Exploring Rubiscosome Biogenesis in Hexaploid Wheat

Louis Caruana

Lancaster University

Lancaster

LA1 4YQ

United Kingdom

Submitted October 2020

The following thesis is submitted in partially fulfilment of the requirements for a Master of Science in Plant Science (by Research)

All research was conducted at the Lancaster Environment Centre under the supervision of Elizabete Carmo-Silva

Contents

Declaration:

I declare that the work contain herein is original and my own and has not been submitted in this form or any other for the award of a higher degree elsewhere

List of Figures:

Abbreviations

RuBP	ribulose-1,5-bisphosphate
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
Raf1	Rubisco Accumulation Factor 1
Raf2	Rubisco Accumulation Factor 2
BSD2	Bundle Sheath Defective 2
Cpn20	Chaperonin 20
Cpn60	Chaperonin 60
CA1P	2-carboxy-D-arabinitol-1-phosphate
CA1Pase	2-carboxy-D-arabinitol-1-phosphate phosphatase
XuBP	xylulose-1,5-bisphosphate
XuBPase	xylulose-1,5-bisphosphate phosphatase
Rca	Rubisco activase

1. Literature Review

Historic and current undernourishment and hunger has been caused by a lack of access to, and disparity of food distribution, the future demand for food is projected to increase, falling behind global food production resulting in millions more people facing a lack of food security (Long, Marshall-Colon and Zhu, 2015a). The Food and Agriculture Organisation of the United Nations (FAO) estimates that 8.9% of the global 2020 population is hungry, meaning that 690 million people do not consume enough calories to sustain a normal, active and healthy life. FAO estimates that the COVID-19 pandemic may have increased this number by an additional 83-132 million in 2020. The prevalence of under nourishment (8.9%) has been increasing in recent years from a low of 8.6% in 2014. Current predictions expect the prevalence of under nourishment to continue increasing, reaching 9.8% by 2030 resulting in 840 million hungry people (FAO, 2020). The increases expected in global hunger will partly be driven by global population growth, which is expected to plateau around 2050. Decreases in population growth are driven by increases in wealth, which subsequently result in increases in the consumption of food per capita and a greater demand of meat and dairy (Godfray *et al.*, 2010). Increased consumption of meat and dairy, compounded with increased competition for agricultural land means that 2005 global crop yields will need to be doubled to feed the 2050 population (Tilman *et al.*, 2011; Ray *et al.*, 2013). In order to achieve this in a sustainable manner and mitigate further large-scale destruction to previously untouched natural ecosystems, crop yields will need to be doubled per unit area (Simkin, López-Calcagno and Raines, 2019).

It is likely that the quantity of land suitable for agriculture in the future will decrease, this is in part due to the trend towards urbanisation, but also due to the intensification of agriculture. Aquifers are being exhausted at a rate that exceeds their recharge rate, which subsequently results in the desertification of currently irrigated lands (Alexandratos and Bruinsma, 2012). Unfortunately, this means that food for human consumption will face increasing competition over diminishing land and resources with increasing fodder and biofuel requirements (Ziska *et al.*, 2012). The remaining land that is suitable for agriculture will face changing climates which will bring increases in the average temperatures, coupled with a greater irregularity in rainfall (IPCC., 2012). Temperature increases will likely have substantial negative impact on the global yields of maize, rice and wheat (Zhao *et al.*, 2017).

Global crop yields have previously increased dramatically to meet the needs of the world population which doubled between 1970 and 2010. These yield increases are termed the ‘Green Revolution’ and have been attributed to the development of high yielding crop varieties in addition

to agronomic advances, including the increased use of fertilisers (Hedden, 2003). The yield increases of wheat, rice, and maize production closely correlate with the production of nitrogenous fertiliser (Evans and Clarke, 2019). Despite the progress of the green revolution, current gains in crop yields are stagnating (Ray *et al.*, 2012). Plant breeding programs of the green revolution have developed modern high yielding crops which have maximised the potential biomass that is allocated to the grain in addition to maximising the interception of sunlight. However, there has been little improvement in the photosynthetic efficiency of these crops. In several crop species including wheat, photosynthetic efficiency operates around just 20% of its theoretical maximum capacity (Long, Marshall-Colon and Zhu, 2015).

Maize (*Zea mays* L.), Rice (*Oryza sativa* L.), Wheat (*Triticum aestivum* L.), and Soybean (*Glycine max* (L.) Merr.) are the four largest and most agronomically important crops, supplying two thirds of the food consumed globally (Ray *et al.*, 2013; Long, Marshall-Colon and Zhu, 2015). Wheat is the staple food source for 30% of the global population, supplying 20% of the total consumed calories and 25% of the total consumed protein (Borrill, Adamski and Uauy, 2015). Therefore increasing yields of wheat is of particular relevance to meeting the growing food demand (Consortium (IWGSC), 2014). It has been predicted that every degree increase in global mean temperature will result in a 6% decrease in wheat yields, therefore genetic improvements of wheat are needed in order to sustain yields in the projected future environment (Zhao *et al.*, 2017).

The nuclear genome of modern bread wheat is allohexaploid, meaning it contains six sets of chromosomes deriving from different species. The nuclear genome of wheat contains three diploid genomes (henceforth subgenomes) originating from two independent allopolyploidization events between three homologous genomes. The first hybridization event occurred 300,000-500,000 years ago with the hybridisation of the diploid genome of *Triticum urartu* (AA) with the diploid genome of a closely related species to *Aegilops speltoides* (BB) forming the tetraploid *Triticum turgidum* (AABB) (Huang *et al.*, 2002). The tetraploid genome of *T. turgidum* (AABB) was subsequently hybridised with the diploid genome of *Aegilops tauschii* (DD) forming the hexaploid genome of *T. aestivum* (AABBDD) around 10,000 years ago (Krasileva *et al.*, 2013). The three subgenomes can be described as homoeologous: pairs of genes or chromosomes in the same species that originated by speciation and were brought back together in the same genome by allopolyploidization (Glover, Redestig and Dessimoz, 2016). The majority of genes encoded in the hexaploid genome have three loci, one on each of the three subgenomes forming homoeologous gene triads which feature over 95% sequence identity across coding sequences (Consortium (IWGSC), 2014).

Despite the homology across the coding sequences, the hexaploid genome functions in a diploid manner, chromosomes will only pair with true homologous chromosomes during meiosis, e.g. chromosome 1A will only pair with chromosome 1A and not with chromosome 1B, or 1D. This results in the maintenance of three distinct subgenomes, and occurs due to the action of the pairing homoeologs gene (*Ph1*) which has been shown to reduce centromere association during meiosis of non-true homologous chromosomes (Martinez-Perez, Shaw and Moore, 2001; Feldman and Levy, 2012). This means that homoeolog genes are maintained as three independent genes despite their redundancy. It is thought that over evolutionary times scales the redundancy of gene triads confers a selective advantage to polyploid species, enabling individual homoeologs to undergo subfunctionalisation or neofunctionalization, thereby enabling polyploid species such as hexaploid wheat to outcompete their diploid progenitors (Comai, 2005). Subfunctionalisation and neofunctionalization of homoeolog gene triads can occur at the genetic level, through alterations of specific subgenome gene sequences. Alternatively, homoeologs can vary at the transcriptional level, with differential gene expression across a triad resulting in variation across plant tissues and environmental conditions (Ramírez-González, Borrill, Lang, Harrington, Brinton, Venturini, Davey, Jacobs, Ex, *et al.*, 2018). Understanding the interactions between the homoeologous genes has been proposed as a promising avenue towards unlocking novel targets for improving the genetic potential of hexaploid wheat (Borrill, Adamski and Uauy, 2015).

It has been reported that increasing CO_2 fixation of photosynthesis in C_3 crops results in increased yields (Ainsworth and Long, 2005). The CO_2 fixation in photosynthesis occurs via the enzymatic action of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) (Wilson and Calvin, 1955). Rubisco fixes CO_2 through carboxylation of the sugar phosphate substrate ribulose-1,5-bisphosphate (RuBP). Rubisco is the entry point of CO_2 into the biosphere and constitutes up to 50% of the soluble protein within a leaf. Despite this, the CO_2-fixing enzyme is characterised by a number of imperfections and has been identified as a major rate limiting step in photosynthesis of field grown crops; therefore, Rubisco is a primary target for increasing photosynthesis and crop yields (Spreitzer and Salvucci, 2002; Long *et al.*, 2006).

Rubisco is one of the most abundant proteins on the planet, it is estimated that for every person on the planet there are 5 kg of Rubisco; it is expressed in such high quantities in the leaves of plants in order to overcome its limitations (Phillips and Milo, 2009; Erb and Zarzycki, 2018). The first major limitation of Rubisco is its two competing catalytic functions; in addition to the reaction with CO_2 it also catalyses a reaction with O_2 which results in phosphoglycolate (2PG) production, which leads to an energy wasteful photorespiratory reaction during which some of the previously fixed CO_2 is released (Bowes, Ogren and Hageman, 1971). Rubisco also features a very poor catalytic rate. The

energy wasteful reaction with O_2 has resulted in selection pressure for increased CO_2 specificity, therefore indirectly selecting for the poor catalytic rate observed in Rubisco of many species (Studer *et al.*, 2014). In order to overcome these limitations, nature has developed CO_2 Concentrating Mechanisms (CCM). CCMs are present in some major crop species including maize, which features C_4 photosynthesis: a biochemical mechanism that results in 10-100 fold increased partial pressure of CO_2 in proximity to Rubisco (Hendrickson *et al.*, 2008). The relatively high concentration of CO_2 around Rubisco in C_4 photosynthesis allows for a lower specificity for CO_2 over O_2, thus enabling an increased catalytic rate with a reduced metabolic cost (Studer *et al.*, 2014).

A further limitation of Rubisco is that its active sites require post translational modifications in order to become active. Carbamylation of the active site occurs when CO_2 binds to a lysine residue within the active site and is subsequently stabilised with the binding of a Mg^{2+} ion, rendering the active site of the enzyme catalytically competent (Lorimer, Badger and Andrews, 1976). Following activation, naturally occurring phosphorylated compounds can bind to the active sites rendering the enzyme inactive. These compounds include 2-carboxy-D-arabinitol-1-phosphate (CA1P), and xylulose-1,5-bisphosphate (XuBP), the binding of the inhibitors play a role in inhibiting Rubisco catalysis (Parry *et al.*, 2008; Lobo *et al.*, 2019). Inactive, inhibitor-bound Rubisco requires the function of its catalytic chaperone Rubisco activase (Rca) which releases the inhibitory compounds from Rubisco in an ATP dependant manner (Robinson and Portis, 1988; Carmo-Silva *et al.*, 2015). Following removal from Rubisco, the inhibitory compounds CA1P and XuBP are subsequently degraded by CA1P phosphatase (CA1Pase) and XuBP phosphatase (XuBPase) respectively (Lobo *et al.*, 2019).

The assembly of Rubisco is complex. In plants, the enzyme is present as a hexadecamer (~550kDa) which is composed of eight large (~55kDa) and eight small subunits (~15kDa). The large subunit (rbcL) is encoded by a single gene within the chloroplast genome, while the small subunit is encoded by a gene family located in the nuclear genome (RbcS) (Morita *et al.*, 2016; Vitlin Gruber and Feiz, 2018). Figure 1.1 summarises the current understanding of Rubisco biogenesis and function in hexaploid wheat. Following transcription, RbcS is synthesised in the cytosol and therefore requires targeting to the stroma of the chloroplast. RbcS is translated into a precursor form of the protein (pre-RbcS), featuring a transit peptide amino acid sequence at its N terminus. Cytosolic chaperones maintain pre-RbcS in an unfolded state prior to its import into the chloroplast (Jarvis, 2008). The chloroplast targeted pre-RbcS is then imported across the chloroplast membrane into the chloroplast stroma via the transmembrane transport proteins Toc (Toc: translocon at the outer envelope membrane of chloroplasts) and Tic (translocon at the inner envelope membrane of

chloroplasts). The transit peptide is then cleaved resulting in the mature form of RbcS (Jarvis and Soll, 2002; Jarvis and Robinson, 2004).

Despite the spatial separation between the two genes, stoichiometry is maintained between the nuclear encoded RbcS and the chloroplast encoded rbcL by several mechanisms at various intermediate assembly stages. Mature RbcS upregulates the transcription of *rbcL* (Suzuki and Makino, 2012). BSD2 has been reported to confer negative transcriptional regulation of rbcL (Doron *et al.*, 2014; Hayer-Hartl and Hartl, 2020), however it has also been shown that *rbcL* transcription does not increase in the absence of BSD2 (Salesse *et al.*, 2017). Unassembled rbcL monomers downregulate the translation of further *rbcL* (Wostrikoff and Stern, 2007).

rbcL monomers are highly prone to aggregation and therefore cannot spontaneously fold into their functional form, instead requiring assistance from the nuclear encoded chloroplast chaperonin complex (Bracher *et al.*, 2017). The chloroplast chaperonin complex is predominantly composed of a tetradecamer of Cpn60 subunits. *Cpn60* has two splice variants, Cpn60α and Cpn60β, both are present in the chaperonin complex. The Cpn60 subunits are arranged into two heptameric rings, forming a cylindrical-like protein. It is not clear if either of the heptameric rings are exclusively composed of a single isoform or if both of the heptameric rings are hetero-oligomeric for both Cpn60 isoforms (Hayer-Hartl and Hartl, 2020). The cylindrical complex is then capped by an additional ring of ATP regulated co-factors, Cpn10 and Cpn20. Cpn20 is a tandem repeat of Cpn10 and is the most highly expressed chaperonin subunit in the chloroplast (Zhao and Liu, 2018). The structure of the chloroplast chaperonin complex is essentially a nano compartment, enabling aggregation prone proteins such as rbcL to be folded independently in isolation. Following binding of ATP, the chaperonin complex undergoes a conformational change enclosing the rbcL subunit into a nano compartment, this enables rbcL to be correctly folded in isolation from other stromal proteins, and folded rbcL is subsequently released following the hydrolysis of the bound ATP molecules (Bracher *et al.*, 2017).

Rubisco holoenzyme ($rbcL_8RbcS_8$) does not occur spontaneously, instead assembly requires assistance from four assembly chaperones, RbcX, Rubisco Accumulation Factor 1 (Raf1), Rubisco Accumulation Factor 2 (Raf2), and Bundle Sheath Defective 2 (BSD2) (Aigner *et al.*, 2017). RbcX (~15kDa) functions as a homodimer and plays a role in stabilising rbcL dimers following folding of subunits within the chloroplast chaperonin complex. It is not clear if rbcL subunits form dimers prior to or following their interaction with RbcX. RbcX binds specifically to the C-terminal peptide of rbcL and disassociates from the $rbcL_8$ core prior to binding of RbcS (Saschenbrecker *et al.*, 2007).

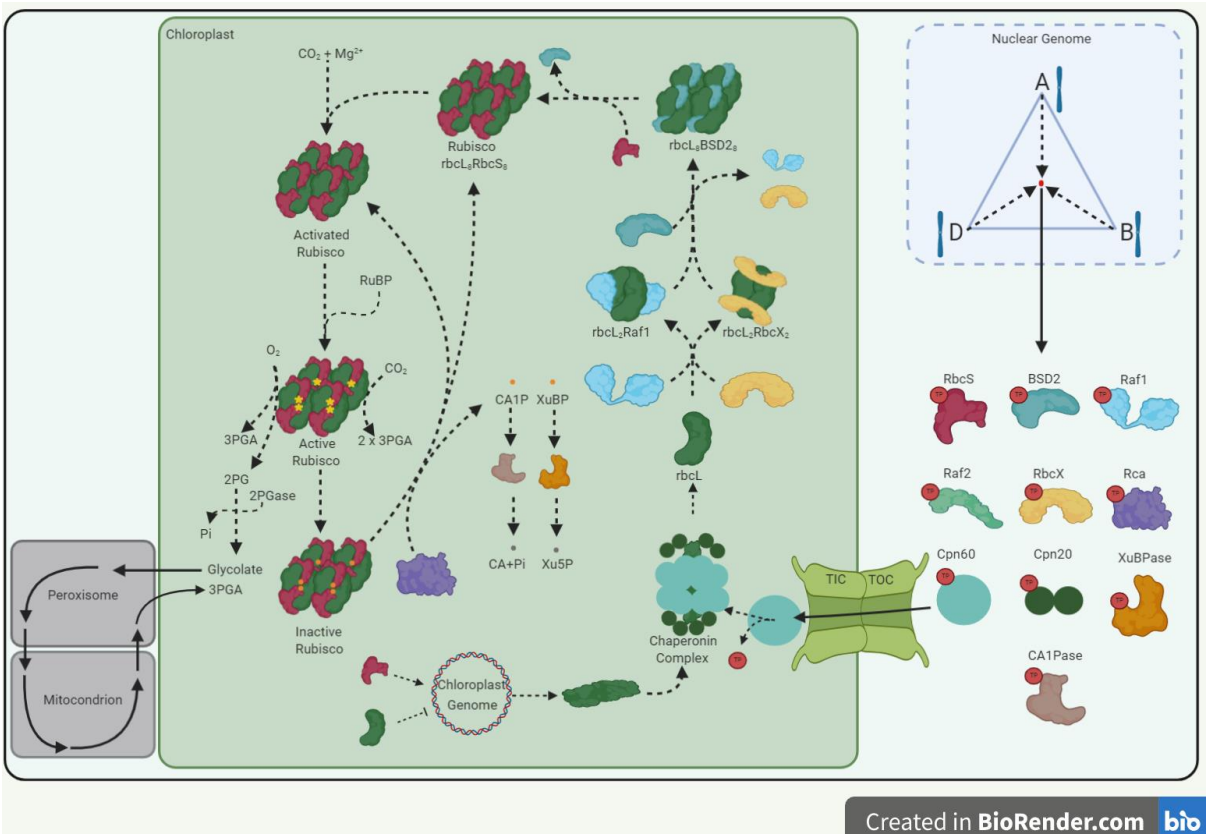


Figure 1.1 A summary of the current understanding of Rubisco biogenesis and function in hexaploid wheat. Top Right: the nuclear genome of hexaploid wheat consists of three subgenomes depicted by A, B, and D. Gene expression is equal to the sum of the expression conferred by the corresponding A, B, and D locus. Bottom Right: the nuclear Rubiscosome, following transcription, precursor RbcS, BSD2, Raf1, Raf2, RbcX, Rca, Cpn60, Cpn20, XuBPase, and CA1Pase are translated in the cytosol, featuring an N terminus transit peptide targeting the proteins to the stroma. Chloroplast: Rubisco biogenesis and function, rbcL is encoded within the chloroplast genome, following translation unfolded rbcL becomes incorporated into the chaperonin complex, enabling rbcL to fold correctly in isolation of the stroma. Following release from the chaperonin complex rbcL monomers interact Raf1 and/or RbcX which both play a role in forming rbcL dimers. rbcL dimers subsequently get arranged into a tetramer forming the core of the holoenzyme. BSD2 binds to the tetramer core increasing stability prior to disassociating upon binding of the RbcS subunits. A lysine residue of the active sites of Rubisco are carbamylated and are subsequently stabilized by the binding of Mg^{2+} rendering the enzyme catalytically competent. Active Rubisco can catalyse two competing reactions with either CO_2 or O_2. Following catalysis, the active sites of Rubisco can bind tightly with the sugar phosphate derivatives CA1P and XuBP rendering the enzyme inactive. Inactive Rubisco then requires the action of Rca to release the sugar phosphate derivatives from the active sites which are subsequently broken down via the Ca1Pase and XuBPase respectively.

Raf1 (~40kDa) was discovered from a maize mutant from the photosynthetic mutant library (Stern, Hanson and Barkan, 2004). The mutant *Raf1-1* was Rubisco deficient, despite normal content of other plastid encoded photosynthetic enzymes, and was seedling lethal (Feiz *et al.*, 2012). Raf1 was reported to associate with Rubisco assembly intermediates, binding to both $RbcL_2$ and $RbcL_8$ and therefore was proposed that the function of Raf1 is to stabilise rbcL monomers into dimers ($rbcL_2Raf1_1$) which are capable of subsequently assembling into the tetramer core ($rbcL_8Raf1_4$) (Hauser *et al.*, 2015). Raf1 and RbcX play seemingly redundant roles, both stabilizing rbcL into dimers, however they interact with rbcL at different sites (Bracher *et al.*, 2017).

In addition to *Raf1-1*, another seedling lethal maize mutant from the photosynthetic mutant library was identified as being rubisco deficient *raf2-1* (Stern, Hanson and Barkan, 2004). *Rubisco accumulation factor 2 (Raf2)* has been shown to be essential for Rubisco assembly (Aigner *et al.*, 2017). Raf2 (~18kDa) has been shown to interact with both rbcL and RbcS in the stroma (Feiz *et al.*, 2014). The role of Raf2 remains to be elucidated, but it has been reported that rbcL in Raf2 mutants associate with the chaperonin complex and therefore it appears that Raf2 functions as a post chaperonin assembly chaperone similar to Raf1 (Aigner *et al.*, 2017; Vitlin Gruber and Feiz, 2018).

Bundle Sheath Defective 2 (BSD2) was identified in maize mutants *bsd2-m1* which displayed no accumulation of either rbcL or RbcS (Brutnell *et al.*, 1999). Similar to Raf1 and Raf2 mutants, rbcL was reported to interact with the chaperonin complex in *bsd2-m1* mutants, therefore suggesting that BSD2 (~8kDa) also operates as a post chaperonin assembly chaperone (Feiz *et al.*, 2014). It has been suggested that BSD2 stabilises the $rbcL_8$ core in the absence of RbcS, and that the Bsd2-rbcL interaction is mediated via the action of Raf2 (Aigner *et al.*, 2017; Vitlin Gruber and Feiz, 2018). The C terminus of BSD2 binds to the active sites of $RbcL_8$ ensuring that the substrate RuBP or other inhibitory compounds bind to the incomplete holoenzyme (Hayer-Hartl and Hartl, 2020). The interaction of RbcX, Raf1, and BSD2, with rbcL appears to be dynamic, each of the three auxiliary factors appear to play redundant roles. However, they have all been shown to be essential for *in vitro* Rubisco assembly. It has been suggested that $RbcL_8Bsd2_8$ is the final assembly intermediate prior to binding of RbcS forming the holoenzyme (Aigner *et al.*, 2017; Hayer-Hartl and Hartl, 2020).

To summarise, in addition to RbcS, Rubisco biogenesis and function is dependent on several nuclear encoded auxiliary factors including Cpn60, Cpn20, RbcX, Raf1, Raf2, Rca, Ca1Pase and XuBPase (Aigner *et al.*, 2017; Bracher *et al.*, 2017; Vitlin Gruber and Feiz, 2018). These molecular machineries can be collectively referred to as the nuclear 'Rubiscosome' (Erb and Zarzycki, 2018). The nuclear encoded auxiliary factors must all be localised and targeted to the stroma of the chloroplast in the same process as described for RbcS. In hexaploid wheat, each of the nuclear

encoded Rubiscosome proteins is likely to be encoded by three loci across each of the three subgenomes, forming homoeolog triads (Borrill, Adamski and Uauy, 2015). However, the expression of nuclear encoded Rubiscosome proteins conferred by a particular gene locus may vary across its respective homoeolog triad (Ramírez-González, Borrill, Lang, Harrington, Brinton, Venturini, Davey, Jacobs, Ex, *et al.*, 2018).

Objectives of the Thesis

The research presented here was divided into two sub-projects. Initially, a lab-based project used wheat transgenic lines to test the hypothesis that overexpression of Raf1 would lead to increased Rubisco content, subsequently resulting in increased biomass and yield. The research focused on characterising the lines as described in the first chapter titled 'Rubisco Accumulation Factor 1 Over Expression in Wheat'. The lack of significant increases in the abundance of Raf1 meant that further characterisation of these plants was abandoned. A further objective was to characterise the expression of the nuclear encoded Rubiscosome across the respective homoeolog triads, as described in the second chapter titled 'Rubiscosome Expression is Symmetric in the Hexaploid Wheat Genome'. The research presented in this chapter aimed to test the hypothesis that all three subgenomes contribute an equal share to the total expression of the nuclear encoded Rubiscosome proteins.

2. Rubisco Accumulation Factor 1 Over Expression in Wheat

2.1 Introduction

Global crop yields need to be increased dramatically to meet the needs of the population in 2050. It has been estimated that global yields from 2005 will need to be doubled in order to meet the demands of the 2050 population, which is predicted to feature an increased consumption of meat, dairy, and, biofuels while simultaneously allocating less land to the production of food (Tilman *et al.*, 2011; Ray *et al.*, 2013). Following rice, wheat is the second most directly consumed source of calories for humans (Long and Ort, 2010). Without agronomic or genetic improvement, global wheat yields have been predicted to decrease by 6% for every degree-Celsius increase in global mean temperature (Zhao *et al.*, 2017). The quantity and efficiency of photosynthesis over a growing season are the primary factors determining final crop yields (Long *et al.*, 2006; Simkin, López-Calcagno and Raines, 2019). Rubisco is the enzyme responsible for assimilating CO_2 in photosynthesis, and improving its function results in increased biomass (Wilson and Calvin, 1955; Carmo-Silva *et al.*, 2015). Therefore, Rubisco is a prime target for improving photosynthetic efficiency and crop yields.

Rubisco exists in various forms throughout nature; algae, cyanobacteria and plants contain form I Rubisco, which features a hexadecamer structure, composed of eight Rubisco large subunits (rbcL), and eight Rubisco small subunits (RbcS) (Andersson, 2008). rbcL subunits are firstly arranged into dimers each containing two active sites. The rbcL dimers are subsequently arranged into a tetramer, which forms the core of the enzyme. The rbcL core is capped by eight RbcS, four at the top and four at the bottom, forming the holoenzyme (Bracher *et al.*, 2017).

Rubisco biogenesis in plants requires the coordinated expression, translation and transportation of the spatially separated subunits and auxiliary factors. In plants, RbcS is encoded by a gene family within the nuclear genome (Spreitzer, 2003), there are at least 25 *RbcS* loci within the nuclear genome of hexaploid wheat (See Chapter 2). RbcS is translated in the cytosol into a precursor form of the protein (pre-RbcS), featuring an additional short amino acid extension (Transit Peptide) that targets pre-RbcS to the chloroplast (Jarvis, 2008). The chloroplast targeted pre-RbcS is then imported into the chloroplast via transmembrane molecular machineries termed the Toc and Tic complex (Toc: translocon at the outer envelope membrane of chloroplasts, Tic: translocon at the inner envelope membrane of chloroplasts) (Jarvis and Soll, 2002). Following import into the stroma of the chloroplast, the transit peptide of pre-RbcS is then cleaved, resulting in the mature form of RbcS (Jarvis and Robinson, 2004). The presence of RbcS in the stroma upregulates the transcription

of *rbcL*, which is encoded in the chloroplast genome, and coupling of *rbcL* expression with RbcS availability maintains stoichiometry between the two subunits (Suzuki and Makino, 2012).

rbcL is encoded, transcribed and translated within the chloroplast. rbcL monomers are highly prone to aggregation thus require chaperonin assisted assembly. The chloroplast chaperonin complex consists of two heptameric rings composed of chaperonin60 (Cpn60) subunits forming a cylindrical like structure. The Cpn60 complex interacts with two additional heptameric rings composed of Chaperonin 20 (Cpn20), which cap the cylindrical complex, forming a nano compartment. The chaperonin complex enables rbcL to fold inside of the noncompetent into its functional form isolated from other stromal rbcL (Bracher *et al.*, 2017; Hayer-Hartl and Hartl, 2020).

Following the interaction with the chaperonin complex, rbcL requires a number of specific assembly chaperones. RbcX, Rubisco Accumulation Factor 1 (Raf1), Rubisco Accumulation Factor 2 (Raf2), and Bundle Sheath Defective 2 (BSD2) have all been reported to form assembly intermediates with rbcL prior to binding of RbcS but downstream of the rbcL-chaperonin complex (Feiz *et al.*, 2014; Hayer-Hartl and Hartl, 2020).

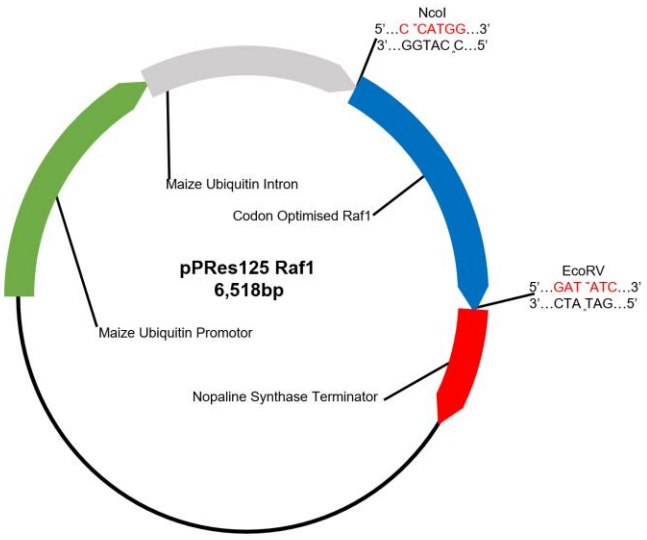
The function of Rubisco has been identified as a rate limiting step in photosynthesis (Spreitzer and Salvucci, 2002). Therefore, efforts have been made to overcome the limitations of Rubisco in crops by increasing Rubisco content. Suzuki *et al*, (2012) increased Rubisco content by 30% in rice by overexpressing *RbcS*, however the increases in Rubisco content did not confer increases in photosynthesis. Wostrikoff *et al*, (2012) attempted to increase Rubisco content in maize by overexpressing both *RbcS* and *rbcL*. Despite observing increases in the protein content of RbcS and rbcL protein there was no increases in the content of assembled rubisco, therefore suggesting that rubisco assembly was limited by the availability of its specific assembly chaperones. More recently it has been reported that Rubisco content can be increased by 36% in maize by over expressing *RbcS* and *Raf1* simultaneously, resulting in increased Rubisco activity and biomass (Salesse-Smith *et al.*, 2018).

This study investigates a number of wheat lines that have previously been independently transformed to over express *Raf1*. The aim of the study was to firstly, to explore if any increases in Raf1 content observed in the leaves of wheat would alter Rubisco content or activity, and secondly if *Raf1* over expression would lead to increases in the biomass of *Raf1* overexpressing lines.

2.2 Materials and Methods

2.2.1 Plant Materials and Culture

Wheat (*Triticum aestivum)* variety Cadenza had been previously been transformed at Rothamsted Research. Wheat lines have been transformed with a codon optimised *Rubisco Accumulation Factor 1 (Raf1)* sequence originating from the wheat D subgenome *Raf1* locus. The overexpression construct contained the codon optimised *Raf1* sequence within an expression cassette, driven by the maize (*Zea mays*) ubiquitin promotor, which is known to confer constitutive expression in wheat (Christensen and Quail, 1996). The expression cassette is terminated by the nopaline synthase terminator. The codon optimised *Raf1* sequence was ligated into the pRRes125 expression vector using the NcoI and EcoRV restriction sites (Figure 2.1). Cadenza wheat plants were transformed via biolistic as described by Sparks and Jones, (2014).



C˘CATGGCTTCTGATGACCTGTTCGAGGGGCTGCCACCACCCGCTGCGCCTGCCGGGGAGGACCGCGCCGCTTCGCCTGCTCC
ACCACCACCCCCACCGGCTGCCCCAGCCCCTAGACCATCTGCTCTGAAGAGCGCCCTGAAGAGGGACAAGCCATCCCCTTCCA
GCGCTGCTACCTCCTCATCTTCCCCAGCCGCTGCTGCCTCCGTGGCCCCTGCTGATGTTGCTGCTGAGGGCCGCGTTCCAGAG
AAGCGCCTGCGCTTCCGGACCACCGTGGACGCGTCCGAGATGCAAATCATCGAGGCCATGCAGAAGATCACCAGCCACATCGG
CAACCCGAGCAAGTTCTCGAAGGCTTCAAAGCTCGCCCTGCAACTCATCGAGGCCGGCTCAGTCAAGCCCGAGACCATCGGCC
AGTTCTTCGGGGCTGCTCGAGGCCGCTATGAGCTCCCCTGGCGCTTGCAACGAGCCATCCGTGAGGGCCGACTACCAAGCTCTG
TTCAACGCCGCGGAGGGGGTCACCGAGTGCTTCAACGAGCAACAGAAGAACCAGTTCGATGTTTGGATGCTGCATGCTGTGGTC
GCCAACGACCTCTGCACGGACGATAGCTTCGTGTTCTCGAAGGCGGTCGGCAAGATCAAGGATGCGATCTCCGCTCTCCCAGTT
GCCACCGTGGACGATGACAACGACGAGGCTACCGCCCTGGCTGCTCTGGCTGAGTCCCAAAGCGGCACCACCGAGAACAAGG
CCGCTGATTCCAACGCTCACGCTGCCGCTTCCAACTCAGGCGAGGAGTCATCTGACCCTTTCGGGCTGGATGACCTCCTGGAGC
ATAAGCCAAAGAAGAGCGAGAAGGCTGCCGAGGCCCTCAACAGGCAGGCTGACGAGGAAGAGTCAAGGCGCCTCCTGAGGTCT
CGGAGAGAGGCTCTCCTGAAGTGCCTGGAGATGGCTGCTAGGCGCTACCGCATCCCGTGGACCCAAACCACGATCGACATCCT
GGGCAGGCACGCTTACGACAGCGTGGGGCGGTTCACGGCCAGACAGAGGGAGGCCGTTGAGAAGCTCTGGAACTCCATCAAG
GAGCAACAGATCCGGAGAAAGCACGGCAAGTCTGCGTCCGGGAAGCTGGACGTGAACGCCTTCGAGAGGCTCCAAGAGAAGTA
CTCCCATGAGAAGATCTCCATCCGGCGGGCGGTTGGCGGGGCTGGCGACAGAAGAGCGACGCAGTGGCTGGGCTGAT ˇ ATC

Figure 2.1 Schematic representation of the codon optimised *Raf1* expression construct within the pPRes125 plasmid, which was used in the transformation of Cadenza wheat. The codon optimised Raf1 sequence is provided by the blue text in the box. Red text highlights the restriction sites of NcoI and EcoRV at the 5' and 3' respectively.

Raf1 transformants were co-transformed with a construct encoding the *bar* selectable marker (Christensen and Quail, 1996) enabling transformed calli to be selected using the herbicide phosphinothricin. The selected plants then underwent genotyping to confirm the presence of the *Raf1* construct, positive plants (T0) were grown to maturity and harvested. The seeds from successfully independently transformed individuals were harvested resulting in T1 independent lines. A total of 220 seeds were sown: 20 seeds from 10 independently transformed lines and 20 seeds of the spring wheat variety Cadenza were grown as a wild type (WT) control. Four seeds failed to germinate, resulting in a total of 216 plants for analysis.

Plants were grown in a semi controlled glasshouse at the Lancaster Environment Centre (Figure 2.2) according to glasshouse speed breeding conditions (Watson *et al.*, 2018). Day and night temperatures were set to 22°C and 17°C respectively. Supplemental sodium lamps (600Watt Plantastar made by Oram Ltd, Newton-Le- Willows, UK) were programmed to a 22-hour photoperiod, supplementing when external light falls below 200 µmol m^{-2} s^{-1}. Prior to sowing, seeds were incubated in water at 4°C for 12 hours. Single seeds were sown into individual 1L deepots (Deepots: D60, Stuewe and Sons) containing wheat mix growth media (Petersfield compost, Hewitt & Son). Plants were maintained well-watered throughout the experiment.

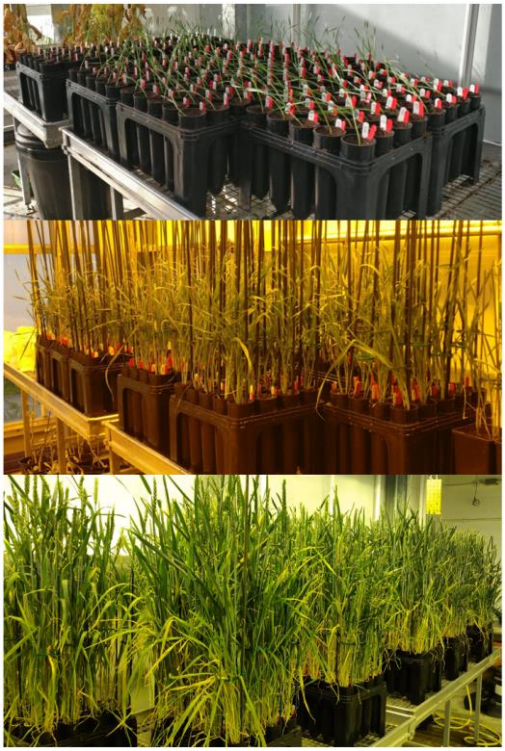


Figure 2.2 Wheat plants growing in the glasshouse. *Raf1* overexpression lines were grown alongside wild type plants in a semi-controlled glasshouse under speed breeding conditions: 22-hour photoperiod, 22°C/17°C day/night temperature regime. Photos taken 1, 3 and 8 weeks after planting.

2.2.2 DNA Extraction

DNA was extracted according to a crude plant DNA extraction protocol, adapted from (Edwards, Johnstone and Thompson, 1991). Leaf samples were taken for DNA extraction from 3-week-old plants of each line and control genotypes, 198 samples were taken from a total of 220 plants (20 biological replicates of 10 independent lines, plus 20 control biological replicates). Leaf segments (~2cm) were sampled directly into a 1.5ml Eppendorf containing DNA extraction buffer. Samples were homogenised using an Eppendorf pestle and centrifuged for 10 minutes. Supernatant was then transferred to equal volume of isopropanol, mixed and incubated at room temperature for 30 minutes. Precipitated DNA was then pelleted, supernatant aspirated, and pellet was left to air dry in a laminar flow hood. The Pellet was then dissolved into sterile water. DNA concentration and quality were determined using a spectrophotometer (SPECTROstar Nano, BMG Labtech). Extracted DNA was amplified in 25 µl reactions according to Taq polymerase instructions (*Taq* DNA Polymerase, New England Biolabs). The expression plasmid and a no template control were included as positive and negative controls, respectively.

2.2.3 Identifying Construct Positive Individuals

PCR amplification was performed on the DNA extractions using the primers, Raf478F (5'-GCCGACTACCAAGCTCTGTT-3', and Nos5rev (5'-ATCGCAAGACCGGCAACAGG-3'). Amplified DNA fragments were separated on a 1% (w/v) agarose gel, containing 1x SYBR Safe DNA Gel Stain (Thermo Fisher Scientific) enabling UV visualisation. Presence of ~800bp bands within amplified samples indicated that the individual plant was positive for the expression construct (Figure 2.3). 8 samples which did not feature a band following amplification were selected as Azygous negative controls. The two independent lines containing the highest proportion of positive individuals (OE14, OE15) were selected for expression and protein analysis.

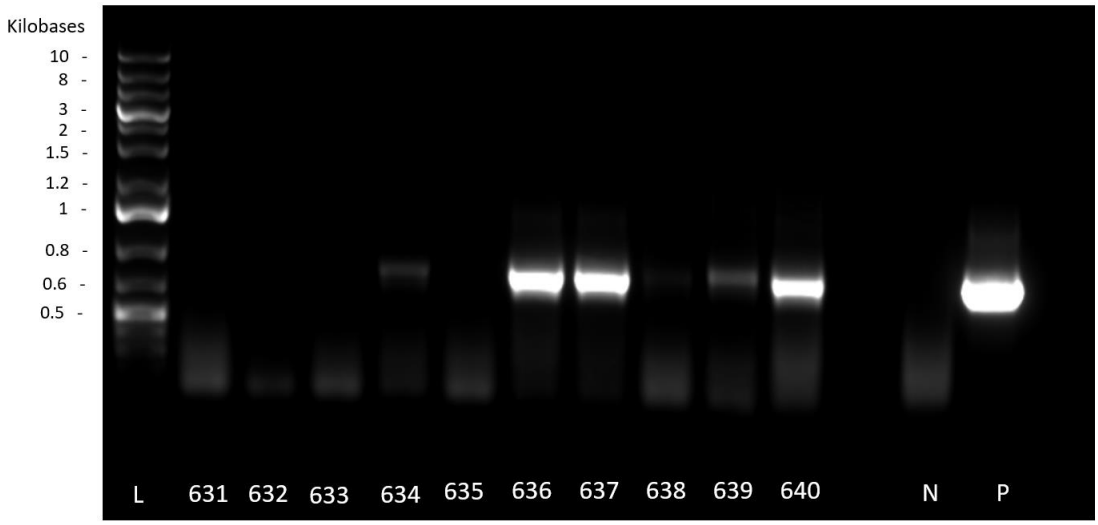


Figure 2.3 Representative gel image of PCR amplification of codon optimised *Raf1* sequence in DNA extraction samples. PCR products were separated on a 1% agarose gel and visualised with SYBR Safe DNA Gel Stain (Thermo Fisher Scientific). L = 1 Kb Plus DNA Ladder. 631-640 are samples from OE19 line. N = Negative PCR control (water). P = Positive PCR control (pPRes125 Raf1).

2.2.4 RNA Isolation

Flag leaves were sampled (~2cm segments) directly into sterile Eppendorf tubes and immediately snap frozen using liquid nitrogen. RNA was isolated from samples using a NucleoSpin RNA Plant Kit (Macherey-Nagel). Samples were firstly homogenised by hand under liquid nitrogen in a pestle and mortar. Homogenised tissue (15-30 µg) was added to a RA1 lysis buffer containing 1% β-mercaptoethanol, and vortexed thoroughly. Lysate was transferred and filtered through a NucleoSpin Filter via centrifugation for 1 min at 11,000x*g*. Filtrate was transferred and mixed into a 70% ethanol solution. Following precipitation, the filtrate ethanol solution was loaded into a NucleoSpin RNA Plant Column and centrifuged for 30s at 11,000x*g*, binding RNA and DNA to the NucleoSpin RNA Plant Column. The membrane of the NucleoSpin RNA Plant Column was desalted with Membrane Desalting Buffer, and centrifuged for 1 min at 11,000x*g* to remove salt. DNase was applied to the membrane of the NucleoSpin RNA Plant Column and incubated for 15 minutes to maximise digestion of bound DNA. Following the incubation, the DNase was inactivated by adding Buffer RAW2 to the NucleoSpin RNA Plant Column and centrifuging for 30s at 11,000x*g*. The NucleoSpin RNA Plant Column was washed twice with Buffer RA3 and centrifuged firstly for 30s at 11,000x*g*, and for an additional 2min following the second wash to completely dry the membrane. Isolated RNA was eluted from the membrane by adding 60µl of RNase free water and centrifuged for 1 min at 11,000x*g*. RNA concentration and quality were determined using a spectrophotometer (SPECTROstar Nano, BMG Labtech).

2.2.5 cDNA Synthesis

Isolated RNA (1 µg) was used as a template for cDNA synthesis using the Precision nanoScript 2 Reverse Transcription Kit (Pimerdesign). The RNA template was added to a PCR tube containing 1µl of Oligo-dT Primer, which preferentially targets the 3' of mRNA. Water was added to the reaction to a final volume of 10µl. The Oligo-dT Primer was annealed to mRNA by incubating the reactions at 65°C for 5 minutes and then immediately cooling on ice. In order to reverse transcribe the mRNA to cDNA, 1µl of nanoScript2 enzyme, and 5µl of nanoScript2 4X Buffer were then added to each reaction. 1µl of Deoxynucleotide triphosphate mix 10mM (dNTPs), was added to each reaction resulting in a final concentration of 0.5mM following addition of additional water to a final reaction volume of 20µl. Reactions were incubated at 42°C for 20 minutes, and then incubated at 75°C for 10 minutes to inactivate the nanoScript2 enzyme.

2.2.6 Reverse-transcription quantitative PCR

cDNA was diluted to a 1:4 dilution. *Ta2291* and *Ta2776* (ADP-ribosylation factor, RNase L inhibitor-like protein, respectively) were selected as reference genes for normalization (Paolacci *et al.*, 2009). Primer efficiencies of Raf1 primers: Raf1_qPCR_F (5'-GAGAGGCTCCAAGAGAAGTAC-3') and Nos5rev (5'-ATC GCAAGACCGGCAACAGG-3'), reference gene primers, and normalised relative quantities (NRQ) were calculated according to (Pfaffl, 2001). RT-qPCR was performed using the PrecisionPLUS qPCR Master Mix Kit (Primerdesign). qPCR reactions were performed in a Mx3005P qPCR system (Stratagene, Agilent Technologies). qPCR conditions: 2min at 95°C, 40 cycles of 95°C for 15s, and 60°C for 60s, followed by 95°C for 60s and 60°C for 30s.

2.2.7 Protein Extraction

Flag Leaves were sampled (~2cm segments) directly into sterile Eppendorf's and immediately snap frozen using liquid nitrogen. Leaf samples were taken from 50 biological replicates across 6 independent lines, an additional 8 azygous and 8 wildtype plants were sampled as controls. Samples were homogenised in 800µl of ice-cold extraction buffer (50mM Tricine-NaOH pH8.0, 10mM EDTA, 1% PVP_{40}, 20mM 2-mercaptoethanol, 1mM PMSF, and 10µM Leupeptin). Homogenate was centrifuged for 3 min at 14,000*g* and 4°C. Aliquots of supernatant were taken for TSP determination and SDS-PAGE.

2.2.8 TSP Determination

Following protein extraction, total soluble protein (TSP) was determined via the Bradford assay (Bradford, 1976), using a calibration curve prepared with known concentrations of bovine serum albumin as a standard. Diluted protein samples were combined with 300 µl of Bradford Reagent (Bio-Rad 500-0006) in wells of a microplate (NUNC, ThermoFisher 442404). The microplate

was incubated in the dark for 10 minutes. TSP was measured using microplate spectrophotometer at 595nm (SPECTROstar Nano, BMG Labtech).

2.2.9 SDS-PAGE

Extracted protein samples were mixed into loading buffer (3.75% SDS, 22.5% Sucrose, 0.5% Bromophenol Blue) at a ratio of 5:4 sample: loading buffer and incubated at 95°C for 4 min. Following TSP determination samples were all diluted using Blank (Extraction Buffer: Loading Buffer, 5:4) to 0.5mg/ml. Gels (12% Mini-PROTEAN TGX Precast Gels; Bio-Rad) were assembled into the electrode assembly of the gel tank (Mini-PROTEAN Tetra cell, Bio-Rad). The lower gel tank was filled with Resolving Gel Buffer (2M Triza-Base, pH9.18). The electrode assembly was filled with Upper Reservoir Buffer (0.4M Trizma-Base, 0.4M Boric Acid, 1%SDS, pH8.4). Diluted samples (5µg) were loaded per lanes, and gels were run at 150V for 105 min.

2.2.10 Western Blotting

Following SDS-PAGE, gels were removed from cassettes and transferred into transfer stacks for blotting (iBlot 2 Dry Blotting System, Thermo Fisher Scientific). Proteins were transferred from gels to membranes by incubating for 7 min under a constant 20V. Membranes were then transferred to Blocking buffer (20mM Trizma-Base, pH7.5, 150mM NaCl, 4% instant non-fat milk) and incubated for 2 hours. Following removal of Blocking buffer, primary antibody raised in rabbit against maize RAF1 (a kind gift of Dr Leila Feiz, Cornell University) at a dilution of 1:1000 in 0.5% blocking solution was applied to membranes and placed on a rocking incubator for 12 hours. Primary antibodies were then removed, membranes were washed in TBST (20mM Trizma-Base, pH7.5, 150mM NaCl, 0.05% Tween-20), secondary antibodies (IRDye Cw800 Anti-Rabbit 1:10,000) were applied to membranes and placed on a rocking incubator for 2 hours. Secondary antibodies were washed off the membrane using TBST and imaged at 700 and 800nm (Odyssey Fc imaging system, LI-COR). Raf1 protein bands (~40kDa, Figure 2.4) were quantified using Image Studio Software (LI-COR).

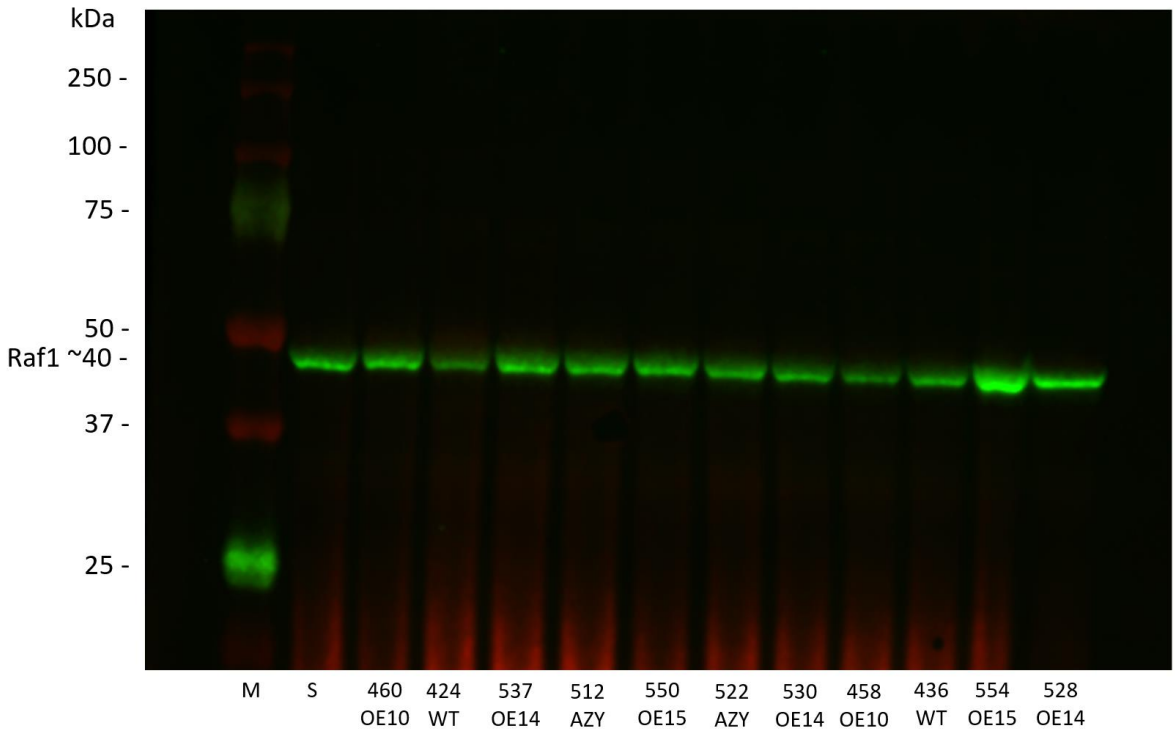


Figure 2.4 Representative quantitative western blot image. Proteins were separated on a 12% Nitrocellulose gel (12% Mini-PROTEAN TGX Precast Gels; Bio-Rad). M= Precision Plus Protein Dual Colour marker (Bio-Rad). S = Standard sample. Numbered lanes denote sample number, and corresponding line. AZY = azygous control; OE = Raf1 over expression line; WT = wild-type control. Primary antibody = anti-Raf1. Secondary antibody = IRDye Cw800 Anti-Rabbit.

2.3 Results

2.3.1 Identification of transgenic *Raf1* positive lines

Screening plants for the presence of the *Raf1* construct discovered that only 55 of the 200 T1 plants that were sowed were positive for the presence of the transgenic *Raf1* construct (Table 2.1). Positive plants were identified in 9 of the 10 independent lines. Positive individuals from two lines (OE14, OE15) both featuring a 35% positive rate, were selected for *Raf1* expression and protein abundance analyses.

Table 2.1 Number of plants genotyped by PCR amplification for the codon optimised *Raf1* sequence, the number of plants positive for the codon optimised *Raf1* sequence, and the percentage of positives.

Transgenic Line	Plants genotyped	Plants positive for codon optimised Raf1	Percentage of positives
OE10	20	8	40%
OE11	20	2	10%
OE12	20	2	10%
OE13	20	0	0%
OE14	20	7	35%
OE15	20	7	35%
OE16	19	2	10.5%
OE17	20	6	30%
OE18	19	10	52%
OE19	20	11	50%

2.3.2 Raf1 Expression analysis

Both transgenic lines (OE14 and OE15) featured greater *Raf1* expression than both the wild type (WT) and azygous (AZY) controls (Figure 2.5). This is expected as the primer pairs used in RT-qPCR are specific to the codon optimised Raf1 sequence transformed into the wheat lines, and not the native *Raf1* gene in the wheat genome. Despite the expression of transgenic *Raf1* in OE14 featuring a much greater spread than OE15, the mean NRQ of both lines is comparable (0.43 and 0.31 respectively).

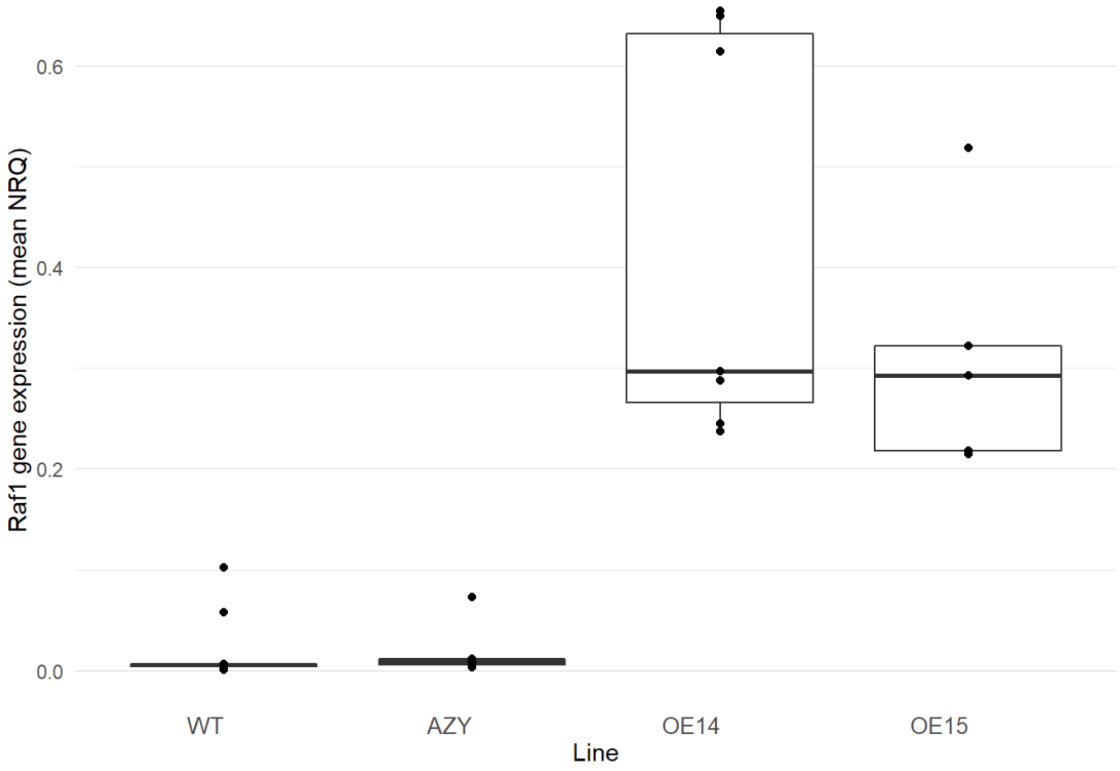


Figure 2.5 Transgenic *Raf1* expression in the flag leaves of two independent wheat lines (OE14, OE15) transformed to over express *Raf1*. Negative controls include WT, Cadenza wheat variety and, AZY, transformed plants that were negative for the transgene (azygous). Wheat was grown in a semi-controlled glass house under speed breeding conditions: 22-hour photoperiod, 22/17°C day/night temperature regime. Each point represents a biological replicate (n=5-9 per genotype), corresponding to the mean of three technical replicates. Centre line of the boxes represent the median. The upper and lower limits of the boxes represent the third and first quartiles respectively. Whiskers represent the range.

2.3.3 Raf1 protein analysis

The Raf1 antibody is not specific to either the native or transgenic Raf1 protein, therefore the relative Raf1 content displayed in Figure 2.6 represents total Raf1 in the flag leaves. The results from the quantitative WB displays that the *Raf1* transformed lines both features marginally greater median values than the WT and AYZ controls, despite the overall data spread not appearing to be greater than either the WT or AYZ control. Statistical analysis in the form of a two tailed test assuming equal variance, displayed that neither OE14 nor OE15 lines were significantly greater than either the WT or AYZ controls.

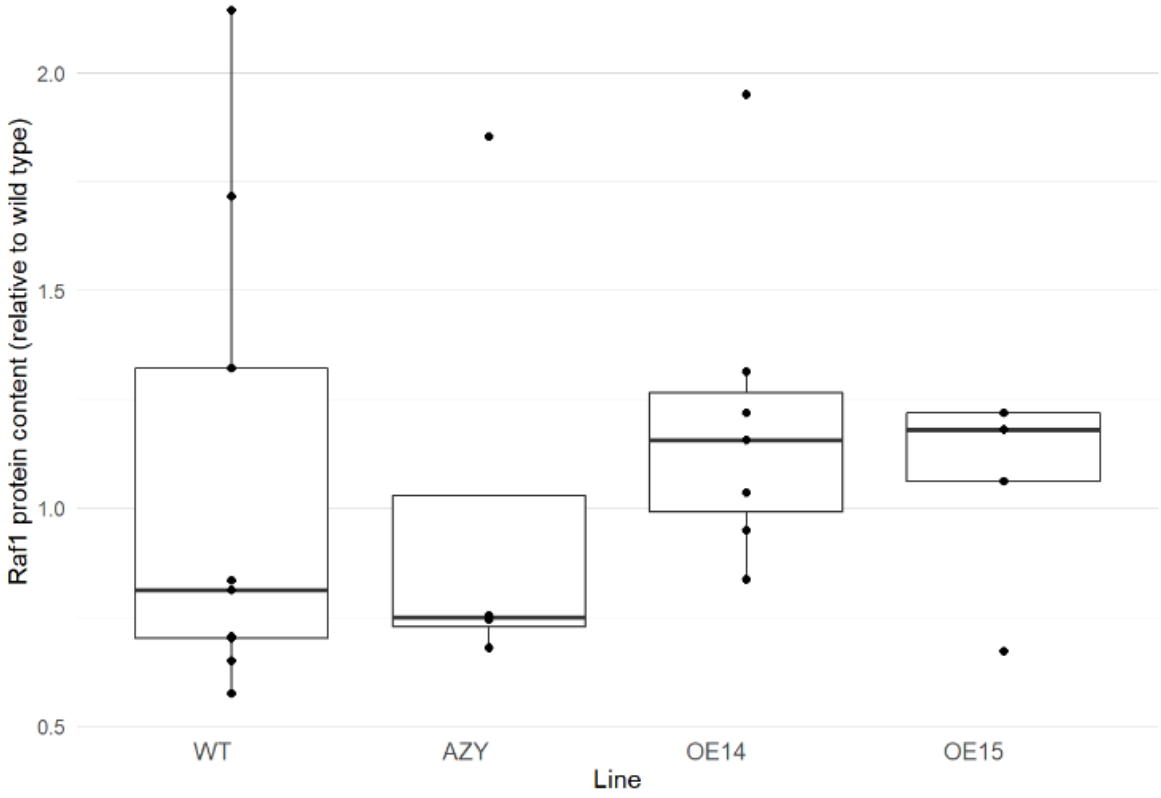


Figure 2.6 Total Raf1 protein content in the flag leaves of independent wheat lines (OE14, OE15) transformed to over express *Raf1*. Negative controls include WT, Cadenza wheat variety and, AZY, transformed plants that were negative for the transgene (azygous). Wheat was grown in a semi-controlled glass house under speed breeding conditions: 22-hour photoperiod, 22/17°C day/night temperature regime. Each point represents a biological replicate (n=5-9 per genotype). Centre line of the boxes represent the median. The upper and lower limits of the boxes represent the third and first quartiles respectively. Whiskers represent the range.

