO:0032784	regulation of DNA-templated transcriptio...	15	3	0.24	0.0146
18	SE 30DPA	GO:0015937	coenzyme A biosynthetic process	12	2	0.19	0.0147
19	SE 30DPA	GO:0006888	ER to Golgi vesicle-mediated transport	33	3	0.52	0.0149
20	SE 30DPA	GO:0007062	sister chromatid cohesion	1	1	0.02	0.0158
21	GRAIN Z75	GO:0006432	phenylalanyl-tRNA aminoacylation	7	3	0.43	0.0067
22	LEAF Z10	GO:0006814	sodium ion transport	21	4	1.00	0.016
23	LEAF Z10	GO:0030244	cellulose biosynthetic process	46	6	2.19	0.021
24	LEAF Z71	GO:0006289	nucleotide-excision repair	24	6	1.78	0.007
25	LEAF Z71	GO:0016226	iron-sulfur cluster assembly	21	5	1.56	0.017
26	LEAF Z71	GO:0009231	riboflavin biosynthetic process	15	4	1.11	0.021
27	SPIKE Z32	GO:0006289	nucleotide-excision repair	24	4	1.01	0.017
28	SPIKE Z32	GO:0016558	protein import into peroxisome matrix	6	2	0.25	0.024
29	SPIKE Z65	GO:0006629	lipid metabolic process	659	49	41.73	0.0067
30	SPIKE Z65	GO:0043086	negative regulation of catalytic activit...	22	5	1.39	0.0107
31	SPIKE Z65	GO:0006289	nucleotide-excision repair	24	5	1.52	0.0156
32	STEM Z32	GO:0016226	iron-sulfur cluster assembly	21	5	1.45	0.013
33	STEM Z65	GO:0008152	metabolic process	8774	657	651.97	0.016
34	STEM Z65	GO:0009082	branched-chain amino acid biosynthetic p...	18	5	1.34	0.021



Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway