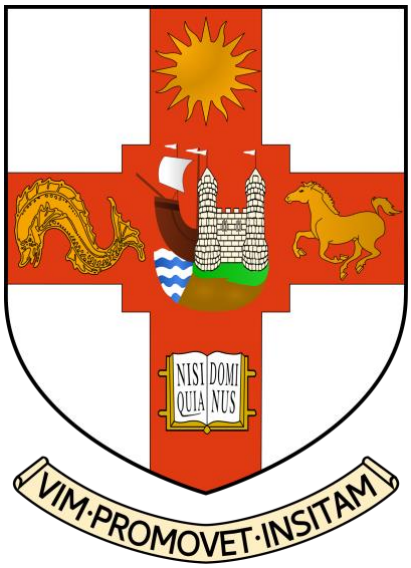
Author:
Biscocho, Ryan R

Title:
To introgress or not to introgress? Quantifying the expression and effectiveness of the 1RS/1BL introgression in a global sample of modern wheat varieties.

To introgress or not to introgress?
Quantifying the expression and effectiveness of the 1RS/1BL introgression in a global sample of modern wheat varieties.

Edmund Ryan Biscocho



Faculty of Life Sciences
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of MSc(R) in Biological Sciences in the Faculty of Life Sciences.

DECEMBER 2020

Word Count: 18,120

Abstract

Bread wheat is an important food source whose low genetic diversity has historically been supplemented with introgressed sequences from a range of different wild grasses. The 1RS/1BL introgression derived from rye is a particularly prevalent introgression that anecdotally provides resistance to several fungal pathogens. However little is known about the expression of these genes and how they behave once introduced into a wheat genetic background, as well as what potential pathways they may be implicated in. To understand this, the transcriptome of various wheat cultivars were studied using an RNA-seq based workflow. Publicly-available RNAseq data, representing a global sample of 38 different varieties, were collated and a subset analysed using a modified RNAseq pipeline to dissect 1RS/1BL expression. As expected, varieties containing the 1RS/1BL introgression showed an upregulation of rye genes and a downregulation of 1B genes, evident of (possible) 1RS introduction and 1BL removal. However these rye genes were expressed at a much lower level relative to wheat genes, which may indicate that these genes are being suppressed, perhaps through homologous co-suppression or by association to nearby transposons. The functional analyses also identified several potential mechanisms by which the 1RS/1BL's documented disease resistance properties may be realised: the use of NBS-LRR receptors and/or a variant jasmonate signalling pathway are possibilities. Overall the study of the 1RS/1BL introgression is able to glean some important generalisations on the use of introgressions, notably that translocated arms should be treated as heterogeneous collections of diverse genes with dynamic expression that may be much lower in a new crop background. Further research, like a similarly designed experimental study, will help fully determine how useful this introgression will be for wheat cultivars, especially its disease resistance properties that will be increasingly important in a future where crop productivity is likely to be under high threat.

Covid-19 Statement

The original aim of this project was to study the expression of the 1RS/1BL introgression in wheat using both Illumina data and MinION data. The latter would have been generated by growing and sequencing select wheat cultivars in laboratory facilities during March/April 2020. As a result of the coronavirus pandemic and lockdown policies during this time sequencing could not go ahead (though plants were grown), and thus this data was not available for analysis and incorporation into this study. Laboratory activities that did not go ahead include RNA extraction, sequencing and library preparation.

If the plans for sequencing had gone ahead an extra 12 Gb of sequence information would have been obtained for analysis. This would have supplemented the approximately 190 Gb of sequence data that was compiled from online databases. Proportionally this is not a huge loss in terms of raw sequence information but this MinION data would have been able to increase the scope of my research project in several ways. Firstly it would have provided RNAseq data on eight extra wheat varieties and two rye varieties which were not already represented in the Illumina data. Having extra varieties to analyse would therefore have made the results of this study more representative of the larger wheat germplasm, and may have improved the quality of the results. Furthermore information on rye varieties would then be available and it would have been possible to study whether or not expression levels change when rye genes are moved from a rye and into a wheat genetic background. Secondly the long reads provided by MinION sequencing would have helped to clarify some results. For example, some observed results may have been due to mis-mapping and due to the better alignments allowed for by longer reads this would have been easier to determine. Thirdly the inclusion of MinION data would have allowed me to explore MinION data more generally. This would have been an exciting opportunity to discuss how this technology could be applied for transcriptomics and crop research, both of which are not well-represented in current uses of MinION sequencing. Overall the limitations caused by the pandemic have reduced the breadth of discussion that was originally intended. However the accessibility of online information has meant that the fundamentals of my analysis and project have remained unchanged.

Dedication and Acknowledgements

I would like to thank the Cereal Genomics Group for their continued help in making this dissertation possible and for their willingness to help whenever I asked, even for small queries that could probably be found within the deep recesses of stackoverflow.com, namely Dr. Gary Barker, Dr. Amanda Burridge, Dr. Paul Wilkinson, Alex Paterson, Gilda Varliero and Virginia Rodriguez-Almansa.

I would also like to thank my friends Olivia Goodwin and Victoria Palumbo, and my immediate family, for their sustained support regarding this project, the former for being practice dummies for presentations and general confidence boosters and the latter for feeding me throughout this process and for allowing me to unofficially convert our dining room into my personal study in my endeavours to complete this dissertation.

Completing a bioinformatics project with effectively zero prior experience on computational biology, on a notoriously difficult genome such as wheat's within a timescale of (approximately) a year wouldn't have been possible without all of your help!

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .....Edmund Ryan Biscocho….. DATE: ....07/12/20.....

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 A Primer on Wheat

The 'big three' cereal crops: bread wheat (*Triticum aestivum*), rice (*Oryza sativa*, *Oryza glaberrima*) and maize (*Zea mays*) provide over half of the calories consumed by humans worldwide, reflecting their sizeable contributions to the global food supply (FAOSTAT, 2020). Out of the three, wheat is the most widely cultivated and occupies the largest amount of land area of any crop plant (FAOSTAT, 2020). This popularity is due to a unique syndrome of characteristics that the plant possesses (Shewry, 2009), which reflect useful features of wheat that make it a prime choice for cultivation as well as biological qualities of its grain that make it ideal as a food source. For example it exhibits a great deal of adaptability and environmental resilience that has allowed it to prosper in a wide range of environments, from as far north as Finland to as far south as Argentina (Lantican, Dubin and Morris, 2005). Wheat also boasts a high yield potential: under optimum conditions wheat can be harvested at yields as high as 9 tonnes ha^{-1} (Lobell, Cassman and Field, 2009). However what truly sets it out from its fellow cereals is its grain. Like the plant itself, the grain is flexible in that it can be processed to make a wide variety of edible end-products like bread and pasta which are extremely adaptable foodstuffs in their own right. The flexibility of wheat as a culinary item can be attributed to characteristics of grain content, as well as the variety of ways in which the wheat grain can be manipulated at different stages of food processing. Post-harvest grains of different varieties exhibit varying levels of hardness, based on the structural properties of a protein known as friabilin (Giroux and Morris, 1998). This results in textural variation in flours with different applications for harder grains (bread) compared to softer ones (cakes). Once flour is processed into dough, wheat exhibits a set of unique physical properties: viscosity, plasticity and elasticity (Wieser, 2007). These fundamental properties of wheat are derived from its high quantities of insoluble gluten proteins (glutenin and gliadin) which, upon contact with water, form a proteinaceous network that is extensible yet elastic (Wieser, 2007). The unique viscoelastic characteristics conferred by glutens can explain wheat's global popularity; by being malleable wheat dough can be manipulated into a variety of shapes and forms to create structure and volume to cooked goods, facilitating its incorporation into a wide diversity of end-products and therefore cultures (Wieser, 2007; Shewry, 2009). With wheat's sheer flexibility, in terms of where and when it can be grown and the potential applications of its grain, combined with its unique physical properties, it is no wonder why it is essentially ubiquitous in all kinds of diets.

As one of the most nutritionally valuable and widespread crops in the world, wheat must be safeguarded against decreases to its yield. This is threatened by external factors in the current world context including climate change, economic instability and the limited availability of arable land. In addition the global population is continuing to grow rapidly, with a projected rise from 7.3 billion people in 2015 to 8.5 billion by 2030, with large implications for world food security (UN, 2015). Even in recent years around 9% of the world was undernourished (FAO, 2020), and this therefore represents an issue that needs to be addressed immediately. It is not enough to maintain wheat productivity – it needs to be improved too. Targets for cereal production require yield increases of over 40% (which represent an increase of almost a billion tonnes per annum) within the next few decades to meet future demands (FAO, 2009) with the caveat that if substantial progress has not been made by 2030 then catching up may not be possible (Rothamsted, 2017).

Despite its importance and urgency, research on wheat has lagged in comparison to other major crops. This can be attributed to aspects of its genetics. For example, wheat has a very complex genome and this complexity has impeded the development of genomic resources, like a reference genome, that would help accelerate its study. A draft reference was generated for rice in 2002 (Goff *et al.*, 2002) and maize in 2009 (Schnable *et al.*, 2009) but one for wheat was not developed until 2012 (Brenchley *et al.*, 2012). A reference genome is a whole genome sequence constructed from one or more individuals of a species and represents a standard against which other genetic sequences can be compared (Ballouz, Dobin and Gillis, 2019). Reference genomes often mark a milestone in the research of an organism, from which rapid jumps in knowledge can be made, and many different research consortia exist or have existed with the singular aim of sequencing whole genomes, like the 100,000 Genomes Project and even the Earth BioGenome Project (Lewin *et al.*, 2018). They are invaluable research tools as they allow important inferences to be made: genes can be annotated and located, SNPs identified and gene expression networks elucidated (Ballouz, Dobin and Gillis, 2019). In addition they allow for important genomic parameters like gene number, gene family copy number and transposon content to be determined, which can glean information on the evolution of a species' genome via comparative genomics. Reference genomes also accelerate the speed of resequencing and further genome assembly by providing a scaffold for reads to map to (Ballouz, Dobin and Gillis, 2019), allowing individual variation to be catalogued and facilitating population genomic studies.

The reasons for wheat's complexity and for the delay in developing a reference genome is two-fold: genome size and polyploidy, each posing their own problems. Where rice and maize are diploids with genome sizes of 0.39 Gb (IRGSP, 2005) and 2.30Gb (Schnable *et al.*, 2009) respectively, wheat is a polyploid with an expansive 17 Gb genome (IWGSC *et al.*, 2018). Such a large genome would have been unfeasible, in terms of both time and money, to sequence using the technology available in the 2000s. This posed a major problem previously but recent jumps in technology, especially in improvements to throughput, have made it possible to sequence such large genome sizes (van Dijk *et al.*, 2014).

The second problem is that of polyploidy. To understand this problem fully one must consider the evolutionary history of wheat. Modern bread wheat is a polyploid, but more specifically an allohexaploid: it has six sets of chromosomes (-hexaploid) derived from different progenitor species (allo-) (Comai, 2005), which are sub-divided into three diploid sub-genomes AABBDD (Gustafson *et al.*, 2009). This is the result of two distinct rounds of hybridisation between the progenitor species of modern bread wheat: wild wheat species and wild grasses known as goatgrasses (Smith *et al.*, 2009). The first occurred between *Triticum urartu* (AA, 7N) and *Aegilops speltoides* (BB, 7N) which gave rise to tetraploid emmer wheat *Triticum turgidum* (AABB, 14N) after a chromosome doubling event that made it fertile (Smith *et al.*, 2009). The second occurred between the resultant emmer and another goatgrass *Aegilops tauschii* (DD, 7N) giving rise to modern bread wheat after another round of chromosome doubling (Smith *et al.*, 2009). The result was a wheat species with a huge genome organised into twenty one pairs of chromosomes with seven pairs per sub-genomes (IWGSC *et al.*, 2018). As its genome is organised into six constituent sets of chromosomes derived from related species, problems arise with accurately identifying genes. All three progenitor species of wheat were close relatives of one another and so genes exist as multiple, sequence-similar copies across the sub-genomes, known as homoeoalleles/homoeologs – a property known as homoeology. Homoeoalleles can share sequence similarities of 97% and over with one another (Uauy,

2017) and so their presence makes it difficult to determine which sub-genome a particular copy is derived from, which poses a computational problem for genome assembly. This is further compounded by the fact that the wheat genome consists of over 80% repetitive elements (Brenchley *et al.*, 2012; IWGSC *et al.*, 2018). These repetitive sequences again make it difficult to assemble reads together. Usually one can rely on synteny – where gene order is conserved across a taxon – among grass species (Feuillet and Keller, 2002) to help with determining where to put which reads, but gene order has been shuffled in certain chromosomal locations in wheat (Langridge, 2012). Further down the line, homoeoalleles also obscure the identification of true SNPs, as variants exist between sub-genomes (homoeologous SNPs) it becomes problematic to distinguish these from variants between different varieties (varietal SNPs), and only the latter are informative for marker-assisted selection (Winfield *et al.*, 2012). The problem with assembling reads also extends to mapping reads: homoeoalleles make it difficult to unambiguously align transcripts to certain regions to quantify gene expression (Kyriakidou *et al.*, 2018) and this can have important effects on RNAseq studies such as this which require accurate mapping. In previous years, and to some extent in the present, the wheat genome has been problematic to research but with recent advancements in sequencing technology it has become much more tractable for genomic study.

Another of wheat's genetic problems is that its germplasm exhibits low genetic diversity. This is problematic as sufficient variation is required to source alleles that can be used for survival and improvement. For example, growing crops of similar genotypes make them extremely susceptible to decimation by pathogens: an epidemic caused by the fungus *Helminthosporium maydis* destroyed almost half of all maize crops in southern US in 1970, an event facilitated by the crop's narrow genetic base (Esquinas-Alcázar, 2005). Using cultivars which all have similar genotypes therefore make them vulnerable to any perturbations in abiotic or biotic conditions, with potentially devastating consequences to yield. This remains true for wheat which has especially low genetic diversity that can be attributed to a series of successive population bottlenecks. Crops in general suffer reduced genetic variation as a result of domestication: only a small proportion of wild individuals become domesticated, reducing effective population size and genetic diversity (Haudry *et al.*, 2007). However this seems to be more pronounced in wheat. Using nucleotide diversity as a measure, most crops exhibit 30% lower genetic diversity than their wild counterparts but bread wheat shows a much greater decrease of 69% (Haudry *et al.*, 2007). This could partially be due to differences in effective population sizes: tetraploid durum wheat (*Triticum durum*) showed an even lower decrease in diversity (84%) than bread wheat, and the former appears to have experienced a more intense population bottleneck, with a smaller resulting effective population size, than the latter (Haudry *et al.*, 2007). Polyploidy has also contributed to lowered genetic variation, acting as another bottleneck in wheat's evolutionary history. Only a relatively small number of individual progenitor plants would have undergone the polyploidisation events that generate hexaploid wheat. This created an effective bottleneck with these new polyploid wheats being unable to interbreed freely with diploid progenitors, preventing gene flow and therefore limiting genetic variation to this new founding population (Willis and McElwain, 2013). Polyploidisation involved two rounds of hybridisation and so this would have involved at least two rounds of bottlenecks. Further losses in genetic diversity are also experienced in the way crops are cultivated. Crops are often bred to generate varieties that produce the highest yield possible, resulting in the formation of a few superior varieties which are favoured and grown widely. To create these varieties a breeder may create inbred lines to enhance useful traits that the lines already possess. The lack of outbreeding here prevents the broadening of the genetic base. These superior crops

may also be grown in large swathes, all having very similar genotypes (Reif *et al.*, 2005). Common agricultural practices therefore act as an obstacle to crop diversity, but this can be alleviated. In the 1990s, after years of decreasing, genetic diversity experienced an upturn as a result of breeders actively using foreign and exotic material to supplement the germplasm of cultivated wheats (Reif *et al.*, 2005). As we will see, outsourcing genomic novelty in such a fashion has become an invaluable method of circumventing the low genetic diversity of cultivated wheats.

1.2 Introgressions

Wild grasses can be excellent sources of genes with adaptive traits; they often represent a much extended gene pool that likely harbour agronomically useful genes which can be drawn upon for wheat breeding (Winfield *et al.*, 2016). This is done via introgression, where hybridisation and repeated backcrossing introduce genetic sequences from the gene pool of a wild grass to wheat (Harrison and Larson, 2014). The resulting product is a crop with a gene or genes of interest from a wild grass within the genetic background of a cultivated wheat variety. Introgressions are readily tolerated by wheat as its polyploidy (the presence of multiple sub-genomes) helps buffer any genomic perturbations caused by substitutions of chromosome portions that it involves (Dubcovsky and Dvorak, 2007). Considering this tolerance and wheat's low genetic diversity, introgressions are effective methods for wheat breeding. Indeed wild grass introgressions are commonplace in wheat, both in cultivars and landraces (Cheng *et al.*, 2019). This is suggestive of them being useful sources of genetic variation both for artificial and natural selection. The wheat germplasm is thought to contain introgressions from 52 species which represent 13 genera (Wulff and Moscou, 2014), such as close relatives like *Aegilops* (Schneider, Molnár and Molnár-Láng, 2008) or environmentally hardy wild grasses like *Thinopyrum* (Ren *et al.*, 2017). The success of wild grasses as a donor of genetic novelty may be due to them being rich in genes for increased survivability and resistance. In contrast to food crops under intense artificial selection pressure for better productivity and nutrition, wild grasses would have been under greater pressure to evolve traits that would have allowed for better survival. Wild grass introgressions may therefore be effective methods of offsetting the relative fragility of cultivated crops. In practice many introgressed genes do serve this function. For example over 50 resistance alleles exist in the wheat germplasm against *Blumeria graminis* f. sp. *tritici*, the causative agent of powdery mildew (Luo *et al.*, 2009; He *et al.*, 2009). Only 25 of these alleles originate from *Triticum aestivum*, the rest being sourced from closely related species of the *Triticum* and *Aegilops* genera, as well as other more distantly related grasses like *Thinopyrum timopheevi*, *Haynaldia villosa* and *Elytrigia intermedium* (Luo *et al.*, 2009; He *et al.*, 2009). Wild grasses can also provide resilience against abiotic stressors, providing phenotypes such as salinity-tolerance (Wang *et al.*, 2014) and drought-tolerance (Placido *et al.*, 2013). Furthermore introgressions may even be used not just for survivability purposes but also to improve end-product qualities of wheat, where chromatin from wild grasses has been shown to bolster grain protein content (Pace *et al.*, 2001; Kumar *et al.*, 2011). From the myriad of examples seen in the literature, introgressions have clearly been invaluable in improving various aspects of wheat biology.

Out of all of the potential sources of introgression, rye (*Secale cereale*) has been one of, if not the most, important for wheat (Ren *et al.*, 2017). Rye chromatin appears frequently in wheat cultivars (Crespo-Herrera, Garkava-Gustavsson and Åhman, 2017), providing a versatile array of advantages for wheat: improved seedling vigour, pathogen resistance and water-use efficiency (Saulescu *et al.*, 2011) make up just a few. In fact, crosses among rye and wheat are so effective that a hybrid between the two even exists: a synthetic grass known as triticale (Mergoum and Gómez-Macpherson, 2004). Triticale combines advantageous aspects of both species – wheat's nutritional properties and rye's environmental resilience – to establish a high-yielding food crop (Mergoum and Gómez-Macpherson, 2004). It is also economically important, filling a niche as animal feed (Mergoum and Gómez-Macpherson, 2004). However as it is not commonly used for human consumption, it is produced on a much lower scale than wheat itself (FAOSTAT, 2020). Triticale exemplifies how compatible and effective rye is as a donor species for wheat improvement. Another example for this is the 1RS introgression, a translocation that

substitutes out a portion of a wheat chromosome for the short arm of rye chromosome 1. This is the most common form in which rye genes are present in wheat varieties (Rabinovich, 1998) and exists as one of two variants: 1RS/1AL or 1RS/1BL, where 1RS substitutes either the short arm of wheat chromosome 1A or 1B, respectively (Graybosch *et al.*, 2019). The 1RS/1BL variant is by far the most prevalent of the two; it exists globally (Graybosch *et al.*, 2019) and within approximately 1000 wheat cultivars (Molnár-Láng, Ceoloni and Doležel, 2015), being present in up to as many as 34% of a country's wheat varieties (Crespo-Herrera, Garkava-Gustavsson and Åhman, 2017). It is also the more prominent variant present in wheat literature. Cultivars containing the 1RS/1BL introgression remain widely used as they improve yield and yield stability, as well as improving crop performance over a wider range of environments (Kumlay *et al.*, 2003). However the main reason for its popularity is in pathogen resistance: it provides a slew of resistance alleles against several wheat pathogens (Ren *et al.*, 2017). These include genes against many high-morbidity fungal pathogens such as stripe rust (*Puccinia striiformis* f. sp *tritici*) (Ren *et al.*, 2009), stem rust (*Puccinia graminis* f. sp. *tritici*) (Koebner, Shepherd and Appels, 1986) and powdery mildew (Ren *et al.*, 2009), though resistance alleles also exist against insect pests like the Russian wheat aphid (*Diuraphis noxia*) (Anderson *et al.*, 2003). In some cases the same rye genotype can even contribute more than one resistance allele to wheat, against different isolates of the same pathogen (Ren *et al.*, 2009). Using the 1RS/1BL introgression therefore allows for resistance traits against many pathogens to be inherited simultaneously, making it a very effective choice for improving wheat crops. Indeed fragments of rye chromatin, but especially this introgression, have been superior choices for wheat breeding and explains why wheat cultivars containing this introgression continue to be grown at a large scale.

The 1RS/1BL introgression however is not without its flaws. Firstly, though the introgression generally improves wheat phenotype this is not always the case. Its effects are dependent on the wheat genetic background as well as the rye source (Ren *et al.*, 2012; Lelley, Eder and Grausgruber, 2004). Some introgressions are more effective than others and, depending on genotype, a wheat line containing this introgression will not necessarily perform better than one without it (Lelley, Eder and Grausgruber, 2004). Even genotypes that do benefit from the introgression may not enjoy this improvement indefinitely. The introgression has a narrow genetic base with most cultivars obtaining their introgression from a singular rye source – Petkus rye (Schlegel and Korzun, 1997) – and so most cultivars share the same resistance alleles (Ren *et al.*, 2012; Lelley, Eder and Grausgruber, 2004). As a result these resistance alleles can easily be overcome by pathogen counter-adaptation and is thought that many of these alleles are no longer effective (Ren *et al.*, 2012). In addition, not all of the genetic changes caused by this introgression are beneficial. Some genes that are introduced are deleterious: the *Sec-1* locus is transferred which codes for rye storage proteins known as secalins and contributes to dough being sticky, an unfavourable trait for breadmaking (Howell *et al.*, 2014). Some beneficial genes have also been excised along with the substituted 1B long arm, including loci encoding for gluten proteins (*Glu-B3* and *Gli-B1*) that strengthen dough (Howell *et al.*, 2014). It is evident then that the increased environmental adaptability provided by the introgression is somewhat offset by negative effects on end-use qualities. A further problem with the 1RS/1BL introgression is that we lack detailed information on the mechanisms by which it brings about its benefits. For one we are unaware of what particular genetic mechanisms are involved in providing disease resistance; the same gene names recur in the literature as being important in delivering resistance to fungal pathogens but we do not know what these genes actually do. Most studies focus on detailing how introgression lines are made and their cytogenetic

characterisation, rather than determining their particular functions. Without this knowledge we have limited insight into the functional basis of the introgression. There are clear drawbacks for the use of the 1RS/1BL introgression and further work is necessary to see how useful it will be for future wheat breeding.

Other potential problems with the use of the 1RS/1BL introgression are related to gene expression. The literature often works under the assumption that all the beneficial genes from an introgression are expressed but this is not always the case. For example the resistance gene for powdery mildew *Pm8* can be suppressed by an orthologous gene found in the 1A genome of wheat (Hurni *et al.*, 2014). Some lines containing this introgression therefore do not actually enjoy increased resistance to this pathogen and resistance is dependent on variety and genetic background (Hurni *et al.*, 2014; Crespo-Herrera, Garkava-Gustavsson and Åhman, 2017). With little to no quantitative data to support the idea that rye transgenes are expressed, one could argue that the effects seen in 1RS/1BL-containing wheat lines are due to the excision of the 1B short arm, not the introduction of the 1R short arm. As allopolyploids like wheat have complex patterns of expression it is not unreasonable to postulate this. Their complex expression is the direct result of allopolyploids fostering several sub-genomes derived from different species, representing multiple distinct modules of ancestral regulation which may conflict with one another. Alleles will be expressed at levels appropriate for their ancestral species and incipient allopolyploids must reconcile different regulation patterns to best fit the new hybrid (Feldman *et al.*, 2012). This is achieved through gene suppression and gene loss, removing competition (or redundancy) between homoeoalleles by reducing the action of one or more of them. Studies show that 30% of loci of hexaploid wheat show unbalanced gene expression where at least one homoeoallele has been suppressed and in 10% of loci only one functional homoeoallele remains functional due to suppression (Feldman *et al.*, 2012). Experimental evidence supports this theory with gene silencing and loss occurring immediately upon artificially-induced polyploidisation of wheat lines (Kashkush, Feldman and Levy, 2002; He *et al.*, 2003). The result is a phenomenon known as genomic asymmetry, where one sub-genome becomes more dominant in expression than the other sub-genomes. While not a global paradigm, genome asymmetry is observed for particular loci causing certain sub-genomes to contribute more for particular phenotypes (Feldman *et al.*, 2012; Pfeifer *et al.*, 2014), and sub-genomes even exhibit a degree of specialisation for said phenotypes. For example the A genome appears to be dominant in the control of morphological traits like caryopsis and grain morphology whereas the B genome is enriched in the expression of genes related to environmental adaptation and tolerance to external stresses (Feldman *et al.*, 2012). Furthermore other nutritionally and agronomically relevant traits like baking quality and pathogen resistance also appear to show sub-genome bias (Pfeifer *et al.*, 2014; Powell *et al.*, 2017). This may have important implications for introgression-centric breeding efforts in the future: will transgenes have to be targeted toward specific sub-genomes so they are expressed?

Another point of concern is the continued suppression of wheat genes. If not immediately suppressed upon introduction, introgressions may still be suppressed given time. Gene suppression in hexaploid wheat appears to be a flexible and continuing process as it has been suggested that the frequency of silenced and lost genes is increasing over time (Bottley, Xia and Koebner, 2006). This may be because wheat, as an evolutionarily young polyploid, may still be in the process of managing the expression of its complex genome. The implication of this is that introgressions introduced in the future may be a target for gene suppression. Some studies even suggest

that gene silencing may be biased against foreign sequences. For example in rice the introgression of sequences from a wild grass resulted in silencing both epigenetically and via retrotransposon action (Liu and Wendel, 2000). Such potential intolerance for transgenes results in very short-lived, if any, effects on the host plant. It is evident that the wheat transcriptome has evolved to be extremely dynamic, using suppression techniques to deal with the stresses of harbouring multiple homoeoalleles (Feldman *et al.*, 2012). However this dynamism poses a threat to the introduction of novel sequences into the wheat germplasm which may represent prime targets for silencing. Investigating the expression of existing introgressions will therefore help shed light on whether or not future introgressions will be tolerated by the wheat genome. This will allow us to evaluate whether the 1RS/1BL introgression, as well as introgressions in general, remain strong strategies for improving wheat cultivars.

1.3 RNAseq

In theory introgressions have been an effective method in improving crop genotypes and phenotypes. A multitude of studies characterise introgressions and their putative benefits to the wheat genotype but very few studies have even looked at whether or not these introgressions are expressed. A recent study showed that only 18 out of 1373 genes found in introgressed blocks showed differential expression between introgression lines and non-introgression lines (Cheng *et al.*, 2019), potentially suggesting transcriptional inactivity of a large proportion of introgressed genes. Another study compared the wheat variety Chinese Spring with variant wheat lines containing artificially introduced chromosomes from *Aegilops longissima* (Dong *et al*., 2020). They showed that most of the genes that were differentially expressed were mostly downregulated or not transcribed – many genes involved in introgression may not be expressed. However, twelve upregulated introgression genes were identified that putatively confer pathogen resistance, proving that introgressions can still be effective. The project outlined here elaborates on both of these previous studies by looking specifically at the 1RS/1BL introgression and by taking a meta-analysis approach, analysing many different wheat varieties to look at the general expression effects of 1RS/1BL and introgressions as a whole.

Studying expression is especially important for wheat given its complex pattern of gene expression that may silence an introgression altogether. For a study of this kind, the 1RS/1BL translocation constitutes a useful model introgression for study: many varieties contain it so there should be plenty of data available for it, and many of these varieties are widely grown so the results of this research should be relevant for wheat breeders. In addition there are some potential pitfalls and uncertainties regarding the use of the 1RS/1BL introgression. Research on how this translocation is expressed may help clarify several questions that will allow us to determine its effects on wheat phenotype more closely, and evaluate how useful it will be for future wheat breeding. For one, like other introgressions, we have no positive confirmation that all of its genes are expressed. Even if they were at some point expressed, they may no longer be: some resistance genes are no longer effective against their pathogen antagonists (Ren *et al.*, 2012; Ren *et al.*, 2017; Crespo-Herrera, Garkava-Gustavsson and Åhman, 2017) and whilst this is likely due to the emergence of resistant isolates it could also be possible that the genes have been silenced. Even if many genes are expressed it would be important to know the identities of those which are and those which are not. Some genes of the 1RS/1BL introgression are beneficial but some are also deleterious and by knowing which are switched on or off we can better evaluate the costs and benefits of its use. Furthermore, it is known that the performance of this introgression is modulated by rye source and wheat background (Ren *et al.*, 2012). As a result one would expect that the introgression would show variant expression among different wheat cultivars and it would be informative to know which varieties expressed it and if there are any evident patterns (e.g. if cultivars from a particular region consistently did not express it). In addition by studying the expression of this introgression we would gain a more complete picture of the mechanisms by which it works. Downstream from obtaining the list of expressed genes one can look at functions and shared pathways that these gene sets may be involved in, providing information on the processes by which the introgression may bring about pathogen resistance and other phenotypic changes.

To obtain information on expression, RNA-sequencing (RNAseq) data can be analysed to study the transcriptomes of wheat. RNAseq is a technique that uses a sequencing-based method to profile a sample's transcriptome, and

has emerged as a result of sequencing technology becoming cheaper and more sophisticated (van Dijk *et al.*, 2014). In RNAseq all the mRNA fragments from a sample are isolated and converted to cDNA (Wang, Gerstein and Snyder, 2009). These cDNA molecules are then sequenced and mapped back to a reference genome and the number of fragments associating to different genes give a measure of that gene's expression (Wang, Gerstein and Snyder, 2009). RNAseq offers various advantages in comparison to other methods like microarrays as it works over a higher dynamic range, provides higher resolution data and can even identify sequence variants in transcripts (Illumina, 2020; Wang, Gerstein and Snyder, 2009). In addition it can be used for species without reference genomes; this information is not required *a priori* as transcriptomes can be generated from scratch (Illumina, 2020; Wang, Gerstein and Snyder, 2009). As a result it has largely supplanted microarrays and has, in recent years, proved invaluable for crop research. For example resources like expVIP now exist which catalogue the information from various RNAseq studies to allow users to easily study the transcriptome of a collection of wheat varieties (Borrill, Ramirez-Gonzalez and Uauy, 2016). This exemplifies the kind of research we are now able to do with such a boom in sequence and transcriptome data; by applying a similar kind of analysis that aggregates large amounts of data we could massively illuminate questions we have on how introgressions are expressed.

In this study I will be leveraging large quantities of information, gathering publicly available RNAseq data and applying a custom bioinformatics pipeline to quantify expression of the 1RS/1BL introgression in an assortment of wheat varieties. By using the 1RS/1BL introgression as a model organism for study, I am aiming to infer how its genes are expressed and how its expression may change across different cultivars, with the hope that this may elucidate general rules on how other introgressions may be treated by the wheat genome. By looking at differential expression I am also hoping to obtain information on what pathways and processes this introgression may be influencing. These results should be able to unveil not just how the 1RS/1BL introgression works, but also some framework for how wheat's regulatory system manages the introduction of introgressions, and how this can inform strategies for designing future wheat cultivars. Originally this would have been supplemented with long-read sequencing, where an extra set of cultivars would have been sequenced using Oxford Nanopore's MinION platform. However due to facilities closing down as a result of the coronavirus pandemic I was unable to follow this through. This would have provided an important comparison point to the Illumina data. In contrast to Illumina, which is cost-effective and produces short reads at high throughput, MinION sequencing is able to produce much longer reads of over several kilobases (Amarasinghe *et al.*, 2020). This has several advantages to Illumina, namely in that it allows for more reliable alignment of reads which would have been a useful reference point to validate any possible erroneous Illumina read mapping. It also has several other benefits such as avoidance of PCR amplification bias and real time sequencing (Amarasinghe *et al.*, 2020), and it would have been an interesting point of discussion to dissect any differences in the results from the Illumina and MinION analyses that may have resulted from their differences in sequencing protocol.

To generate the data to answer these questions an RNAseq pipeline must be used. Expression studies like this leverage sequential use of programs that perform the same basic steps: pre-processing, read alignment, counting and functional analysis. Many variant RNAseq pipelines exist which collect together different individual programs which yield the same expression information. These have different workflows, use different methods and may provide different results (Costa-Silva, Domingues and Lopes, 2017) and so a particular pipeline must be

chosen carefully to provide optimal results. The pipeline in this study uses fastp, HISAT2, htseq-count and DESeq2. This pipeline was chosen as it has a simple workflow and has been explained and validated in a paper as an appropriate means for studying gene expression (Yalamanchili, Wan and Liu, 2018). In addition, these programs together do not form a pre-defined 'protocol' like the Tuxedo or BallGown protocols so may be more amenable to substitutions of individual programs (as such protocols may require some level of backward dependency). This is beneficial and important in this case as it allows for a similar pipeline to be used with the MinION data that would have been collected, which would facilitate comparisons between the two kinds of data. MinION data necessitates the use of a different read mapper (as a result of long-read sequencing having a higher error rate than Illumina) and so any pipeline used must be robust to changes. No current read mapper is optimal in both scenarios and so different read mappers should be used for the different data types. The use of this apparently robust pipeline therefore allows for this and facilitates comparisons between the two.

With regard to some specifics of the pipeline, fastp is used for pre-processing, HISAT2 for read alignment, htseq-count for counting and DESeq2 for differential expression analysis. After the pipeline the Gene Ontology web resource can be used for gene enrichment analysis. All of the programs are necessary to produce high-quality data. fastp is necessary to gauge the quality of the reads and process when appropriate; this involves automatically removing low quality data e.g. removing low quality reads, trimming adapters and low quality bases at the 5' and 3' ends. As different studies may have used different pre-processing methods before submitting their reads to bioinformatics repositories, it is important to ensure that they had all been processed in the same fashion before putting them through the pipeline. The reads remaining after clipping and filtering can then be aligned to a reference genome using HISAT2. HISAT2 is a splice-aware aligner which makes it suitable for an RNAseq study like this; as mRNA reads can span introns and exons that may be spliced out, aligners need to be aware of such splicing events for them to accurately map these reads back to a genome. In this study a custom reference genome (containing both the wheat reference and a rye draft assembly) is used. This provided the opportunity for wheat reads derived from the 1RS/1BL introgression to correctly map to rye, therefore allowing its expression to be studied. Without this genome the rye-derived reads may erroneously map to sequence-similar wheat regions due to homology, or not map at all. After alignment, reads mapping to certain gene regions must be counted in order to get data regarding expression by htseq-count. This produces raw, non-normalised counts for each gene which can be used as a proxy for gene expression once normalised. A custom GFF is also used here containing both wheat and rye information, in order to match the custom reference genome and alignment information from the previous step. After applying the pipeline the count data can be analysed using DESeq2, a statistics package on R. DESeq2 effectively takes an input of raw count data and outputs a list of differentially expressed genes (DEGs) between specified conditions. In this case varieties are split into two conditions: introgression-containing and non-introgression-containing. Once functions have been found for these DEGs, gene enrichment analysis can then be used to provide further context by outputting a list of biological processes that may be perturbed as a result of the introgression. The DEGs and their downstream gene enrichment results make up the crux of the process and provide information on what functions may have been changed as a result of introgression, and as a result what phenotypes one may predict in cultivars when introducing this introgression as part of wheat breeding strategies.

My main research questions are as follows:

1. **Are the 1RS genes expressed in any wheat varieties?**

   Which varieties show expression of 1RS genes?

2. **What genes are expressed and what are their functions?**

   Does this support the notion that the 1RS genes are the cause of adaptive phenotypes?

3. **How can my results inform future wheat breeding?**

2 Materials and Methods

2.1 Section 1: Illumina Data

RNAseq reads representing different hexaploid wheat varieties from various sources/studies were obtained as shown in Table 1. The full dataset containing all of the metadata is seen in the Supplementary Material (S1). Data was split into two conditions: introgression and non-introgression, based on whether or not they contained the 1RS/1BL introgression, using several references (Winfield *et al.*, 2016; Schlegel and Korzun, 2020; Rabinovich, 1998).

Obtaining RNA-Seq Data

A comprehensive manual literature search was performed to obtain sources of information that provided RNAseq data on hexaploid wheat varieties and rye varieties. Searches were made in literature search engines like Google Scholar and through data repositories such as the NCBI SRA and EBI ENA. Sources which had relevant information were included in this study and their read data downloaded and included in the analysis. Information was deemed relevant if it passed three criteria:

1. Does it provide data on a hexaploid wheat (*Triticum aestivum*) variety?
2. Does it provide RNAseq data?
3. Do we know if it contains the 1RS/1BL introgression?

Many sources did not pass all criteria e.g. information on tetraploid wheat was given, RNAseq data was inaccessible, microarray data was provided instead, etc. The last criterion excluded the greatest number of sources and many varieties had no known information on if it contained the 1RS/1BL introgression, and thus could not be included given the nature of this study's design. To determine whether or not a particular variety contained the 1RS/1BL introgression a publicly-available, independently-curated and reference-supported database of wheat varieties was referenced (Schlegel and Korzun, 2020). My lab group's own exome-capture data (which provides similar data on a smaller number of varieties) was also referenced (Winfield *et al.*, 2016).

Raw fastq files of relevant sources were downloaded from the EBI ENA database's FTP server using *wget*. The EBI ENA database was used as the source for all reads due to it having the easiest interface to download files from. Once downloaded, the relevant reads were compiled together using the *cat* command so that one variety would be represented by a single file of RNAseq reads – this single file would represent all of the runs, tissue types, etc available for that particular read, unless a particular tissue such as leaf-tissue needed to be analysed. If reads were paired-end then a single variety would be represented by two files of paired-end reads, again using *cat* whilst taking care that files were concatenated in the correct order. Using one file (or one pair of files) for one variety facilitates easier manipulation of data in the downstream analysis.

Table 1
Information on wheat varieties

Table shows the varieties with which RNAseq data was collected along with their sources (NCBI accession and paper DOI) and related metadata (tissue and country). Varieties shaded in orange contain the introgression and those in blue do not contain the introgression. This is a sample of a larger dataset that contains more metadata information such as study type and data quantity. This full dataset can be accessed in the Supplementary Material (S1).

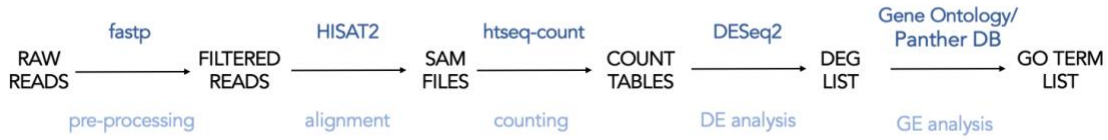
Variety	Tissue	Country	NCBI Accession	Reference
Aimengniu	Spike	China	PRJNA348655	(Wang *et al.*, 2017)
Bacanora	Leaf	Mexico	PRJEB5290	(Harper *et al.*, 2016)
Banks	Spike	Australia	PRJEB23118	(Rangan, Furtado and Henry, 2020)
Beaver	Leaf	UK	PRJEB5290	(Harper *et al.*, 2016)
Bobwhite	Leaf	Mexico	PRJNA497810	(Borrill *et al.*, 2019)
Equinox	Leaf	UK	PRJEB5290	(Harper *et al.*, 2016)
Florida	Leaf	Germany	PRJEB5290	(Harper *et al.*, 2016)
Haven	Leaf	UK	PRJEB5290	(Harper *et al.*, 2016)
Lovrin 10	Spike	Romania	PRJNA348655	(Wang *et al.*, 2017)
Lumai 15	Leaf	China	PRJNA351906	(Ni *et al.*, 2017)
Nautica	Leaf	Netherlands	PRJEB5290	(Harper *et al.*, 2016)
Palur	Leaf	Germany	PRJEB5290	(Harper *et al.*, 2016)
PBW 343	Leaf	India	PRJNA613349	Unpublished
ProIntaFederal	Leaf	Argentina	PRJNA490015	(de Haro *et al.*, 2019)
Rialto	Leaf	UK	PRJEB5290	(Harper *et al.*, 2016)
Savannah	Leaf	UK	PRJEB5290	(Harper *et al.*, 2016)
Svilena	Microspore	Bulgaria	PRJNA297977	(Seifert *et al.*, 2016)
Alba	Leaf	Poland	PRJEB5290	(Harper *et al.*, 2016)
Atlas 66	Leaf	USA	PRJNA563057	(Cheuk, Ouellet and Houde, 2020)
Avalon	Leaf	UK	PRJEB5290	(Harper *et al.*, 2016)
Borenos	Leaf	Germany	PRJEB5290	(Harper *et al.*, 2016)
C 306	Root	India	PRJNA529036	(Kaur *et al.*, 2019)
Capo	Leaf	Austria	PRJEB5290	(Harper *et al.*, 2016)
Chinese Spring	Leaf	China	PRJEB5290	(Harper *et al.*, 2016)
Chuanmai 25	Leaf	China	PRJNA555667	(Bhoite *et al.*, 2020)
Fortuna	Leaf	Russia	PRJNA514367	(Jobson *et al.*, 2019)
Holdfast	Grain	Australia	PRJEB7795	(Pearce *et al.*, 2015)
Jagger	Leaf	USA	PRJNA485724	(Peng *et al.*, 2019)
Jimai 19	Root	China	PRJNA355905	(Jiang *et al.*, 2017)
Obelisk	Leaf	Netherlands	PRJNA415716	(Haueisen *et al.*, 2019)
Raj 3765	Root	India	PRJNA435777	(Dalal *et al.*, 2018)
Saratovskaya 29	Leaf	Russia	PRJNA630059	(Ermakov *et al.*, 2019)
Sevin	Leaf	Denmark	PRJNA196595	(Yang, Li and Jørgensen, 2013)
Stoa	Shoot/Leaf	USA	PRJNA397654	(Bajgain, Russell and Mohammadi, 2018)
Sumai 3	Shoot/Leaf	China	PRJNA397654	(Bajgain, Russell and Mohammadi, 2018)
Triumph	Spike	USA	PRJNA348655	(Wang *et al.*, 2017)
Yecora Rojo	Leaf	Mexico	PRJNA629995	(Wu *et al.*, 2019)
Yumai 18	Spike	China	PRJNA491844	(Tang *et al.*, 2020)

Bioinformatics Pipeline

A custom bash pipeline was written containing all of the analysis steps and was used to process the compiled data (see appendix). The pipeline analyses the data from the individual varieties one by one, reading the file's name, applying different parameters depending on if the read data is single-end or paired-end and outputting files that are named by variety.

The pipeline uses the programs fastp, HISAT2 and HTseq-count. After this the pipeline's output is then manually analysed and passed onto DESeq2 (run on an R environment) and Gene Ontology's gene enrichment analysis tool (run on a browser). fastp is first used to preprocess raw reads. HISAT2 then aligns the filtered reads to a custom reference genome, resulting in SAM files. htseq-count then counts the number of reads mapping to gene regions, with the assistance of a custom GFF, resulting in raw, un-normalised count tables. These count tables can then be imported to R, where DESeq2 can be run. Using the count table information, DESeq2 determines which genes are differentially expressed between the conditions, resulting in a list of differentially expressed genes (DEGs). The list of DEGs can then be passed on to the Gene Ontology gene enrichment analysis tool (which uses the Panther database), resulting in a list of biological processes that are overrepresented and underrepresented in one condition in comparison to the other.

This can be visualised in the following diagram, which shows the programs in dark blue, their functions in light blue and the input/output files in black:



Further information on the individual steps is provided below.

fastp:

fastp version 0.20.0 was used with default settings (Chen *et al.*, 2018). An example of the command line used (for a single-end RNAseq file) in the pipeline is shown below:

```
fastp -i [INPUT FILE] -o [OUTPUT FILE]
```

HISAT2:

HISAT2 version 2.1.0 was used with default settings (Kim *et al.*, 2019). An example of the command line used (for a single-end RNAseq file) in the pipeline is shown below:

```
hisat2      -p [THREAD NUMBER] -x [PATH TO REFERENCE GENOME]
      -U [INPUT FILE] -S [OUTPUT FILE]
```

htseq-count:

htseq-count (which was part of the HTseq framework) version 0.11.0 was used with the following command line options (Anders, Pyl and Huber, 2015):

```
--type=gene
--idattr=gene_id
--strandedness=no
```

The first two options allow the GFF to be parsed correctly. The last option forces all data to be treated as unstranded which allows for standardisation as some studies have stranded information and some do not. All remaining parameters were set to default. An example of the command line used (for a single-end RNAseq file) in the pipeline is shown below:

```
htseq-count --type=gene --idattr=gene_id --stranded=no
      [INPUT FILE] [GFF FILE] > [OUTPUT FILE]
```

DESeq2

DESeq2 version 1.28.1 (Love, Huber and Anders, 2014) was used and steps were followed according to the DESeq2 vignette (Love, Anders and Huber, 2020).

Functional annotation and GO Analysis

Functional annotation of the top DEGs is necessary to determine which biological processes may have changed due to introgression. Here functional annotation refers to manually finding the functions (or putative functions) of the genes by BLASTing sequences as well as seeing if existing repositories of functional information have data on them. Various results for functions are compiled together in a dataset (Supplementary Material, S2) with the most informative result for function used in downstream analysis. Below describes how this information was obtained for wheat genes and rye contigs – different measures are used as different kinds of data are available to them.

For wheat genes:

Repositories of functional information are searched to find putative gene functions. First UniProt is searched and any function found is inputted into the database. The IWGSC also have a set of gene annotations which is searched for functions. Again any function found is inputted into the database. If functions found here appeared to contradict then the gene would be searched via blastp to find the functions of top hits and determine which function would be used in downstream analysis. The amino acid sequence of the gene would be obtained from UniProt by searching for the gene ID e.g. TraesCS1B02G020600, which is then used as the input for blastp. BLAST 2.10.1 is used and set to search through the 'non-redundant protein sequences/nr' database, with the search limited to the grass taxon Poaceae (taxid: 4479). All other parameters are set to default. The best or top-reported match of this search is inputted into the database. If the top hit has no apparent functional information (e.g. an 'uncharacterised protein') and other genes with matching functions appear at least twice in the top 20 matches (e.g. homologous serine kinases from different organisms) then this is also inputted as this may provide a clue to the gene's function. Match parameters (bit score, query cover, e-value) are also inputted along with information on whether or not the

match originates from the taxon Triticeae. This information provides a gauge of how good the match is: using a system of quality-control taken from AHRD (Hallab *et al.*, 2017), the best matches will have a bit score > 50, query cover > 60%, e-value < e-10 and a hit to a Triticeae species. This is coded in the database as ****, with a * coding representing that a particular condition had been fulfilled and – coding that it has not, just like with the AHRD system.

For rye contigs:
Different methods/more BLAST-heavy approaches are used for characterising the functions of rye contigs as functional information on rye genes is not as freely available. First the nucleotide sequence of the contig is isolated from the reference sequence using a custom Python script (see appendix). This is used as the input for blastx. BLAST 2.10.1 is used and set to search through the 'non-redundant protein sequences/nr' database, with the search limited to the grass taxon Poaceae (taxid: 4479). All other parameters are set to default. The same protocol for recording the outputs of BLAST runs as above for the wheat genes is used here. Another repository of functional information is also searched, in this case gene annotations from the paper that supplied the reference genome (Bauer *et al.*, 2017). Any function found here is inputted into the dataset.

Once all of the information from above is compiled into the 'Functional Information' dataset (Supplementary Material, S2), there will be various putative gene annotations allocated to each wheat gene or rye contig as a result of the various searches. For each DEG the most informative gene annotation is chosen and used as the input for downstream analysis, with priority for annotations from the gene models provided by the IWGSC and *Bauer et al*., 2017 (as these involved more sophisticated search steps). This list of 'prime' gene annotations is then inputted into Gene Ontology's gene enrichment analysis tool (2020-08-10 release, database 2020-07-16 release). These lists are found in the Supplementary Material (S3). In this list raw gene names are used e.g. peroxidase or putative gene names e.g. BHLH domain-containing protein. The GO analysis tool is set to search for changes to 'Biological Process' and 'Molecular Function' with the organism set to '*Triticum aestivum*'. All other parameters are set to default. The full outputs of the enrichment analyses are found in the Supplementary Material (S4).

Custom Files

Reference Genome

A custom reference genome was required to positively identify rye reads (from the 1RS/1BL introgression) and to avoid rye reads from erroneously mapping to wheat regions due to homology. This involved appending a draft genome of rye (Bauer *et al.*, 2017) onto the IWGSC reference assembly for wheat (IWGSC *et al.*, 2018). First the rye genome was stripped of contig names (which otherwise interrupted the sequence) using a custom Python script (see appendix). Then it was appended to the end of the wheat genome using the *cat* command.

GFF

A custom rye GFF was similarly required to identify all gene regions (both wheat and rye) specified by the custom reference genome. The custom GFF involved the creation of a preliminary rye GFF which was appended to the wheat GFF supplied by IWGSC (IWGSC *et al.*, 2018). No GFF is readily available for rye so this preliminary GFF was created using the contigs available in the rye reference genome as putative gene regions which may or may not be representative of the actual rye gene space. These contigs cover the entirety of the reference genome and are non-overlapping so are suitable for this use. The preliminary rye GFF was created by using a custom Python script (see appendix). This script ran through the reference genome and outputted a text file listing each contig, their base start position and base end position. Various awk and sed commands were then used to fill in the other columns of information with default values to correctly modify it into the GFF format. The rye GFF was then appended to the IWGSC GFF for wheat (IWGSC, 2018) using the *cat* command. The overall GFF contains 1,686,907 gene regions: 105,200 validated regions from the IWGSC region for wheat plus 1,581,707 putative gene regions created from the contig sequences of the rye reference for the rye GFF.

2.2 Section 2: MinION Data

This section details the work that was completed regarding the MinION data collection side of the project, as well as work that was intended to be done but was unable to be completed due to the coronavirus pandemic.

Study Design Summary

Obtaining and analysing the MinION data involved roughly the same workflow as the Illumina data. The main difference was that for the MinION data, it was intended that plants would be grown, RNA would be extracted and then sequenced using Oxford Nanopore's MinION sequencing platform, in comparison to Illumina data which would be extracted from online data repositories. The downstream data analysis would largely have been the same as detailed in Section 1, aside from the use of a different read mapper.

Growth of Plants

Originally five introgression varieties, five non-introgression varieties and two rye varieties were grown in triplicate (36 seeds sowed in total). If varieties did not germinate they were removed, soaked in gibberellic acid to promote growth and replanted. Some varieties still did not grow (represented in strikethrough in Table 2) so a further two introgression and three non-introgression varieties were grown approximately three weeks after (bracketed varieties in the table). Plants were grown in small containers with peat-based compost (Levington F2) in a growth chamber (GroDome) under constant temperature (18°C) and constant photoperiod conditions (16 hours, 5am – 9pm). Plants were allowed to grow for at least two weeks (to surpass the seedling stage). After this point a small portion of the youngest leaf from each of the plants was cut and stored in a centrifuge tube. Each sample was immediately put into liquid nitrogen to prevent degradation by RNase and later transferred to a -80°C freezer for indefinite storage.

Table 2
Grown varieties

Table shows the names of the hexaploid wheat varieties grown, subdivided into those that putatively do not contain the introgression (Non-Introgression) and those that do (Introgression). The names of rye varieties grown are also shown. Varieties which did not germinate originally are written in strikethrough and varieties that were grown later are in brackets. Varieties in which leaf tissue was obtained are bolded. The remaining varieties did not germinate and so leaf tissue was not obtained.

Non-Introgression	Introgression	Rye
Chinese Spring	**Brompton**	**578092**
Skyfall	**Gatsby**	**Blanco**
Revelation	**Humber**	
Viscount	**KWS Kielder**	
~~Crusoe~~	~~Lynx~~	
(Capelle-Desprez)	(Opata)	
(Recital)	(Relay)	

Reasoning for Choosing Varieties

The introgression varieties were chosen as they definitely have an introgression in the 1B chromosome based on the lab's genotyping array data, though require proper validation on whether or not this introgression is the 1RS translocation specifically. These varieties are also not already represented in the Illumina data so they increase the number of varieties available for analysis in the study.

The non-introgression varieties were chosen as they do not have an introgression in the 1BL region (based on information from rye-gene-map.de). These are also varieties that are widely grown and are therefore of particular interest to wheat breeders. In addition four of these varieties (Skyfall, Revelation, Crusoe and Viscount) are part of the AHDB's Recommended List for Cereals for 2020/2021 and so are especially relevant for wheat breeding (Hallab *et al.*, 2017). Chinese Spring is also a very important variety and is often taken as the standard for experimental studies on wheat; it is the variety sequenced for the 2012 wheat draft assembly (Brenchley *et al.*, 2012) and the 2018 IWGSC reference assembly (IWGSC *et al.*, 2018). Another variety known as Costello was grown (not shown in Table 1) and it is unknown whether or not it contains the 1RS/1BL introgression. However as seed was available for it and it is another AHDB recommended variety it would be a useful addition to the study.

From this point the laboratory work that would have been done but was unable to be completed due to the coronavirus pandemic will be discussed.

Extraction of Material

Total RNA would have been extracted from each of the samples. First the plant material (leaf cuttings) would have been placed inside an Eppendorf tube with two silver ball bearings each and grinded to break open cell walls using a GenoGrinder. The Qiagen RNeasy Plant Mini Kit would have been used to obtain total RNA from this material, following instructions. The RNA would then have been purified via polyA+ selection to obtain only mRNA using the Qiagen Oligotex mRNA Mini Kit, again following instructions.

MinION Sequencing

MinION sequencing would have been performed for each of the samples. Firstly library preparation would have been carried out, adding leading and trailing adapter sequences to the RNA molecules before sequencing. Twelve samples (reflecting the 12 varieties) would have been multiplexed over two MinION flow cells, ultimately producing 12 GB of data. This would have resulted in about 1 GB of data for each variety with an average cover of 5x for each gene (under an assumption of equal expression).

Data Preparation

The resulting FAST5 files would have been converted to FASTQ files using a tool such as poretools. The files, if not already, would have been compiled into one file per variety ready for downstream analysis.

MinION Pipeline

This pipeline would largely have been the same as the previous Illumina pipeline, except with the use of an alternative read mapper to cope with the higher error rates of MinION sequencing. Several programs could have been used to do this, for example GMAP (Križanovic *et al.*, 2018) or LAST (Seki *et al.*, 2019), both of which should be appropriate for RNAseq studies. The remaining analysis steps would have been the same as with the Illumina data (htseq-count, DESeq2, GO analysis) to enable easier comparison of the two methods. The two streams of analysis (Illumina vs MinION) would have been treated as separate and analysed as such.

3 Results

3.1 Clustering of Data

To study the 1RS/1BL introgression publicly-available RNAseq data reads were compiled, representing a total of 38 different hexaploid wheat (*Triticum aestivum*) varieties from various sources (Table 1). The collected RNAseq data was put through a custom bioinformatics pipeline which involved the following steps: pre-processing, alignment and counting. Data was split into two conditions: 'introgression' and 'non-introgression', depending on whether a variety contained the 1RS/1BL translocation or not. Given that this introgression is the main difference between the conditions, comparisons between the two should yield information on how it is expressed. To categorise varieties into their correct conditions, two sources (Schlegel and Korzun, 2020; Winfield *et al.*, 2016) were referenced. From this, seventeen varieties were shown to putatively contain the 1RS/1BL introgression, with the other twenty one not containing it. Data from these two conditions were compared via differential expression analysis and gene enrichment analysis which would inform how genes in the introgression condition (like the 1RS/1BL genes) behave in comparison to the non-introgression condition as a control. The compiled data contains metadata such as tissue type that may explain variation between different expression profiles and so Principal Component Analysis (PCA) was used to gauge the level with which different explanatory variables contributed to variation. Any clusters shown in the PCA should reflect groups of biological interest with data points within the same cluster sharing similar properties. Data points should therefore be sampled from the same clusters when possible when wanting to remove extraneous variation. In this case it is expected that clusters would form for variables such as introgression condition and/or tissue type – both of which are likely to have large effects on expression – and as tissue type is not of interest, sampling should take place within the same tissue types if clusters are evident. To create the PCA raw count data is first transformed using DESeq2's internal normalisation process. This calculates a geometric mean for each gene and uses this to calculate a size factor individual to each sample which each raw count is then divided by. These normalised counts are then passed through a variance-stabilising transformation which ensures that data is homoscedastic for visualising – in other words the variance is kept the same across different values of the mean (Love, Anders and Huber, 2020). This is important as RNAseq data consists of counts and so is naturally heteroscedastic. After normalisation and transformation, the total set of varieties were plotted as individual data points on the two axes representing the two principal components and were annotated by introgression condition and by tissue type (Figure 1A). In Figure 1B the points representing leaf data cluster together. As most of the variation is on the X axis (75% > 6%) and as this cluster is separate from the other points in terms of horizontal distance one can be confident that this cluster is sufficiently distinct to those of other tissues. The other tissues themselves do not form distinct clusters. There are therefore two evident clusters here: leaf tissue (encircled in blue) and the other tissues (encircled in pink). When looking at the whole original collection of data (all 38 varieties) there is no visible clustering by introgression condition (Figure 1A). However when data is filtered just for leaf tissue and another PCA is ran (Figure 1C) this clustering pattern is evident: introgression varieties cluster on one corner and non-introgressions cluster on the opposite corner. Note that there are some anomalous varieties which join the cluster of the opposite condition: Alba, Sevin, Chuanmai 25 and Bacanora. These varieties were further investigated (see next section). Given the overall results of the PCA: leaf data clusters together, leaf data shows distinct introgression vs non-introgression groups, as well as the fact that the majority of collected data comes from leaf tissue, suggests that the best course for analysis is to analyse only the leaf tissue. Whilst applying the data to the entire dataset may yield informative results, the effect of different

tissue types on gene expression is likely to confound any findings on the differences between introgression and non-introgression-containing varieties, and so a leaf-only analysis should give the most accurate results on these expression differences.

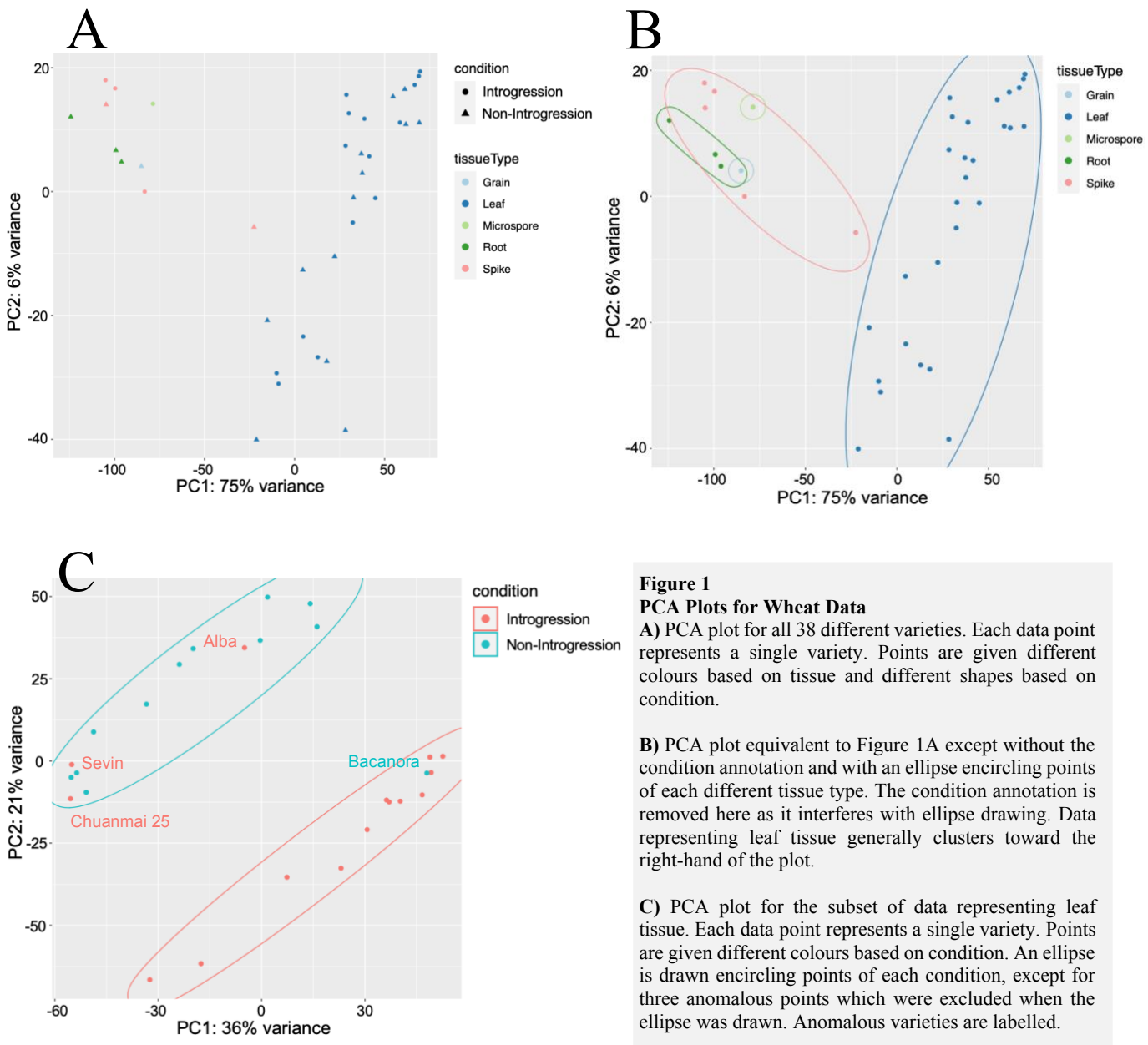


Figure 1
PCA Plots for Wheat Data
A) PCA plot for all 38 different varieties. Each data point represents a single variety. Points are given different colours based on tissue and different shapes based on condition.

B) PCA plot equivalent to Figure 1A except without the condition annotation and with an ellipse encircling points of each different tissue type. The condition annotation is removed here as it interferes with ellipse drawing. Data representing leaf tissue generally clusters toward the right-hand of the plot.

C) PCA plot for the subset of data representing leaf tissue. Each data point represents a single variety. Points are given different colours based on condition. An ellipse is drawn encircling points of each condition, except for three anomalous points which were excluded when the ellipse was drawn. Anomalous varieties are labelled.

3.2 Exploration of Anomalous Varieties

The anomalous varieties that clustered in a contradictory fashion were explored further to see if they should be placed in a different condition. If kept in these conditions then they may interfere with the results of differential expression analysis so it is essential that they are correctly classified. Here gene expression for the 1B chromosome was investigated, where an average 'reference' line was calculated for introgression varieties and non-introgression varieties (calculated without information from the anomalies). The varieties were then plotted separately to see which condition's expression profile they matched more closely to. Consistent with the PCA plots, the anomalies do clearly match the expression of varieties from the opposing condition (Figure 2). The graphs show that Alba, Chuanmai 25 and Sevin express genes in the 1BS arm which would not be the case if they contained the introgression. Inversely, Bacanora does not express genes in this region which suggests that the 1BS arm has been removed, consistent with an introgression. Based on these results these anomalies were then placed in their opposing, putatively-correct conditions (Table 3). Other studies contradict these results but this may be a case of different pedigrees being used in different studies, where some pedigrees have the introgression and some do not, but they all share the same name.

Table 3
List of Varieties Used
A list of all of the varieties represented in the collected RNAseq library, split and colour-coded into two different conditions: introgression and non-introgression, based on whether or not they contain the 1RS/1BL introgression. Varieties representing leaf tissue (and were included in the final analysis) are in black, the rest are greyed out. Varieties that originally grouped in a different condition (anomalies) are coloured in the opposite colour to show that they initially belonged to the opposing condition.

Introgression Varieties		Non-Introgression Varieties	
Aimengniu	Lumai 15	Alba	Jimai 19
Bacanora	Nautica	Atlas 66	Obelisk
Banks	Palur	Avalon	Raj 3765
Beaver	PBW 343	Borenos	Saratovskaya 29
Bobwhite	ProIntaFederal	C 306	Sevin
Equinox	Rialto	Capo	Stoa
Florida	Savannah	Chinese Spring	Sumai 3
Haven	Svilena	Chuanmai 25	Triumph
Lovrin 10		Fortuna	Yecora Rojo
		Holdfast	Yumai 18
		Jagger	

Note that in Figure 2 counts are used which are proxies for gene expression – counts are proportional but not equivalent to true gene expression. These counts are normalised using DESeq2's aforementioned internal normalisation process. This method accounts for differences in library size and RNA composition, allowing the same gene across different samples to be compared between one another. However it does not account for gene length and so some genes may show greater expression due to them having longer coding sequences. (Love, Huber and Anders, 2014). This prevents comparison between different genes. This is an imperfect way of visualising but this matters less in this case as the main point of this graph is to visualise expression differences between varieties (or condition) and not between genes. Also note that the variance-stabilising transformation was not used like before: as the count values for the condition 'reference' lines are a mean average across each variety in a condition, this reduces the effect of heteroscedasticity where higher means have higher variance (Love, Huber and Anders, 2014). In summary the normalisation allows data from different varieties (and by extension different conditions) to be compared between one another.

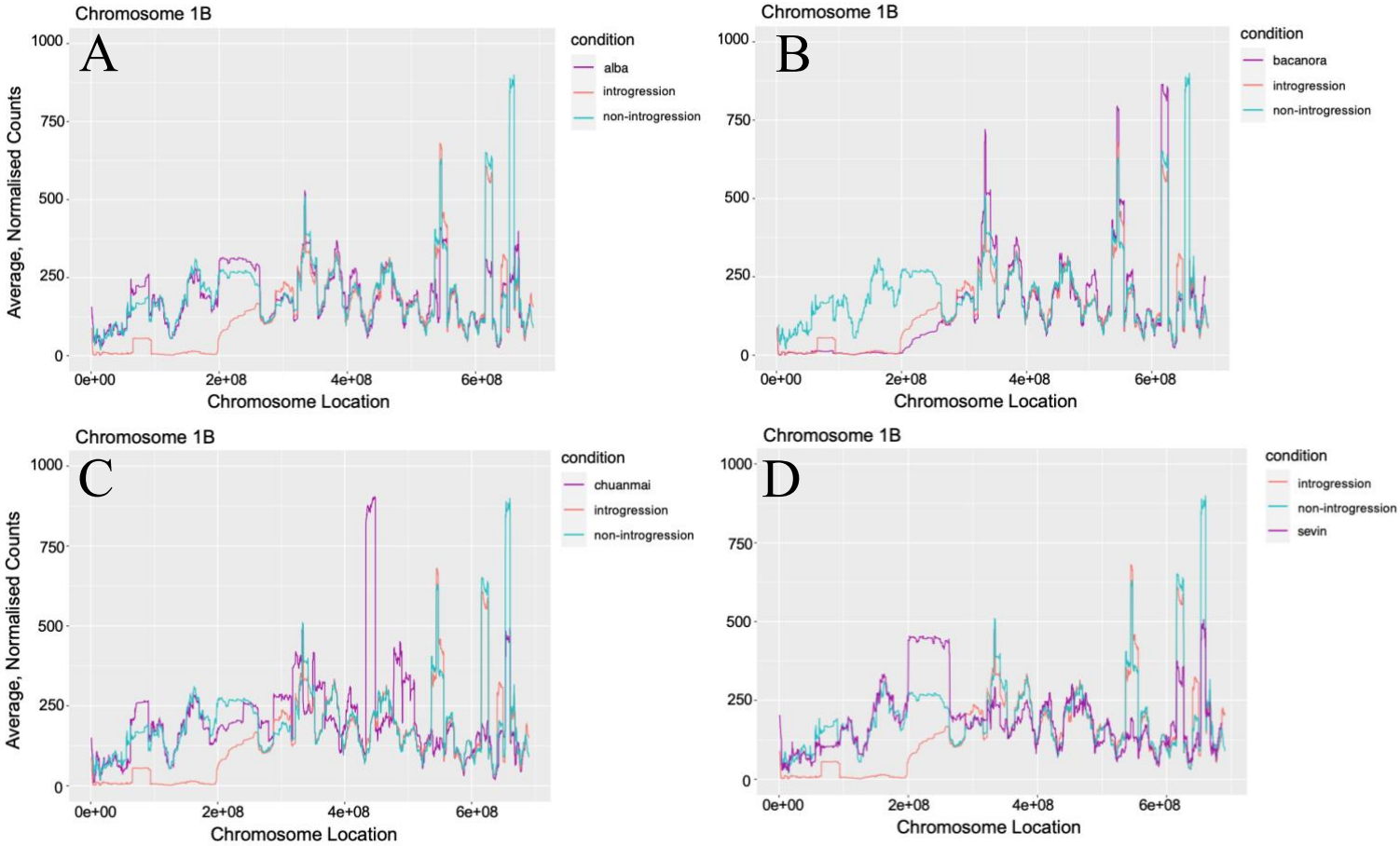


Figure 2
Expression Graphs for Anomalous Varieties
A) Rolling average graph showing gene expression for chromosome 1B for the two conditions and for the variety Alba. Count data is used which is first normalised and averaged across all varieties for a particular condition. A rolling average of 100 genes is then calculated using this normalised count data. This is plotted along chromosome location. The averages for the two conditions are calculated without the values for the anomalous varieties. Lines are colour-coded by condition/variety.
B) Equivalent graph for Bacanora. **C)** Equivalent graph for Chuanmai 25. **D)** Equivalent graph for Sevin.

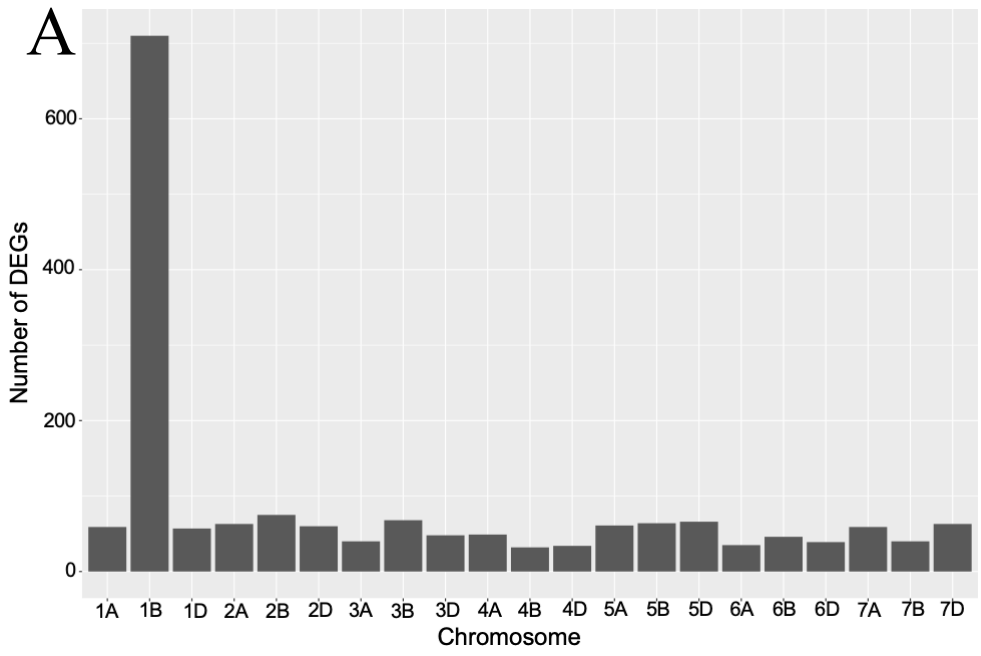
3.3 Distribution of DEGs Across Chromosomes

Given the results of the PCA, the total set of data was filtered to contain only accessions generated from leaf tissue, resulting in 28 accessions that would be analysed (13 introgression vs 15 non-introgression) (Table 4). The pipeline was applied to each of these accessions from which count data and a list of differentially expressed genes was obtained. The former can be used as a proxy for gene expression, like previously shown, and the latter can be used to identify functional differences between the two conditions. Both of these can make up the basis for further analysis and/or visualisation downstream.

Table 4
Mapping Statistics
Table showing the varieties that were analysed with their respective data types and data quantities. For each variety the total number of reads (for single-end data) and read-pairs (for paired-end data) are shown, as well as the number available after mapping, number mapped and unmapped and percentage mapped. Introgression-containing varieties are shaded in orange and non-introgression varieties in blue. Varieties that originally grouped in a different condition (anomalies) are coloured in the opposite colour to show that they initially belonged to the opposing condition.

Variety	Data Type	Total	Mappable	Mapped	Unmapped	% Mapped
Bacanora	Single	27034735	26040321	22748951	3291370	87.4
Beaver	Single	24909118	23475128	22055511	1419617	94.0
Equinox	Single	26583664	24652073	22357749	2294324	90.7
Florida	Single	33975792	32074550	30314156	1760394	94.5
Haven	Single	19003144	17075084	16017784	1057300	93.8
Nautica	Single	35369544	33829297	31836307	1992990	94.1
Palur	Single	42341554	40318624	37793250	2525374	93.7
ProIntaFederal	Single	359368239	359366553	337150544	22216009	93.8
Rialto	Single	22562564	20384803	19081781	1303022	93.6
Savannah	Single	23570762	19344177	14667902	4676275	75.8
Bobwhite	Paired	362778518	325728235	308331782	17396453	94.7
Lumai15	Paired	111328604	109005138	105848496	3156642	97.1
PBW343	Paired	31358314	31039550	29728583	1310967	95.8
Alba	Single	69560169	64102245	54844112	9258133	85.6
Atlas66	Single	49565258	49088422	45672982	3415440	93.0
Avalon	Single	25476815	24862425	22827975	2034450	91.8
Borenos	Single	19489316	17398453	16255469	1142984	93.4
Capo	Single	23049962	19697725	18122717	1575008	92.0
Chinese Spring	Single	101134481	94578490	85960241	8618249	90.1
Fortuna	Single	69746697	69480439	67077602	2402837	96.5
Obelisk	Single	434967728	419870228	288947084	130923144	68.8
Saratovskaya 29	Single	23365771	22638224	16619722	6018502	73.4
Chuanmai 25	Paired	82706752	82703274	79035311	3667963	95.6
Jagger	Paired	49814619	49814502	47456045	2358457	95.3
Sevin	Paired	49619979	49616691	44294279	5322412	89.3
Stoa	Paired	24061955	23774951	23057073	717878	97.0
Sumai 3	Paired	22992469	22790766	22277112	513654	97.7
Yecora Rojo	Paired	169827291	168146892	156749169	11397723	93.2

After running the pipeline, DESeq2 outputs a list of differentially expressed genes (DEGs) between the varieties containing the introgression and those that do not. These genes are those that are significantly upregulated or downregulated in the introgression varieties, in comparison to the non-introgression varieties. The results of my analysis yielded 4300 significant ($p < 0.05$) differentially expressed genes. By looking at the number of DEGs attributed to each chromosome a general, low-resolution measure of the differences in expression between the two conditions can be obtained and one can see which chromosomes have had their expression most perturbed by the introduction of the introgression. Given that the introgression involves the translocation of a substantial portion of the 1B chromosome, there should at least be a large-scale change in 1B expression that should be detectable using this method. Out of the wheat chromosomes, 1B had the greatest change in expression in terms of the number of DEGs attributed to it which is as expected (Figure 3A). 710 of the 4300 DEGs are found in the 1B chromosome, in comparison to the other chromosomes (each of which have 32-75 DEGs attributed to them). In addition the majority of genes that are differentially expressed are actually rye in origin (2532 genes) as opposed to wheat in origin (1768 genes) (Figure 3B) – these rye genes can be considered to arise from the 1RS/1BL introgression.



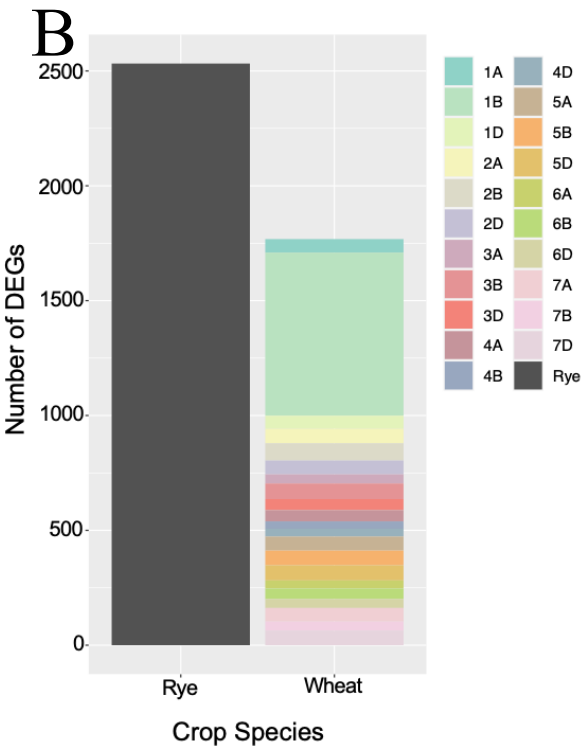


Figure 3
DEG Bar Charts
A) Shows the number of DEGs from each wheat chromosome.

B) Shows the number of DEGs that represent wheat genes or rye genes (putatively from the 1RS/1BL introgression). The bar for wheat is colour-coded with each segment representing the number of DEGs from each wheat chromosome. The same data from A was used to generate B.

3.4 Gene Expression Along the Chromosome

To further investigate expression and to look deeper at the changes that occurred to chromosome 1B of wheat, changes in gene expression along chromosomal length was studied. This should provide higher resolution data on how gene expression differs between the two conditions. The same methods for generating Figure 2 was used here. Regarding the 1B chromosome, introgression and non-introgression varieties generally show the same pattern of expression along the chromosome, except for the proximal part (Figure 4A) – this approximately corresponds to location 0 to ~2.6e-08. Here introgression varieties show a noticeable dip in expression that is consistent along this stretch of chromosome. Using chromosomes 1A (Figure 4B) and 2B (Figure 4C) as points of comparison it is evident that generally introgression and non-introgression varieties have equivalent gene expression, and that this dip in 1B expression is the most noticeable change in all three chromosomes. Such evident changes in gene expression is consistent with the previous finding, where chromosome 1B showed the greatest perturbations in gene expression, as a result of introgression. By cross-referencing one of the genes present on this region this proximal part of the chromosome can be parsed to be 1BS. This was done by looking for the presence of the gene Gli-B1: this is known to be found on the 1B short arm (IWGSC, 2018) and corresponds to the gene ID TraesCS1B02G010600. This gene is found in this proximal region of the 1B reference sequence which shows the dip in expression, suggesting that this region corresponds to the 1BS arm. This is consistent with the 1RS/1BL introgression as it is the 1RS arm that is excised (Graybosch *et al.*, 2019).

This approach to gene expression was repeated for the rye chromosomes to see if introgression varieties were upregulated for rye genes. For the rye genes physical location information was available but plotting this resulted in a plot that was difficult to interpret due to many genes occupying the same physical location 'bins'. In addition the contigs of the reference are not organised into a sequence that represents their actual location in the genome, as it is a draft sequence and so the same x-axis cannot be used as in the previous plot. Instead arbitrary units are used to more easily visualise this plot – this arbitrary unit corresponds to the order of numbering of the rye contigs e.g. the first contig in the list is arbitrary unit 1. This system simply allows the same graph to be produced in the absence of location information, and allows each gene to be plotted uniformly on the x-axis. For chromosome 1R introgression varieties generally show higher gene expression than their non-introgression counterparts (Figure 4D). However this difference is not as pronounced as in the 1B result. It is also not consistent across all genes (for some genes non-introgression varieties have slightly higher expression) and without location information it is not possible to see if genes from a whole stretch of chromosome (e.g. the 1RS arm) are all consistently upregulated, like in the 1B example. Note that here the y-axis (count values acting as expression proxy) is on a smaller scale than in the wheat genes, but this is partially due to many rye genes having 0 expression which likely weighs down the rolling average that is plotted. For comparison, the range of average count values seen for the 1B genes is 0 to 112,265 and 22% of genes have 0 expression. For 1R the range of count values is 0 to 1334 and 85% of genes have 0 expression. Rye genes are therefore expressed to a much lower level but not to as low a level as the axes on the graphs suggest. Chromosomes 2R and 3R also show similar patterns of expression to 1R, characterised by a slight upregulation in introgression varieties with some larger peaks (Figure 4E, 4F). This supports the idea that rye genes are expressed in the wheat genome but this is unexpected: 1R genes should be present in the introgression varieties, and all other rye chromosomes should have 0 or much lower expression in comparison to chromosome 1R.

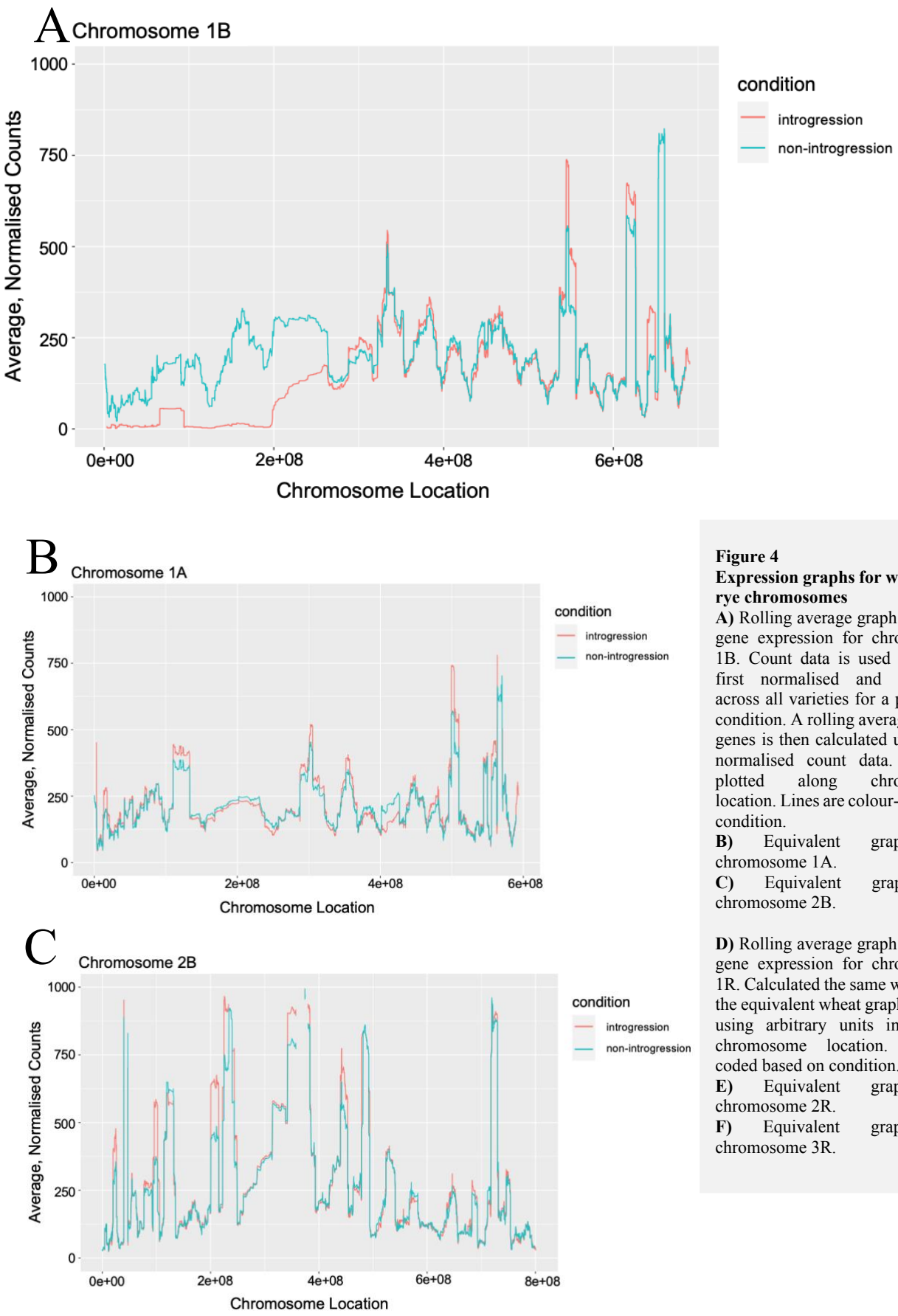


Figure 4
Expression graphs for wheat and rye chromosomes

A) Rolling average graph showing gene expression for chromosome 1B. Count data is used which is first normalised and averaged across all varieties for a particular condition. A rolling average of 100 genes is then calculated using this normalised count data. This is plotted along chromosome location. Lines are colour-coded by condition.

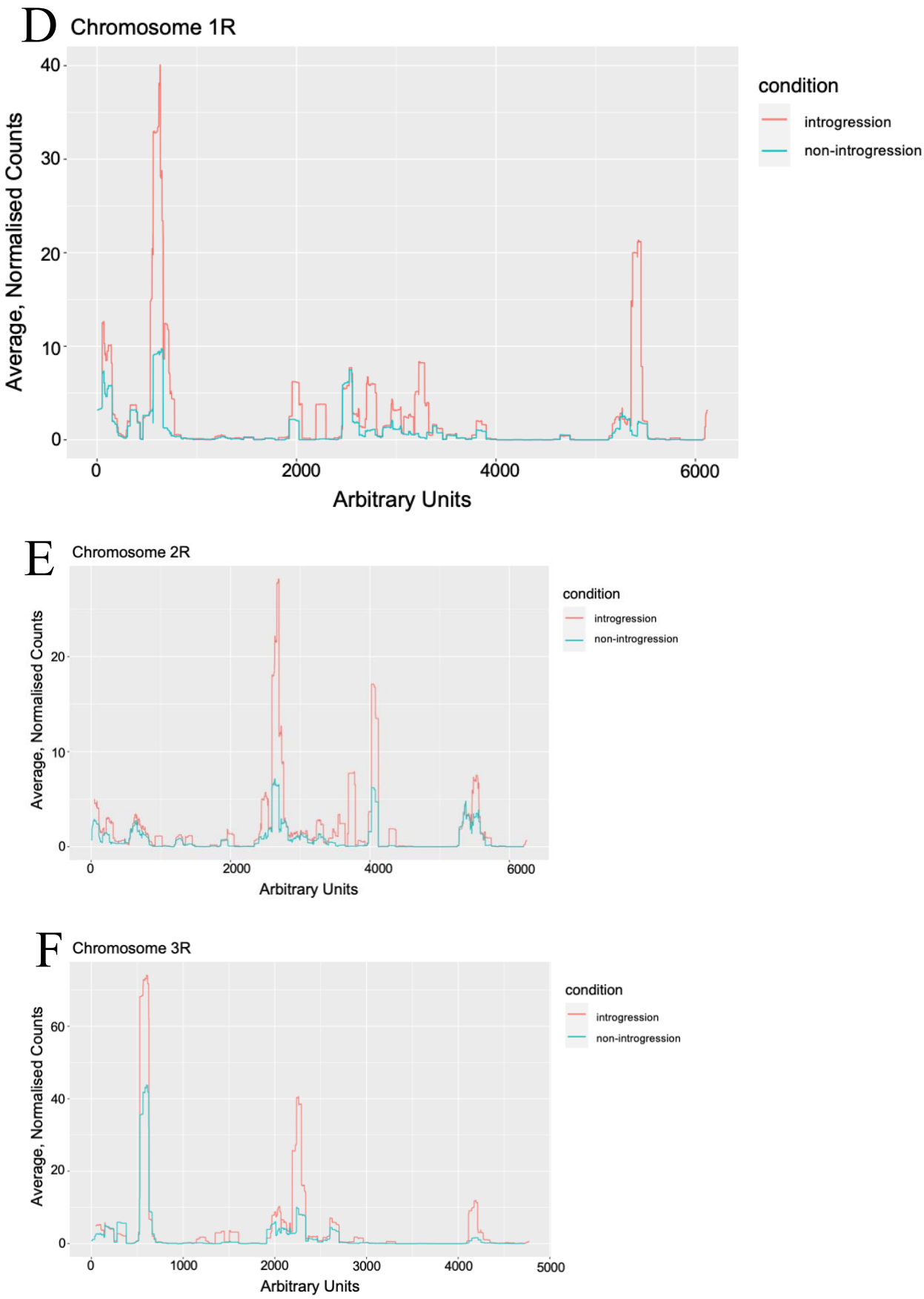
B) Equivalent graph for chromosome 1A.

C) Equivalent graph for chromosome 2B.

D) Rolling average graph showing gene expression for chromosome 1R. Calculated the same way as for the equivalent wheat graphs except using arbitrary units instead of chromosome location. Colour-coded based on condition.

E) Equivalent graph for chromosome 2R.

F) Equivalent graph for chromosome 3R.


D
Chromosome 1R
Average, Normalised Counts
40
30
20
10
0
0
2000
4000
6000
Arbitrary Units
condition
introgression
non-introgression
E
Chromosome 2R
Average, Normalised Counts
20
10
0
0
2000
4000
6000
Arbitrary Units
condition
introgression
non-introgression
F
Chromosome 3R
Average, Normalised Counts
60
40
20
0
0
1000
2000
3000
4000
5000
Arbitrary Units
condition
introgression
non-introgression

3.5 Top Differentially Expressed Genes

After obtaining a list of differentially expressed genes it is important to isolate which of these are likely to actually contribute to functional changes between the two conditions. For example, those with the largest fold-changes and/or the lowest p-values are likely to be the best indicators of biological difference between the two conditions and should be taken under higher consideration. Figure 5A shows a heat map visualising the top 100 DEGs. Here two clusters form: one consisting of rye contigs upregulated in introgression varieties and the other consisting of wheat 1B genes downregulated in introgression varieties. There is a distinct separation of these clusters and it is clear that introgression varieties have much greater expression of rye contigs than non-introgression varieties, at the cost of 1B expression. This is consistent with a 1RS/1BL introgression. Furthermore, these changes to gene expression are consistent within conditions (i.e. all varieties of a condition generally will either upregulate or downregulate a specific gene) and varieties have more similar expression profiles to others in the same condition as them. The results of the heat map and these findings suggest that all varieties have been placed in the correct conditions, and it was correct to change the original allocations of the anomalous varieties described previously. Figure 5B shows an extension of this heat map, visualising 1000 top DEGs. The same patterns found in Figure 5A are seen here, with two main clusters: rye and wheat 1B genes (though row names are not shown), which are upregulated and downregulated respectively in the introgression varieties. Only two anomalies are seen here: TraesCS1D02G114400 and TraesCS1B02G036300, which are two wheat genes found in the rye cluster that are upregulated in introgression varieties. Overall, both heat maps provide support for the 1RS/1BL introgression and for the correct allocation of conditions, at different scales.

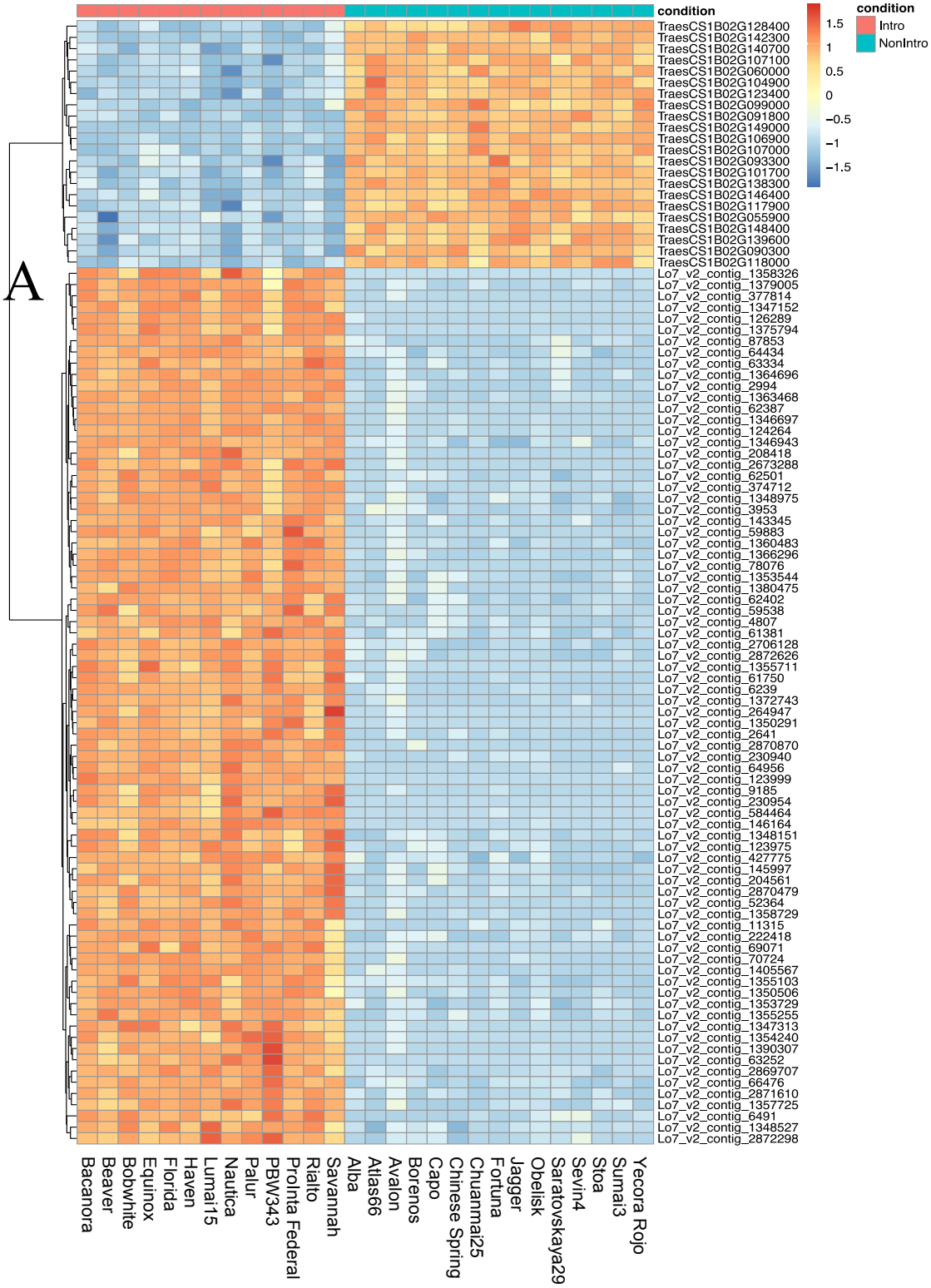


Figure 5
Heat map of 100 DEGs
A) The top 100 DEGs with the smallest p-values. Each column represents a specific variety, colour-coded by condition (introgression vs non-introgression) and each row represents a specific gene (hierarchically clustered). These genes are not ordered by p-value. Wheat genes are named with the prefix 'TraesCS…' and rye contigs are named with the prefix 'Lo7_v2_contig_...'. The syntax for wheat genes show which chromosomes they came from e.g. TraesCS1B comes from chromosome 1B. No such information is available for the rye contigs. Each cell is coloured based on their z-score showing the number of standard deviations that specific value is away from the overall mean. This z-score is calculated from count values that were normalised by DESeq2 and then transformed via a variance-stabilising transformation.

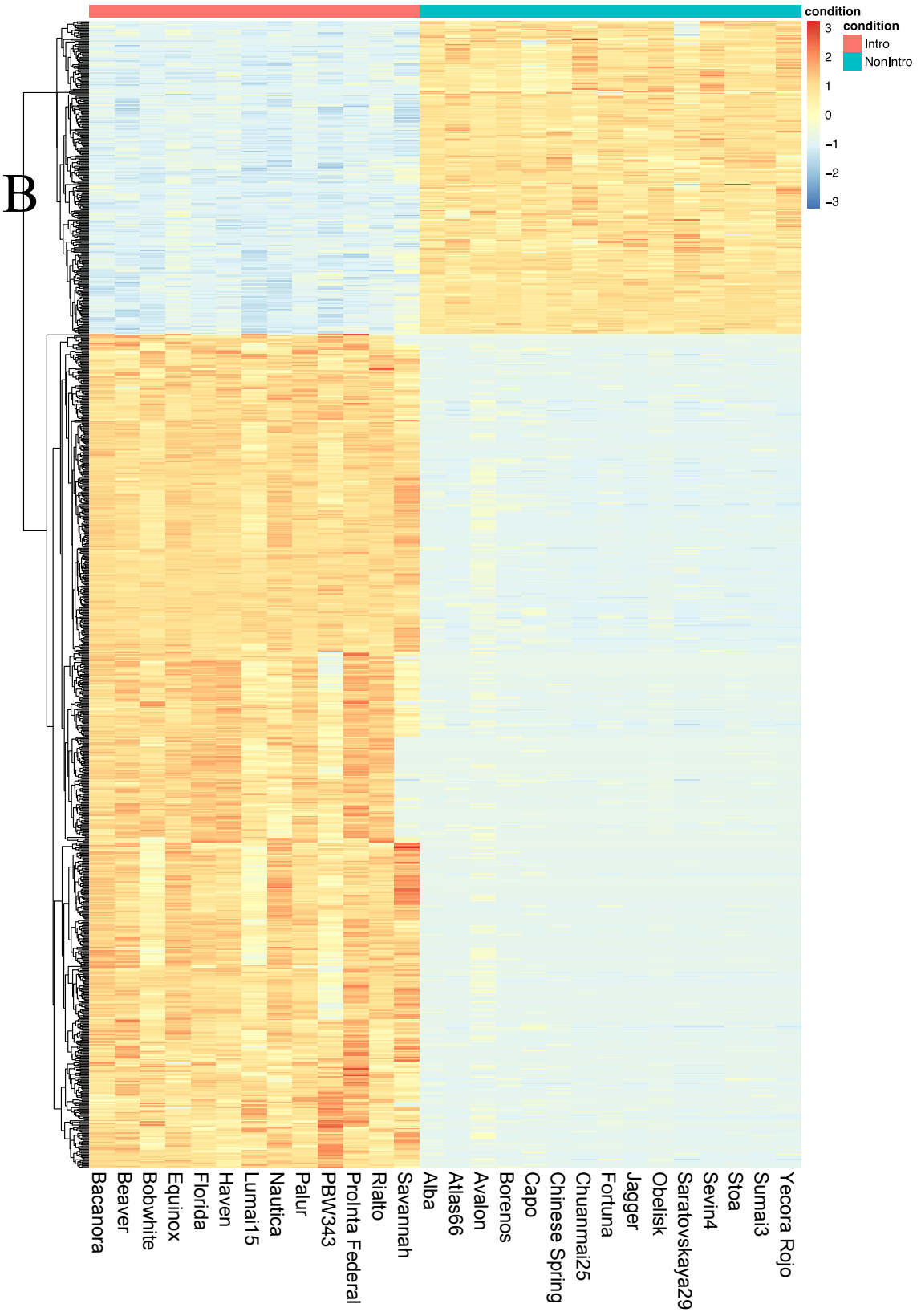


Figure 5
Heat map of 1000 DEGs

B) Equivalent graph to A but visualising 1000 DEGs. All of the genes downregulated in the introgression-containing varieties are wheat 1B genes and almost all of the upregulated genes in the introgression varieties are rye contigs – the two exceptions are TraesCS1D02G114400 and TraesCS1B02G036300.

3.6 Variety-Dependent Gene Expression of Rye Contigs

The heat map shows that there is some variation in the expression of rye contigs between different varieties. This was explored in further detail by visualising the expression of the top 10 rye DEGs between different conditions, to see if rye gene expression can be modulated by wheat genetic background like the literature says (Ren *et al.*, 2012). In theory this means that some varieties will better express the 1RS/1BL introgression than others. It is evident that there is an appreciable level of variation in gene expression for every rye contig (Figure 6). In most cases one variety will express a contig at around double the level of expression as another variety. This variation is most pronounced for contig 1350291 where ProIntaFederal shows over 5-fold greater expression for the contig than varieties Palur, Rialto and Haven. However in some contigs like 230940 such an effect is much less evident and this could be attributed to natural variation in a biological trait. As a result variation is evident in the expression of the same rye contigs with some having much greater variance than others. In addition there is no apparent consensus on a specific variety consistently showing greater or lower expression for rye sequences than other varieties. For example Savannah shows the greatest expression for contig 52364 but the lowest for contig 87853.

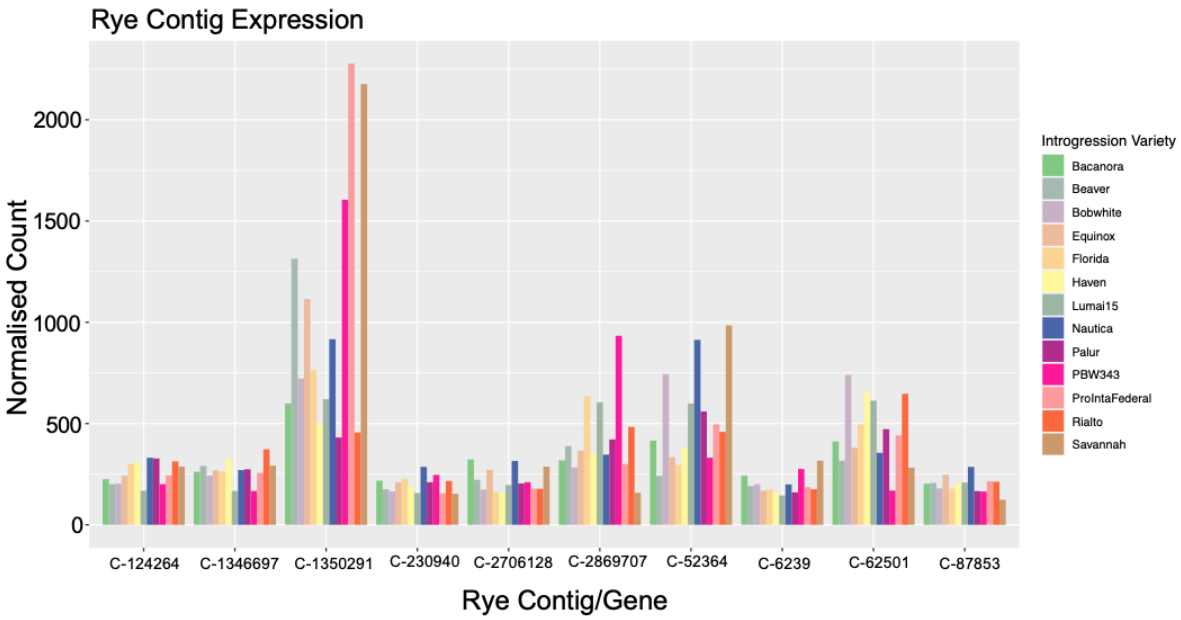


Figure 6
Variant Gene Expression Bar Charts
Bar chart showing gene expression levels of the top 10 differentially expressed rye contigs in each of the 13 different introgression-containing varieties. The bars are grouped by contig with the x-axis showing the number contig that the bars correspond to, for easier comparison, and are colour-coded by variety.

3.7 Gene Expression of 1B Homoeologs

Further exploration of the 1B genes was also undertaken, instead looking at the potential effects that their excision and downregulation may have had on other genes. Each 1B gene has homoeologs (homologs present in equivalent sub-genomes) and changes to a gene's expression may have subsequent changes on its homoeologs. Almost all of the top DEGs are from chromosome 1B and the large-scale change caused by many 1B gene losses may have been followed by upregulations in homoeologous genes to compensate. To visualise this a plot showing the expression of the top 10 1B DEGs and their respective homoeologs was generated (homoeologs were identified using the web resource wheat-exp.com). Figure 7 clearly shows that there is no appreciable response, positive nor negative, from other chromosomes when 1B genes were downregulated. After looking at the p-values generated from DESeq2, none of the small fold-changes in the homoeologs were significant whereas all of the 1B changes were highly significant.

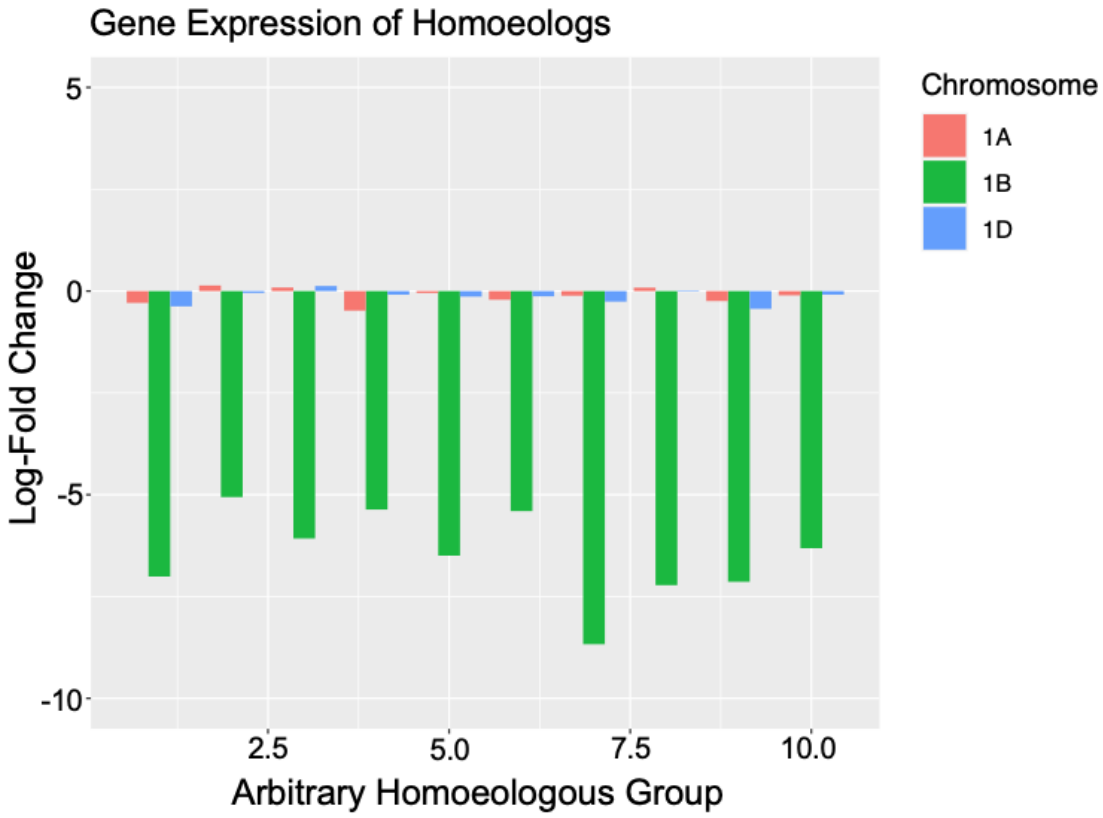


Figure 7
Homoeologous Gene Expression Bar Charts
Bar chart showing the log-fold change expression of the top 10 differentially expressed wheat 1B genes and their 1A and 1D homoeologs. The change in expression is between introgression-containing varieties compared to non-introgression varieties i.e. negative values mean that the gene is downregulated in introgression varieties. The bars are grouped by arbitrary numbers to distinguish between the different homoeologous gene groups (which otherwise have no name) and are colour-coded by chromosome.

3.8 Functional Annotation of DEGs

To determine the biological significance of the introgression i.e. to see how varieties containing the introgression and those that do not actually differ in terms of phenotype, it is necessary to functionally annotate the significant DEGs. This involves BLASTing their sequences as well as manually searching through repositories of functional information. Various methods were used to find putative functions for each of the wheat and rye genes and this information is all compiled in the Supplementary Material (S2). Here the top 100 wheat DEGs (which were all 1B and downregulated) and the top 100 rye DEGs (which were all upregulated) were studied further. An extra 31 rye DEGs were also studied – these were derived from the 1R chromosome and could be important candidate genes for understanding the effects of the 1RS/1BL introgression. All 100 wheat genes were annotated with putative functions but only 87 out of 131 genes (66%) from the rye clusters could be annotated (Table 5). This is expected as rye not being a model organism or major food crop has fewer genomic resources available for its study.

For the wheat genes 39 were allocated to various enzyme functions e.g. synthases and peroxidase. 4 genes were also allocated to 'Binding Protein' functions such as calmodulin. This functional class was not represented in the rye gene set. Many (43%) were also allocated to functions in the 'Miscellaneous' category; these genes putatively represent a wide diversity of functions such as peptide chain release factor, as well as proteins that don't have a clear function but contain specific domains such as CULLIN 2-domain containing protein. For the rye genes, a similar number of contigs were allocated to enzyme functions (33, 25%) and miscellaneous functions (28, 21%). In addition 12 (9%) of genes were putatively known to be transposon in origin, in comparison to the wheat gene set which had none. Similar numbers of transcription factors and transposons seem to be downregulated and upregulated. Overall, there are a wide variety of functions that are represented in the list of top DEGs and there does not seem to be a consensus on if genes representing any particular functional class was largely lost or acquired due to the 1RS/1BL introgression. The only exception is transposon-related sequences and an appreciable proportion of active transposons may have been introduced as a result of the introgression.

Table 5
Summarised results of functional annotation
Table shows the number of genes annotated with specific classes of functions – each function appeared multiple times during the annotation process. Wheat genes are those that are downregulated in introgression varieties and rye genes are those that are upregulated in the same varieties. Examples of functions in the 'Miscellaneous' class are shown. The full dataset showing the specific annotations to each gene is found in the Supplementary Material (S2).

	Wheat (downregulated)	Rye (upregulated)
Function Annotated		
Binding Protein	4	0
Enzyme (Metabolism)	26	16
Enzyme (Kinases)	5	9
Enzyme (Other)	8	8
Transcription Factor (or nucleotide binding protein)	10	9
Transporter	4	5
Transposon	0	12
Miscellaneous/Unknown	43	28
No Information	0	44
Total	**100**	**131**

Miscellaneous Functions
Actin-like protein
BTB/POZ and MATH domain-containing protein
Cyclin-like protein
Formin-like protein
Gamma prolamin
Histone H2B
INO80 complex subunit
Kinesin-like protein
Peptide chain release factor
Regulator of nonsense transcripts 1-like protein
Replication protein A subunit
Seed storage protein

3.9 Gene Enrichment Analysis

Gene enrichment analysis can be performed to provide greater context to the functional information obtained previously by looking at which pathways these genes may be involved in. This provides information on what biological processes may have changed as a result of the 1RS/1BL introgression and thus what phenotypic changes it may confer. The Gene Ontology's enrichment analysis tool was used to do this, using the total list of gene functions found from the annotation step to generate a list of biological processes that are either over- or under-represented in the introgression varieties. Over- or under-represented functions are those that are seen in greater or lower frequencies than expected, and represent potentially important functions that may have been changed as a result of the introgression. The list of rye DEGs upregulated and wheat DEGs downregulated in the introgression varieties were used as the input to the gene enrichment analysis, separately. Tables 6 and 7 show the summarised outputs of these two separate analyses, noting the important and interesting results, all of which happened to be over-represented. For rye over-represented functions reflect those that would have been conferred to introgression varieties as a result of the introduction of the 1RS arm (Table 6). For wheat over-represented functions reflect those that would have been lost from introgression varieties as a result of the excision of the 1BS arm (Table 7). Both results therefore represent potentially important changes that the introgression may have caused, by introduction and removal of genes respectively.

Table 6
Over-represented terms from GO enrichment analysis (rye)
Table shows a portion of the results from GO Enrichment Analysis using the set of gene functions obtained from manually annotating the upregulated rye DEGs. Only over-represented terms are shown. Over-represented 'biological processes' are shaded in darker orange and over-represented 'molecular functions' associated with these processes are shaded in lighter orange as sub-headings. P-values are calculated using the Bonferroni correction. Note that not all significantly over/under-represented processes from the analysis are included – the ones included have the greatest fold-changes and/or are of interest in explaining phenotypic changes between conditions. The full datasets showing the entirety of the results of the enrichment analyses are found in the Supplementary

Biological Process/Molecular Function	Fold Change	P-value
Ubiquitin-dependent protein catabolism	>100	6.4E-09
Ubiquitin protein ligase activity	34.91	7.2E-114
Calcium ion transmembrane transport	83.74	3.7E-55
Calcium transmembrane transporter activity	>100	8.3E-67
Calmodulin binding	25.86	2.9E-25
Histone H3 deacetylation	>100	1.5E-26
H3-K14-specific histone deacetylase activity	>100	1.9E-27
Oxylipin biosynthetic process	>100	1.4E-81
Triglyceride biosynthetic process	28.18	3.3E-04
Steroid biosynthetic process	8.18	2.0E-02

With regard to the rye analysis, some of the top over-represented processes with the greatest fold-changes and lowest p-values include ubiquitin-dependent protein catabolism and calcium transport. In addition, several processes related to lipid anabolism are over-represented, including oxylipin biosynthesis, triglyceride biosynthesis and steroid biosynthesis. These are functions that introgression-containing varieties may be better at performing and/or regulating. With regard to the wheat analysis, several top over-represented processes are involved in oxidative stress pathways: hydrogen peroxide catabolism, cellular oxidant detoxification and response to oxidative stress. Other over-represented processes are involved in signalling (ethylene and brassinosteroid signalling) and carbohydrate metabolism (malate and pyruvate metabolism). In addition, ubiquitin-dependent protein catabolism appears again here. These are functions that introgression-containing varieties may be worse at performing and/or regulating as they would have lost the 1BS arm that contains the genes with these functions. Overall, gene enrichment analysis helps clarify the results from the previous steps and identifies some important pathways that may explain phenotypic differences between varieties containing the introgression and those that do not.

Table 7
Over-represented terms from GO enrichment analysis (wheat)
Same as for A) except this table shows the results of GO Enrichment Analysis using the set of gene functions obtained from manually annotating the downregulated wheat genes. Over-represented 'biological processes' are shaded in darker blue and over-represented 'molecular functions' associated with these processes are shaded in lighter blue as sub-headings.

Biological Process/Molecular Function	Fold Change	P-value
Ubiquitin-dependent protein catabolism	89.55	3.4E-06
Ubiquitin-specific protease binding	55.97	3.5E-04
Hydrogen peroxide catabolism	80.93	0
Peroxidase activity	66.21	0
Cellular oxidant detoxification	59.50	0
Response to oxidative stress	57.17	0
Ethylene-activated signalling	39.80	2E-11
Ethylene binding	59.70	4E-13
Ethylene receptor activity	59.70	4E-13
Brassinosteroid signalling	12.57	1.4E-03
Malate metabolic process	37.70	1.4E-15
Malate dehydrogenase activity	84.28	7.7E-20
Pyruvate metabolic process	5.45	2.7E-04

4 Discussion

4.1 Expression of 1RS/1BL

My results suggest that the 1RS/1BL translocation is expressed in all of the introgression varieties sampled here as they show a downregulation of 1B gene expression and slight upregulation of many rye contigs, consistent with introgression. The introgression should therefore be active in most, if not all, of the varieties it is found in. However, it is important to note that my results are unable to provide conclusive evidence that it is the 1RS/1BL introgression that is expressed, and not another similar translocation. The 1RS/1BL introgression involves the wheat 1BS arm being substituted out for the rye 1RS arm (Lukaszewski, 2000), which leads to the expectation of an absence of 1BS gene expression and the presence of 1RS gene expression in lieu of this. The evidence for the former is strong: Figure 4A clearly shows a region of downregulated gene expression of the proximal region of the 1B chromosome of introgression varieties. This corresponds to roughly the whole short arm of 1B, suggesting that its complete removal as part of the translocation causes this dip in expression. It is possible that gene expression by introgression varieties here could be due to them being more outbred than the original introgressed population. However this would require the introgression to be broken up by recombination and it is more likely that this is the result of mis-mapping. The complementary evidence that the 1RS genes are introduced and expressed is not as strong. While Figure 4D shows that introgression varieties show greater expression for rye genes, this is on a substantially lower scale than the wheat genes. It is also not as consistent of a signal: this pattern is only evident for a few genes that show up as smaller peaks. Note that in Figure 3, most of the DEGs were derived from rye but this can occur from a small amount of transcriptional activity by rye genes in the introgression varieties, as non-introgression varieties should have a null value (or very low counts) for rye gene expression. The DEG analysis will therefore be more sensitive to differences in rye gene expression. In addition some gene expression from chromosomes 2R and 3R are also observed, though the 1RS/1BL introgression should only involve 1R, all of which have similar scales of gene expression. Two possibilities are likely to explain this: other rye regions have been translocated into wheat and/or that some contigs have been assigned to the incorrect chromosome. The results here cannot conclusively say that the 1RS arm is the main source of the rye signal in the tested wheat varieties but as 1RS is the most common form of rye chromatin in wheat we can be sufficiently confident this is so (Graybosch *et al.*, 2019). When a subset of varieties (five introgression and five non-introgression) were further investigated using the Cereal Genomics group's 35K axiom array data (Wilkinson *et al.*, 2016), introgression varieties generally had lower SNP calls for the 1B chromosome, which suggest the loss of 1B genes in introgression varieties, providing further support for the loss of a 1B segment.

It is important to note that positional information was not available to many of these rye genes: only 23% of the top 100 rye DEGs could be traced back to a specific chromosome. As a result it is only possible to visualise a small proportion of 1R genes, and the lack of a strong 1R signal could be the result of the absence of this chromosome information. Regardless it is clear that rye genes are present in the germplasm of the wheat varieties sampled here with introgression varieties exhibiting greater rye expression, suggesting that the difference between introgression and non-introgression varieties can be attributed to their rye content. Furthermore if the observed alignment to rye genes is not due to mis-mapping and is evident of true gene expression then this could highlight that rye sequences and therefore the history of outbreeding with rye may be much more prevalent in wheat that previously thought.

In further exploring and quantifying the expression of the 1RS/1BL introgression, it is clear that the expression of introgressed rye sequences is markedly lower than that of native wheat genes. In Figure 4 the scales of normalised counts are much higher for wheat genes, on average at least 20x so, than for rye genes. A similar discrepancy is observed when looking at the average count values of genes: for the 1B chromosome the mean average of normalised counts mapping to a gene is 232 whereas for the 1R chromosome it is 10 (this average excludes genes which are not expressed at all). The range for the former is 0 to 112,265 and 0 to 1334 for the latter. This suggests that rye genes are expressed in wheat varieties but on a much lower scale, which brings about the possibility for targeted gene suppression of genes from 1RS chromosomes. In plants the presence of native and foreign homologs can co-suppress each other's expression (Jorgensen, 1990) and this could be the case here with the 1RS and 1BS arms being homologous. Furthermore 85% of 1RS genes appear to show 0 expression, in comparison with 22% in 1BS and this large discrepancy could be due to rye-preferential gene silencing. Alternatively they could just represent non-coding sequence elements on the rye chromosome as the GFF used here contains putative, not confirmed, gene regions. Overall it is evident that rye genes are expressed at a low level; they are not expressed to the same extent as wheat genes but at least a small proportion remains transcriptionally active, and likely still exert effects on wheat phenotype. However the results here show a greater change in gene expression due to 1BS removal as opposed to 1RS introduction, which highlights that chromosome loss due to introgression may be just as important as chromosome introduction, if not more.

Other patterns of gene expression were also identified in this study. Firstly the expression of rye genes appear to show a degree of heterogeneity: rye genes are expressed to different extents in different varieties, with some genes showing greater variance than others. This could be the result of the introgression performing differently in different wheat backgrounds as is known in the literature (Ren *et al.*, 2012), though no particular variety seems to be better at or worse at expressing rye genes than others. This could therefore just be a result of natural variation in gene expression or could reflect the different conditions the varieties were in at the time of sampling. In addition, there does not seem to be any compensation of gene expression by homoeologs and the substitution of the 1BS arm for the 1RS arm seems to be a relatively independent event that does not perturb other sub-genomes.

4.2 Functional Characterisation of 1RS/1BL

It is clear that the 1RS/1BL introgression has impacted the transcriptional landscape of the varieties containing it and this likely resulted in a wide variety of various phenotypic changes to wheat. These changes would have been two-fold: the additive changes brought about from the introduction of genes from the 1RS arm and the reductive changes resulting from the excision of genes from the 1BS arm. The former is explained by looking at the rye genes upregulated in introgression varieties and the over-represented terms observed in this list, while the latter is explained by the wheat 1B genes downregulated in introgression varieties and the over-represented terms observed in this list. Note that the results identified here reflect the lower-end of the diversity of functions that would have been perturbed as only a total of 231 DEGs were functionally annotated, with 44 of these not having any accessible functional information. It is also important to address that the functions provided here are by no means perfectly accurate: multiple strategies were used to find functions for these genes, to compensate for the lack of functional information for wheat and rye, and so some annotations may be much better or worse than others. A detailed explanation of this is in the Methods section and the full dataset containing quality measurements of annotations is included in the Supplementary Material (S2).

Benefits of 1RS introduction

The introduction of the 1RS arm is generally accepted in the literature as having a main advantage of providing various disease resistance loci to the wheat germplasm. For example it contains resistance genes such as *Lr26* and *Yr9*. However the functions of many of these genes are still unknown and so the results here may represent a useful point for exploring the potential mechanisms with which they may work (Crespo-Herrera, Garkava-Gustavsson and Åhman, 2017). In this study, differential expression analysis identifies several DEGs with putative functions that may contribute to such disease resistance. For example rye contig 1360483 shows homology to a disease resistance protein found in *Arabidopsis thaliana*, known as RPM1. Homology matches and functional characterisation of these rye contigs (and wheat DEGs) can be found in Supplementary Material 2. This is a protein that contains an NBS (nucleotide binding site) and LRR (leucine rich repeat) domain (Boyes, Nam and Dangl, 1998). Such NBS-LRR proteins are known to be important in disease resistance, making up the largest class of known plant resistance genes (Martin, Bogdanove and Sessa, 2003; Shao *et al*., 2019). They are thought to be effective at identifying the presence of pathogens and triggering signalling pathways that result in the induction of plant defences. The LRR domain is used to detect pathogen molecules known as effector proteins (McHale *et al.*, 2006; Ng and Xavier, 2011), and activates the NBS domain which then begins the signalling cascade. Downstream plant defences eventually ensue such as the hypersensitive response (HR), which causes cell death at the site of infection, reducing further pathogen spread (van Ooijen *et al.*, 2008). In the case of RPM1, effector proteins produced by *Pseudomonas syringae* modify the protein RIN4, with this change being detected by RPM1 as an indicator of pathogen invasion (Mackey *et al.*, 2003). Its activation then ultimately triggers HR-induced cell death (Grant *et al.*, 2000). Another example of a potentially important NBS-LRR protein identified in the results is a homolog for RPP8 (contig 1358729). In *Arabidopsis thaliana*, RPP8 identifies the presence of *Hyaloperonospora arabidopsidis*, (which causes powdery mildew) and provides resistance against it (McDowell *et al.*, 1998; Mohr *et al.*, 2010), and another homolog of this protein, HRT, provides resistance to turnip crinkle virus (Cooley *et al.*, 2000). NBS-LRR proteins are therefore integral in plant immunity due to their ability to specifically recognise pathogens, with homologous variants able to recognise different pathogens with the same

level of specificity. The NBS-LRR proteins identified here thus represent possible candidates for explaining the rye-derived resistance seen in 1RS/1BL-containing varieties.

Another potential candidate could be an LRR-receptor serine/threonine kinase (contig 143345). In this case the LRR domain, useful for specific pathogen identification via protein-protein interactions (Bella *et al.*, 2008), is used alongside a kinase that triggers the signalling cascade instead of the NBS domain. However without any further information this is only speculative. Other potential candidates could be the various rye-derived kinases (9) identified in the analyses. Serine/threonine kinases can play important roles in resistance by identifying pathogens and subsequently propagating signals through protein phosphorylation (Cao *et al.*, 2011), and they represent a separate, but important, class of disease resistance proteins separate to NBS-LRR proteins (Martin, Bogdanove and Sessa, 2003). Specific examples are even seen in wheat: *Yr36* or *WKS1* uses a serine/threonine kinase and lipid-binding domain to provide resistance against stripe rust (Wang *et al.*, 2019) and *Pm21* or *Stpk-V* is a serine/threonine kinase that provides resistance against powdery mildew (Cao *et al.*, 2011). The latter is also the product of an introgression from the short arm of chromosome 6 of *Haynaldia villosa* (6VS) (Cao *et al.*, 2011). As a result it is evident that, not only can serine/threonine kinases provide disease resistance functions in wheat, but also that introgressions from other grasses can perform this same function in wheat, when introgressed. The DEGs characterised here to be possible kinases may play important roles in disease resistance, and though this is only speculative may constitute good starting points for further research.

An alternative, but not mutually exclusive, hypothesis is that the 1RS arm provides genes which do not identify specific pathogens but instead exert an effect downstream to facilitate the immune response. Little is known about the specific processes that occur downstream of pathogen detection and receptor activation (Dodds and Rathjen, 2010), representing various understudied pathways that these rye genes could be involved in to generate disease resistance. For example rye contig 61381 putatively codes for 'pathogenesis-related protein 5-like', which has been proven to be necessary for disease resistance against leaf rust mediated by the gene *Lr35* in wheat (Zhang *et al.*, 2018). This is thought to provide a glucanase function that helps break down fungal cell walls (Zhang *et al.*, 2018; Liang *et al.*, 2019). Proteins that work downstream in more of a defence function may therefore be just as integral to resistance against specific pathogens as the receptors that identify them.

One theory is that the 1RS arm's pathogen resistance could be mediated via changes to jasmonate (JA) signalling. JAs are compounds known as oxylipins that play integral roles in stress- and defence-related responses via changes to gene expression (Howe and Schilmiller, 2002). The effects of JA signalling are sufficient to cause wide and specific disease resistance (Ellis, Karafyllidis and Turner, 2002) and so is an integral part of plant immunity. The results of the GO enrichment analysis identified 'oxylipin biosynthesis' as one of the top over-represented processes (fold change > 100, p-value = 1.4E-81) observed in the list of upregulated rye genes which makes it likely that the rye genes in the introgression exert some effect on this pathway. Other oxylipins do exist and this could feed into their synthesis instead, though JA is the principal oxylipin known in plants (Farmer, Alméras and Krishnamurthy, 2003). Further evidence for the potential impact of JA signalling in rye-mediated disease resistance in wheat is that potential homologs of several core components of this pathway are observed as top differentially expressed genes from this analysis: contig 62402 shows homology for an E3 ubiquitin-ligase-like

protein, 62998 for histone deacetylase-like and 1358116 for ethylene-responsive transcription factor (ERF) ERF110-like. In *Arabidopsis* such genes work in concert to change gene expression under pathogen attack: JA-derivatives are perceived by an E3 ubiquitin-ligase like SCF^{COI}, which can tag histone deacetylase for degradation, de-repressing certain genes which can then be activated by an ERF, resulting in the expression of defence genes (Farmer, Alméras and Krishnamurthy, 2003; Devoto *et al.*, 2002; Nagels Durand, Pauwels and Goossens, 2016). GO terms related to these genes and processes were also observed to be over-represented: 'ubiquitin-dependent protein catabolism' and 'histone H3 deacetylation' were both shown to have enrichment values of >100. It is therefore possible that this rye-derived JA signalling pathway is used in a wheat system. The implication of this is that the 1RS arm may allow the use of a variant JA pathway, notably a different ERF. This could result in the expression of a different assortment of defence genes and provide resistance to a new pathogen or isolate that the wheat variety was susceptible to before, as different ERFs are known to activate and repress different JA-responsive genes with varying outcomes on resistance (Thirugnanasambantham *et al.*, 2015). If this theory is correct then this provides evidence that the introgression provides not only singular disease resistance genes but potentially entire suites of self-contained immune-functioning toolkits. Theoretically this could allow for smoother functioning of signalling pathways if all, or at least some, of the signalling components are derived from rye, especially if coherent signalling is strictly dependent on specific protein variants.

There may also be other candidate pathways for the observed rye-derived resistance observed in introgression-containing wheat varieties. One of the most over-represented processes in the 1RS arm is 'calcium ion transmembrane transport' (FC > 83.74, p = 3.7E-55). With calcium ion influxes known to be integral as an apoptosis signal (Levine *et al.*, 1996), the upregulation of genes like a 'calcium transporting ATPase' (contig 2641) could be used as a means to better facilitate programmed cell death in limiting pathogen spread. Similarly several contigs show homology to ABC transporters. These are important ATP-dependent membrane proteins that can import or export substrates (Rees, Johnson and Lewinson, 2009) and can theoretically provide pathogen resistance via efflux of anti-pathogen compounds or pathogen effectors. In wheat an important resistance gene, *Lr34*, encodes an ABC transporter that provides resistance against several fungal pathogens by exporting out metabolites that deter pathogen growth (Krattinger *et al.*, 2009). Two of the contigs identified in my results are homologous to 'ABC transporter G 37' which is overexpressed after pathogen exposure (Gräfe and Schmitt, 2020), which may make it important in the immune response. Overall various potential pathways are identified which may be able to explain the disease resistance traits conferred by the 1RS arm, which may work in identification, signalling and/or defence-related traits.

Potential disadvantages of rye

Not all of the genes introduced by the 1RS arm may be beneficial: some may be disadvantageous and some may not have any effect at all. The 1RS/1BL introgression is widely known to have negative effects on end-use quality – this is the result of the *Sec-1* locus that introduces omega secalins that worsen the bread-making properties of the grain (Li *et al.*, 2016). While the gene coding for this secalin was not identified in the select proportion of DEGs annotated other genes were characterised that may contribute to the negative end-use qualities caused by this locus. Contig 2808155 shows homology for a gamma prolamin from *Secale cereale* (secalins are a type of prolamin) and contig 1357725 shows homology for a seed storage prolamin (glutelin). These may form part of

the *Sec-1* locus as other prolamins, notably gamma prolamins, are known to be genetically linked to omega secalins (Shewry *et al.*, 1984). With their relatively high expression in the introgression-containing varieties they could also play a role in negatively impacting wheat grain quality. In addition gamma secalins can cross-react with gliadins as an allergen (Palosuo *et al.*, 2001), which may make introgression-containing varieties even worse choices for bread-making when considering their effects on those with allergies and insensitivities to gluten.

The genes present on the 1RS arm may also not be functional at all: transposons/transposable elements (TEs) and/or retroviral sequences make up the second most commonly represented functional class of the rye DEGs. This is in contrast to the 1B DEGs, none of which were putative TEs. These may have been tolerated by the wheat genome as they already contain a large proportion of TEs. In fact wheat has one of the greatest numbers of transposons of any crop species (Zhao *et al.*, 2017) and 85% of its genome is made up of TEs (IWGSC *et al.*, 2018). The presence of these transposons suggest that a sizeable portion of the introgression may compose of non-functional and even selfish elements that use up cellular resources for their own propagation. In addition these transposons may have substantial impacts on the expression of other genes in the introgression. Transposons may exist as part of a gene's coding sequence or by themselves (Fu *et al.*, 2009) and this may also be the case here as the BLAST hits for these rye contigs have query covers that range from 14% to 91%. The contigs showing lower query covers could be associated with rye genes, and this association could potentially lower the gene's expression. In a study of *Aegilops tauschii* (the progenitor species of the wheat D genome), it was shown that around half of its genes contained TEs (Zhao *et al.*, 2017). These genes showed reduced expression in comparison to their non-TE-containing counterparts, likely due to the increased methylation of TE sequences which cause downregulation of the TE and associated genes (Zhao *et al.*, 2017). The TEs in the 1RS arm could act as a self-contained mechanism for reducing rye gene expression which could explain why rye genes are expressed at a lower level than wheat genes, thus depreciating the benefits of the introgression – this is an aspect of the introgression that has yet to be discussed in the literature. This finding could also explain why wheat has so many transposons and why its genome is so large: it has been subject to many different introgressions over time and each event may have introduced many active transposons that would proliferate rapidly until silenced. Interestingly these transposons could also provide raw sequence material for adaptation, as transposons can be co-opted into proteins with novel functions (Joly-Lopez and Bureau, 2018) as well as for generating novel regulatory motifs (Feschotte, 2008). However with transposons already being extremely pervasive in the wheat genome the introduction of further transposon sequences by the introgression may be negligible, and this is unlikely to be fruitful regarding current timescales of crop improvement.

Disadvantages of 1BS excision

The removal of the 1BS arm is also likely to result in functional changes. These changes are represented as the wheat 1B genes which are significantly downregulated in introgression varieties and the over-represented terms that arise when these genes are analysed, representing genes that introgression-containing varieties no longer have access to.

From the enrichment analysis of the downregulated wheat genes, terms like 'malate metabolic process' and 'pyruvate metabolic process' and 'brassinosteroid' and 'ethylene signalling' are over-represented. This suggests

that introgression-containing varieties may suffer reduced functionality in aspects of metabolism and also signalling. However any disadvantages may be negligible as no such effects are described in experimental studies of the 1RS/1BL introgression. Instead reported disadvantages are mostly concerned with negative effects on bread-making quality (Graybosch, 2001), though the analysis of wheat genes did not identify any proteins or pathways that could contribute to this due to loss of 1B genes. However interestingly the results of this analysis identified genes downregulated in the introgression varieties that show homology to important stress-related proteins. More specifically, two putative chaperone proteins and a heat-shock protein were downregulated, as well as a putative peroxidase and metal tolerance protein. These would likely play roles in tolerance to heat, oxidative and heavy metal stress, respectively. In addition the enrichment analysis identified that in the downregulated wheat 1B genes, terms relating to 'cellular response to stress', 'cellular response to toxic substance' and 'response to oxidative stress', were over-represented, suggesting that introgression varieties are worse able to deal with these stressors. The B sub-genome of wheat is also known to have a relatively high number of genes relating to environmental adaptation due to sub-genome specialisation (Feldman *et al.*, 2012), so this is likely to be representative of the 1B chromosome as a whole. However this contradicts with the literature: the 1RS/1BL introgression is known to improve abiotic stress tolerance, which is consistent with rye being an environmentally hardy grass (Howell *et al.*, 2014). This suggests that the loss of these genes must have been matched and surpassed by an introduction of even more genes relating to stress tolerance from the 1RS arm – the introgression doesn't just provide more genes to wheat but it also compensates for certain gene losses. These stress-tolerant genes were not identified in the results of the functional analysis, but may arise if further genes were annotated.

Continuing on from this point, it appears that genes lost due to the 1BS arm may be compensated by an introduction of similar or equivalent genes within the 1RS arm, or genes involved in the same pathways. This is in contrast to 1A and 1D gene expression, which does not appear to compensate for 1B loss. This is expected as Triticeae chromosomes are homologous and syntenic with one another, though the group 1 chromosomes show the least synteny (Feuillet and Keller, 2002). For example several ligase enzymes and actin-related proteins are both lost and gained as a result of 1RS/1BL introgression. Also some shared terms are over-represented in both the analysis of the upregulated rye genes and downregulated wheat genes such as 'ubiquitin-dependent protein catabolic process'. This suggests that, in some cases, wheat genes are being swapped out with rye equivalents and the presence of having novel variants as opposed to entirely novel genes may be important in the introgression's benefits. Having slightly similar genes (with similarity resulting from homology) makes them useful but also usable in the genetic background of a new host.

4.3 Relevance to Future Wheat Breeding

My results suggest that the 1RS/1BL introgression should be expressed in most varieties it is found in, though at a very low level. However the genes are not all silenced and many still remain transcriptionally active. In conjunction with what is known in the literature it would be safe to surmise that the introgression remains useful and may still be a good choice for wheat breeding, though its effects may be dependent on the wheat genetic background it is used in. However it may have negative effects such as reduced metabolic capability, though my study is unable to make comment or quantify the relative costs and benefits of this trade-off. Also as its expression in wheat seems to be quite low it may not have as noticeable an impact on wheat phenotype as previously thought, and its importance in bolstering wheat yield should not be overstated. Though it may still be a good option for improving wheat, when choosing between cultivars to grow it may be better to consider not whether a variety contains this introgression but rather whether a variety comes from a better pedigree, and use the presence of the introgression as a tie-breaker when deciding which to use. Unfortunately more in-depth recommendations cannot be made as further research is required that can supplement the findings in this study and address its limitations.

Limitations

Though the scope of the results here are sufficiently large, in that the effects of the 1RS/1BL introgression can be discussed on several fronts, there are many unknowns that prevent high-resolution conclusions from being made as a result of several limitations. Firstly the lack of functional data for both wheat and rye makes it difficult to maximise the amount of information that can be gleaned from this study. Only 66% of the rye DEGs could be functionally annotated and the remaining unknowns represent a lot of information that cannot be analysed. Assuming they have functions, incorporation of the unannotated 34% would likely have a large effect on the results, allowing more confident conclusions to be made or changing them altogether. There are also other limitations regarding information. For example many wheat varieties did not have known introgression information and so much RNAseq data existing in data repositories could not have been used in this study, which would have improved the clarity of the results. The genomic resources available for rye also lacked sufficient positional information to provide chromosome assignments or locations for every single rye contig. As a result the chromosomal origin of many rye contigs were unknown, and so 1R gene expression could not be properly determined, especially in comparison to other rye genes, and as a consequence makes it difficult to conclude if it is the 1R chromosome that is responsible for the rye signal obtained here. Furthermore the reference genome used in this study is sub-optimal and may have obscured some information. Whilst the majority of 1RS/1BL introgressions are derived from the rye variety ‘Petkus’ (Schlegel and Korzun, 1997), the rye reference was sequenced from the variety ‘Lo7’ (Bauer *et al.*, 2017). These varieties may be different in terms of genomic composition and so this reference may not be able to fully capture the expression dynamics of the 1RS/1BL genes. Finally there are limitations to consider with using this framework of study to a meta-analysis which may mask true patterns of variation. Here many varieties are compiled together from various studies and as a result of their different origins, samples would have been sequenced under different conditions. As a result it is difficult to determine which of the changes in expression observed between conditions were actually due to biological differences between varieties, or just differences in the conditions they were sampled under. In addition the use of different varieties as ‘pseudoreplicates’ for DESeq2 mean that the results it generates are not as powerful as if true biological replicates are used. This is because the program uses estimates of dispersion for each gene to

determine differential expression and use of the latter provides more accurate estimations of this for analysis (Love, Huber and Anders, 2014).

This study could be drastically improved if it was repeated using the same structure – many different introgression varieties vs many non-introgression varieties – with the difference that all plants would be grown and sequenced under the same conditions. This would help clarify what changes were due to biology or due to external factors. In addition if phenotypic characteristics were measured it would be possible to confirm the putative functional changes identified in this study with actual changes observed in the plants.

4.4 Summary and Conclusion

Overall my results are able to reinforce current knowledge regarding the 1RS/1BL introgression as well as unveil new areas of knowledge that may represent promising areas of further research. For one there is evidence to support the expression of the 1RS/1BL introgression but this is followed by certain caveats. Firstly genes from other rye chromosomes may be expressed, not just those from 1R, which prevents confirmation that the source of rye expression is this specific chromosome, though this is likely. Secondly the expression of rye genes is much lower than native wheat genes (around 20x lower) suggesting that there may be a mechanism leading to their suppression or non-preferential transcription. This also casts doubt on how influential these rye genes actually are in improving wheat phenotype. There may be an equal or greater effect of 1BS removal given that the signal for a drop in 1B gene expression was much stronger than the signal representing an increase in 1R expression. Changes that result from chromosome loss during introgression should therefore be regarded with the same amount of importance as that of novel chromosome introduction. Thirdly expression of rye genes may be dependent on wheat genetic background, which is a finding consistent with the literature. Overall it can be said that the 1RS/1BL introgression remains transcriptionally active and its expression in a wheat background is dynamic with non-uniform effects.

Whilst the changes conferred by the introgression are well-described in the literature, the means by which these are realised are mostly unknown and the results here identify potential pathways that can explain this, unveiling promising veins of further research. Several homologs to important disease resistance proteins in *Arabidopsis* may have been donated from rye to wheat and could be important candidates for the disease resistance the introgression is known for. In addition JA signalling may play an important role in such rye-derived resistance and the use of rye-derived components for this signalling may organise the expression of a novel set of defence genes that provide disease resistance to existing wheat pathogens. The disease resistance provided by the 1RS/1BL introgression appears to be complex and multi-faceted, and the results here identify promising starting points for future research to understand it in greater detail, especially given that many of these disease resistance loci have yet to be cloned functionally.

This study also highlights some potential generalisations on introgressions which could be clarified by further research. Firstly introgressions may involve ‘homologous swaps’ where genes from a crop species are replaced with variant homologs from a donor species (1BS vs 1RS in this case), and shared homology may allow novel variants to remain functional in a new genetic background. Secondly introgressions may not just introduce independently functioning genes but also self-contained gene sets that contain several, functional components of a pathway that may allow that pathway to function optimally in a novel environment. Thirdly the effects of introgressions can be conceptualised as chromosome loss vs chromosome gain, and their respective effects on phenotype may be asymmetric like with the loss of the 1BS arm resulting in a greater transcriptional footprint than the introduction of 1RS. Finally it is important to consider the non-functional elements of introgressions such as transposons – here the identification of many transposon-derived sequences highlight the need to view introgressions as having a varied genetic landscape. Treating introgressions as mere vehicles for a few beneficial loci ignore the other consequences that it may have on its new host’s genome, such as the potential suppressant effects of transposons.

Whilst this study is not meant to be an in-depth functional analysis of the 1RS/1BL introgression it is able to touch base on several important properties of the 1RS/1BL introgression (and introgressions in general) which represent promising starting points for elaboration. Indeed further research is required to be able to fully exploit the genomic novelties available for crop breeding. However this is gated by the current availability of genomics resources where for many introgression sources (like rye) a lack of information hampers the ability to fully make use of data available. However major strides have been made in recent years which makes a study like this possible, and even greater leaps are likely once the rising necessity for food security becomes realised by the wider populace. Hopefully this study will be a starting point and a standard with which other introgressions can be studied, so that in the future the beneficial genes in the Triticeae gene pool can be catalogued and the various kinds of introgressions characterised. This could ideally allow for all usable genomic novelties for wheat to be compiled, providing a diverse and accessible library that could be used to improve wheat germplasm against the uncertainties of the future.

Appendix

This section details all of the custom scripts written and used for various points of the data analysis.

BASH SCRIPT: pipeline

This is the main pipeline used for analysing all of the reads.

```
# Works for single-end and paired-end reads on the basis that single-end reads have
Reads.fastq.gz suffix and paired-end reads have _1/_2.fastq.gz

# First create a list variable of all the variety names
# find: prints out files ending with .gz with a newline character in between each
# sed: removes Reads.fastq.gz from the file name to get the base variety names
singleList=$(find . -name \*Reads.fastq.gz -printf "%f\n" | sed
's/Reads.fastq.gz//g')
echo $singleList

pairedList=$(find . -name \*_1.fastq.gz -printf "%f\n" | sed 's/_1.fastq.gz//g')
echo $pairedList

# Perform a for loop to carry out each of the processes
# if clause: paired-end reads will have two instances of the base name and single-
end reads will just have one

# THIS IS FOR SINGLE-END READS
for i in $singleList
do
echo "NOW PROCESSING $i"
fastp -i $i'Reads.fastq.gz' -o $i'Clipped'
hisat2 -p 8 -x ~/data/referenceGenome/refIndices/fullIndex -U $i'Clipped' -S
$i'Align'
htseq-count --type=gene --idattr=gene_id --stranded=no $i'Align'
~/data/GFFs/fullChimeraGFF > $i'Count'
done

# THIS IS FOR THE PAIRED-END READS
for i in $pairedList
do
echo "NOW PROCESSING $i"
fastp -i $i'_1.fastq.gz' -I $i'_2.fastq.gz' -o $i'_1Clipped' -O $i'_2Clipped'
hisat2 -p 8 -x ~/data/referenceGenome/refIndices/fullIndex -1 $i'_1Clipped' -2
$i'_2Clipped' -S $i'Align'
htseq-count --type=gene --idattr=gene_id --stranded=no $i'Align'
~/data/GFFs/fullChimeraGFF > $i'Count'
done
```

PYTHON SCRIPT: removeHeader.py

Used for removing contig names from the rye reference genome. Contigs had to be removed to create a contiguous reference genome where contigs were not broken up by headers.

```
#!/usr/bin/python

# Program to remove headers from FASTA files.

# Store the FASTA file into object fasta and read the first line.

path = input("Please enter the absolute path of your FASTA file. \n")
fname = input("What will the new file be named? \n")
fasta = open(path)

# Create empty strings that lines of the sequence will be added to.
title = ""
sequence = []

# Store the title line and remove whitespace.
for line in fasta:
    if line.startswith(">"):
        title = line
    else:
        sequence.append(line)

# Convert the list into a string
string = "".join(sequence)

# Creates a new file and writes to it.
# This will create the file in whatever directory you are in.

f = open(fname, "w+")
f.write(string)
f.close()
```

PYTHON SCRIPT: isolateBounds.py

Used for isolating contig bounds for GFF. Output of this is the input for modifyGFF.sh

```
#!/usr/bin/python

#sys.argv[1] is the reference genome, sys.argv[2] is the file you'll write to.

import sys

genome = open(sys.argv[1])

contigCount = 0
baseCount = 0

# need three separate lists that will then be joined together later.
contigList = []
startList = []
endList = []

for line in genome:
        if line.startswith(">"):
                contigCount += 1
                splitLine = line.split()
                contigList.append(splitLine[6])

# store the start value (but don't actually modify the count as it messes up the
loop later)
                startList.append(baseCount+1)

# store the end value (but not for the first line)
                if contigCount > 1:
                        endList.append(baseCount)

# if line doesn't start with a header, iterate through the lines and count the
number of bases
        else:
                for base in line:
                        if base != "/N":
                                baseCount += 1

# append the last end list value when the for loop has finished
endList.append(baseCount)

# create a new nested list based on the other three lists
# zip merges the lists together but in the form of a zip/tuple data form - use
list(value) to convert into lists
nestedList = [list(value) for value in zip(contigList, startList, endList)]

# nested for loop to write the list to output
with open((sys.argv[2]), "w+") as f:
        for list in nestedList:
                        f.write("\n")
                        for item in list:
                                f.write("{}\t".format(item))
```

AWK/SED COMMANDS: modifyGFF.sh

Used to create a full GFF file from contig bounds. Input for this is the output from isolateBounds.py.

```
#!/bin/bash

# this has the awk (and sed commands) to create a manual GFF from the contigs and
bounds derived from the rye genome.

# first re-order the columns so it goes START; END; CONTIG
awk '{$4=$1}{print}' isolateBounds > copy2

# change the first column (duplicate contig column) to 1R
awk '{$1="1R"}{print}' copy2 > copy3

# add the second and third columns by appending it to 1R
awk '{$1=$1 FS "manual" FS "gene"}{print}' copy3 > copy4

# add the sixth, seventh and eighth columns by appending it to END
awk '{$5=$5 FS "." FS "+" FS "."}{print}' copy4 > copy5

# put gene_id in front of the contig name (ninth column)
awk '{$9="gene_id="$9}{print} copy5 > copy6

# substitute the spaces for tabs
sed 's/ /\t/g' copy6 > copy7

# then manually write in the first three lines: ##gff-version 3, etc
```

PYTHON SCRIPT: findContigSeq.py

Used to isolate the sequence of a particular contig. This was necessary when a contig sequence needed to be put through BLAST to find its putative function.

```
import sys

# sys.argv[1] is the genome, sys.argv[2] is the contig number you want to find.

genome = open(sys.argv[1])
query = sys.argv[2]
searchContig = "Lo7_v2_contig_" + str(query)

queryList = []

# define the function: iterate through the genome until you find the contig based
on whether or not the query matches up with the 6th field of the header.

def findContig():
    for line in genome:
        if line.startswith(">"):
            splitLine = line.split()

            if searchContig == splitLine[6]:
                for line in genome:
                    if line.startswith(">"):
                        return
                    else:
                        for base in line:
                            queryList.append(base)

            else:
                continue
        else:
            continue

# join method used to join together the bases from the loop to make a contiguous
sequence: the contig's sequence.

findContig()
sequence = "".join(queryList)

# print sequence to standard output
print(searchContig)
print(sequence, end="")
```

Bibliography

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E. and Gouil, Q. (2020) 'Opportunities and challenges in long-read sequencing data analysis', *Genome Biol,* 21(1), pp. 30.
Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq--a Python framework to work with high-throughput sequencing data', *Bioinformatics,* 31(2), pp. 166-9.

Anderson, G. R., Papa, D., Peng, J., Tahir, M. and Lapitan, N. L. (2003) 'Genetic mapping of Dn7, a rye gene conferring resistance to the Russian wheat aphid in wheat', *Theor Appl Genet,* 107(7), pp. 1297-303.

Bajgain, P., Russell, B. and Mohammadi, M. (2018) 'Phylogenetic analyses and in-seedling expression of ammonium and nitrate transporters in wheat', *Sci Rep,* 8(1), pp. 7082.

Ballouz, S., Dobin, A. and Gillis, J. A. (2019) 'Is it time to change the reference genome?', *Genome Biol,* 20(1), pp. 159.

Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M. M., Twardziok, S. O., Hackauf, B., Gordillo, A., Wilde, P., Schmidt, M., Korzun, V., Mayer, K. F., Schmid, K., Schön, C. C. and Scholz, U. (2017) 'Towards a whole-genome sequence for rye (Secale cereale L.)', *Plant J,* 89(5), pp. 853-869.

Bella, J., Hindle, K. L., McEwan, P. A. and Lovell, S. C. (2008) 'The leucine-rich repeat structure', *Cell Mol Life Sci,* 65(15), pp. 2307-33.

Bhoite, R., Si, P., Siddiquie, K. H. M. and Yan, G. 2020. Genes, mechanisms and novel EST-SSR markers associated with metribuzin tolerance in wheat (T*riticum aestivum L.*): targets for improving photosynthetic capacity and yield.

Borrill, P., Harrington, S. A., Simmonds, J. and Uauy, C. (2019) 'Identification of Transcription Factors Regulating Senescence in Wheat through Gene Regulatory Network Modelling', *Plant Physiol,* 180(3), pp. 1740-1755.

Borrill, P., Ramirez-Gonzalez, R. and Uauy, C. (2016) 'expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform', *Plant Physiol,* 170(4), pp. 2172-86.

Bottley, A., Xia, G. M. and Koebner, R. M. (2006) 'Homoeologous gene silencing in hexaploid wheat', *Plant J,* 47(6), pp. 897-906.

Boyes, D. C., Nam, J. and Dangl, J. L. (1998) 'The Arabidopsis thaliana RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response', *Proc Natl Acad Sci U S A,* 95(26), pp. 15849-54.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M. C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W. R., Hall, A., Mayer, K. F., Edwards, K. J., Bevan, M. W. and Hall, N. (2012) 'Analysis of the bread wheat genome using whole-genome shotgun sequencing', *Nature,* 491(7426), pp. 705-10.

Cao, A., Xing, L., Wang, X., Yang, X., Wang, W., Sun, Y., Qian, C., Ni, J., Chen, Y., Liu, D. and Chen, P. (2011) 'Serine/threonine kinase gene Stpk-V, a key member of powdery mildew resistance gene Pm21, confers powdery mildew resistance in wheat', *Proc Natl Acad Sci U S A,* 108(19), pp. 7727-32.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics,* 34(17), pp. i884-i890.

Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N., Li, Z., Wang, Q., Zheng, Z., Li, M., Cui, L., Liu, Z., Bian, J., Wang, Z., Xu, S., Yang, Q., Appels, R., Han, D., Song, W., Sun, Q. and Jiang, Y. (2019) 'Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat', *Genome Biol,* 20(1), pp. 136.

Cheuk, A., Ouellet, F. and Houde, M. (2020) 'The barley stripe mosaic virus expression system reveals the wheat C2H2 zinc finger protein TaZFP1B as a key regulator of drought tolerance', *BMC Plant Biol,* 20(1), pp. 144.

Comai, L. (2005) 'The advantages and disadvantages of being polyploid', *Nat Rev Genet,* 6(11), pp. 836-46.

Cooley, M. B., Pathirana, S., Wu, H. J., Kachroo, P. and Klessig, D. F. (2000) 'Members of the Arabidopsis HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens', *Plant Cell,* 12(5), pp. 663-76.

Costa-Silva, J., Domingues, D. and Lopes, F. M. (2017) 'RNA-Seq differential expression analysis: An extended review and a software tool', *PLoS One,* 12(12), pp. e0190152.

Crespo-Herrera, L. A., Garkava-Gustavsson, L. and Åhman, I. (2017) 'A systematic review of rye (', *Hereditas,* 154, pp. 14.

Dalal, M., Sahu, S., Tiwari, S., Rao, A. R. and Gaikwad, K. (2018) 'Transcriptome analysis reveals interplay between hormones, ROS metabolism and cell wall biosynthesis for drought-induced root growth in wheat', *Plant Physiol Biochem,* 130, pp. 482-492.

de Haro, L. A., Arellano, S. M., Novák, O., Feil, R., Dumón, A. D., Mattio, M. F., Tarkowská, D., Llauger, G., Strnad, M., Lunn, J. E., Pearce, S., Figueroa, C. M. and Del Vas, M. (2019) 'Mal de Río Cuarto virus infection causes hormone imbalance and sugar accumulation in wheat leaves', *BMC Plant Biol,* 19(1), pp. 112.

Devoto, A., Nieto-Rostro, M., Xie, D., Ellis, C., Harmston, R., Patrick, E., Davis, J., Sherratt, L., Coleman, M. and Turner, J. G. (2002) 'COI1 links jasmonate signalling and fertility to the SCF ubiquitin-ligase complex in Arabidopsis', *Plant J,* 32(4), pp. 457-66.

Dodds, P. N. and Rathjen, J. P. (2010) 'Plant immunity: towards an integrated view of plant-pathogen interactions', *Nat Rev Genet,* 11(8), pp. 539-48.

Dong, Z., Ma, C., Tian, X., Zhu, C., Wang, G., Lv, Y., Friebe, B., Li, H. and Liu, W. (2020) 'Genome-wide impacts of alien chromatin introgression on wheat gene transcriptions', *Sci Rep,* 10(1), pp. 4801.

Dubcovsky, J. and Dvorak, J. (2007) 'Genome plasticity a key factor in the success of polyploid wheat under domestication', *Science,* 316(5833), pp. 1862-6.

Ellis, C., Karafyllidis, I. and Turner, J. G. (2002) 'Constitutive activation of jasmonate signaling in an Arabidopsis mutant correlates with enhanced resistance to Erysiphe cichoracearum, Pseudomonas syringae, and Myzus persicae', *Mol Plant Microbe Interact,* 15(10), pp. 1025-30.

Ermakov, A., Bobrovskikh, A., Zubairova, U., Konstantinov, D. and Doroshkov, A. (2019) 'Stress-induced changes in the expression of antioxidant system genes for rice (', *PeerJ,* 7, pp. e7791.
Esquinas-Alcázar, J. (2005) 'Science and society: protecting crop genetic diversity for food security: political, ethical and technical challenges', *Nat Rev Genet,* 6(12), pp. 946-53.
FAO 2009. How to Feed the World in 2050. FAO.

FAO, I., UNICEF, WFP and WHO 2020. The State of Food Security and Nutrition in the World 2020. FAO, IFAD, UNICEF, WFP and WHO.

FAOSTAT (2020) *Statistical Division of the UN Food and Agriculture Organization*. Available at: http://faostat.fao.org/.

Farmer, E. E., Alméras, E. and Krishnamurthy, V. (2003) 'Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory', *Curr Opin Plant Biol,* 6(4), pp. 372-8.

Feldman, M., Levy, A. A., Fahima, T. and Korol, A. (2012) 'Genomic asymmetry in allopolyploid plants: wheat as a model', *J Exp Bot,* 63(14), pp. 5045-59.

Feschotte, C. (2008) 'Transposable elements and the evolution of regulatory networks', *Nat Rev Genet,* 9(5), pp. 397-405.

Feuillet, C. and Keller, B. (2002) 'Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution', *Ann Bot,* 89(1), pp. 3-10.
Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T. and Dubcovsky, J. (2009) 'A kinase-START gene confers temperature-dependent resistance to wheat stripe rust', *Science,* 323(5919), pp. 1357-60.

Giroux, M. J. and Morris, C. F. (1998) 'Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b', *Proc Natl Acad Sci U S A,* 95(11), pp. 6262-6.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W. L., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. (2002) 'A draft sequence of the rice genome (Oryza sativa L. ssp. japonica)', *Science,* 296(5565), pp. 92-100.

Grant, M., Brown, I., Adams, S., Knight, M., Ainslie, A. and Mansfield, J. (2000) 'The RPM1 plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death', *Plant J,* 23(4), pp. 441-50.

Graybosch, R. A. (2001) 'Mini Review: Uneasy Unions: Quality Effects of Rye Chromatin Transfers to Wheat', *Journal of Cereal Science,* 33.

Graybosch, R. A., Bai, G., St. Amand, P. and Sarath, G. (2019) 'Persistence of rye (Secale cereale L.) chromosome arm 1RS in wheat (Triticum aestivum L.) breeding programs of the Great Plains of North America', *Genetic Resources and Crop Evolution,* 66, pp. 941-950.

Gräfe, K. and Schmitt, L. (2020) 'The ABC transporter G subfamily in Arabidopsis thaliana', *J Exp Bot*.

Gustafson, P., Raskina, O., Ma, X. and Nevo, E. (2009) 'Wheat Evolution, Domestication, and Improvement', in Carver, B.F. (ed.) *Wheat: Science and Trade*: Wiley-Blackwell.

Hallab, A., Klee, K., Boecker, F., Srinivas, G. and Schoof, H. 2017. Automated Assignment of Human Readable Descriptions (AHRD).

Harper, A. L., Trick, M., He, Z., Clissold, L., Fellgett, A., Griffiths, S. and Bancroft, I. (2016) 'Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat', *Plant Biotechnol J,* 14(5), pp. 1207-14.

Harrison, R. G. and Larson, E. L. (2014) 'Hybridization, introgression, and the nature of species boundaries', *J Hered,* 105 Suppl 1, pp. 795-809.

Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., Hochu, I., Poirier, S., Santoni, S., Glémin, S. and David, J. (2007) 'Grinding up wheat: a massive loss of nucleotide diversity since domestication', *Mol Biol Evol,* 24(7), pp. 1506-17.

Haueisen, J., Möller, M., Eschenbrenner, C. J., Grandaubert, J., Seybold, H., Adamiak, H. and Stukenbrock, E. H. (2019) 'Highly flexible infection programs in a specialized wheat pathogen', *Ecol Evol,* 9(1), pp. 275-294.

He, P., Friebe, B. R., Gill, B. S. and Zhou, J. M. (2003) 'Allopolyploidy alters gene expression in the highly stable hexaploid wheat', *Plant Mol Biol,* 52(2), pp. 401-14.

He, R., Chang, Z., Yang, Z., Yuan, Z., Zhan, H., Zhang, X. and Liu, J. (2009) 'Inheritance and mapping of powdery mildew resistance gene Pm43 introgressed from Thinopyrum intermedium into wheat', *Theor Appl Genet,* 118(6), pp. 1173-80.

Howe, G. A. and Schilmiller, A. L. (2002) 'Oxylipin metabolism in response to stress', *Curr Opin Plant Biol,* 5(3), pp. 230-6.

Howell, T., Hale, I., Jankuloski, L., Bonafede, M., Gilbert, M. and Dubcovsky, J. (2014) 'Mapping a region within the 1RS.1BL translocation in common wheat affecting grain yield and canopy water status', *Theor Appl Genet,* 127(12), pp. 2695-709.

Hurni, S., Brunner, S., Stirnweis, D., Herren, G., Peditto, D., McIntosh, R. A. and Keller, B. (2014) 'The powdery mildew resistance gene P*m8* derived from rye is suppressed by its wheat ortholog P*m3*', *The Plant Journal,* 79.

Illumina (2020) *RNA Sequencing*. Available at: https://emea.illumina.com/techniques/sequencing/rna-sequencing.html.

IRGSP (2005) 'The map-based sequence of the rice genome', *Nature,* 436(7052), pp. 793-800.

IWGSC and Appels, R. and Eversole, K. and Feuillet, C. and Keller, B. and Rogers, J. and Stein, N. and Pozniak, C. J. and Choulet, F. and Distelfeld, A. and Poland, J. and Ronen, G. and Sharpe, A. G. and Pozniak, C. and Barad, O. and Baruch, K. and Keeble-Gagnère, G. and Mascher, M. and Ben-Zvi, G. and Josselin, A. A. and Himmelbach, A. and Balfourier, F. and Gutierrez-Gonzalez, J. and Hayden, M. and Koh, C. and Muehlbauer, G. and Pasam, R. K. and Paux, E. and Rigault, P. and Tibbits, J. and Tiwari, V. and Spannagl, M. and Lang, D. and Gundlach, H. and Haberer, G. and Mayer, K. F. X. and Ormanbekova, D. and Prade, V. and Šimková, H. and Wicker, T. and Swarbreck, D. and Rimbert, H. and Felder, M. and Guilhot, N. and Kaithakottil, G. and Keilwagen, J. and Leroy, P. and Lux, T. and Twardziok, S. and Venturini, L. and Juhász, A. and Abrouk, M. and Fischer, I. and Uauy, C. and Borrill, P. and Ramirez-Gonzalez, R. H. and Arnaud, D. and Chalabi, S. and Chalhoub, B. and Cory, A. and Datla, R. and Davey, M. W. and Jacobs, J. and Robinson, S. J. and Steuernagel, B. and van Ex, F. and Wulff, B. B. H. and Benhamed, M. and Bendahmane, A. and Concia, L. and Latrasse, D. and Alaux, M. and Bartoš, J. and Bellec, A. and Berges, H. and Doležel, J. and Frenkel, Z. and Gill, B. and Korol, A. and Letellier, T. and Olsen, O. A. and Singh, K. and Valárik, M. and van der Vossen, E. and Vautrin, S. and Weining, S. and Fahima, T. and Glikson, V. and Raats, D. and Číhalíková, J. and Toegelová, H. and Vrána, J. and Sourdille, P. and Darrier, B. and Barabaschi, D. and Cattivelli, L. and Hernandez, P. and Galvez, S. and Budak, H. and Jones, J. D. G. and Witek, K. and Yu, G. and Small,

I. and Melonek, J. and Zhou, R. and Belova, T. and Kanyuka, K. and King, R. and Nilsen, K. and Walkowiak, S. and Cuthbert, R. and Knox, R. and Wiebe, K. and Xiang, D. and Rohde, A. and Golds, T. and Čížková, J. and Akpinar, B. A. and Biyiklioglu, S. and Gao, L. and N'Daiye, A. and Kubaláková, M. and Šafář, J. and Alfama, F. and Adam-Blondon, A. F. and Flores, R. and Guerche, C. and Loaec, M. and Quesneville, H. and Condie, J. and Ens, J. and Maclachlan, R. and Tan, Y. and Alberti, A. and Aury, J. M. and Barbe, V. and Couloux, A. and Cruaud, C. and Labadie, K. and Mangenot, S. and Wincker, P. and Kaur, G. and Luo, M. and Sehgal, S. and Chhuneja, P. and Gupta, O. P. and Jindal, S. and Kaur, P. and Malik, P. and Sharma, P. and Yadav, B. and Singh, N. K. and Khurana, J. and Chaudhary, C. and Khurana, P. and Kumar, V. and Mahato, A. and Mathur, S. and Sevanthi, A. and Sharma, N. and Tomar, R. S. and Holušová, K. and Plíhal, O. and Clark, M. D. and Heavens, D. and Kettleborough, G. and Wright, J. and Balcárková, B. and Hu, Y. and Salina, E. and Ravin, N. and Skryabin, K. and Beletsky, A. and Kadnikov, V. and Mardanov, A. and Nesterov, M. and Rakitin, A. and Sergeeva, E. and Handa, H. and Kanamori, H. and Katagiri, S. and Kobayashi, F. and Nasuda, S. and Tanaka, T. and Wu, J. and Cattonaro, F. and Jiumeng, M. and Kugler, K. and Pfeifer, M. and Sandve, S. and Xun, X. and Zhan, B. and Batley, J. and Bayer, P. E. and Edwards, D. and Hayashi, S. and Tulpová, Z. and Visendi, P. and Cui, L. and Du, X. and Feng, K. and Nie, X. and Tong, W. and Wang, L. (2018) 'Shifting the limits in wheat research and breeding using a fully annotated reference genome', *Science,* 361(6403).

Jiang, Q., Niu, F., Sun, X., Hu, Z., Li, X., Ma, Y. and Zhang, H. (2017) 'RNA-seq analysis of unintended effects in transgenic wheat overexpressing the transcription factor GmDREB1,' *The Crop Journal,* 5(3).

Jobson, E. M., Johnston, R. E., Oiestad, A. J., Martin, J. M. and Giroux, M. J. (2019) 'The Impact of the Wheat *Rht-B1b* Semi-Dwarfing Allele on Photosynthesis and Seed Development Under Field Conditions', *Front Plant Sci,* 10, pp. 51.

Joly-Lopez, Z. and Bureau, T. E. (2018) 'Exaptation of transposable element coding sequences', *Curr Opin Genet Dev,* 49, pp. 34-42.

Jorgensen, R. (1990) 'Altered gene expression in plants due to trans interactions between homologous genes', *Trends Biotechnol,* 8(12), pp. 340-4.

Kashkush, K., Feldman, M. and Levy, A. A. (2002) 'Gene loss, silencing and activation in a newly synthesized wheat allotetraploid', *Genetics,* 160(4), pp. 1651-9.

Kaur, G., Shukla, V., Kumar, A., Kaur, M., Goel, P., Singh, P., Shukla, A., Meena, V., Kaur, J., Singh, J., Mantri, S., Rouached, H. and Pandey, A. K. (2019) 'Integrative analysis of hexaploid wheat roots identifies signature components during iron starvation', *J Exp Bot,* 70(21), pp. 6141-6161.

Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L. (2019) 'Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype', *Nat Biotechnol,* 37(8), pp. 907-915.

Koebner, R. M., Shepherd, K. W. and Appels, R. (1986) 'Controlled introgression to wheat of genes from rye chromosome arm 1RS by induction of allosyndesis : 2. Characterisation of recombinants', *Theor Appl Genet,* 73(2), pp. 209-17.

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., McFadden, H., Bossolini, E., Selter, L. L. and Keller, B. (2009) 'A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat', *Science,* 323(5919), pp. 1360-3.

Križanovic, K., Echchiki, A., Roux, J. and Šikic, M. (2018) 'Evaluation of tools for long read RNA-seq splice-aware alignment', *Bioinformatics,* 34(5), pp. 748-754.

Kumar, J., Jaiswal, V., Kumar, A., Kumar, N., Mir, R. R., Kumar, S., Dhariwal, R., Tyagi, S., Khandelwal, M., Prabhu, K. V., Prasad, R., Balyan, H. S. and Gupta, P. K. (2011) 'Introgression of a major gene for high grain protein content in some Indian bread wheat cultivars', *Field Crops Research,* 123, pp. 226-233.

Kumlay, A. M., Baenziger, P. S., Gill, K. S., Shelton, D. R., Graybosch, R. A., Lukaszewski, A. J. and Wesenberg, D. M. (2003) 'Understanding the Effect of Rye Chromatin in Bread Wheat', *Crop Science,* 43.

Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D. and Strömvik, M. V. (2018) 'Current Strategies of Polyploid Plant Genome Sequence Assembly', *Front Plant Sci,* 9, pp. 1660.

Langridge, P. (2012) 'Genomics: Decoding our daily bread', *Nature,* 491(7426), pp. 678-80.
Lantican, M. A., Dubin, H. J. and Morris, M. L. (2005) *Impacts of International Wheat Breeding Research in the Developing World, 1988-2002.* CIMMYT.

Lelley, T., Eder, C. and Grausgruber, H. (2004) 'Influence of 1BL.1RS wheat-rye chromosome translocation on genotype by environment interaction', *Journal of Cereal Science,* 39, pp. 313-320.

Levine, A., Pennell, R. I., Alvarez, M. E., Palmer, R. and Lamb, C. (1996) 'Calcium-mediated apoptosis in a plant hypersensitive disease resistance response', *Curr Biol,* 6(4), pp. 427-37.

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., van Sluys, M. A., Soltis, P. S., Xu, X., Yang, H. and Zhang, G. (2018) 'Earth BioGenome Project: Sequencing life for the future of life', *Proc Natl Acad Sci U S A,* 115(17), pp. 4325-4333.

Li, Z., Ren, T., Yan, B., Tan, F., Yang, M. and Ren, Z. (2016) 'A Mutant with Expression Deletion of Gene Sec-1 in a 1RS.1BL Line and Its Effect on Production Quality of Wheat', *PLoS One,* 11(1), pp. e0146943.

Liang, F., Du, X., Zhang, J., Li, X., Wang, F., Wang, H. and Liu, D. (2019) 'Wheat TaLr35PR2 gene is required for Lr35-mediated adult plant resistance against leaf rust fungus', *Funct Plant Biol,* 47(1), pp. 26-37.

Lobell, D. B., Cassman, K. G. and Field, C. B. (2009) 'Crop Yield Gaps: Their Importance, Magnitudes, and Causes', *Annual Review of Environment and Resources,* 34, pp. 179-204.

Love, M. I., Anders, S. and Huber, W. 2020. Analyzing RNA-seq data with DESeq2.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol,* 15(12), pp. 550.

Lukaszewski, A. J. (2000) 'Manipulation of the 1RS.1BL Translocation in Wheat by Induced Homoeologous Recombination', *Crop Science,* 40.

Luo, P. G., Luo, H. Y., Chang, Z. J., Zhang, H. Y., Zhang, M. and Ren, Z. L. (2009) 'Characterization and chromosomal location of Pm40 in common wheat: a new gene for resistance to powdery mildew derived from Elytrigia intermedium', *Theor Appl Genet,* 118(6), pp. 1059-64.
Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R. and Dangl, J. L. (2003) 'Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance', *Cell,* 112(3), pp. 379-89.

Martin, G. B., Bogdanove, A. J. and Sessa, G. (2003) 'Understanding the functions of plant disease resistance proteins', *Annu Rev Plant Biol,* 54, pp. 23-61.

McDowell, J. M., Dhandaydham, M., Long, T. A., Aarts, M. G., Goff, S., Holub, E. B. and Dangl, J. L. (1998) 'Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis', *Plant Cell,* 10(11), pp. 1861-74.

McHale, L., Tan, X., Koehl, P. and Michelmore, R. W. (2006) 'Plant NBS-LRR proteins: adaptable guards', *Genome Biol,* 7(4), pp. 212.

Mergoum, M. and Gómez-Macpherson, H. 2004. Triticale improvement and production. Food and Agricultural Organisation of the United Nations.

Mohr, T. J., Mammarella, N. D., Hoff, T., Woffenden, B. J., Jelesko, J. G. and McDowell, J. M. (2010) 'The Arabidopsis downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements', *Mol Plant Microbe Interact,* 23(10), pp. 1303-15.

Molnár-Láng, M., Ceoloni, C. and Doležel, J. (2015) *Alien Introgression in Wheat Cytogenetics, Molecular Biology, and Genomics.* Springer.

Nagels Durand, A., Pauwels, L. and Goossens, A. (2016) 'The Ubiquitin System and Jasmonate Signaling', *Plants (Basel),* 5(1).

Ng, A. and Xavier, R. J. (2011) 'Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity', *Autophagy,* 7(9), pp. 1082-4.

Ni, F., Qi, J., Hao, Q., Lyu, B., Luo, M. C., Wang, Y., Chen, F., Wang, S., Zhang, C., Epstein, L., Zhao, X., Wang, H., Zhang, X., Chen, C., Sun, L. and Fu, D. (2017) 'Wheat Ms2 encodes for an orphan protein that confers male sterility in grass species', *Nat Commun,* 8, pp. 15121.

Pace, C. D., Snidaro, D., Ciaffi, M., Vittori, D., Ciofo, A., Cenci, A., Tanzarella, O. A., Qualset, C. O. and Scarascia Mugnozza, G. T. (2001) 'Introgression of Dasypyrum villosum chromatin into common wheat improves grain protein quality', *Euphytica,* 117, pp. 67-75.

Palosuo, K., Alenius, H., Varjonen, E., Kalkkinen, N. and Reunala, T. (2001) 'Rye γ-70 and γ-35 secalins and barley γ-3 hordein cross-react with ω-5 gliadin, a major allergen in wheat-dependent, exercise-induced anaphylaxis', *Clinical & Experimental Allergy,* 31.

Pearce, S., Huttly, A. K., Prosser, I. M., Li, Y. D., Vaughan, S. P., Gallova, B., Patil, A., Coghill, J. A., Dubcovsky, J., Hedden, P. and Phillips, A. L. (2015) 'Heterologous expression and transcript analysis of gibberellin biosynthetic genes of grasses reveals novel functionality in the GA3ox family', *BMC Plant Biol,* 15, pp. 130.

Peng, Z., Hu, Y., Zhang, J., Huguet-Tapia, J. C., Block, A. K., Park, S., Sapkota, S., Liu, Z., Liu, S. and White, F. F. (2019) '*Xanthomonas translucens* commandeers the host rate-limiting step in ABA biosynthesis for disease susceptibility', *Proc Natl Acad Sci U S A,* 116(42), pp. 20938-20946.

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., Mayer, K. F., Olsen, O. A. and Consortium, I. W. G. S. (2014) 'Genome interplay in the grain transcriptome of hexaploid bread wheat', *Science,* 345(6194), pp. 1250091.

Placido, D. F., Campbell, M. T., Folsom, J. J., Cui, X., Kruger, G. R., Baenziger, P. S. and Walia, H. (2013) 'Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat', *Plant Physiol,* 161(4), pp. 1806-19.

Powell, J. J., Fitzgerald, T. L., Stiller, J., Berkman, P. J., Gardiner, D. M., Manners, J. M., Henry, R. J. and Kazan, K. (2017) 'The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias', *Plant Biotechnol J,* 15(4), pp. 533-543.

Rabinovich, S. V. (1998) 'Importance of wheat-rye translocations for breeding modern cultivar of Triticum aestivum L.', *Euphytica,* 100, pp. 323-340.

Rangan, P., Furtado, A. and Henry, R. (2020) 'Transcriptome profiling of wheat genotypes under heat stress during grain-filling', 91.

Rees, D. C., Johnson, E. and Lewinson, O. (2009) 'ABC transporters: the power to change', *Nat Rev Mol Cell Biol,* 10(3), pp. 218-27.

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., van Ginkel, M., Hoisington, D., Bohn, M. and Melchinger, A. E. (2005) 'Wheat genetic diversity trends during domestication and breeding', *Theor Appl Genet,* 110(5), pp. 859-64.

Ren, T., Tang, Z., Fu, S., Yan, B., Tan, F., Ren, Z. and Li, Z. (2017) 'Molecular Cytogenetic Characterization of Novel Wheat-rye T1RS.1BL Translocation Lines with High Resistance to Diseases and Great Agronomic Traits', *Front Plant Sci,* 8, pp. 799.

Ren, T.-H., Chen, F., Yan, B.-J., Zhang, H.-Q. and Ren, Z.-L. (2012) 'Genetic diversity of wheat–rye 1BL.1RS translocation lines derived from different wheat and rye sources ', *Euphytica,* 183, pp. 133-146.

Ren, T.-H., Yang, Z.-J., Yan, B.-J., Zhang, H.-Q., Fu, S.-L. and Ren, Z.-L. (2009) 'Development and characterization of a new 1BL.1RS translocation line with resistance to stripe rust and powdery mildew of wheat', *Euphytica,* 169, pp. 207-213.

Rothamsted 2017. Rothamsted Research Annual Report 2016-2017. Rothamsted Research.
Saulescu, N. N., Ittu, G., Ciuca, M., Ittu, M., Serban, G. and Mustatea, P. (2011) 'Transferring Useful Rye Genes to Wheat, Using Triticale as a Bridge', *Czech Journal of Genetics and Plant Breeding,* 47, pp. 56-62.

Schlegel, R. and Korzun, V. (1997) 'About the origin of 1RS.1BL wheat-rye chromosome translocations from Germany', *Euphytica,* 116.

Schlegel, R. and Korzun, V. (2020) *GENES, MARKERS AND LINKAGE DATA OF RYE*. Available at: http://www.rye-gene-map.de/. Available at: http://www.rye-gene-map.de/rye-introgression/index.html.

Schnable, P. S. and Ware, D. and Fulton, R. S. and Stein, J. C. and Wei, F. and Pasternak, S. and Liang, C. and Zhang, J. and Fulton, L. and Graves, T. A. and Minx, P. and Reily, A. D. and Courtney, L. and Kruchowski, S. S. and Tomlinson, C. and Strong, C. and Delehaunty, K. and Fronick, C. and Courtney, B. and Rock, S. M. and Belter, E. and Du, F. and Kim, K. and Abbott, R. M. and Cotton, M. and Levy, A. and Marchetto, P. and Ochoa, K. and Jackson, S. M. and Gillam, B. and Chen, W. and Yan, L. and Higginbotham, J. and Cardenas, M. and Waligorski, J. and Applebaum, E. and Phelps, L. and Falcone, J. and Kanchi, K. and Thane, T. and Scimone, A. and Thane, N. and Henke, J. and Wang, T. and Ruppert, J. and Shah, N. and Rotter, K. and Hodges, J. and Ingenthron, E. and Cordes, M. and Kohlberg, S. and Sgro, J. and Delgado, B. and Mead, K. and Chinwalla, A. and Leonard, S. and Crouse, K. and Collura, K. and Kudrna, D. and Currie, J. and He, R. and Angelova, A. and Rajasekar, S. and Mueller, T. and Lomeli, R. and Scara, G. and Ko, A. and Delaney, K. and Wissotski, M. and Lopez, G. and Campos, D. and Braidotti, M. and Ashley, E. and Golser, W. and Kim, H. and Lee, S. and Lin, J. and Dujmic, Z. and Kim, W. and Talag, J. and Zuccolo, A. and Fan, C. and Sebastian, A. and Kramer, M. and Spiegel, L. and Nascimento, L. and Zutavern, T. and Miller,

B. and Ambroise, C. and Muller, S. and Spooner, W. and Narechania, A. and Ren, L. and Wei, S. and Kumari, S. and Faga, B. and Levy, M. J. and McMahan, L. and Van Buren, P. and Vaughn, M. W. and Ying, K. and Yeh, C. T. and Emrich, S. J. and Jia, Y. and Kalyanaraman, A. and Hsia, A. P. and Barbazuk, W. B. and Baucom, R. S. and Brutnell, T. P. and Carpita, N. C. and Chaparro, C. and Chia, J. M. and Deragon, J. M. and Estill, J. C. and Fu, Y. and Jeddeloh, J. A. and Han, Y. and Lee, H. and Li, P. and Lisch, D. R. and Liu, S. and Liu, Z. and Nagel, D. H. and McCann, M. C. and SanMiguel, P. and Myers, A. M. and Nettleton, D. and Nguyen, J. and Penning, B. W. and Ponnala, L. and Schneider, K. L. and Schwartz, D. C. and Sharma, A. and Soderlund, C. and Springer, N. M. and Sun, Q. and Wang, H. and Waterman, M. and Westerman, R. and Wolfgruber, T. K. and Yang, L. and Yu, Y. and Zhang, L. and Zhou, S. and Zhu, Q. and Bennetzen, J. L. and Dawe, R. K. and Jiang, J. and Jiang, N. and Presting, G. G. and Wessler, S. R. and Aluru, S. and Martienssen, R. A. and Clifton, S. W. and McCombie, W. R. and Wing, R. A. and Wilson, R. K. (2009) 'The B73 maize genome: complexity, diversity, and dynamics', *Science,* 326(5956), pp. 1112-5.

Schneider, A., Molnár, I. and Molnár-Láng, M. (2008) 'Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat', *Euphytica,* (163), pp. 1-19.

Seifert, F., Bössow, S., Kumlehn, J., Gnad, H. and Scholten, S. (2016) 'Analysis of wheat microspore embryogenesis induction by transcriptome and small RNA sequencing using the highly responsive cultivar "Svilena"', *BMC Plant Biol,* 16, pp. 97.

Seki, M., Katsumata, E., Suzuki, A., Sereewattanawoot, S., Sakamoto, Y., Mizushima-Sugano, J., Sugano, S., Kohno, T., Frith, M. C., Tsuchihara, K. and Suzuki, Y. (2019) 'Evaluation and application of RNA-Seq by MinION', *DNA Res,* 26(1), pp. 55-65.

Shao, Z. Q., Xue, J. Y., Wang, Q., Wang, B. and Chen, J. Q. (2019) 'Revisiting the Origin of Plant NBS-LRR Genes', Trends Plant Sci, 24(1), pp. 9-12.

Shewry, P. R. (2009) 'Wheat', *J Exp Bot,* 60(6), pp. 1537-53.

Shewry, P. R., Bradberry, D., Franklin, J. and White, R. P. (1984) 'The chromosomal locations and linkage relationships of the structural genes for the prolamin storage proteins (secalins) of rye', *Theor Appl Genet,* 69(1), pp. 63-9.

Smith, A. M., Coupland, G., Dolan, L., Harberd, N., Jones, J., Martin, C., Sablowski, R. and Amey, A. (2009) *Plant Biology.* Garland Science.

Tang, C., Li, M., Cao, M., Lu, R., Zhang, H., Liu, C., Huang, S., Zhang, P., Hu, H., Zhao, W. and Wu, L. (2020) 'Transcriptome analysis suggests mechanisms for a novel flowering type: Cleistogamous wheat', *The Crop Journal,* 8(2).

Thirugnanasambantham, K., Durairaj, S., Saravanan, S., K., K., Muralidaran, S. and Islam, V. I. H. (2015) 'Role of Ethylene Response Transcription Factor (ERF) and Its Regulation in Response to Stress Encountered by Plants', *Plant Molecular Biology Reporter,* 33, pp. 347-357.

Uauy, C. (2017) 'Wheat genomics comes of age', *Curr Opin Plant Biol,* 36, pp. 142-148.
UN 2015. Population 2030 Demographic challenges and opportunities for sustainable development planning. UN.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014) 'Ten years of next-generation sequencing technology', *Trends Genet,* 30(9), pp. 418-26.
van Ooijen, G., Mayr, G., Kasiem, M. M., Albrecht, M., Cornelissen, B. J. and Takken, F. L. (2008) 'Structure-function analysis of the NB-ARC domain of plant disease resistance proteins', *J Exp Bot,* 59(6), pp. 1383-97.