2.4 Discussion

Raf1 is known to play an essential role in Rubisco assembly in plants, maize mutants from The Photosynthetic Mutant Library lacking Raf1 were shown to be Rubisco deficient and seedling lethal (Feiz *et al.*, 2012). It has been previously reported that maize plants over expressing *Raf1* featured greater Rubisco content and activity, and that plants simultaneously over expressing *Raf1* and the small subunit of Rubisco displayed a 36% increase in Rubisco content and increased biomass relative to wildtype maize plants. However, the authors reported that increases in Rubisco content resulted in decreased Rubisco activation, attributed to the lower proportion of Rca to Rubisco (Salesse-Smith *et al.*, 2018). This study aimed to investigate previously transformed wheat lines that had been independently transformed to overexpress *Raf1*. The aim of this study was to determine the effect of increased levels of Raf1 on Rubisco content and activity in wheat.

The results clearly show that both transformed lines feature expression from the *Raf1* construct that is not present in either the wildtype or azygous controls, this is because the primers used in RT-qPCR are specific to the codon optimised *Raf1* sequence used in the expression construct and therefore do not anneal to the native *Raf1* sequence. Despite the clear expression of the codon optimised *Raf1* at the transcript level, the total Raf1 protein content in the flag leaves of transformed *Raf1* lines is not different from the total Raf1 content in flag leaves of the wildtype or azygous controls, possibly due to the great variability observed in the data for these genotypes. Analysis of biomass and yield traits were not explored due to the lack of significant changes in Raf1 content observed between the transformed lines and wild type controls.

It is not clear why the increased *Raf1* expression in the transformed lines did not result in increased Raf1 protein content. It may potentially be explained by the photosynthetic nitrogen budget (CO_2 assimilation per unit leaf nitrogen (Ghannoum, Evans and von Caemmerer, 2011) of C_3 photosynthesis. C_3 species compensate for the poor kinetic properties of Rubisco by investing in substantial quantities of the protein. Rubisco constitutes 20% of nitrogen budget in the leaves of C_3 plants, with Rca constituting a further 3%, representing a substantial nitrogen burden (Evans and Clarke, 2019). Unassembled rbcL monomers repress further translation of *rbcL* transcripts (Wostrikoff and Stern, 2007). Increasing the levels of Raf1 protein would result in increased stabilisation of rbcL into dimers and therefore deregulating the translational repression of *rbcL,* resulting in a further investment of leaf nitrogen to Rubisco. This would potentially feature as positive feedback loop quickly maximising the photosynthetic nitrogen budget of C_3 plants, and therefore limiting further protein synthesis.

In contrast, C_4 species such as maize feature improved Rubisco catalytic properties as a result of the CO_2 concentrating mechanism, increasing CO_2 assimilation, and therefore feature a better photosynthetic nitrogen use efficiency (Ghannoum, Evans and von Caemmerer, 2011). The relatively smaller investment of leaf nitrogen to Rubisco in maize have enabled the 36% increase in Rubisco content reported in Salesse-Smith *et al*, (2018).

Overexpressing *RbcS* results in increases in Rubisco content, it has been reported that the availability of RbcS upregulates expression of the *rbcL* transcript (Suzuki and Makino, 2012). Increasing Rubisco content in wheat via *Raf1* may be possible in wheat by simultaneously overexpressing *Raf1* and *RbcS*. However, in order maximise the carbon assimilation potential of any increases in Rubisco content, it would also be necessary to proportionally increase Rca content, which is a potential challenge due to the limited nitrogen budget in C_3 leaves.

3. Rubiscosome Expression is Symmetric in the hexaploid Wheat Genome

3.1 Introduction

In order to sustainably meet the calorific requirements of the global population in the year 2050, yields from existing agricultural land need to be increased substantially (Ray *et al.*, 2013). This challenge is compounded by climate change, as temperatures rise photosynthesis rates in crops will decline (Way and Yamori, 2014). Crop yields can be increased by increasing CO_2 fixation of crops (Ainsworth and Long, 2005). As the carbon fixing enzyme of photosynthesis, Rubisco is central to engineering efforts that aim to increase CO_2 fixation (Andralojc *et al.*, 2018).

Rubisco biogenesis in plants is complex, requiring the coordinated expression and transportation of the nuclear encoded small subunit of Rubisco (*RbcS*) to the chloroplast where large subunit of Rubisco (*rbcL*) is encoded (Suzuki and Makino, 2012; Vitlin Gruber and Feiz, 2018). Assembly of Rubisco ($rbcL_8RbcS_8$) does not occur spontaneously, instead requiring the transcription, translation, and transportation of several nuclear encoded 'Rubiscosome' proteins into the chloroplast. These include Rubisco specific assembly chaperones: Rubisco accumulation Factor 1 (*Raf1*), Rubisco accumulation factor 2 (*Raf2*), Bundle sheath defective 2 (*Bsd2*), and *RbcX* (Bracher *et al.*, 2017). Additionally, several ancillary factors are required including, Rubisco activase (*Rca*), Xylulose-1,5-bisphosphate Phosphatase (*XuBPase*), and 2-carboxy-D-arabinitol-1-phosphate Phosphatase (*Ca1Pase*), which are essential for maintaining active Rubisco (Parry *et al.*, 2008; Carmo-Silva *et al.*, 2015), and therefore should be considered as part of the Rubiscosome.

Understanding Rubisco biogenesis and function is vital to identifying targets which can be exploited to gain increased CO_2 fixation. Despite decades of research surrounding the assembly of this critical enzyme, plant Rubisco has not been amenable to assembly in a bacterial host, this is due the bacterial GroEL chaperonin subunit being incapable of folding RbcL of plant Rubisco correctly, instead requiring the chloroplast specific chaperonin subunit Cpn60. *In vitro* assembly of plant Rubisco in *Escherichia coli* was first achieved in 2017 by the co-expression of *Arabidopsis thaliana* Rubisco subunits with the cognate chaperonin complex genes *Cpn60* and *Cpn20* in addition to *A. thaliana Raf1, Raf2, RbcX, and Bsd2*(Aigner *et al.*, 2017). *In vitro* expression of functional plant Rubisco represents a significant advancement, which will expedite the identification of engineering targets (Hayer-Hartl and Hartl, 2020). However, for targets outlined by *in vitro* work to be implemented in the genome of crop species, Rubisco biogenesis must be well understood at the genomic and transcriptional level *in vivo*. A comprehensive understanding of the expression of Rubiscosome genes is of particular importance in polyploid crop species, such as wheat.

Understanding the relative contribution of the multiple nuclear genomes within polyploids is important in the context of food security as more than 50% of angiosperms are polyploid (Borrill, Adamski and Uauy, 2015). The significance of polyploidy in evolution is not fully understood, it is possible that the high prevalence in plants is simply due to mild phenotypic effects (Otto and Whitton, 2000). There are three significant advantages conferred by the increased copy number of genes and the associated redundancy (Madlung, 2013). Firstly, increased gene copy number protects gene function by masking deleterious mutations of individual copies (Gu *et al.*, 2003). Secondly, increased gene copy number can lead to subfunctionalisation of individual copies increasing the plasticity of an organism (Moore and Purugganan, 2005). Thirdly, polyploids feature increased heterosis relative to their diploid counterparts, which is thought to be due in part to the gene dosage effect (Birchler *et al.*, 2010). Incidences of polyploidy are higher in crop species relative to their wild relatives (Salman-Minkov, Sabath and Mayrose, 2016), suggesting that the increased plasticity associated with polyploidy has contributed to the domestication of many crop species.

Bread wheat (*Triticum aestivum* L.) contains a hexaploid genome, which is the product of two independent hybridization events, specifically allopolyploidization; a process in which nuclear chromosomes from a closely related yet distinct species are combined into a single nucleus. Tetraploid wheat *T. turgidum* (AABB) first occurred 300,000 - 500,000 years ago, following the hybridisation of the AA (diploid) genome from *T. urartu* and the BB (diploid) genome from a species closely related to *Aegilops speltoides* (Huang *et al.*, 2002). Hexaploid *T. aestivum* (AABBDD) occurred much more recently, only c.10,000 years ago, as the product of the subsequent hybridisation between *T. turgidum* (AABB) and the DD (diploid) genome from *Ae. Tauschii* (Krasileva *et al.*, 2013).

The hexaploid genome of *T. aestivum* contains 21 chromosomes, consisting of the three distinct diploid genomes originating from the closely related donor species (*T. urartu, Ae. speltoides, and Ae. tauschii*). Each genome (henceforth subgenome) contains a near identical set of homoeolog genes forming homoeolog triads (Consortium (IWGSC), 2014). Genes that had previously been separated by speciation (orthologous genes) become homoeologs during allopolyploidization as they have been re-incorporated into a single genome (Glover, Redestig and Dessimoz, 2016).

Despite homoeologs being 97.2% identical across coding sequences (Krasileva *et al.*, 2017), variation exists within the non-coding and repetitive sequences including intronic sequences of homoeolog genes, enabling the subgenome origin of transcripts to be determined (Feldman and Levy, 2012). Total gene expression of homoeolog triads can be balanced, meaning that each of the three homoeologs contributes equally to the expression of the respective gene. Alternatively, gene expression of homoeolog triads can be asymmetric, where different subgenomes make varied

contributions to the total expression of the respective gene (Ramírez-González, Borrill, Lang, Harrington, Brinton, Venturini, Davey, Jacobs, Ex, *et al.*, 2018). Genomic asymmetry in wheat can be further categorised as subgenome dominant or suppressed for either the A, B, or D subgenomes. Ramírez-González *et al.* (2018) studied triad expression of 53,259 genes in wheat across various tissues, and found that despite tissue variation, overall the majority of triads were balanced (c.72.5%), and within the remining assymetric triads single subgenome supression was more common (c.20.5%) than single subgenome dominance (c.7.1%). D subgenome expression had a minor yet significant higher relative abundance than the B and A subgenomes (33.65%, 33.29%, 33.06% respectively). The authors suggested this could be related to a relatively lower insidence in the D subgenome of histone marks across the gene body, which correspond to gene supression by DNA methylation. It is possible that the disparity between the subgenomes is due to genetic diploidization, the inactivation or subfunctionalisation of homoeologus genes over evolutionary time scales (Feldman and Levy, 2012).

Following polyploidization genomes undergo diploidization as a result of genome shock, which may eventually restrict gene expression of specific genes to just one of the subgenomes by silencing homoeologs to combat the effects of doubling/tripling of gene expression (Renny-Byfield and Wendel, 2014). In an evolutionary context, *T. aestivum* is relatively young, which may explain the seemingly small degree of diploidization present in the hexaploid wheat genome. Additionally, despite their high homology, the AA, BB, and DD subgenomes of allohexaploid wheat remain distinct and resist diploidization due in part to the action of the *Ph1* gene, which only permits recombination between true homologous chromosomes (Martinez-Perez, Shaw and Moore, 2001). Unlike genes encoding transcription factors, suppressors and microRNAs which feature a relatively high degree of subfunctionalisation, homoeologs of genes that encode enzymes resist diploidization and therefore have a high degree of retention (Feldman and Levy, 2012). This suggests that the Rubisco pool is composed of subunits which have been transcribed from loci spanning all three subgenomes.

The role that each subgenome plays in the expression of the Rubiscosome has not been explored. For example, it has been suggested that Rca1 and Rca2 are most highly expressed by the B subgenome (Carmo-Silva *et al.*, 2015). However, the results of the study utilising expressed sequence tags (EST) data have yet to be published. Additionally, it has been stated that the A subgenome preferentially controls morphological traits, while the B and D subgenomes preferentially control the reaction to biotic and abiotic factors (Feldman and Levy, 2012). Understanding the relative subgenome contribution to the expression of each of the Rubiscosome genes would be of value to future biotechnological efforts to improve Rubisco function in hexaploid wheat.

The main aim of this study is to detail the expression of the Rubiscosome conferred by each of the three subgenomes in hexaploid wheat, thereby resolving any uncertainty surrounding the relative subgenome contribution to Rca, and other essential genes required for the biogenesis of a functional Rubiscosome. The aim of this study will be achieved in two stages, firstly the identification of all homoeolog loci encoding *RbcS, Cpn60, Cpn20, Raf1, Raf2, Bsd2, RbcX, Rca1, Rca2, XuBPase*, and *Ca1Pase*. Secondly this study will utilise publicly available expression data from the Wheat Expression Browser (www.wheat-expression.com) to explore the relative subgenome contribution to Rubiscosome gene expression in hexaploid wheat.

3.2 Methods

3.2.1 Identification of Rubiscosome genes within the hexaploid wheat genome

For the purpose of this study 'Rubiscosome' genes include *RbcS, Cpn60, Cpn20, Raf1, Raf2, Bsd2, RbcX, Rca1, Rca2, XuBPase*, and *Ca1Pase*, names and functions are listed in Table 3.1. *rbcL* is omitted due to being encoded on the chloroplast genome and therefore disparate from the hexaploid nuclear genome. The nuclear genome Rubiscosome genes were identified using the BLAST search feature on EnsemblPlants (Howe *et al.*, 2020). Nucleic and amino acid sequences of Rubiscosome homologs from soybean (*Glycine max*) and cowpea (*Vigna unguiculata*) which kindly provided for use as query sequences by Dr Doug Orr and Dr Mike Page (Lancaster University) and Sequences for maize (*Zea mays*), tobacco (*Nicotiana tabacum*), and Arabidopsis (*Arabidopsis thaliana*) which were available for some genes from the literature, were used for query sequences (Feiz *et al.*, 2012; Aigner *et al.*, 2017; Lin *et al.*, 2020).

Rubiscosome Gene_IDs that were identified from the BLAST analysis were collected in a .csv file and populated with relevant metadata including the encoded gene, gene locus coordinates, and all corresponding Transcript_IDs. Gene_IDs correspond to a gene locus within the wheat genome. A gene locus may contain several Transcript_IDs, each corresponding to a unique predicted transcript. Transcript_IDs are denoted by a decimal number at the terminus of a Gene_ID, for example TraesCS4A02G177500.1 and TraesCS4A02G177500.2 are Transcript_IDs which correspond to the alpha and beta isoforms (respectively) of TraesCS4A02G177500, the A subgenome homoeolog locus of *Rca2.* To further ensure that the identified genes corresponded to the query genes, transcript and protein sequences for all Transcript_IDs were downloaded in FASTA format for comparative analysis to a homolog of a different species to the one used as the query sequence. Comparative analysis of transcript and peptide sequences were all performed using the Geneious Alignment feature of Geneious 9.1.8 (https://www.geneious.com).

Table 3.1. Names and functions of the Rubiscosome proteins explored in this study.

Protein	Name	Function
BSD2	Bundle Sheath Defective 2	Rubisco Assembly Chaperone
Ca1Pase	2-carboxy-D-arabinitol-1-phosphate Phosphatase	Auxiliary Factor
Cpn20	Chaperonin 20	Chaperonin Subunit
Cpn60	Chaperonin 60	Chaperonin Subunit
Raf1	Rubisco Accumulation Factor 1	Rubisco Assembly Chaperone
Raf2	Rubisco Accumulation Factor 2	Rubisco Assembly Chaperone
Rca1/Rca2	Rubisco Activase	Rubisco Regulation
RbcS	Rubisco Small Subunit	Rubisco Subunit
XuBPase	Xylulose-1,5-bisphosphate Phosphatase	Auxiliary factor

3.2.2 Rubiscosome Gene_IDs

Table 3.2 contains the Gene_IDs of all loci encoding Rubiscosome proteins. Gene_IDs were grouped together, by their subgenome location and by the Rubiscosome protein that they encode. The majority of the Rubiscosome proteins are encoded by an even number of loci which have been mapped to the A, B and D subgenomes, with a couple of exceptions detailed below.

Only the *Raf2* A (TraesCS5A02G545700) and B (TraesCS4B02G379500) homoeologs have been mapped to chromosomes successfully in the reference genome used in this study. A blast search query of the A and B sequences also returned a Gene_ID (TraesCSU02G129700) which had been mapped to an unassigned chromosome category in the reference genome. A sequence alignment of the mature protein sequence of these three Gene_IDs returned a 95.9% pairwise identity. Therefore, the unassigned TraesCSU02G129700 was assumed to be the D subgenome homoeolog of *Raf2*.

The *RbcS* loci identified are not balanced equally across the three subgenome with the A, B, and D subgenomes containing 9, 8, and 8 homoeologs respectively. It is not possible to determine which of the loci are homoeologous.

Table 3.2 Gene identifiers for known components of the Rubiscosome in wheat. Nomenclature of the A subgenome homoeolog of *Bsd2* Gene ID explained: '**Traes**' refers to the species **Tr**iticum **aes**tivum: **CS** refers to the accession, **C**hinese **S**pring: **7A** refers to chromosome **7**, subgenome **A**: **02** refers to RefSeq v1.1: **G** refers to refers to this loci encoding a **G**ene: **341000** is the unique identifier for this loci.

Gene	A Subgenome	B Subgenome	D Subgenome
Bsd2	TraesCS7A02G341000	TraesCS7B02G242200	TraesCS7D02G338600
Ca1Pase	TraesCS4A02G184100	TraesCS4B02G134600	TraesCS4D02G129300
Cpn20	TraesCS6A02G340300	TraesCS6B02G371500	TraesCS6D02G320800
	TraesCS5A02G212500	TraesCS5B02G211200	TraesCS5D02G219500
	TraesCS7A02G161000	TraesCS7B02G066000	TraesCS7D02G162300
	TraesCS2A02G146000	TraesCS2B02G171400	TraesCS2D02G150600
Cpn60	TraesCS4A02G315500	TraesCS5B02G563900	TraesCS5D02G550700
	TraesCS5A02G366800	TraesCS5B02G368900	TraesCS5D02G376000
Raf1	TraesCS1A02G142000	TraesCS1B02G159700	TraesCS1D02G141100
Raf2	TraesCS5A02G545700	TraesCS4B02G379500	TraesCSU02G129700
RbcS	TraesCS2A02G066700	TraesCS2B02G079100	TraesCS2D02G065100
	TraesCS2A02G066800	TraesCS2B02G079200	TraesCS2D02G065200
	TraesCS2A02G066900	TraesCS2B02G079300	TraesCS2D02G065300
	TraesCS2A02G067000	TraesCS2B02G079400	TraesCS2D02G065400
	TraesCS2A02G067100	TraesCS2B02G079500	TraesCS2D02G065500
	TraesCS2A02G067200	TraesCS2B02G078900	TraesCS2D02G065600
	TraesCS2A02G067300		
	TraesCS5A02G165400	TraesCS5B02G162600	TraesCS5D02G169600
	TraesCS5A02G165700	TraesCS5B02G162800	TraesCS5D02G169900
RbcX	TraesCS2A02G198700	TraesCS2B02G226100	TraesCS2D02G206500
	TraesCS5A02G459200	TraesCS5B02G468800	TraesCS5D02G470300
Rca1	TraesCS4A02G177600	TraesCS4B02G140200	TraesCS4D02G134900
Rca2	TraesCS4A02G177500	TraesCS4B02G140300	TraesCS4D02G135000
XuBPase	TraesCS7A02G335600	TraesCS7B02G247200	TraesCS7D02G343300

3.2.3 Expression Data Collection

The wheat expression browser (www.wheat-expression.com) contains expression data from 36 independent studies (as of September 2020), incorporating a broad range of biotic and abiotic stress conditions (Borrill, Ramirez-Gonzalez and Uauy, 2016). To establish the expression of Rubiscosome genes under relatively consistent conditions, and in order to prevent the results from this study being influenced by any stress imposed on the plants, six studies were selected which stated similar photoperiod and temperature regimes of their plant growth conditions (Table 3.3). Thereby ensuring that the results of this study reflect gene expression of wheat under stable conditions. Heat stress analysis (Figure 3.3) exclusively used data from study number 7, Drought and heat stress time course in seedlings (Liu *et al.*, 2015a).

Table 3.3 Reported photoperiod and temperature regime from seven studies available from the wheat expression browser (www.wheat-expression.com). *Data from Liu et al., (2015) was exclusively used for heat stress analysis.

Study	Day Length: Night Length	Temperature Day: Night (°C)	Heat Stress Temperature (°C)	Study Number
Developmental time-course of Chinese Spring (Ramírez-González., *et al.*, 2018)	16:8	25:15	NA	1
Chinese Spring seedling and spikes at anthesis (Ramírez-González., *et al.*, 2018)	12:12	20	NA	2
Chinese Spring leaves and roots from seven leaf stage (Ramírez-González., *et al.*, 2018)	12:12	20	NA	3
Chinese Spring early meiosis, early prophase (Martín *et al.*, 2018)	16:8	20:15	NA	4
Developmental time-course of Azhurnaya (Ramírez-González., *et al.*, 2018)	16:8	25:15	NA	5
Gene expression during a time course of flag leaf senescence (Borrill *et al.*, 2019)	16:8	20:15	NA	6
*Drought and heat stress time course in seedlings (Liu *et al.*, 2015)	16:8	22:18	40	7

3.2.4 Expression Data Wrangling and Visualisation

Sample specific expression data per Gene_ID (Table 3.2) was accessed from the wheat expression browser as transcripts per million (tpm). The mean of the samples per Gene_ID was firstly calculated. In the case of proteins that were encoded by multiple loci per sub genome, the mean tpm per Gene_ID were summed to give total tpm per gene per subgenome:

$$Total\ A\ subgenome\ expression\ of\ Cpn60 = TraesCS4A02G315500 + TraesCS5A02G366800$$

In order to ensure that the relative expression of each of the Rubiscosome proteins was standardised across subgenomes, the relative expression per subgenome of each protein was expressed as a fraction of 1:

$$Relative\ A\ subgenome\ expression = \frac{TPM(Total\ A)}{TPM(Total\ A) + TPM(Total\ B) + TPM(Total\ D)}$$

$$Relative\ B\ subgenome\ expression = \frac{TPM(Total\ B)}{TPM(Total\ A) + TPM(Total\ B) + TPM(Total\ D)}$$

$$Relative\ D\ subgenome\ expression = \frac{TPM(Total\ D)}{TPM(Total\ A) + TPM(Total\ B) + TPM(Total\ D)}$$

Finally, in order to visualise the total expression per Rubiscosome protein the sum of total tpm per subgenome was calculated, and Log2 transformed:

$$Log2(TPM(Total\ A) + TPM(Total\ B) + TPM(Total\ D))$$

All data wrangling was completed using the R Language (R Core Team, 2020), tidyr and dplyr packages as part of the Tidyverse. Circular diagram (Figure 3.1) was generated using the R language adaption of BioCircos.js (Cui *et al.*, 2016). Ternary diagrams (Figures 3.2-3.3) were generated using ggtern package (Hamilton and Ferry, 2018). Code of analysis available at https://github.com/LouisCaruana/Wheat-Rubiscosome-Expression-Balance-

3.3 Results

3.3.1 Identification of Rubiscosome homoeolog loci within the hexaploid wheat genome

The blast search of the wheat genome returned a total of 70 gene loci within the hexaploid nuclear genome that encode Rubiscosome proteins. Figure 3.1 shows that nuclear encoded Rubiscosome genes are well distributed across the chromosomes of the nuclear genome, with only the chromosome 3 triplicate not encoding any Rubiscosome genes.

The majority of Rubiscosome genes have a 1:1:1 correspondence of homoeologs across the three subgenomes. However, this is not true of *RbcS* gene loci which are located in tandem with the chromosome 2 triplicate each encoding six copies plus an additional copy on chromosome 2A, and the chromosome 5 triplicate each encoding an additional two copies.