Wang, M., Qin, L., Xie, C., Li, W., Yuan, J., Kong, L., Yu, W., Xia, G. and Liu, S. (2014) 'Induced and constitutive DNA methylation in a salinity-tolerant wheat introgression line', *Plant Cell Physiol,* 55(7), pp. 1354-65.

Wang, S., Li, Q. P., Wang, J., Yan, Y., Zhang, G. L., Zhang, H., Wu, J., Chen, F., Wang, X., Kang, Z., Dubcovsky, J. and Gou, J. Y. (2019) 'YR36/WKS1-Mediated Phosphorylation of PsbO, an Extrinsic Member of Photosystem II, Inhibits Photosynthesis and Confers Stripe Rust Resistance in Wheat', *Mol Plant,* 12(12), pp. 1639-1650.

Wang, Y., Yu, H., Tian, C., Sajjad, M., Gao, C., Tong, Y., Wang, X. and Jiao, Y. (2017) 'Transcriptome Association Identifies Regulators of Wheat Spike Architecture', *Plant Physiol,* 175(2), pp. 746-757.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nat Rev Genet,* 10(1), pp. 57-63.

Wieser, H. (2007) 'Chemistry of gluten proteins', *Food Microbiol,* 24(2), pp. 115-9.
Wilkinson, P. A., Winfield, M. O., Barker, G. L., Tyrrell, S., Bian, X., Allen, A. M., Burridge, A., Coghill, J. A., Waterfall, C., Caccamo, M., Davey, R. P. and Edwards, K. J. (2016) 'CerealsDB 3.0: expansion of resources and data integration', *BMC Bioinformatics,* 17, pp. 256.
Willis, K. J. and McElwain, J. C. (2013) *The Evolution of Plants.* 2nd Edition edn.: Oxford University Press.

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L., Benbow, H. R., Wilkinson, P. A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., King, J., West, C., Griffiths, S., King, I., Bentley, A. R. and Edwards, K. J. (2016) 'High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool', *Plant Biotechnol J,* 14(5), pp. 1195-206.

Winfield, M. O., Wilkinson, P. A., Allen, A. M., Barker, G. L., Coghill, J. A., Burridge, A., Hall, A., Brenchley, R. C., D'Amore, R., Hall, N., Bevan, M. W., Richmond, T., Gerhardt, D. J., Jeddeloh, J. A. and Edwards, K. J. (2012) 'Targeted re-sequencing of the allohexaploid wheat exome', *Plant Biotechnol J,* 10(6), pp. 733-42.

Wu, J., Gao, J., Bi, W., Zhao, J., Yu, X., Li, Z., Liu, D., Liu, B. and Wang, X. (2019) 'Genome-Wide Expression Profiling of Genes Associated with the', *Int J Mol Sci,* 20(18).

Wulff, B. B. and Moscou, M. J. (2014) 'Strategies for transferring resistance into wheat: from wide crosses to GM cassettes', *Front Plant Sci,* 5, pp. 692.

Yalamanchili, H. K., Wan, Y.-W. and Liu, Z. (2018) 'Data Analysis Pipeline for RNA-seq Experiments: From Differential Expression to Cryptic Splicing', *Current Protocols in Bioinformatics*.
Yang, F., Li, W. and Jørgensen, H. J. (2013) 'Transcriptional reprogramming of wheat and the hemibiotrophic pathogen Septoria tritici during two phases of the compatible interaction', *PLoS One,* 8(11), pp. e81606.

Zhang, J., Wang, F., Liang, F., Zhang, Y., Ma, L., Wang, H. and Liu, D. (2018) 'Functional analysis of a pathogenesis-related thaumatin-like protein gene TaLr35PR5 from wheat induced by leaf rust fungus', *BMC Plant Biol,* 18(1), pp. 76.

Zhao, G., Zou, C., Li, K., Wang, K., Li, T., Gao, L., Zhang, X., Wang, H., Yang, Z., Liu, X., Jiang, W., Mao, L., Kong, X., Jiao, Y. and Jia, J. (2017) 'The Aegilops tauschii genome reveals multiple impacts of transposons', *Nat Plants,* 3(12), pp. 946-955.

Supplementary Material

S1) Metadata table of varieties used in the study

INTROGRESSION-CONTAINING VARIETIES

Variety	Paper DOI	Tissue	Study Type	Country of Origin	NCBI Accession	Sequencing Type	Sequencing Platform	Amount of Data (GB)	Notes
Aimengniu	https://doi.org/10.1104/pp.17.00694	Reproductive tissue (spikes?)	Investigating spike architecture	China	PRJNA348655	Paired	Illumina HiSeq 2500	4	
Bacanora	https://doi.org/10.1111/pbi.12486	Leaf?	Normal	Mexico	PRJEB5290	Single	Illumina Genome Analyzer II	1.8	
Banks	https://doi.org/10.1016/j.jcs.2019.102895	Spikes	Drought, time-series	Australia	PRJEB23118	Paired	Illumina HiSeq 4000	1.5	
Beaver	https://doi.org/10.1111/pbi.12486	Leaf	Normal	UK	PRJEB5290	Single	Illumina HiSeq 2000	1.6	
Bobwhite	https://doi.org/10.1104/pp.19.00380	Flag leaf	10-point time-series	Mexico	PRJNA497810	Paired	Illumina HiSeq 2500	60	
Equinox	https://doi.org/10.1111/pbi.12486	Leaf	Normal	UK	PRJEB5290	Single	Illumina HiSeq 2000	1.8	
Florida	https://doi.org/10.1111/pbi.12486	Leaf	Normal	Germany	PRJEB5290	Single	Illumina HiSeq 2000	2.9	
Haven	https://doi.org/10.1111/pbi.12486	Leaf	Normal	UK	PRJEB5290	Single	Illumina HiSeq 2000	1.6	
Lovrin 10	https://doi.org/10.1104/pp.17.00694	Spike	Investigating spike architecture	Romania	PRJNA348655	Paired	Illumina HiSeq 2500	4	
Lumai 15	https://doi.org/10.1038/ncomms15121	Flag leaf, (fertile anther, sterile anther, pistil)	Investigating male sterility gene	China	PRJNA351906	Paired	Illumina HiSeq 2500	64	Only flag leaf reads were used.
Nautica	https://doi.org/10.1111/pbi.12486	Leaf	Normal	Netherlands	PRJEB5290	Single	Illumina HiSeq 2000	2.4	
Palur	https://doi.org/10.1111/pbi.12486	Leaf	Normal	Germany	PRJEB5290	Single	Illumina HiSeq 2000	3.6	
PBW 343	UNPUBLISHED	Leaf	Inoculated with stripe rust	India	PRJNA613349	Paired	Illumina HiSeq 4000	4	Control reads downloaded.

ProIntaFederal	https://doi.org/10.1186/s12870-019-1709-y	Leaf	Inoculated with MRCV	Argentina	PRJNA490015	Single	Illumina HiSeq 3000	9.8	Control reads downloaded.
Rialto	https://doi.org/10.1111/pbi.12486	Leaf	Normal	UK	PRJEB5290	Single	Illumina HiSeq 2000	2	
Savannah	https://doi.org/10.1111/pbi.12486	Leaf	Normal	UK	PRJEB5290	Single	Illumina HiSeq 2000	1.4	
Svilena	https://doi.org/10.1186/s12870-016-0782-8	Microspore	Time-series, treatment of Microspore	Bulgaria	PRJNA297977	Single	Illumina HiSeq 2000	9.3	

NON INTROGRESSION-CONTAINING VARIETIES

Variety	Paper	Tissue	Study Type	Country of Origin	NCBI Accession	Sequencing Type	Sequencing Platform	Amount of Data (GB)	Notes
Alba	https://doi.org/10.1111/pbi.12486	Leaf	Normal	Poland	PRJEB5290	Single	Illumina HiSeq 2000	5.5	
Atlas 66	https://doi.org/10.1186/s12870-020-02355-x	Leaf	Overexpression of a TF to see relevance in drought-resistance	USA	PRJNA563057	Single	Illumina NovaSeq 6000	3.8	
Avalon	https://doi.org/10.1111/pbi.12486	Leaf?	Normal	UK	PRJEB5290	Single	Illumina Genome Analyzer II	1.7	
Borenos	https://doi.org/10.1111/pbi.12486	Leaf?	Normal	Germany	PRJEB5290	Single	Illumina HiSeq 2000	1.7	
C 306	https://doi.org/10.1093/jxb/erz358	Roots	Iron deprivation	India	PRJNA529036	Paired	Illumina NextSeq 5000	8	
Capo	https://doi.org/10.1111/pbi.12486	Leaf?	Normal	Austria?	PRJEB5290	Single	Illumina Genome Analyzer II	1.4	
Chinese Spring	https://doi.org/10.1111/pbi.12486	Leaf?	Normal	China	PRJEB5290	Single	Illumina HiSeq 2000	7	
Chuan Mai 25	https://doi.org/10.21203/rs.3.rs-29235/v1	Leaf	Metribuzin (herbicide) treated	China	PRJNA555667	Paired	Illumina NovaSeq 6000	11	Control reads could not be downloaded.

Fortuna	https://doi.org/10.3389/fpls.2019.00051	Leaf,	Transcriptomes of normal and semi-dwarfing gene-containing cultivars	Russia	PRJNA514367	Single	Illumina HiSeq 2500	8	Only leaf reads taken.
Holdfast	https://doi.org/10.1186/s12870-015-0520-7	Grain	Different tissue layers of developing grain	Australia	PRJEB7795	Single	Illumina Genome Analyzer II	4	
Jagger	https://doi.org/10.1073/pnas.1911660116	Leaf	Inoculation with X. translucens	USA	PRJNA485724	Paired	Illumina HiSeq 2000	6	No mention of if reads were inoculated or not.
Jimai 19	https://doi.org/10.1016/j.cj.2016.12.001	Roots	Drought stress, salt stress	China	PRJNA355905	Paired	Illumina HiSeq 2500	4	
Obelisk	https://doi.org/10.1002/ece3.4724	Leaf	Inoculation with Z. tritici	Netherlands	PRJNA415716	Single	Illumina HiSeq 2500	38	
Raj 3765	https://doi.org/10.1016/j.plaphy.2018.07.035	Roots	Control vs stress	India	PRJNA435777	Paired	Illumina HiSeq 1000	3	
Saratovskaya 29	https://doi.org/10.7717/peerj.7791	Leaf	Water deficiency, cold stress	Russia	PRJNA630059	Single	Illumina MACE/Next Seq 5000	1	
Sevin	https://doi.org/10.1371/journal.pone.0081606	Leaf	Inoculated with S. tritici	Denmark	PRJNA196595	Paired	Illumina HiSeq 2000	7.6	Control reads downloaded.
Stoa	https://doi.org/10.1038/s41598-018-25430-8	Shoot/Leaf (seedling stage)	5 different lines to see differential expression of transporters	USA	PRJNA397654	Paired?	Illumina HiSeq 2500	7.5	Only shoot reads downloaded.
Sumai3	https://doi.org/10.1038/s41598-018-25430-8	Shoot/Leaf (seedling stage)	5 different lines to see differential expression of transporters	China	PRJNA397654	Paired?	Illumina HiSeq 2500	7	Only shoot reads downloaded.
Triumph	https://doi.org/10.1104/pp.17.00694	Spike	Sequencing various cultivars	USA	PRJNA348655	Paired	Illumina HiSeq 2500	5	
Yecora Rojo	https://doi.org/10.3390/ijms20184498	Leaf	Inoculated with P. triticina	Mexico	PRJNA629995	Paired	Illumina HiSeq 4000	28.5	
Yumai 18	https://doi.org/10.1016/j.cj.2019.08.009	Spikelet	Exploring cleistogamous phenotype	China	PRJNA491844	Paired	Illumina HiSeq 2000	8.5	

S2) FUNCTIONAL INFORMATION DATASET

DIFFERENTIALLY EXPRESSED WHEAT GENES

contig	blast-accession	human-readable-description	quality-code	quality-values	blast2	uniprot	iwgsc
TraesCS1B02G148400	N/A	N/A	N/A	N/A	N/A	ER membrane protein complex subunit 1	ER membrane protein complex subunit 1, ***
TraesCS1B02G104900	N/A	N/A	N/A	N/A	N/A	Mitogen-activated protein kinase, EC 2.7.11.24	Mitogen-activated protein kinase, ***
TraesCS1B02G142300	N/A	N/A	N/A	N/A	N/A	Non-specific serine/threonine protein kinase, EC 2.7.11.1	Kinase family protein, ***
TraesCS1B02G060000	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	E3 UFM1-protein ligase 1-like protein, ***
TraesCS1B02G090300	N/A	N/A	N/A	N/A	N/A	Prolycopene isomerase, EC 5.2.1.13	Carotenoid isomerase, putative, expressed, ***
TraesCS1B02G107100	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Ankyrin repeat domain-containing protein 2, *-*
TraesCS1B02G106900	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	DNA/RNA helicase protein, *-*
TraesCS1B02G107000	N/A	N/A	N/A	N/A	N/A	Serine/threonine-protein phosphatase, EC 3.1.3.16	Serine/threonine-protein phosphatase, ***
TraesCS1B02G091800	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	RNA polymerase II-associated protein 3, ***
TraesCS1B02G146400	N/A	N/A	N/A	N/A	N/A	UBC core domain-containing protein	Ubiquitin-conjugating enzyme E2, ***
TraesCS1B02G101700	N/A	N/A	N/A	N/A	N/A	RING-type domain-containing protein	RING/U-box superfamily protein, putative, ***
TraesCS1B02G139600	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	70 kDa heat shock protein, ***
TraesCS1B02G123400	N/A	N/A	N/A	N/A	N/A	HECT domain-containing protein	E3 ubiquitin-protein ligase, *-*
TraesCS1B02G140700	N/A	N/A	N/A	N/A	N/A	DUF1664 domain-containing protein	bZIP transcription factor, putative (DUF1664), ***
TraesCS1B02G099000	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	ABC transporter B family protein, ***

TraesCS1B02G149000	N/A	N/A	N/A	N/A	N/A	KAT8 regulatory NSL complex subunit 2	INO80 complex subunit D, ***
TraesCS1B02G138300	N/A	N/A	N/A	N/A	N/A	RING-type domain-containing protein	RING/U-box superfamily protein, ***
TraesCS1B02G055900	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Coiled-coil domain-containing protein, putative, ***
TraesCS1B02G118000	N/A	N/A	N/A	N/A	N/A	CULLIN_2 domain-containing protein	Cullin-1, ***
TraesCS1B02G093300	N/A	N/A	N/A	N/A	N/A	Protein SQS1	G-patch domain containing protein, ***
TraesCS1B02G128400	N/A	N/A	N/A	N/A	N/A	Golgi apparatus membrane protein TVP23	Golgi apparatus membrane protein TVP23, ***
TraesCS1B02G117900	N/A	N/A	N/A	N/A	N/A	CULLIN_2 domain-containing protein	Cullin-1, ***
TraesCS1B02G098200	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Zinc ion binding protein, ***
TraesCS1B02G099100	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Diacylglycerol acyltransferase, *-*
TraesCS1B02G119300	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Starch synthase family protein, ***
TraesCS1B02G147600	N/A	N/A	N/A	N/A	N/A	NFACT-R_1 domain-containing protein	Coiled-coil domain-containing protein 25, ***
TraesCS1B02G112700	N/A	N/A	N/A	N/A	N/A	Na_H_Exchanger domain-containing protein	Sodium/hydrogen exchanger, ***
TraesCS1B02G126500	N/A	N/A	N/A	N/A	N/A	Importin subunit alpha	Importin subunit alpha, ***
TraesCS1B02G140400	N/A	N/A	N/A	N/A	N/A	FACT complex subunit SSRP1	FACT complex subunit SSRP1, ***
TraesCS1B02G123100	N/A	N/A	N/A	N/A	N/A	RING-type domain-containing protein	RING/U-box superfamily protein, ***
TraesCS1B02G088700	N/A	N/A	N/A	N/A	N/A	J domain-containing protein	Chaperone protein dnaJ, ***
TraesCS1B02G141700	N/A	N/A	N/A	N/A	N/A	Malic enzyme	Malic enzyme, ***
TraesCS1B02G129300	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Acyl-CoA dehydrogenase, putative, ***
TraesCS1B02G126700	N/A	N/A	N/A	N/A	N/A	Nfu_N domain-containing protein	Fe/S biogenesis protein nfuA, *-*

TraesCS1B02G125100	N/A	N/A	N/A	N/A	N/A	J domain-containing protein	Chaperone protein dnaJ, putative, ***
TraesCS1B02G140800	N/A	N/A	N/A	N/A	N/A	Myb_DNA-bind_3 domain-containing protein	Myb/SANT-like DNA-binding domain protein, --*
TraesCS1B02G069300	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Mammalian uncoordinated homology 13, domain 2, ***
TraesCS1B02G146600	N/A	N/A	N/A	N/A	N/A	RRM domain-containing protein	RNA-binding family protein, *-*
TraesCS1B02G134900	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Protein apaG, ***
TraesCS1B02G099600	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Aspartic proteinase, ***
TraesCS1B02G059800	N/A	N/A	N/A	N/A	N/A	Nudix hydrolase domain-containing protein	Nudix hydrolase 9, ***
TraesCS1B02G108200	N/A	N/A	N/A	N/A	N/A	Fibronectin type-III domain-containing protein	Protein VERNALIZATION INSENSITIVE 3, ***
TraesCS1B02G019100	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Ras-like protein, ***
TraesCS1B02G071700	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Zinc finger family protein, ***
TraesCS1B02G001100	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Nucleic acid-binding, OB-fold-like protein, ***
TraesCS1B02G101000	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Agenet domain containing protein, ***
TraesCS1B02G098300	N/A	N/A	N/A	N/A	N/A	RF_PROK_I domain-containing protein	Peptide chain release factor 1, ***
TraesCS1B02G001600	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Adenylyl cyclase-associated protein 1, --*
TraesCS1B02G089900	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	phosphoglycolate phosphatase, ***
TraesCS1B02G141900	N/A	N/A	N/A	N/A	N/A	Proteasome subunit beta, EC 3.4.25.1	Proteasome subunit beta type, ***
TraesCS1B02G105200	N/A	N/A	N/A	N/A	N/A	WD_REPEATS_REGION domain-containing protein	Transducin/WD-like repeat-protein, ***
TraesCS1B02G141300	N/A	N/A	N/A	N/A	N/A	Ephrin_rec_like domain-containing protein	Glycine-rich protein, ***
TraesCS1B02G132900	N/A	N/A	N/A	N/A	N/A	Autophagy-related protein 18a	WD repeat phosphoinositide-interacting-like protein, ***

TraesCS1B02G105300	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	O-acyltransferase, --*
TraesCS1B02G123200	N/A	N/A	N/A	N/A	N/A	Kinesin-like protein	Kinesin-like protein, ***
TraesCS1B02G100800	N/A	N/A	N/A	N/A	N/A	RRM domain-containing protein	RNA-binding family protein, ***
TraesCS1B02G024500	N/A	N/A	N/A	N/A	N/A	Actin	Actin, ***
TraesCS1B02G090400	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Metal tolerance protein, ***
TraesCS1B02G148200	N/A	N/A	N/A	N/A	N/A	PA domain-containing protein	Vacuolar-sorting receptor-like protein, ***
TraesCS1B02G132200	N/A	N/A	N/A	N/A	N/A	F-box domain-containing protein	F-box family protein, --*
TraesCS1B02G095200	N/A	N/A	N/A	N/A	N/A	Non-specific serine/threonine protein kinase, EC 2.7.11.1	Protein kinase, ***
TraesCS1B02G145500	N/A	N/A	N/A	N/A	N/A	alpha-1,2-Mannosidase, EC 3.2.1.-	alpha-1,2-Mannosidase, ***
TraesCS1B02G127000	N/A	N/A	N/A	N/A	N/A	Histidine kinase domain-containing protein	Ethylene receptor, ***
TraesCS1B02G138200	N/A	N/A	N/A	N/A	N/A	Pre-mRNA-splicing factor 38	Pre-mRNA-splicing factor, ***
TraesCS1B02G083900	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	E3 SUMO-protein ligase SIZ1, ***
TraesCS1B02G058000	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	UDP galactose transporter-related protein, ***
TraesCS1B02G148000	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Calmodulin, ***
TraesCS1B02G102200	N/A	N/A	N/A	N/A	N/A	Replication protein A subunit	Replication protein A 70 kDa DNA-binding subunit, ***
TraesCS1B02G145200	N/A	N/A	N/A	N/A	N/A	N-acetyltransferase domain-containing protein	N-acetyltransferase-like protein, ***
TraesCS1B02G097700	N/A	N/A	N/A	N/A	N/A	Protein kinase domain-containing protein	Kinase family protein, ***
TraesCS1B02G119000	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Double stranded RNA binding protein 3, ***

TraesCS1B02G089700	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	ATP-dependent zinc metalloprotease FtsH, *-*
TraesCS1B02G139900	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Late embryogenesis abundant protein, ***
TraesCS1B02G065900	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Arginine/serine-rich splicing factor, putative, ***
TraesCS1B02G092200	N/A	N/A	N/A	N/A	N/A	DUF4033 domain-containing protein	Beta-carotene isomerase d27, chloroplastic, ***
TraesCS1B02G019200	N/A	N/A	N/A	N/A	N/A	CAP-Gly domain-containing protein	Tubulin-specific chaperone cofactor E-like protein, ***
TraesCS1B02G087300	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Methyl-CpG-binding domain protein 4, ***
TraesCS1B02G017800	N/A	N/A	N/A	N/A	N/A	Methionine S-methyltransferase, EC 2.1.1.12	Methionine S-methyltransferase, ***
TraesCS1B02G058600	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Dual specificity phosphatase, ***
TraesCS1B02G144000	N/A	N/A	N/A	N/A	N/A	O-acyltransferase	O-acyltransferase, ***
TraesCS1B02G126100	N/A	N/A	N/A	N/A	N/A	Protein disulfide-isomerase, EC 5.3.4.1	Disulfide-isomerase, ***
TraesCS1B02G102700	N/A	N/A	N/A	N/A	N/A	Aminotran_1_2 domain-containing protein	Aminotransferase, ***
TraesCS1B02G118900	N/A	N/A	N/A	N/A	N/A	Chromatin-remodeling complex ATPase	Chromatin remodeling factor, putative ***
TraesCS1B02G147300	N/A	N/A	N/A	N/A	N/A	PMR5N domain-containing protein	Trichome birefringence-like protein, ***
TraesCS1B02G051700	N/A	N/A	N/A	N/A	N/A	Morc6_S5 domain-containing protein	MORC family CW-type zinc finger protein 4, ***
TraesCS1B02G105100	XP_003568961.1	ribosome biogenesis protein NOP53 [Brachypodium distachyon]	***-	587, 100%, 0.0, Y	N/A	Ribosome biogenesis protein NOP53	Glioma tumor suppressor-like protein, ***
TraesCS1B02G048800	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	BRI1-KD interacting protein, ***
TraesCS1B02G102600	N/A	N/A	N/A	N/A	N/A	N-acetyltransferase domain-containing protein	N-acetyltransferase, putative, ***
TraesCS1B02G119100	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Transport inhibitor response 1, ***

TraesCS1B02G121500	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Pre-rRNA-processing protein TSR2, ***
TraesCS1B02G142200	N/A	N/A	N/A	N/A	N/A	Derlin	Derlin, ***
TraesCS1B02G130500	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	centrosomal protein of 135 kDa-like protein, ***
TraesCS1B02G108700	N/A	N/A	N/A	N/A	N/A	SUN domain-containing protein	SAD1/UNC-84 domain protein, putative, --*
TraesCS1B02G128100	N/A	N/A	N/A	N/A	N/A	Serine/threonine-protein phosphatase, EC 3.1.3.16	Serine/threonine-protein phosphatase, ***
TraesCS1B02G100700	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	Peptide chain release factor 1, --*
TraesCS1B02G060100	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	E3 UFM1-protein ligase 1-like protein, *-*
TraesCS1B02G130800	N/A	N/A	N/A	N/A	N/A	GTP diphosphokinase, EC 2.7.6.5	RelA/SpoT-like protein, *-*
TraesCS1B02G099800	N/A	N/A	N/A	N/A	N/A	4HBT domain-containing protein	Thioesterase family protein, ***
TraesCS1B02G142800	N/A	N/A	N/A	N/A	N/A	Peroxidase, EC 1.11.1.7	Peroxidase, ***
TraesCS1B02G096700	N/A	N/A	N/A	N/A	N/A	Uncharacterized protein	phosphoribosylformylglycinamidine synthase, ***

DIFFERENTIALLY EXPRESSED RYE CONTIGS

contig	blast-accession	human-readable-description	quality-code	quality-values	blast2	bauer	chromosome
Lo7_v2_contig_62501	KAF6983978.1	hypothetical protein CFC21_002053 [Triticum aestivum]	*-**	2109, 25%, 0.0, Y	N/A	DNA replication and repair recF, ***	1R
Lo7_v2_contig_2869707	VAI93343.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	140, 13%, 2e-48, Y	actin-related protein 2/3 complex subunit 3 [Brachypodium distachyon]	Actin-related protein 2/3 complex subunit 3, ***	6R
Lo7_v2_contig_87853	SPT16341.1	unnamed protein product [Triticum aestivum]	*-**	337, 20%, 2e-102, Y	N/A	N/A	N/A
Lo7_v2_contig_230940	XP_020198414.1	nucleolin-like [Aegilops tauschii subsp. tauschii]	*-**	73.9, 10%, 2e-12, Y	N/A	N/A	N/A
Lo7_v2_contig_1346697	XP_020150077.1	inversin-B-like [Aegilops tauschii subsp. tauschii]	*-**	330, 15%, 3e-161, Y	ankyrin repeat-containing protein ITN1 [Brachypodium distachyon]	Ankyrin repeat family protein, putative, ***	6R
Lo7_v2_contig_124264	XP_020154354.1	uncharacterized protein LOC109739699 [Aegilops tauschii subsp. tauschii]	*-**	94.4, 13%, 1e-21, Y	transposon protein, putative, unclassified [Oryza sativa Japonica Group]	N/A	N/A
Lo7_v2_contig_2706128	XP_020173519.1	uncharacterized protein LOC109759094 [Aegilops tauschii subsp. tauschii]	****	267, 97%, 6e-83, Y	Rim2 protein [Oryza sativa Indica Group]/TPA: transposase [Oryza sativa]	N/A	N/A
Lo7_v2_contig_1350291	EMS59497.1	ABC transporter G family member 37 [Triticum urartu]	*-**	89.7, 11%, 2e-33, Y	N/A	N/A	7R
Lo7_v2_contig_52364	N/A	N/A	N/A	N/A	N/A	N/A	7R

Lo7_v2_contig_6239	VAI36291.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	145, 5%, 7e-35, Y	BTB/POZ domain-containing protein At1g30440 [Aegilops tauschii subsp. tauschii]	N/A	N/A
Lo7_v2_contig_1360483	EMS53266.1	Disease resistance protein RPM1 [Triticum urartu]	*-**	702, 15%, 0.0, Y	N/A	Ankyrin repeat domain-containing protein 2, ***	N/A
Lo7_v2_contig_126289	KAF7067932.1	hypothetical protein CFC21_073747 [Triticum aestivum]	*-**	570, 32%, 0.0, Y	N/A	N/A	N/A
Lo7_v2_contig_62387	KAE8812924.1	Receptor-like protein kinase HSL1 [Hordeum vulgare]	*-**	86.3, 15%, 3e-59, Y	N/A	N/A	N/A
Lo7_v2_contig_1358729	VAI69795.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	73.6, 19%, 1e-16, Y	putative disease resistance RPP8-like protein 4 [Triticum urartu]	N/A	N/A
Lo7_v2_contig_4807	EMS58960.1	hypothetical protein TRIUR3_19133 [Triticum urartu]	*--*	50.8, 3%, 3e-04, Y	N/A	N/A	4R
Lo7_v2_contig_143345	KAF7012669.1	hypothetical protein CFC21_026836 [Triticum aestivum]	*-**	1547, 56%, 0.0, Y	probable LRR receptor-like serine/threonine-protein kinase At3g47570 [Aegilops tauschii subsp. tauschii]	N/A	N/A
Lo7_v2_contig_1366296	XP_020153283.1	uncharacterized protein LOC109738600 [Aegilops tauschii subsp. tauschii]	*-**	163, 21%, 1e-107, Y	transposon protein, putative, Mutator sub-class [Oryza sativa Japonica Group]	N/A	1R
Lo7_v2_contig_374712	XP_020165672.1	uncharacterized protein LOC109751192 [Aegilops tauschii subsp. tauschii]	*-**	622, 12%, 0.0, Y	F-box protein At5g49610-like [Aegilops tauschii subsp. tauschii]	N/A	N/A
Lo7_v2_contig_1379005	VAH54650.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	113, 22%, 4e-26, Y	putative N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase [Hordeum vulgare]	N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase, ***	2R
Lo7_v2_contig_2871610	KAF6989608.1	hypothetical protein CFC21_006926 [Triticum aestivum]	*-**	474, 32%, 0.0, Y	U-box domain-containing protein 57 [Triticum urartu]	Kinase family protein, ***	N/A

Lo7_v2_contig_1348975	XP_020195405.1	uncharacterized protein LOC109781215 [Aegilops tauschii subsp. tauschii]	*-**	286, 10%, 2e-108, Y	Serine carboxypeptidase-like 18 [Triticum urartu]	N/A	4R
Lo7_v2_contig_2870479	KAF7068198.1	hypothetical protein CFC21_073973 [Triticum aestivum]	*-**	612, 47%. 0.0, Y	N/A	N/A	N/A
Lo7_v2_contig_1353544	BAH79979.1	putative unclassified retrotransposon protein [Oryza sativa Indica Group]	*-*-	206, 28%, 0.0, N	putative non-LTR retroelement reverse transcriptase [Sorghum bicolor]	N/A	N/A
Lo7_v2_contig_9185	XP_020190265.1	wiskott-Aldrich syndrome protein homolog 1-like [Aegilops tauschii subsp. tauschii]	*-**	67.8, 21%, 1e-12, Y	N/A	N/A	N/A
Lo7_v2_contig_1364696	KAF7088973.1	hypothetical protein CFC21_092039 [Triticum aestivum]	****	584, 60%, 0.0, Y	4-hydroxyphenylpyruvate dioxygenase [Aegilops tauschii subsp. tauschii]	N/A	N/A
Lo7_v2_contig_222418	hypothetical protein C2845_PM09G02540 [Panicum miliaceum]	hypothetical protein C2845_PM09G02540 [Panicum miliaceum]	*--*	52.0, 5%, 2e-06, N	retrotransposon protein, putative, Ty1-copia subclass, expressed [Oryza sativa Japonica Group]	N/A	N/A
Lo7_v2_contig_1353729	KAE8771443.1	Heterogeneous nuclear ribonucleoprotein 27C [Hordeum vulgare]	*-**	459, 14%, 3e-146, Y	peptide chain release factor PrfB3, chloroplastic [Oryza sativa Japonica Group]	Peptide chain release factor 2, ***	N/A
Lo7_v2_contig_61381	VAI86876.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	463, 7%, 2e-147, Y	pathogenesis-related protein 5-like [Aegilops tauschii subsp. tauschii]	Pathogenesis-related thaumatin-like protein, ***	N/A
Lo7_v2_contig_3953	KAF7073925.1	hypothetical protein CFC21_078845 [Triticum aestivum]	*-**	831, 16%, 0.0, Y	N/A	N/A	5R
Lo7_v2_contig_1363468	AKE47417.1	hypothetical protein TAANSRALLhA_1740J17.g00004 [Triticum aestivum]	*-**	92.0, 4%, 1e-18, Y	N/A	N/A	N/A
Lo7_v2_contig_78076	KAE8805228.1	putative cyclin-dependent kinase F-2 [Hordeum vulgare]	*-**	661, 25%, 0.0, Y	N/A	Kinase-like, ***	N/A

Lo7_v2_contig_1372743	KAE8781419.1	hypothetical protein D1007_45278 [Hordeum vulgare]	*-**	145, 9%, 3e-38 , Y	Hydroxymethylglutaryl-CoA lyase, mitochondrial [Hordeum vulgare]/NHL repeat-containing protein 2 [Hordeum vulgare]	N/A	N/A
Lo7_v2_contig_1348151	AAQ56285.1	putative gag-pol protein [Oryza sativa Japonica Group]	*-*-	116, 6%, 9e-40, N	N/A	MADS-box transcription factor 1, ***	5R
Lo7_v2_contig_1348527	XP_024316431.1	uncharacterized protein LOC100839419 isoform X2 [Brachypodium distachyon]	*-*-	135, 4%, 6e-52, N	N/A	N/A	N/A
Lo7_v2_contig_2641	VAI02585.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	1232, 21%, 0.0, Y	calcium-transporting ATPase 2, plasma membrane-type-like isoform X2 [Aegilops tauschii subsp. tauschii]	N/A	4R
Lo7_v2_contig_64434	SPT15596.1	unnamed protein product [Triticum aestivum]	*-**	438, 12%, 1e-134, Y	probable membrane-associated kinase regulator 4 [Aegilops tauschii subsp. tauschii]	N/A	1R
Lo7_v2_contig_1347313	XP_020190442.1	uncharacterized protein LOC109776199 [Aegilops tauschii subsp. tauschii]	*-**	337, 7%, 3e-102, Y	N/A	Unknown protein	N/A
Lo7_v2_contig_59883	XP_020163190.1	uncharacterized protein LOC109776199 [Aegilops tauschii subsp. tauschii]	*-**	622, 21%, 0.0, Y	N/A	F-box domain containing protein, expressed, ***	6R
Lo7_v2_contig_377814	N/A	N/A	N/A		N/A	N/A	N/A
Lo7_v2_contig_208418	XP_020180543.1	uncharacterized protein LOC109766178 [Aegilops tauschii subsp. tauschii]	*-**	331, 25%, 6e-103, Y	N/A	N/A	N/A
Lo7_v2_contig_1390307	XP_020195337.1	actin cytoskeleton-regulatory complex protein pan1-like [Aegilops tauschii subsp. tauschii]	*-**	66.2, 12%, 9e-13, Y	N/A	N/A	N/A
Lo7_v2_contig_1346943	XP_020187078.1	uncharacterized protein LOC109772792 [Aegilops tauschii subsp. tauschii]	*-**	485, 14%, 0.0, Y	N/A	N/A	N/A

Lo7_v2_contig_264947	VAH22123.1	unnamed protein product [Triticum turgidum subsp. durum]	*--*	51.2, 2%, 1e-05 , Y	Histone H2B.2 [Hordeum vulgare]	N/A	1R
Lo7_v2_contig_62402	XP_020193939.1	BOI-related E3 ubiquitin-protein ligase 1-like [Aegilops tauschii subsp. tauschii]	*-**	339, 8%, 7e-103, Y	N/A	N/A	N/A
Lo7_v2_contig_2870870	KAE8794488.1	Homeobox protein KNOX3 [Hordeum vulgare]	*-**	180, 29%, 3e-50, Y	N/A	N/A	N/A
Lo7_v2_contig_123975	EMS49933.1	hypothetical protein TRIUR3_25873 [Triticum urartu]	\---*	46.2, 4%, 3e-04, Y	N/A	N/A	N/A
Lo7_v2_contig_1355103	SPT21164.1	unnamed protein product [Triticum aestivum]	*-**	679, 26%, 0.0, Y	retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]	N/A	N/A
Lo7_v2_contig_69071	XP_020189365.1	uncharacterized protein LOC109775022 [Aegilops tauschii subsp. tauschii]	*-**	196, 35%, 5e-56, Y	N/A	N/A	N/A
Lo7_v2_contig_63334	XP_020177626.1	uncharacterized protein LOC109763185 [Aegilops tauschii subsp. tauschii]	*-**	580, 12%, 0.0, Y	N/A	N/A	N/A
Lo7_v2_contig_2994	BAJ99011.1	predicted protein [Hordeum vulgare subsp. vulgare]	*-**	331, 9%, 5e-100, Y	N/A	Cyclin-like protein, putative, ***	7R
Lo7_v2_contig_70724	KAE8805831.1	Alpha-amylase [Hordeum vulgare]	*-**	333, 16%, 1e-103, Y	N/A	Protein kinase G11A, ***	N/A
Lo7_v2_contig_1347152	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Lo7_v2_contig_2872298	N/A	N/A	N/A	N/A	N/A	N/A	1R
Lo7_v2_contig_1355711	VAH41411.1	unnamed protein product [Triticum turgidum subsp. durum]	*--*	60.8, 11%, 2e-08, Y	N/A	N/A	N/A

Lo7_v2_contig_230954	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Lo7_v2_contig_2673288	VAI58481.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	66.2, 33%, 6e-11, Y	N/A	N/A	N/A
Lo7_v2_contig_59538	VAI67989.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	286, 8%, 2e-117, Y	GDSL esterase/lipase At5g45950-like [Aegilops tauschii subsp. tauschii]	GDSL esterase/lipase, ***	N/A
Lo7_v2_contig_1405567	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Lo7_v2_contig_1375794	XP_020148526.1	uncharacterized protein LOC109733718 [Aegilops tauschii subsp. tauschii]	*-**	883, 47%, 0.0, Y	putative reverse transcriptase [Sorghum bicolor]	N/A	N/A
Lo7_v2_contig_1354240	KAF7050430.1	hypothetical protein CFC21_058801 [Triticum aestivum]	*--*	54.3, 5%, 8e-06, Y	N/A	N/A	N/A
Lo7_v2_contig_11315	XP_020147081.1	uncharacterized protein LOC109732305 [Aegilops tauschii subsp. tauschii]	*-**	530, 18%, 4e-169, Y	N/A	N/A	N/A
Lo7_v2_contig_1357725	KAF7111876.1	hypothetical protein CFC21_111831 [Triticum aestivum]	*-**	564, 22%, 0.0, Y	glutelin type-A 1-like [Hordeum vulgare]	11S seed storage protein, ***	N/A
Lo7_v2_contig_2872626	N/A	N/A	N/A	N/A	N/A	N/A	1R
Lo7_v2_contig_61750	ACN88793.1	putative polyprotein [Secale cereale]	*-**	295, 6%, 1e-139, Y	Putative retroelement [Oryza sativa Japonica Group]	N/A	N/A
Lo7_v2_contig_145997	XP_020160305.1	uncharacterized protein LOC109745593 [Aegilops tauschii subsp. tauschii]	*-**	98.6, 8%, 2e-19, Y	N/A	N/A	N/A
Lo7_v2_contig_64956	KAF7027069.1	hypothetical protein CFC21_039141 [Triticum aestivum]	*-**	1008, 41%, 0.0, Y	putative ubiquitin carrier protein E2 23 [Hordeum vulgare]	N/A	N/A
Lo7_v2_contig_6491	EMS68313.1	putative polyamine oxidase 4 [Triticum urartu]	*-**	251, 27%,	N/A	Lysine-specific histone demethylase 1, ***	N/A

				6e-138, Y			
Lo7_v2_contig_1350506	KAF7012839.1	hypothetical protein CFC21_026988 [Triticum aestivum]	*-**	341, 25%, 2e-101, Y	(E)-beta-farnesene synthase [Setaria italica]	Terpenoid cyclases/Protein prenyltransferases superfamily protein LENGTH=603, ***	
Lo7_v2_contig_427775	SPT17925.1	unnamed protein product [Triticum aestivum]	*-**	154, 20%, 1e-40, Y	N/A	N/A	N/A
Lo7_v2_contig_63252	VAH55546.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	362, 17%, 2e-108, Y	N/A	O-acyltransferase WSD1, ***	3R
Lo7_v2_contig_204561	AAA66167.1	unknown protein [Triticum urartu]	*-**	81.3, 20%, 6e-17, Y	N/A	N/A	N/A
Lo7_v2_contig_66476	CAB3493741.1	unnamed protein product [Digitaria exilis]	*-**	116, 15%, 2e-27, Y	ADP-ribosylation factor-like protein 2 [Zea mays]	ADP-ribosylation factor, ***	N/A
Lo7_v2_contig_584464	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Lo7_v2_contig_1358326	KAE8805029.1	hypothetical protein D1007_19052 [Hordeum vulgare]	*-**	82.4, 9%, 2e-21, Y	putative methionyl-tRNA synthetase [Hordeum vulgare]	N/A	N/A
Lo7_v2_contig_123999	SPT18624.1	unnamed protein product [Triticum aestivum]	*-**	80.1, 30%, 9e-25, Y	putative F-box/FBD/LRR-repeat protein At5g52460 [Aegilops tauschii subsp. tauschii]	N/A	N/A
Lo7_v2_contig_1355255	AAV80394.1	polyprotein [Hordeum vulgare subsp. vulgare]	*-**	378, 55%, 0.0, Y	retrotransposon protein, putative, Ty3-gypsy subclass [Oryza sativa Japonica Group]	N/A	N/A
Lo7_v2_contig_1380475	KAE8805271.1	hypothetical protein D1007_18675 [Hordeum vulgare]	*-**	425, 27%, 6e-140, Y	N/A	N/A	N/A

Lo7_v2_contig_146164	XP_020156837.1	G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130 [Aegilops tauschii subsp. tauschii]	*-**	1353, 33%, 0.0, Y	N/A	N/A	N/A
Lo7_v2_contig_2871655	VAH38114.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	552, 14%, 8e-177, Y	F-box protein At3g07870-like [Aegilops tauschii subsp. tauschii]/Programmed cell death protein 2-like protein [Hordeum vulgare]	Programmed cell death protein 2-like protein, ***	2R
Lo7_v2_contig_1344244	XP_020196522.1	uncharacterized protein LOC109782328 [Aegilops tauschii subsp. tauschii]	*-**	95.9, 8%, 2e-45, Y	putative nuclease HARBI1 [Aegilops tauschii subsp. tauschii]	N/A	N/A
Lo7_v2_contig_141118	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Lo7_v2_contig_1386227	ABA96959.2	retrotransposon protein, putative, Ty1-copia subclass [Oryza sativa Japonica Group]	*-**	320, 14%, 4e-90, Y	N/A	N/A	N/A
Lo7_v2_contig_1405569	ABE60891.1	putative polyprotein [Oryza sativa Japonica Group]	*-**	407, 59%, 3e-171, Y	N/A	N/A	N/A
Lo7_v2_contig_2492	VAI55040.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	1176, 32%, 0.0, Y	putative histone acetyltransferase HAC-like 1 [Triticum urartu]	Histone acetyltransferase, ***	6R
Lo7_v2_contig_262970	KAE8783768.1	hypothetical protein D1007_42748 [Hordeum vulgare]	*-**	546, 18%, 8e-176, Y	N/A	Protein of unknown function (DUF793) LENGTH=382, ***	N/A
Lo7_v2_contig_61289	SPT18691.1	unnamed protein product [Triticum aestivum]	*-**	368, 19%, 0.0, Y	N/A	N/A	N/A
Lo7_v2_contig_1358116	XP_020151355.1	protein trichome birefringence-like 9 [Aegilops tauschii subsp. tauschii]	*-**	122, 15%, 2e-34, Y	ethylene-responsive transcription factor ERF110-like [Aegilops tauschii subsp. tauschii]	N/A	N/A
Lo7_v2_contig_1364344	KAF7099606.1	hypothetical protein CFC21_101227 [Triticum aestivum]	*--*	65.9, 17%,	N/A	N/A	N/A

				2e-09, Y			
Lo7_v2_contig_124266	EMS48257.1	Protein IN2-1-like protein B [Triticum urartu]	*-**	67.8, 5%, 5e-10, Y	N/A	N/A	N/A
Lo7_v2_contig_70563	VAI12735.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	280, 9%, 2e-111, Y	N/A	Unknown protein	N/A
Lo7_v2_contig_139988	EMS48183.1	hypothetical protein TRIUR3_10502 [Triticum urartu]	*-**	73.6, 10%, 4e-12, Y	retrotransposon protein, putative, Ty3-gypsy subclass [Oryza sativa Japonica Group]	N/A	N/A
Lo7_v2_contig_1358102	XP_020182672.1	probable disease resistance RPP8-like protein 4 [Aegilops tauschii subsp. tauschii]	*-**	790, 39%, 0.0, Y	N/A	N/A	N/A
Lo7_v2_contig_6407	XP_020148001.1	uncharacterized protein LOC109733211 [Aegilops tauschii subsp. tauschii]	*-**	57.0, 3%, 3e-06, Y	Serine/threonine-protein kinase SMG1 [Hordeum vulgare]	N/A	N/A
Lo7_v2_contig_112373	AAO66559.1	putative copia-type pol polyprotein [Oryza sativa Japonica Group]	*-**	307, 26%, 3e-153, Y	N/A	N/A	N/A
Lo7_v2_contig_11782	KAF7080142.1	hypothetical protein CFC21_084265 [Triticum aestivum]	*-**	199, 22%, 3e-70, Y	enhancer of mRNA-decapping protein 4-like [Hordeum vulgare]	N/A	6R
Lo7_v2_contig_59943	VAH20602.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	1486, 20%, 0.0, Y	G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130 [Aegilops tauschii subsp. tauschii]	Serine/threonine-protein kinase, ***	1R
Lo7_v2_contig_145050	BAJ92268.1	predicted protein [Hordeum vulgare subsp. vulgare]	*-**	103, 7%, 3e-23, Y	N/A	N/A	3R
Lo7_v2_contig_85958	XP_020194961.1	uncharacterized protein LOC109780791 [Aegilops tauschii subsp. tauschii]	\---*	43.1, 20%, 0.022, Y	N/A	N/A	N/A

Lo7_v2_contig_2808155	AEW46799.1	gamma prolamin [Secale cereale subsp. afghanicum]	*-**	64.3, 28%, 3e-12, Y	gamma-gliadin [Triticum aestivum]	N/A	N/A
Lo7_v2_contig_2875859	EMS58901.1	hypothetical protein TRIUR3_27333 [Triticum urartu]	*-**	86.3, 5%, 2e-17, Y	putative methyltransferase At1g22800 [Aegilops tauschii subsp. tauschii]	N/A	N/A

contig	blast-accession	human-readable-description	quality-code	quality-values	blast2	bauer	chromosome
Lo7_v2_contig_62501	KAF6983978.1	hypothetical protein CFC21_002053 [Triticum aestivum]	*-**	2109, 25%, 0.0, Y	N/A	DNA replication and repair recF, ***	1R
Lo7_v2_contig_1366296	XP_020153283.1	uncharacterized protein LOC109738600 [Aegilops tauschii subsp. tauschii]	*-**	163, 21%, 1e-107, Y	transposon protein, putative, Mutator sub-class [Oryza sativa Japonica Group]	N/A	1R
Lo7_v2_contig_64434	SPT15596.1	unnamed protein product [Triticum aestivum]	*-**	438, 12%, 1e-134, Y	probable membrane-associated kinase regulator 4 [Aegilops tauschii subsp. tauschii]	N/A	1R
Lo7_v2_contig_264947	VAH22123.1	unnamed protein product [Triticum turgidum subsp. durum]	*--*	51.2, 2%, 1e-05 , Y	Histone H2B.2 [Hordeum vulgare]	N/A	1R
Lo7_v2_contig_2872298	N/A	N/A	N/A	N/A	N/A	N/A	1R
Lo7_v2_contig_2872626	N/A	N/A	N/A	N/A	N/A	N/A	1R

Lo7_v2_contig_59943	VAH20602.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	1486, 20%, 0.0, Y	G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130 [Aegilops tauschii subsp. tauschii]	Serine/threonine-protein kinase, ***	1R
Lo7_v2_contig_1347926	KAF6986909.1	hypothetical protein CFC21_004604 [Triticum aestivum]	*-**	694, 27%, 0.0, Y	N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages. LENGTH=690, *-*	1R
Lo7_v2_contig_1358115	XP_015613671.1	BTB/POZ and MATH domain-containing protein 1 [Oryza sativa Japonica Group]	*-*-	233, 23%, 1e-67, N	N/A	BTB/POZ domain containing protein, ***	1R
Lo7_v2_contig_1350507	VAH22617.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	423, 8%, 4e-134, Y	N/A	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein, *-*	1R
Lo7_v2_contig_377813	KAF6992766.1	hypothetical protein CFC21_009729 [Triticum aestivum]	*-**	297, 32%, 6e-90, Y	formin-like protein 14 [Hordeum vulgare]	Formin-like protein, ***	1R
Lo7_v2_contig_2871483	XP_020193577.1	uncharacterized protein LOC109779371 isoform X2 [Aegilops tauschii subsp. tauschii]	*-**	292, 6%, 7e-82, Y	N/A	Unknown protein	1R
Lo7_v2_contig_1365155	XP_020186178.1	uncharacterized protein LOC109771898 [Aegilops tauschii subsp. tauschii]	*-**	217, 9%, 3e-62, Y	lactoylglutathione lyase [Hordeum vulgare]	Lactoylglutathione lyase / glyoxalase I family protein, ***	1R
Lo7_v2_contig_3439	VAH04317.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	165, 15%, 4e-39, Y	Transportin-3 [Triticum urartu]	N/A	1R

Lo7_v2_contig_234473	N/A	N/A	N/A	N/A	N/A	N/A	1R
Lo7_v2_contig_1361945	XP_02001983 17.1	ABC transporter F family member 4-like [Aegilops tauschii subsp. tauschii]	*-**	1111, 20%, 0.0, Y	N/A	ATP-binding cassette transporter, ***	1R
Lo7_v2_contig_21711	N/A	N/A	N/A	N/A	N/A	N/A	1R
Lo7_v2_contig_60157	KAF6 983787.1	hypothetical protein CFC21_001903 [Triticum aestivum]	*-**	590, 6%, 0.0, Y	zinc finger HIT domain-containing protein 2 isoform X1 [Aegilops tauschii subsp. tauschii]	Zinc finger HIT domain-containing 2, ***	1R
Lo7_v2_contig_60319	VAH1 3453.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	1464, 40%, 0.0, Y	E3 ubiquitin-protein ligase UPL4-like [Aegilops tauschii subsp. tauschii]	Tudor/PWWP/MBT superfamily protein, *-*	1R
Lo7_v2_contig_62998	VAH2 0705.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	253, 30%, 6e-72, Y	histone deacetylase 5-like [Aegilops tauschii subsp. tauschii]	Histone deacetylase 6, ***	1R
Lo7_v2_contig_566693	VAH2 1070.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	158, 16%, 5e-39, Y	carnosine N-methyltransferase-like [Aegilops tauschii subsp. tauschii]	N/A	1R
Lo7_v2_contig_278194	VAH9 5365.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	117, 8%, 2e-26, Y	lipoxygenase 3 [Triticum aestivum]	N/A	1R
Lo7_v2_contig_62904	XP_02001644 50.1	mixed-linked glucan synthase 2-like [Aegilops tauschii subsp. tauschii]	*-**	1048, 23%, 0, Y	N/A	N/A	1R
Lo7_v2_contig_63454	KAF6 994190.1	hypothetical protein CFC21_010947 [Triticum aestivum]	*-**	135, 9%, 7e-34, Y	40S ribosomal protein S7 [Aegilops tauschii subsp. tauschii]	40S ribosomal protein S7, ***	1R

Lo7_v2_contig_154997	EMS60767.1	hypothetical protein TRIUR3_28838 [Triticum urartu]	*-**	56.6, 6%, 6e-07, Y	N/A	N/A	1R
Lo7_v2_contig_179456	SPT17767.1	unnamed protein product [Triticum aestivum]	*-**	87.4, 42%, 8e-20, Y	Iron sulfur cluster assembly protein 1, mitochondrial [Triticum urartu]	N/A	1R
Lo7_v2_contig_1360047	VAH02851.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	264, 12%, 9e-80, Y	PHD finger protein ALFIN-LIKE 1 [Aegilops tauschii subsp. tauschii]	PHD finger protein ALFIN-LIKE 2, ***	1R
Lo7_v2_contig_1489150	XP_02014916 4.1	uncharacterized protein DDB_G0286299-like [Aegilops tauschii subsp. tauschii]	*** *	362, 79%, 2e-121, Y	titin-like [Aegilops tauschii subsp. tauschii]	N/A	1R
Lo7_v2_contig_1345370	KAF6997447.1	hypothetical protein CFC21_013666 [Triticum aestivum]	*-**	177, 6%, 6e-48, Y	N/A	N/A	1R
Lo7_v2_contig_262762	VAH18435.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	77, 19%, 1e-69, Y	glyoxysomal fatty acid beta-oxidation multifunctional protein MFP-a-like [Aegilops tauschii subsp. tauschii]	Fatty acid oxidation complex subunit alpha, ***	1R
Lo7_v2_contig_6835	KAF6996721.1	hypothetical protein CFC21_013032 [Triticum aestivum]	*** *	1095, 54%, 0.0, Y	N/A	Unknown protein	1R
Lo7_v2_contig_87727	XP_02017174 4.1	uncharacterized protein LOC109757330 [Aegilops tauschii subsp. tauschii]	*-**	238, 15%, 4e-65, Y	ABC transporter G family member 37 [Hordeum vulgare]	Unknown protein	1R
Lo7_v2_contig_60329	KAE880055 1.1	hypothetical protein D1007_24051 [Hordeum vulgare]	*-**	296, 7%,	Isopenicillin N epimerase [Hordeum vulgare]	N/A	1R

				2e-85, Y			
Lo7_v2_contig_2883402	XP_024318271.1	DNA mismatch repair protein MSH5 isoform X5 [Brachypodium distachyon]	*-**	64.3, 24%, 2e-20, Y	N/A	N/A	1R
Lo7_v2_contig_1370013	VAI33703.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	535, 26%, 7e-171, Y	N/A	N/A	1R
Lo7_v2_contig_2868832	KAF6997729.1	ethylene-responsive transcription factor ERF061 [Aegilops tauschii subsp. tauschii]	*-**	355, 6%, 4e-109, Y	AP2/ERF domain-containing transcription factor [Triticum turgidum subsp. durum]	N/A	1R
Lo7_v2_contig_2871363	PUZ68739.1	hypothetical protein GQ55_2G052800 [Panicum hallii var. hallii]	*-**	528, 33%, 2e-164, Y	wall-associated receptor kinase 1-like [Panicum hallii]	N/A	1R
Lo7_v2_contig_90867	VAH09720.1	unnamed protein product [Triticum turgidum subsp. durum]	*-**	427, 16%, 0.0, Y	putative glycerophosphoryl diester phosphodiesterase 3 [Triticum urartu]	Glycerophosphoryl diester phosphodiesterase 2-like protein, ***	1R

S3) List of input genes used for GO Enrichment Analysis

DOWNREGULATED WHEAT GENES
ER membrane protein complex subunit
Mitogen-activated protein kinase
Non-specific serine/threonine protein kinase
Ligase-like protein
Prolycopene isomerase
Ankyrin repeat domain-containing protein
DNA/RNA helicase
Serine/threonine-protein phosphatase
RNA polymerase II-associated protein
Ubiquitin-conjugating enzyme
RING/U-box superfamily protein
heat shock protein
E3 ubiquitin-protein ligase
bZIP transcription factor
ABC transporter protein
KAT8 regulatory NSL complex subunit
RING/U-box superfamily protein
Coiled-coil domain-containing protein
CULLIN_2 domain-containing protein
G-patch domain containing protein
Golgi apparatus membrane protein
CULLIN_2 domain-containing protein
Zinc ion binding protein
Diacylglycerol acyltransferase
Starch synthase family protein
NFACT-R_1 domain-containing protein
Sodium/hydrogen exchanger
Importin subunit alpha
FACT complex subunit
RING/U-box superfamily protein
Chaperone protein
Malic enzyme
Acyl-CoA dehydrogenase
Fe/S biogenesis protein
Chaperone protein
Myb/SANT-like DNA-binding domain protein

RNA-binding family protein
Protein apaG
Aspartic proteinase Nudix hydrolase
Fibronectin type-III domain-containing protein
Ras-like protein
Zinc finger family protein
Nucleic acid-binding OB-fold-like protein
Agenet domain containing protein
Peptide chain release factor
Adenylyl cyclase-associated protein
Phosphoglycolate phosphatase
Proteasome subunit beta type
Transducin/WD-like repeat-protein
Glycine-rich protein
Autophagy-related protein
O-acyltransferase
Kinesin-like protein
RNA-binding family protein
Actin
Metal tolerance protein
Vacuolar-sorting receptor-like protein
F-box domain-containing protein
Non-specific serine/threonine protein kinase
alpha-12-Mannosidase
Ethylene receptor
Pre-mRNA-splicing factor
E3 SUMO-protein ligase
UDP galactose transporter-related protein
Calmodulin
Replication protein A subunit
N-acetyltransferase domain-containing protein
Protein kinase domain-containing protein
Double stranded RNA binding protein
ATP-dependent zinc metalloprotease
Late embryogenesis abundant protein
Arginine/serine-rich splicing factor
Beta-carotene isomerase
Tubulin-specific chaperone cofactor
Methyl-CpG-binding domain protein

Methionine S-methyltransferase
Dual specificity phosphatase
O-acyltransferase
Protein disulfide-isomerase
Aminotransferase
Chromatin-remodeling complex ATPase
Trichome birefringence-like protein
MORC family CW-type zinc finger protein
Ribosome biogenesis protein
BRI1-KD interacting protein
N-acetyltransferase domain-containing protein
Transport inhibitor response
Pre-rRNA-processing protein
Derlin
Centrosomal protein
SUN domain-containing protein
Serine/threonine-protein phosphatase
Peptide chain release factor
protein ligase 1-like protein
GTP diphosphokinase
Thioesterase family protein
Peroxidase
Phosphoribosylformylglycinamidine synthase

UPREGULATED RYE GENES
DNA replication and repair
Actin-related protein
nucleolin-like
Ankyrin repeat family protein
transposon protein
Rim2
ABC transporter G
BTB/POZ domain-containing protein
Disease resistance protein RPM1
Ankyrin repeat domain-containing protein
Receptor-like protein kinase
disease resistance protein RPP8
LRR receptor-like serine/threonine-protein kinase

Transposon protein
F-box protein-like
N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase
U-box domain-containing kinase
Serine carboxypeptidase-like
non-LTR retroelement reverse transcriptase
wiskott-Aldrich syndrome protein homolog
1-like4-hydroxyphenylpyruvate dioxygenase
retrotransposon protein
Peptide chain release factor
Pathogenesis-related thaumatin-like protein
Cyclin-dependent kinase
Hydroxymethylglutaryl-CoA lyase
MADS-box transcription factor
calcium-transporting ATPase
membrane-associated kinase regulator
F-box domain containing protein
Actin cytoskeleton-regulatory complex protein
Histone H2B.2
BOI-related E3 ubiquitin-protein ligase 1-like
Homeobox protein KNOX3
retrotransposon protein
Cyclin-like protein
Protein kinase G11A
GDSL esterase/lipase
Reverse transcriptase
11S seed storage protein
Polyprotein
Ubiquitin carrier protein
Lysine-specific histone demethylase
Terpenoid cyclase
O-acyltransferase
ADP-ribosylation factor
Methionyl-tRNA synthetase
F-box/FBD/LRR-repeat protein
Polyprotein
G-type lectin S-receptor-like serine/threonine-protein kinase
Programmed cell death protein 2-like protein
nuclease HARBI1

Retrotransposon protein
Polyprotein
Histone acetyltransferase
ethylene-responsive transcription factor ERF110-like
Protein IN2-1-like protein B
Retrotransposon protein
Disease resistance RPP8-like protein
Serine/threonine-protein kinase SMG1
Polyprotein
Enhancer of mRNA-decapping protein 4-like
G-type lectin S-receptor-like serine/threonine-protein kinase
gamma gliadin
methyltransferase
BTB/POZ domain containing protein
Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
Formin-like protein
Lactoylglutathione lyase
Transportin
ATP-binding cassette transporter
Zinc finger HIT domain-containing
E3 ubiquitin-protein ligase
Histone deacetylase
carnosine N-methyltransferase-like
lipoxygenase
mixed-linked glucan synthase 2-like
40S ribosomal protein
Iron sulfur cluster assembly protein
PHD finger protein ALFIN-LIKE
titin-like
Fatty acid oxidation complex subunit alpha
ABC transporter G family protein
Isopenicillin N epimerase
DNA mismatch repair protein MSH5 isoform
ethylene-responsive transcription factor ERF061
wall-associated receptor kinase-like
Glycerophosphoryl diester phosphodiesterase-like

S4) Output of GO Enrichment Analysis

DOWNREGULATED WHEAT GENES - BIOLOGICAL PROCESS:

Analysis Type: PANTHER Overrepresentation Test (Released 20200728)
Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4033054 Released 2020-09-10
Analyzed List: upload_1 (Triticum aestivum)
Reference List: Triticum aestivum (all genes in database)
Test Type: FISHER
Correction: BONFERRONI
Bonferroni count: 2196

GO biological process complete	Triticum aestivum - REFLIST (102802)	upload_1 (1148)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (P-value)
ubiquitin-dependent protein catabolic process via the N-end rule pathway (GO:0071596)	6	6	.07	+	89.55	3.43E-06
hydrogen peroxide catabolic process (GO:0042744)	675	610	7.54	+	80.93	0.00E00
hydrogen peroxide metabolic process (GO:0042743)	675	610	7.54	+	80.93	0.00E00
reactive oxygen species metabolic process (GO:0072593)	719	610	8.03	+	75.97	0.00E00
cellular oxidant detoxification (GO:0098869)	918	610	10.25	+	59.50	0.00E00
response to oxidative stress (GO:0006979)	957	611	10.69	+	57.17	0.00E00
cellular response to toxic substance (GO:0097237)	960	610	10.72	+	56.90	0.00E00
cellular detoxification (GO:1990748)	960	610	10.72	+	56.90	0.00E00
detoxification (GO:0098754)	970	610	10.83	+	56.31	0.00E00
response to toxic substance (GO:0009636)	981	610	10.95	+	55.68	0.00E00
malate metabolic process (GO:0006108)	38	16	.42	+	37.70	1.38E-15
NLS-bearing protein import into nucleus (GO:0006607)	19	8	.21	+	37.70	8.74E-07
MAPK cascade (GO:0000165)	105	44	1.17	+	37.53	3.36E-46
sodium ion transmembrane transport (GO:0035725)	29	12	.32	+	37.05	4.02E-11
sodium ion transport (GO:0006814)	30	12	.34	+	35.82	5.57E-11
sodium ion import across plasma membrane (GO:0098719)	27	10	.30	+	33.17	1.52E-08
telomere maintenance via telomerase (GO:0007004)	41	14	.46	+	30.58	2.33E-12
microtubule-based movement (GO:0007018)	166	56	1.85	+	30.21	4.62E-55
inorganic cation import across plasma membrane (GO:0098659)	30	10	.34	+	29.85	3.58E-08
inorganic ion import across plasma membrane (GO:0099587)	30	10	.34	+	29.85	3.58E-08
signal transduction by protein phosphorylation (GO:0023014)	151	50	1.69	+	29.65	1.35E-48
import across plasma membrane (GO:0098739)	31	10	.35	+	28.89	4.69E-08
cellular response to chemical stimulus (GO:0070887)	1995	635	22.28	+	28.50	0.00E00
telomere maintenance via telomere lengthening (GO:0010833)	44	14	.49	+	28.49	5.27E-12
cellular response to nitrogen starvation (GO:0006995)	22	6	.25	+	24.42	1.20E-03
DNA unwinding involved in DNA replication (GO:0006268)	52	14	.58	+	24.11	3.71E-11
cellular response to ethylene stimulus (GO:0071369)	48	12	.54	+	22.39	5.88E-09

histone H3-K4 methylation (GO:0051568)	24	6	.27	+	22.39	1.86E-03	
ethylene-activated signaling pathway (GO:0009873)	48	12	.54	+	22.39	5.88E-09	
movement of cell or subcellular component (GO:0006928)	226	56	2.52	+	22.19	9.30E-49	
cellular response to nitrogen levels (GO:0043562)	25	6	.28	+	21.49	2.29E-03	
response to chemical (GO:0042221)	2762	639	30.84	+	20.72	0.00E00	
RNA-dependent DNA biosynthetic process (GO:0006278)	65	14	.73	+	19.29	5.22E-10	
root development (GO:0048364)	29	6	.32	+	18.53	4.86E-03	
root system development (GO:0022622)	29	6	.32	+	18.53	4.86E-03	
import into cell (GO:0098657)	50	10	.56	+	17.91	2.61E-06	
cellular catabolic process (GO:0044248)	4286	749	47.86	+	15.65	0.00E00	
response to ethylene (GO:0009723)	71	12	.79	+	15.14	3.22E-07	
cellular response to stimulus (GO:0051716)	5210	842	58.18	+	14.47	0.00E00	
catabolic process (GO:0009056)	5046	749	56.35	+	13.29	0.00E00	
response to stress (GO:0006950)	4443	654	49.62	+	13.18	0.00E00	
multicellular organism development (GO:0007275)	898	127	10.03	+	12.66	1.63E-87	
regulation of intracellular pH (GO:0051453)	71	10	.79	+	12.61	5.27E-05	
regulation of cellular pH (GO:0030641)	71	10	.79	+	12.61	5.27E-05	
cellular response to organic cyclic compound (GO:0071407)	57	8	.64	+	12.57	1.37E-03	
brassinosteroid mediated signaling pathway (GO:0009742)	57	8	.64	+	12.57	1.37E-03	
cellular response to steroid hormone stimulus (GO:0071383)	57	8	.64	+	12.57	1.37E-03	
cellular response to brassinosteroid stimulus (GO:0071367)	57	8	.64	+	12.57	1.37E-03	
steroid hormone mediated signaling pathway (GO:0043401)	57	8	.64	+	12.57	1.37E-03	
response to steroid hormone (GO:0048545)	57	8	.64	+	12.57	1.37E-03	
anatomical structure homeostasis (GO:0060249)	100	14	1.12	+	12.54	9.17E-08	
telomere maintenance (GO:0000723)	100	14	1.12	+	12.54	9.17E-08	
telomere organization (GO:0032200)	100	14	1.12	+	12.54	9.17E-08	
response to brassinosteroid (GO:0009741)	59	8	.66	+	12.14	1.74E-03	
intracellular signal transduction (GO:0035556)	786	105	8.78	+	11.96	1.85E-69	
cellular monovalent inorganic cation homeostasis (GO:0030004)	76	10	.85	+	11.78	9.46E-05	
anatomical structure development (GO:0048856)	1010	130	11.28	+	11.53	4.35E-85	
response to organic cyclic compound (GO:0014070)	72	8	.80	+	9.95	6.81E-03	
response to stimulus (GO:0050896)	7734	845	86.37	+	9.78	0.00E00	
microtubule-based process (GO:0007017)	541	56	6.04	+	9.27	1.49E-30	
regulation of pH (GO:0006885)	125	12	1.40	+	8.60	1.07E-04	
nucleotide-excision repair (GO:0006289)	148	14	1.65	+	8.47	9.93E-06	
multicellular organismal process (GO:0032501)	1395	127	15.58	+	8.15	1.17E-66	
ubiquitin-dependent protein catabolic process (GO:0006511)	1450	132	16.19	+	8.15	1.92E-69	
protein dephosphorylation (GO:0006470)	519	47	5.80	+	8.11	5.26E-23	
monovalent inorganic cation homeostasis (GO:0055067)	133	12	1.49	+	8.08	2.01E-04	

cell redox homeostasis (GO:0045454)	145	13	1.62	+	8.03	6.33E-05					
DNA duplex unwinding (GO:0032508)	161	14	1.80	+	7.79	2.68E-05					
DNA geometric change (GO:0032392)	161	14	1.80	+	7.79	2.68E-05					
developmental process (GO:0032502)	1529	130	17.07	+	7.61	4.57E-65					
oxidation-reduction process (GO:0055114)	7562	626	84.45	+	7.41	0.00E00					
modification-dependent protein catabolic process (GO:0019941)	1628	132	18.18	+	7.26	7.67E-64					
DNA biosynthetic process (GO:0071897)	175	14	1.95	+	7.16	7.11E-05					
protein import into nucleus (GO:0006606)	126	10	1.41	+	7.11	6.91E-03					
modification-dependent macromolecule catabolic process (GO:0043632)	1681	132	18.77	+	7.03	2.65E-62					
import into nucleus (GO:0051170)	129	10	1.44	+	6.94	8.41E-03					
response to endoplasmic reticulum stress (GO:0034976)	235	18	2.62	+	6.86	1.54E-06					
dicarboxylic acid metabolic process (GO:0043648)	209	16	2.33	+	6.86	1.35E-05					
protein localization to nucleus (GO:0034504)	134	10	1.50	+	6.68	1.15E-02					
signal transduction (GO:0007165)	2764	202	30.87	+	6.54	5.11E-93					
signaling (GO:0023052)	2821	202	31.50	+	6.41	1.60E-91					
protein ubiquitination (GO:0016567)	1973	140	22.03	+	6.35	3.27E-61					
protein modification by small protein conjugation (GO:0032446)	2086	140	23.29	+	6.01	2.01E-58					
proteolysis involved in cellular protein catabolic process (GO:0051603)	2025	132	22.61	+	5.84	1.76E-53					
cellular protein catabolic process (GO:0044257)	2028	132	22.65	+	5.83	2.07E-53					
protein catabolic process (GO:0030163)	2076	132	23.18	+	5.69	2.55E-52					
cell communication (GO:0007154)	3356	208	37.48	+	5.55	1.11E-83					
peptidyl-serine phosphorylation (GO:0018105)	222	13	2.48	+	5.24	6.00E-03					
pyruvate metabolic process (GO:0006090)	277	16	3.09	+	5.17	5.25E-04					
double-strand break repair via homologous recombination (GO:0000724)	243	14	2.71	+	5.16	3.01E-03					
peptidyl-serine modification (GO:0018209)	226	13	2.52	+	5.15	7.22E-03					
protein modification by small protein conjugation or removal (GO:0070647)	2434	140	27.18	+	5.15	8.08E-51					
recombinational repair (GO:0000725)	260	14	2.90	+	4.82	6.35E-03					
cellular macromolecule catabolic process (GO:0044265)	2516	132	28.10	+	4.70	1.64E-43					
organonitrogen compound catabolic process (GO:1901565)	2584	132	28.86	+	4.57	2.56E-42					
dephosphorylation (GO:0016311)	1035	51	11.56	+	4.41	3.05E-14					
cellular homeostasis (GO:0019725)	473	23	5.28	+	4.35	2.89E-05					
DNA-dependent DNA replication (GO:0006261)	290	14	3.24	+	4.32	2.08E-02					
macromolecule catabolic process (GO:0009057)	2900	132	32.38	+	4.08	3.12E-37					
homeostatic process (GO:0042592)	871	39	9.73	+	4.01	3.29E-09					
cellular metabolic process (GO:0044237)	26857	1037	299.91	+	3.46	0.00E00					
cellular protein modification process (GO:0006464)	10192	387	113.82	+	3.40	1.83E-101					
protein modification process (GO:0036211)	10192	387	113.82	+	3.40	1.83E-101					
protein phosphorylation (GO:0006468)	5552	200	62.00	+	3.23	9.41E-44					
macromolecule modification (GO:0043412)	10849	387	121.15	+	3.19	1.73E-93					

proteolysis (GO:0006508)	4070	134	45.45	+	2.95	2.80E-24				
organic substance catabolic process (GO:1901575)	4248	132	47.44	+	2.78	1.18E-21				
phosphorylation (GO:0016310)	6558	200	73.23	+	2.73	9.42E-34				
cellular protein metabolic process (GO:0044267)	12945	392	144.56	+	2.71	1.49E-74				
cellular process (GO:0009987)	37512	1135	418.90	+	2.71	0.00E00				
metabolic process (GO:0008152)	35025	1045	391.13	+	2.67	0.00E00				
phosphate-containing compound metabolic process (GO:0006796)	8754	253	97.76	+	2.59	2.89E-40				
phosphorus metabolic process (GO:0006793)	8828	253	98.58	+	2.57	8.59E-40				
protein metabolic process (GO:0019538)	14647	394	163.56	+	2.41	5.56E-61				
regulation of biological quality (GO:0065008)	1571	42	17.54	+	2.39	1.67E-03				
cellular macromolecule metabolic process (GO:0044260)	16379	407	182.91	+	2.23	2.66E-54				
cellular response to stress (GO:0033554)	1628	39	18.18	+	2.15	3.49E-02				
organonitrogen compound metabolic process (GO:1901564)	17646	397	197.05	+	2.01	6.90E-42				
biological_process (GO:0008150)	53482	1144	597.24	+	1.92	6.15E-310				
regulation of cellular process (GO:0050794)	10969	228	122.49	+	1.86	2.53E-16				
macromolecule metabolic process (GO:0043170)	20597	409	230.01	+	1.78	7.68E-31				
regulation of biological process (GO:0050789)	11900	229	132.89	+	1.72	1.04E-12				
biological regulation (GO:0065007)	13579	255	151.64	+	1.68	2.49E-13				
nitrogen compound metabolic process (GO:0006807)	22445	412	250.65	+	1.64	8.16E-24				
primary metabolic process (GO:0044238)	26618	427	297.25	+	1.44	1.29E-13				
organic substance metabolic process (GO:0071704)	28058	443	313.33	+	1.41	4.10E-13				
regulation of gene expression (GO:0010468)	6951	42	77.62	-	.54	1.78E-02				
regulation of macromolecule metabolic process (GO:0060255)	8364	42	93.40	-	.45	2.62E-06				
regulation of metabolic process (GO:0019222)	8501	42	94.93	-	.44	1.19E-06				
small molecule metabolic process (GO:0044281)	4313	20	48.16	-	.42	1.14E-02				
cellular nitrogen compound biosynthetic process (GO:0044271)	4275	17	47.74	-	.36	9.79E-04				
biosynthetic process (GO:0009058)	8721	34	97.39	-	.35	8.65E-11				
transport (GO:0006810)	7778	29	86.86	-	.33	6.70E-10				
cellular macromolecule biosynthetic process (GO:0034645)	4336	16	48.42	-	.33	1.61E-04				
establishment of localization (GO:0051234)	7872	29	87.91	-	.33	2.62E-10				
localization (GO:0051179)	8145	29	90.96	-	.32	2.97E-11				
macromolecule biosynthetic process (GO:0009059)	4502	16	50.27	-	.32	4.92E-05				
transmembrane transport (GO:0055085)	4374	15	48.84	-	.31	4.67E-05				
cellular biosynthetic process (GO:0044249)	7925	27	88.50	-	.31	2.00E-11				
organic substance biosynthetic process (GO:1901576)	8094	27	90.39	-	.30	3.93E-12				
organic substance transport (GO:0071702)	3403	11	38.00	-	.29	9.36E-04				
nucleic acid metabolic process (GO:0090304)	5372	16	59.99	-	.27	3.93E-08				
heterocycle metabolic process (GO:0046483)	7191	20	80.30	-	.25	9.67E-13				
nucleobase-containing compound metabolic process (GO:0006139)	6563	18	73.29	-	.25	1.42E-11				

organic cyclic compound metabolic process (GO:1901360)	7760	20	86.66	-	.23	6.06E-15	
cellular aromatic compound metabolic process (GO:0006725)	7640	19	85.32	-	.22	3.45E-15	
cellular nitrogen compound metabolic process (GO:0034641)	8997	20	100.47	-	.20	3.49E-20	
cellular component biogenesis (GO:0044085)	3172	7	35.42	-	.20	2.65E-05	
carbohydrate derivative metabolic process (GO:1901135)	1937	4	21.63	-	.18	1.82E-02	
cellular amide metabolic process (GO:0043603)	2413	3	26.95	-	.11	2.28E-05	
organophosphate metabolic process (GO:0019637)	1874	2	20.93	-	.10	6.80E-04	
positive regulation of biological process (GO:0048518)	1877	2	20.96	-	.10	6.81E-04	
positive regulation of nucleobase-containing compound metabolic process (GO:0045935)	1268	1	14.16	-	.07	4.11E-02	
positive regulation of macromolecule biosynthetic process (GO:0010557)	1286	1	14.36	-	.07	4.29E-02	
positive regulation of cellular biosynthetic process (GO:0031328)	1287	1	14.37	-	.07	4.31E-02	
positive regulation of biosynthetic process (GO:0009891)	1287	1	14.37	-	.07	4.31E-02	
organonitrogen compound biosynthetic process (GO:1901566)	4033	3	45.04	-	.07	9.38E-13	
ncRNA metabolic process (GO:0034660)	1361	1	15.20	-	.07	2.11E-02	
regulation of transcription by RNA polymerase II (GO:0006357)	1500	1	16.75	-	.06	4.85E-03	
positive regulation of nitrogen compound metabolic process (GO:0051173)	1600	1	17.87	-	.06	1.57E-03	
positive regulation of cellular metabolic process (GO:0031325)	1606	1	17.93	-	.06	1.62E-03	
positive regulation of macromolecule metabolic process (GO:0010604)	1610	1	17.98	-	.06	1.03E-03	
positive regulation of metabolic process (GO:0009893)	1646	1	18.38	-	.05	7.10E-04	
small molecule biosynthetic process (GO:0044283)	1691	1	18.88	-	.05	4.95E-04	
positive regulation of cellular process (GO:0048522)	1752	1	19.56	-	.05	2.33E-04	
amide biosynthetic process (GO:0043604)	1901	1	21.23	-	.05	5.32E-05	
regulation of transcription, DNA-templated (GO:0006355)	5717	3	63.84	-	.05	7.79E-21	
regulation of RNA biosynthetic process (GO:2001141)	5754	3	64.26	-	.05	5.24E-21	
regulation of nucleic acid-templated transcription (GO:1903506)	5754	3	64.26	-	.05	5.24E-21	
RNA metabolic process (GO:0016070)	3972	2	44.36	-	.05	1.20E-13	
regulation of RNA metabolic process (GO:0051252)	5979	3	66.77	-	.04	4.91E-22	
regulation of nucleobase-containing compound metabolic process (GO:0019219)	6091	3	68.02	-	.04	1.51E-22	
regulation of cellular macromolecule biosynthetic process (GO:2000112)	6242	3	69.71	-	.04	3.19E-23	
regulation of macromolecule biosynthetic process (GO:0010556)	6286	3	70.20	-	.04	1.33E-23	
regulation of cellular biosynthetic process (GO:0031326)	6309	3	70.45	-	.04	1.41E-23	
regulation of biosynthetic process (GO:0009889)	6312	3	70.49	-	.04	1.43E-23	
RNA processing (GO:0006396)	2153	1	24.04	-	.04	3.98E-06	
regulation of nitrogen compound metabolic process (GO:0051171)	7681	3	85.77	-	.03	1.74E-30	
gene expression (GO:0010467)	5125	2	57.23	-	.03	3.03E-19	
regulation of primary metabolic process (GO:0080090)	7714	3	86.14	-	.03	1.15E-30	
regulation of cellular metabolic process (GO:0031323)	7797	3	87.07	-	.03	5.18E-31	
carbohydrate metabolic process (GO:0005975)	3152	1	35.20	-	.03	6.28E-11	
Unclassified (UNCLASSIFIED)	49320	4	550.76	-	< 0.01	0.00E00	

DOWNREGULATED WHEAT GENES - MOLECULAR FUNCTION:

Analysis Type: PANTHER Overrepresentation Test (Released 20200728)
Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4033054 Released 2020-09-10
Analyzed List: upload_1 (Triticum aestivum)
Reference List: Triticum aestivum (all genes in database)
Test Type: FISHER
Correction: BONFERRONI
Bonferroni count: 1853

GO molecular function complete	Triticum aestivum - REFLIST (102802)	upload_1 (1148)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (P-value)
methionine S-methyltransferase activity (GO:0030732)	8	8	.09	+	89.55	4.76E-09
malate dehydrogenase (decarboxylating) (NAD+) activity (GO:0004471)	17	16	.19	+	84.28	8.01E-20
malic enzyme activity (GO:0004470)	17	16	.19	+	84.28	8.01E-20
malate dehydrogenase (decarboxylating) (NADP+) activity (GO:0004473)	13	12	.15	+	82.66	2.63E-14
oxidoreductase activity, acting on peroxide as acceptor (GO:0016684)	825	610	9.21	+	66.21	0.00E00
peroxidase activity (GO:0004601)	825	610	9.21	+	66.21	0.00E00
antioxidant activity (GO:0016209)	914	610	10.21	+	59.76	0.00E00
ethylene binding (GO:0051740)	18	12	.20	+	59.70	4.15E-13
alkene binding (GO:0072328)	18	12	.20	+	59.70	4.15E-13
ethylene receptor activity (GO:0038199)	18	12	.20	+	59.70	4.15E-13
ubiquitin-specific protease binding (GO:1990381)	8	5	.09	+	55.97	3.61E-04
protease binding (GO:0002020)	9	5	.10	+	49.75	5.56E-04
MAP kinase activity (GO:0004707)	98	44	1.09	+	40.21	2.48E-47
sodium ion transmembrane transporter activity (GO:0015081)	29	12	.32	+	37.05	3.39E-11
sodium:proton antiporter activity (GO:0015385)	29	12	.32	+	37.05	3.39E-11
sequence-specific single stranded DNA binding (GO:0098847)	37	14	.41	+	33.88	6.09E-13
single-stranded telomeric DNA binding (GO:0043047)	37	14	.41	+	33.88	6.09E-13
malate dehydrogenase activity (GO:0016615)	45	16	.50	+	31.84	1.04E-14
microtubule motor activity (GO:0003777)	166	56	1.85	+	30.21	3.90E-55
potassium ion antiporter activity (GO:0022821)	30	10	.34	+	29.85	3.02E-08
potassium:proton antiporter activity (GO:0015386)	30	10	.34	+	29.85	3.02E-08
intramolecular oxidoreductase activity, transposing S-S bonds (GO:0016864)	40	13	.45	+	29.10	3.53E-11
protein disulfide isomerase activity (GO:0003756)	40	13	.45	+	29.10	3.53E-11
ATP-dependent microtubule motor activity, plus-end-directed (GO:0008574)	23	7	.26	+	27.25	5.95E-05
ATP-dependent microtubule motor activity (GO:1990939)	25	7	.28	+	25.07	9.65E-05
heme binding (GO:0020037)	2471	610	27.59	+	22.11	0.00E00
motor activity (GO:0003774)	230	56	2.57	+	21.80	1.81E-48

metal ion:proton antiporter activity (GO:0051139)	50	12	.56	+	21.49	7.50E-09	
tetrapyrrole binding (GO:0046906)	2600	610	29.03	+	21.01	0.00E00	
monovalent cation:proton antiporter activity (GO:0005451)	53	12	.59	+	20.28	1.36E-08	
telomeric DNA binding (GO:0042162)	63	14	.70	+	19.90	3.03E-10	
cation:cation antiporter activity (GO:0015491)	60	12	.67	+	17.91	4.82E-08	
nuclear import signal receptor activity (GO:0061608)	56	10	.63	+	15.99	5.80E-06	
microtubule binding (GO:0008017)	377	56	4.21	+	13.30	3.73E-38	
nuclear localization sequence binding (GO:0008139)	56	8	.63	+	12.79	1.03E-03	
tubulin binding (GO:0015631)	401	56	4.48	+	12.51	7.30E-37	
ubiquitin protein ligase activity (GO:0061630)	749	100	8.36	+	11.96	4.76E-66	
nucleocytoplasmic carrier activity (GO:0140142)	76	10	.85	+	11.78	7.98E-05	
ubiquitin-like protein ligase activity (GO:0061659)	766	100	8.55	+	11.69	3.30E-65	
phosphotransferase activity, nitrogenous group as acceptor (GO:0016775)	46	6	.51	+	11.68	4.38E-02	
protein histidine kinase activity (GO:0004673)	46	6	.51	+	11.68	4.38E-02	
phosphorelay sensor kinase activity (GO:0000155)	46	6	.51	+	11.68	4.38E-02	
small molecule sensor activity (GO:0140299)	46	6	.51	+	11.68	4.38E-02	
damaged DNA binding (GO:0003684)	124	14	1.38	+	10.11	1.02E-06	
solute:proton antiporter activity (GO:0015299)	116	12	1.30	+	9.26	4.23E-05	
solute:cation antiporter activity (GO:0015298)	123	12	1.37	+	8.74	7.69E-05	
protein serine/threonine phosphatase activity (GO:0004722)	315	30	3.52	+	8.53	1.30E-14	
phosphoprotein phosphatase activity (GO:0004721)	503	47	5.62	+	8.37	1.29E-23	
oxidoreductase activity (GO:0016491)	7180	626	80.18	+	7.81	0.00E00	
cytoskeletal protein binding (GO:0008092)	691	59	7.72	+	7.65	3.26E-28	
intramolecular oxidoreductase activity (GO:0016860)	156	13	1.74	+	7.46	1.19E-04	
zinc ion binding (GO:0008270)	1906	149	21.28	+	7.00	9.84E-71	
metal ion binding (GO:0046872)	10305	803	115.08	+	6.98	0.00E00	
cation binding (GO:0043169)	10372	803	115.83	+	6.93	0.00E00	
ubiquitin-protein transferase activity (GO:0004842)	1343	101	15.00	+	6.73	2.38E-45	
ubiquitin-like protein transferase activity (GO:0019787)	1377	101	15.38	+	6.57	1.92E-44	
NAD binding (GO:0051287)	230	16	2.57	+	6.23	4.03E-05	
single-stranded DNA binding (GO:0003697)	237	14	2.65	+	5.29	1.92E-03	
protein serine/threonine kinase activity (GO:0004674)	3305	171	36.91	+	4.63	6.10E-57	
phosphatase activity (GO:0016791)	1004	51	11.21	+	4.55	7.84E-15	
ion binding (GO:0043167)	20725	1050	231.44	+	4.54	0.00E00	
phosphoric ester hydrolase activity (GO:0042578)	1202	51	13.42	+	3.80	7.76E-12	
heterocyclic compound binding (GO:1901363)	24029	903	268.33	+	3.37	0.00E00	
organic cyclic compound binding (GO:0097159)	24069	903	268.78	+	3.36	0.00E00	
protein kinase activity (GO:0004672)	5562	197	62.11	+	3.17	4.71E-42	
transition metal ion binding (GO:0046914)	4326	149	48.31	+	3.08	1.86E-29	

catalytic activity, acting on a protein (GO:0140096)	10753	360	120.08	+	3.00	2.42E-78						
phosphotransferase activity, alcohol group as acceptor (GO:0016773)	6005	197	67.06	+	2.94	1.81E-37						
kinase activity (GO:0016301)	6261	197	69.92	+	2.82	4.92E-35						
catalytic activity (GO:0003824)	34957	1091	390.37	+	2.79	0.00E00						
binding (GO:0005488)	34554	1069	385.87	+	2.77	0.00E00						
transferase activity, transferring phosphorus-containing groups (GO:0016772)	6911	197	77.18	+	2.55	1.88E-29						
ATP binding (GO:0005524)	9039	250	100.94	+	2.48	9.67E-37						
adenyl ribonucleotide binding (GO:0032559)	9166	250	102.36	+	2.44	8.43E-36						
adenyl nucleotide binding (GO:0030554)	9198	250	102.71	+	2.43	1.38E-35						
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides (GO:0016818)	2435	66	27.19	+	2.43	3.27E-07						
hydrolase activity, acting on acid anhydrides (GO:0016817)	2449	66	27.35	+	2.41	3.97E-07						
nucleoside-triphosphatase activity (GO:0017111)	2251	60	25.14	+	2.39	3.95E-06						
purine ribonucleoside triphosphate binding (GO:0035639)	9677	253	108.06	+	2.34	2.18E-33						
purine ribonucleotide binding (GO:0032555)	9804	253	109.48	+	2.31	2.08E-32						
purine nucleotide binding (GO:0017076)	9856	253	110.06	+	2.30	3.88E-32						
ribonucleotide binding (GO:0032553)	9956	253	111.18	+	2.28	2.68E-31						
carbohydrate derivative binding (GO:0097367)	10066	253	112.41	+	2.25	1.20E-30						
pyrophosphatase activity (GO:0016462)	2406	60	26.87	+	2.23	4.45E-05						
nucleotide binding (GO:0000166)	10950	269	122.28	+	2.20	2.37E-31						
nucleoside phosphate binding (GO:1901265)	10950	269	122.28	+	2.20	2.37E-31						
small molecule binding (GO:0036094)	11572	278	129.23	+	2.15	5.69E-31						
transferase activity (GO:0016740)	14789	353	165.15	+	2.14	2.97E-41						
anion binding (GO:0043168)	11616	262	129.72	+	2.02	9.15E-25						
molecular_function (GO:0003674)	56344	1141	629.20	+	1.81	1.03E-277						
transmembrane transporter activity (GO:0022857)	4098	15	45.76	-	.33	3.15E-04						
nucleic acid binding (GO:0003676)	11115	30	124.12	-	.24	2.89E-22						
DNA binding (GO:0003677)	6755	18	75.43	-	.24	2.26E-12						
RNA binding (GO:0003723)	3666	7	40.94	-	.17	1.86E-07						
peptidase activity (GO:0008233)	2546	2	28.43	-	.07	5.51E-07						
UDP-glycosyltransferase activity (GO:0008194)	1300	1	14.52	-	.07	2.40E-02						
double-stranded DNA binding (GO:0003690)	1467	1	16.38	-	.06	5.96E-03						
endopeptidase activity (GO:0004175)	1720	1	19.21	-	.05	2.86E-04						
transferase activity, transferring glycosyl groups (GO:0016757)	2410	1	26.91	-	.04	1.56E-07						
DNA-binding transcription factor activity (GO:0003700)	2812	1	31.40	-	.03	2.52E-09						
transcription regulator activity (GO:0140110)	3264	1	36.45	-	.03	1.73E-11						
molecular function regulator (GO:0098772)	5087	1	56.81	-	.02	1.43E-20						
Unclassified (UNCLASSIFIED)	46458	7	518.80	-	.01	0.00E00						

UPREGULATED RYE GENES - BIOLOGICAL PROCESS:

Analysis Type: PANTHER Overrepresentation Test (Released 20200728)
Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4033054 Released 2020-09-10
Analyzed List: upload_1 (Triticum aestivum)
Reference List: Triticum aestivum (all genes in database)
Test Type: FISHER
Correction: BONFERRONI
Bonferroni count: 2196

GO biological process complete	Triticum aestivum - REFLIST (102802)	upload_1 (399)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (P-value)
ubiquitin-dependent protein catabolic process via the N-end rule pathway (GO:0071596)	6	6	.02	+	> 100	6.40E-09
histone H3 deacetylation (GO:0070932)	39	30	.15	+	> 100	8.28E-51
histone deacetylation (GO:0016575)	63	30	.24	+	> 100	5.30E-46
protein deacetylation (GO:0006476)	66	30	.26	+	> 100	1.64E-45
oxylipin biosynthetic process (GO:0031408)	104	47	.40	+	> 100	1.21E-72
oxylipin metabolic process (GO:0031407)	104	47	.40	+	> 100	1.21E-72
protein deacylation (GO:0035601)	75	30	.29	+	> 100	3.80E-44
macromolecule deacylation (GO:0098732)	81	30	.31	+	95.43	2.59E-43
calcium ion transmembrane transport (GO:0070588)	120	39	.47	+	83.74	3.73E-55
calcium ion transport (GO:0006816)	122	39	.47	+	82.36	6.47E-55
nucleosome assembly (GO:0006334)	186	59	.72	+	81.73	1.93E-84
lipid oxidation (GO:0034440)	151	47	.59	+	80.20	2.83E-66
nucleosome organization (GO:0034728)	227	59	.88	+	66.97	5.07E-80
chromatin assembly (GO:0031497)	230	59	.89	+	66.09	1.00E-79
chromatin assembly or disassembly (GO:0006333)	238	59	.92	+	63.87	5.87E-79
DNA packaging (GO:0006323)	263	59	1.02	+	57.80	1.06E-76
protein-DNA complex assembly (GO:0065004)	342	59	1.33	+	44.45	1.12E-70
protein-DNA complex subunit organization (GO:0071824)	389	59	1.51	+	39.08	1.09E-67
divalent metal ion transport (GO:0070838)	286	39	1.11	+	35.13	4.11E-42
multicellular organism development (GO:0007275)	898	120	3.49	+	34.43	1.08E-137
divalent inorganic cation transport (GO:0072511)	301	39	1.17	+	33.38	2.53E-41
DNA conformation change (GO:0071103)	466	59	1.81	+	32.62	1.80E-63
lipid modification (GO:0030258)	374	47	1.45	+	32.38	8.89E-50
anatomical structure development (GO:0048856)	1010	120	3.92	+	30.61	4.67E-132
fatty acid biosynthetic process (GO:0006633)	418	47	1.62	+	28.97	1.07E-47
chromatin organization (GO:0006325)	892	91	3.46	+	26.28	1.37E-92
triglyceride biosynthetic process (GO:0019432)	61	6	.24	+	25.34	5.78E-04
acylglycerol biosynthetic process (GO:0046463)	61	6	.24	+	25.34	5.78E-04

neutral lipid biosynthetic process (GO:0046460)	61	6	.24	+	25.34	5.78E-04				
triglyceride metabolic process (GO:0006641)	65	6	.25	+	23.78	8.18E-04				
acylglycerol metabolic process (GO:0006639)	65	6	.25	+	23.78	8.18E-04				
neutral lipid metabolic process (GO:0006638)	65	6	.25	+	23.78	8.18E-04				
ubiquitin-dependent protein catabolic process (GO:0006511)	1450	127	5.63	+	22.57	7.40E-125				
multicellular organismal process (GO:0032501)	1395	120	5.41	+	22.16	1.73E-116				
monocarboxylic acid biosynthetic process (GO:0072330)	549	47	2.13	+	22.06	1.38E-42				
fatty acid metabolic process (GO:0006631)	574	47	2.23	+	21.10	9.43E-42				
developmental process (GO:0032502)	1529	120	5.93	+	20.22	4.70E-112				
modification-dependent protein catabolic process (GO:0019941)	1628	127	6.32	+	20.10	6.27E-119				
modification-dependent macromolecule catabolic process (GO:0043632)	1681	127	6.52	+	19.47	2.74E-117				
histone modification (GO:0016570)	441	32	1.71	+	18.70	4.26E-26				
covalent chromatin modification (GO:0016569)	441	32	1.71	+	18.70	4.26E-26				
protein ubiquitination (GO:0016567)	1973	140	7.66	+	18.28	4.50E-127				
chromosome organization (GO:0051276)	1329	91	5.16	+	17.64	4.93E-78				
protein modification by small protein conjugation (GO:0032446)	2086	140	8.10	+	17.29	6.34E-124				
proteolysis involved in cellular protein catabolic process (GO:0051603)	2025	127	7.86	+	16.16	9.28E-108				
cellular protein catabolic process (GO:0044257)	2028	127	7.87	+	16.13	1.10E-107				
protein catabolic process (GO:0030163)	2076	127	8.06	+	15.76	1.73E-106				
sterol biosynthetic process (GO:0016126)	136	8	.53	+	15.16	2.43E-04				
protein modification by small protein conjugation or removal (GO:0070647)	2434	140	9.45	+	14.82	3.27E-115				
cellular macromolecule catabolic process (GO:0044265)	2516	127	9.77	+	13.01	1.10E-96				
organonitrogen compound catabolic process (GO:1901565)	2584	127	10.03	+	12.66	2.50E-95				
cellular protein-containing complex assembly (GO:0034622)	1295	59	5.03	+	11.74	1.86E-39				
metal ion transport (GO:0030001)	880	39	3.42	+	11.42	1.03E-24				
macromolecule catabolic process (GO:0009057)	2900	127	11.26	+	11.28	1.77E-89				
monocarboxylic acid metabolic process (GO:0032787)	1151	48	4.47	+	10.74	7.16E-30				
protein-containing complex assembly (GO:0065003)	1426	59	5.53	+	10.66	3.15E-37				
lipid biosynthetic process (GO:0008610)	1565	61	6.07	+	10.04	3.47E-37				
organic acid biosynthetic process (GO:0016053)	1245	47	4.83	+	9.73	2.26E-27				
carboxylic acid biosynthetic process (GO:0046394)	1245	47	4.83	+	9.73	2.26E-27				
protein-containing complex subunit organization (GO:0043933)	1655	59	6.42	+	9.19	8.20E-34				
inorganic cation transmembrane transport (GO:0098662)	1106	39	4.29	+	9.09	2.96E-21				
inorganic ion transmembrane transport (GO:0098660)	1225	39	4.75	+	8.20	9.89E-20				
steroid biosynthetic process (GO:0006694)	252	8	.98	+	8.18	2.04E-02				
proteolysis (GO:0006508)	4070	127	15.80	+	8.04	1.67E-72				
cation transmembrane transport (GO:0098655)	1257	39	4.88	+	7.99	2.38E-19				
cellular component assembly (GO:0022607)	1923	59	7.46	+	7.90	2.05E-30				
organic substance catabolic process (GO:1901575)	4248	128	16.49	+	7.76	2.04E-71				

sterol metabolic process (GO:0016125)	267	8	1.04	+	7.72	3.05E-02	
cellular catabolic process (GO:0044248)	4286	128	16.64	+	7.69	5.66E-71	
small molecule biosynthetic process (GO:0044283)	1691	47	6.56	+	7.16	7.02E-22	
cellular lipid metabolic process (GO:0044255)	1948	53	7.56	+	7.01	1.33E-24	
catabolic process (GO:0009056)	5046	128	19.58	+	6.54	6.34E-63	
cation transport (GO:0006812)	1575	39	6.11	+	6.38	4.66E-16	
organelle organization (GO:0006996)	3875	91	15.04	+	6.05	5.76E-40	
ion transmembrane transport (GO:0034220)	1771	39	6.87	+	5.67	2.18E-14	
lipid metabolic process (GO:0006629)	2846	61	11.05	+	5.52	2.09E-23	
cellular component biogenesis (GO:0044085)	3172	60	12.31	+	4.87	2.77E-20	
carboxylic acid metabolic process (GO:0019752)	2600	48	10.09	+	4.76	2.93E-15	
oxoacid metabolic process (GO:0043436)	2629	48	10.20	+	4.70	4.50E-15	
organic acid metabolic process (GO:0006082)	2638	48	10.24	+	4.69	5.14E-15	
cellular protein modification process (GO:0006464)	10192	178	39.56	+	4.50	7.54E-68	
protein modification process (GO:0036211)	10192	178	39.56	+	4.50	7.54E-68	
macromolecule modification (GO:0043412)	10849	178	42.11	+	4.23	9.74E-64	
cellular component organization (GO:0016043)	5672	91	22.01	+	4.13	1.80E-27	
ion transport (GO:0006811)	2637	39	10.23	+	3.81	5.22E-09	
cellular component organization or biogenesis (GO:0071840)	6640	92	25.77	+	3.57	3.06E-23	
cellular protein metabolic process (GO:0044267)	12945	178	50.24	+	3.54	2.22E-52	
protein metabolic process (GO:0019538)	14647	178	56.85	+	3.13	1.04E-44	
small molecule metabolic process (GO:0044281)	4313	49	16.74	+	2.93	7.67E-08	
cellular macromolecule metabolic process (GO:0044260)	16379	178	63.57	+	2.80	5.25E-38	
organonitrogen compound metabolic process (GO:1901564)	17646	178	68.49	+	2.60	1.11E-33	
primary metabolic process (GO:0044238)	26618	239	103.31	+	2.31	8.19E-43	
transmembrane transport (GO:0055085)	4374	39	16.98	+	2.30	5.13E-03	
cellular process (GO:0009987)	37512	330	145.59	+	2.27	2.36E-76	
macromolecule metabolic process (GO:0043170)	20597	178	79.94	+	2.23	4.39E-25	
cellular metabolic process (GO:0044237)	26857	232	104.24	+	2.23	7.80E-38	
organic substance metabolic process (GO:0071704)	28058	240	108.90	+	2.20	2.94E-39	
nitrogen compound metabolic process (GO:0006807)	22445	178	87.11	+	2.04	1.77E-20	
organic substance biosynthetic process (GO:1901576)	8094	62	31.41	+	1.97	7.50E-04	
biosynthetic process (GO:0009058)	8721	62	33.85	+	1.83	1.16E-02	
metabolic process (GO:0008152)	35025	240	135.94	+	1.77	7.89E-23	
biological_process (GO:0008150)	53482	340	207.58	+	1.64	7.04E-41	
Unclassified (UNCLASSIFIED)	49320	59	191.42	-	.31	0.00E00	
organic cyclic compound metabolic process (GO:1901360)	7760	9	30.12	-	.30	1.39E-02	
phosphate-containing compound metabolic process (GO:0006796)	8754	9	33.98	-	.26	7.08E-04	
phosphorus metabolic process (GO:0006793)	8828	9	34.26	-	.26	5.04E-04	

regulation of gene expression (GO:0010468)	6951	5	26.98	-	.19	7.41E-04				
regulation of transcription, DNA-templated (GO:0006355)	5717	4	22.19	-	.18	9.99E-03				
regulation of RNA biosynthetic process (GO:2001141)	5754	4	22.33	-	.18	6.65E-03				
regulation of nucleic acid-templated transcription (GO:1903506)	5754	4	22.33	-	.18	6.65E-03				
regulation of RNA metabolic process (GO:0051252)	5979	4	23.21	-	.17	3.28E-03				
regulation of nucleobase-containing compound metabolic process (GO:0019219)	6091	4	23.64	-	.17	2.31E-03				
regulation of cellular macromolecule biosynthetic process (GO:2000112)	6242	4	24.23	-	.17	1.11E-03				
regulation of macromolecule biosynthetic process (GO:0010556)	6286	4	24.40	-	.16	1.11E-03				
regulation of cellular biosynthetic process (GO:0031326)	6309	4	24.49	-	.16	1.13E-03				
regulation of biosynthetic process (GO:0009889)	6312	4	24.50	-	.16	1.13E-03				
regulation of macromolecule metabolic process (GO:0060255)	8364	5	32.46	-	.15	4.21E-06				
biological regulation (GO:0065007)	13579	8	52.70	-	.15	5.61E-12				
regulation of metabolic process (GO:0019222)	8501	5	32.99	-	.15	2.93E-06				
protein phosphorylation (GO:0006468)	5552	3	21.55	-	.14	2.34E-03				
regulation of nitrogen compound metabolic process (GO:0051171)	7681	4	29.81	-	.13	9.42E-06				
regulation of primary metabolic process (GO:0080090)	7714	4	29.94	-	.13	6.14E-06				
regulation of cellular metabolic process (GO:0031323)	7797	4	30.26	-	.13	4.21E-06				
phosphorylation (GO:0016310)	6558	3	25.45	-	.12	5.61E-05				
regulation of biological process (GO:0050789)	11900	5	46.19	-	.11	9.38E-12				
regulation of cellular process (GO:0050794)	10969	4	42.57	-	.09	2.76E-11				
RNA metabolic process (GO:0016070)	3972	1	15.42	-	.06	1.17E-02				
organonitrogen compound biosynthetic process (GO:1901566)	4033	1	15.65	-	.06	7.78E-03				
cellular macromolecule biosynthetic process (GO:0034645)	4336	1	16.83	-	.06	2.47E-03				
macromolecule biosynthetic process (GO:0009059)	4502	1	17.47	-	.06	1.78E-03				
gene expression (GO:0010467)	5125	1	19.89	-	.05	1.11E-04				
nucleic acid metabolic process (GO:0090304)	5372	1	20.85	-	.05	5.16E-05				
nucleobase-containing compound metabolic process (GO:0006139)	6563	1	25.47	-	.04	4.29E-07				
heterocycle metabolic process (GO:0046483)	7191	1	27.91	-	.04	3.82E-08				
cellular aromatic compound metabolic process (GO:0006725)	7640	1	29.65	-	.03	7.89E-09				
response to stimulus (GO:0050896)	7734	1	30.02	-	.03	5.15E-09				
cellular nitrogen compound metabolic process (GO:0034641)	8997	1	34.92	-	.03	2.11E-11				

UPREGULATED RYE GENES - MOLECULAR FUNCTION:

Analysis Type: PANTHER Overrepresentation Test (Released 20200728)
Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4033054 Released 2020-09-10
Analyzed List: upload_1 (Triticum aestivum)
Reference List: Triticum aestivum (all genes in database)
Test Type: FISHER

Correction: BONFERRONI
Bonferroni count: 1853

GO molecular function complete	Triticum aestivum - REFLIST (102802)	upload_1 (399)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (P-value)
histone deacetylase activity (H3-K14 specific) (GO:0031078)	36	30	.14	+	> 100	1.23E-51
NAD-dependent histone deacetylase activity (H3-K14 specific) (GO:0032041)	36	30	.14	+	> 100	1.23E-51
NAD-dependent histone deacetylase activity (GO:0017136)	36	30	.14	+	> 100	1.23E-51
NAD-dependent protein deacetylase activity (GO:0034979)	39	30	.15	+	> 100	6.98E-51
calcium transmembrane transporter activity, phosphorylative mechanism (GO:0005388)	51	39	.20	+	> 100	8.66E-67
histone deacetylase activity (GO:0004407)	57	30	.22	+	> 100	4.07E-47
protein deacetylase activity (GO:0033558)	60	30	.23	+	> 100	1.38E-46
deacetylase activity (GO:0019213)	68	30	.26	+	> 100	2.86E-45
ion transmembrane transporter activity, phosphorylative mechanism (GO:0015662)	109	39	.42	+	92.19	1.31E-56
calcium ion transmembrane transporter activity (GO:0015085)	114	39	.44	+	88.14	5.75E-56
oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen (GO:0016702)	142	47	.55	+	85.28	2.03E-67
oxidoreductase activity, acting on single donors with incorporation of molecular oxygen (GO:0016701)	185	47	.72	+	65.46	9.37E-63
ATPase-coupled cation transmembrane transporter activity (GO:0019829)	192	39	.75	+	52.33	2.82E-48
ATPase-coupled ion transmembrane transporter activity (GO:0042625)	192	39	.75	+	52.33	2.82E-48
divalent inorganic cation transmembrane transporter activity (GO:0072509)	201	39	.78	+	49.99	1.39E-47
protein heterodimerization activity (GO:0046982)	427	73	1.66	+	44.05	3.10E-88
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides (GO:0016811)	180	30	.70	+	42.94	2.69E-34
dioxygenase activity (GO:0051213)	292	47	1.13	+	41.47	1.90E-54
ubiquitin protein ligase activity (GO:0061630)	749	100	2.91	+	34.40	2.89E-113
ubiquitin-like protein ligase activity (GO:0061659)	766	100	2.97	+	33.64	2.26E-112
diacylglycerol O-acyltransferase activity (GO:0004144)	52	6	.20	+	29.73	2.04E-04
calmodulin binding (GO:0005516)	353	39	1.37	+	28.47	6.28E-39
acylglycerol O-acyltransferase activity (GO:0016411)	61	6	.24	+	25.34	4.87E-04
ATPase-coupled transmembrane transporter activity (GO:0042626)	445	39	1.73	+	22.58	2.50E-35
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds (GO:0016810)	353	30	1.37	+	21.90	2.73E-26
active ion transmembrane transporter activity (GO:0022853)	461	39	1.79	+	21.80	8.86E-35
primary active transmembrane transporter activity (GO:0015399)	488	39	1.89	+	20.59	6.80E-34
ubiquitin-protein transferase activity (GO:0004842)	1343	101	5.21	+	19.38	3.04E-91
zinc ion binding (GO:0008270)	1906	141	7.40	+	19.06	1.57E-130
ubiquitin-like protein transferase activity (GO:0019787)	1377	101	5.34	+	18.90	3.14E-90
metal ion transmembrane transporter activity (GO:0046873)	568	39	2.20	+	17.69	1.56E-31
protein dimerization activity (GO:0046983)	1528	75	5.93	+	12.65	1.72E-53
inorganic cation transmembrane transporter activity (GO:0022890)	1112	39	4.32	+	9.04	3.01E-21
transition metal ion binding (GO:0046914)	4326	141	16.79	+	8.40	2.87E-84

cation transmembrane transporter activity (GO:0008324)			1210	39	4.70	+	8.30	5.48E-20		
active transmembrane transporter activity (GO:0022804)			1415	39	5.49	+	7.10	1.11E-17		
inorganic molecular entity transmembrane transporter activity (GO:0015318)			1739	39	6.75	+	5.78	1.02E-14		
ion transmembrane transporter activity (GO:0015075)			1821	39	7.07	+	5.52	4.53E-14		
protein binding (GO:0005515)	5436	116	21.10	+	5.50	1.08E-48				
metal ion binding (GO:0046872)	10305	211	40.00	+	5.28	1.04E-97				
cation binding (GO:0043169)	10372	211	40.26	+	5.24	3.52E-97				
catalytic activity, acting on a protein (GO:0140096)			10753	137	41.74	+	3.28	5.92E-34		
ion binding (GO:0043167)	20725	251	80.44	+	3.12	3.66E-73				
DNA binding (GO:0003677)	6755	76	26.22	+	2.90	1.55E-13				
transferase activity (GO:0016740)	14789	156	57.40	+	2.72	2.41E-30				
transmembrane transporter activity (GO:0022857)			4098	39	15.91	+	2.45	8.62E-04		
binding (GO:0005488)	34554	328	134.11	+	2.45	3.91E-85				
transporter activity (GO:0005215)	4338	39	16.84	+	2.32	3.86E-03				
catalytic activity (GO:0003824)	34957	245	135.68	+	1.81	2.07E-25				
nucleic acid binding (GO:0003676)	11115	76	43.14	+	1.76	2.73E-03				
molecular_function (GO:0003674)	56344	359	218.69	+	1.64	2.54E-49				
Unclassified (UNCLASSIFIED)	46458	40	180.31	-	.22	0.00E00				
RNA binding (GO:0003723)	3666	1	14.23	-	.07	3.05E-02				