The chromosomal position of each gene triad used in this study is visualised by the connecting lines in Figure 3.1. Overall, with a few exceptions, the A, B and D, homoeologs of each gene triad are positioned in a similar position on their respective chromosomes. *Cpn20* is encoded by four discrete gene triads, encoded on the triplicates of chromosomes 2, 4, 6, and 7. Additionally, there are two discrete *Cpn60* gene triads, one of which (*Cpn60_2*) is encoded exclusively on the chromosome 5 triplicate, while the B and D subgenome homoeologs of the other (*Cpn60_1*) have been mapped to chromosomes 5B and 5D respectively, and the A subgenome homoeolog has been mapped to chromosome 4A. The assumed D subgenome homoeolog of *Raf2* has not been mapped to a chromosome in the reference genome used, and therefore is displayed in the unassigned chromosome of the reference genome. The A subgenome homoeolog of *Raf 2* has been mapped to chromosome 5A while the B subgenome homoeolog has been mapped to chromosome 4B. Despite the homoeolog loci of these gene triads spanning separate chromosome triplicates, they are consistent with known translocation events within the wheat genome (Appels *et al.*, 2018).

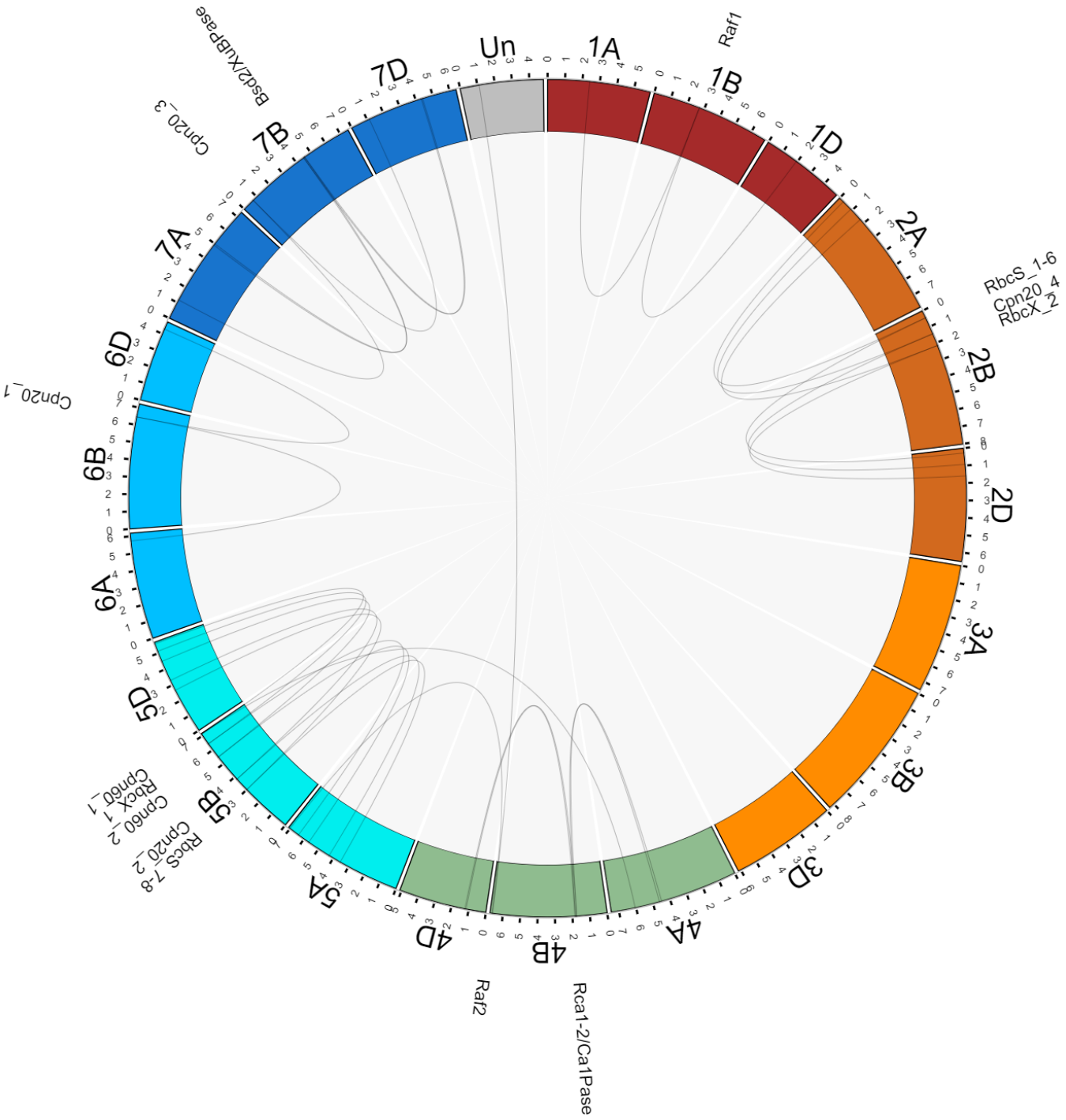


Figure 3.1 Circular visualisation of the hexaploid wheat genome and the position of the homoeolog triads used in this study. The tracks from the outside to the centre specify: names of each homoeolog triad; chromosome name and size (100Mb tick size). Connecting lines represent homoeologous relationships between genes across chromosomes. Chromosome 'Un' indicates homoeologs unallocated to a chromosome position, i.e. within the 'unassigned chromosome' of the RefSeq1.1 reference genome used for this study.

3.3.2 Relative subgenome expression of the Rubiscosome

The A, B, and the D loci of the majority of the Rubiscosome genes contribute equally to the total gene expression of their respective genes. *Bsd2, Ca1pase*, *Cpn20*, *Cpn60*, *Raf1*, *Raf2, RbcS, rbcx*, Rca2, and *Xubpase* genes are all also encoded equally by their respective loci in the leaves and shoots of hexaploid wheat, as can be seen by the cluster of points in the centre of a ternary plot of expression balance (Fig. 3.2A).

The expression balance of the Rubiscosome proteins observed in the leaves and shoot tissues of hexaploid wheat (Fig. 3.2A) is consistent with the expression balance of homoeolog triads observed in the spike tissue of hexaploid wheat (Fig. 3.2B). The two ternary plots display a nearly identical data spread with most of the points clustering in the centre of the plots, indicative of highly balanced expression between the 3 subgenomes. This shows that the relative expression of most of the Rubiscosome homoeologs is not tissue dependant and suggests that the relative gene expression by each of their respective loci is underpinned by a constitutive mechanism.

The results also highlight the comparatively asymmetric expression within leaves of the *Rca1*. Total *Rca1* expression in the leaves and shoots of hexaploid wheat is comprised of 22%, 23%, and 55% from the A, B, and D subgenomes respectively, however despite appearing distinct from the other Rubiscosome proteins in the leaves and shoots of wheat, expression of *Rca1* still falls within the balanced expression category, and thus should not be considered asymmetric. *Rca1* expression in the spike tissues displays a more pronounced asymmetry, comprised of 19%, 16% and 65% from the A, B, and D subgenomes respectively. *Rca1* expression in the spike falls within the B subgenome suppressed category.

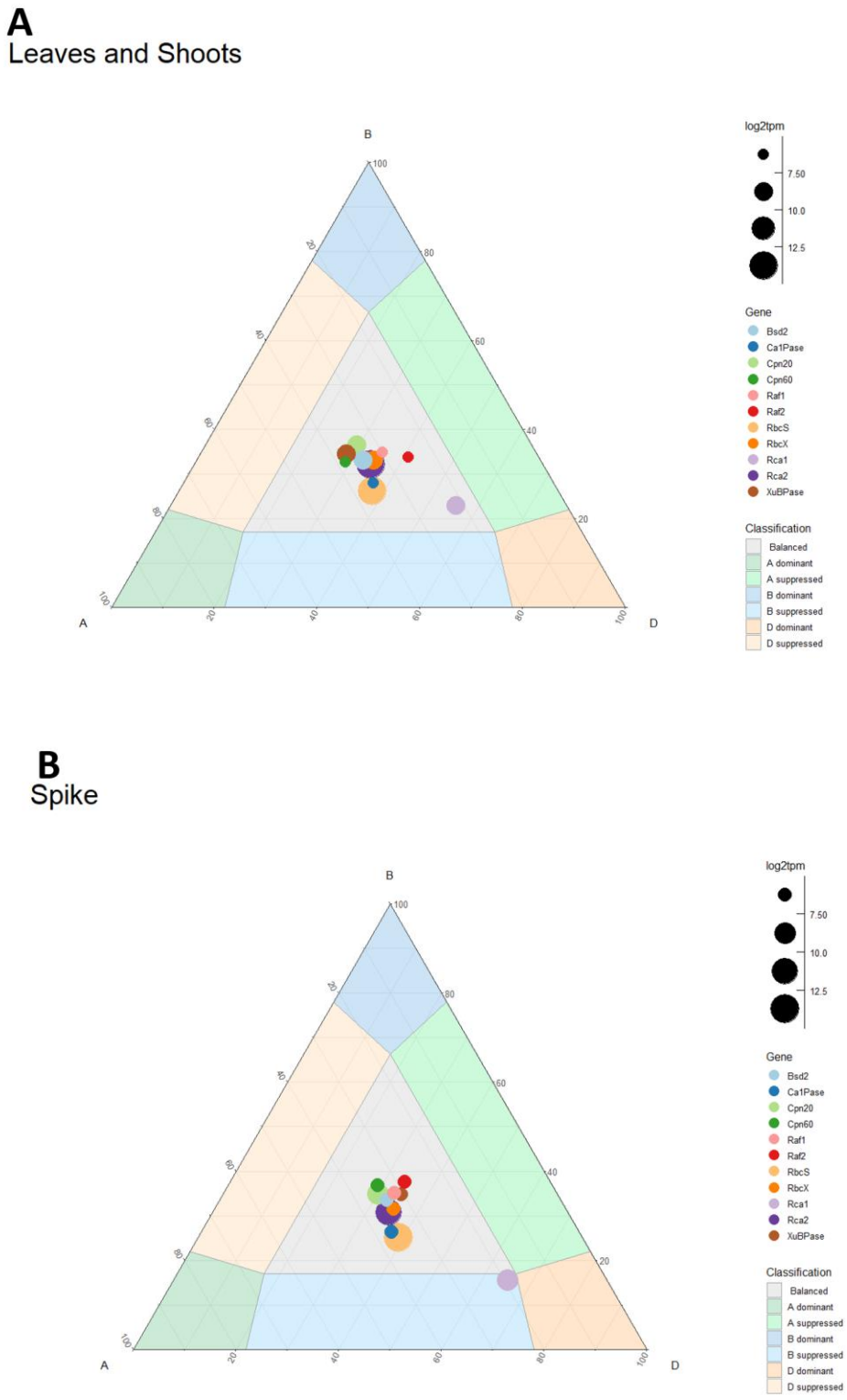


Figure 3.2 Relative expression and expression balance of Rubiscosome triads in the A) leaves and shoots and B) spike of hexaploid wheat from six comparable studies (Table 3.3). The three axes each correspond to a subgenome indicated by the letter. The position of each circle represents the relative contribution of each subgenome specific homoeolog to the overall expression of the respective gene. The size of each symbol is representative of the total expression of each gene triad (Log2 TPM).

3.3.3 Rca homoeolog response to heat stress

There are three isoforms of Rca in wheat, two of which are splice variants of *Rca2* and one additional isoform encoded by *Rca1* (Carmo-Silva *et al.*, 2015). They have been shown to differ significantly in their thermostability (Degen, Worrall and Carmo-Silva, 2020) and their relative protein abundance in flag leaves changed in response to heat stress (Degen, Orr and Carmo-Silva, 2021). To explore the potential connection between these prior observations and *Rca* gene triads, the homoeolog expression balance under heat stress was analysed utilising data from an available heat and drought stress study utilising the TAM107 wheat variety (Liu *et al.*, 2015).

Under heat stress the expression of the *Rca2* in TAM107 does not change in relative subgenome balance or in total expression and in is therefore does not appear to respond to heat stress conditions (Fig. 3.3). However, the expression balance of Rca1 in TAM107 shifts dramatically between control and heat stress conditions. Rca1 expression under control conditions is comprised of 17%, 53%, and 28% from the A, B, and D subgenomes respectively. Under heat stress conditions Rca1 expression is comprised of 16%, 34%, and 48% by the A, B and D subgenomes respectively. Rca1 total expression increases from 102 transcripts per million under control conditions to 3152 transcripts per million under heat stress conditions. Under both control and heat stress conditions Rca1 expression falls on the boundary between the balanced category and the A subgenome supressed category.

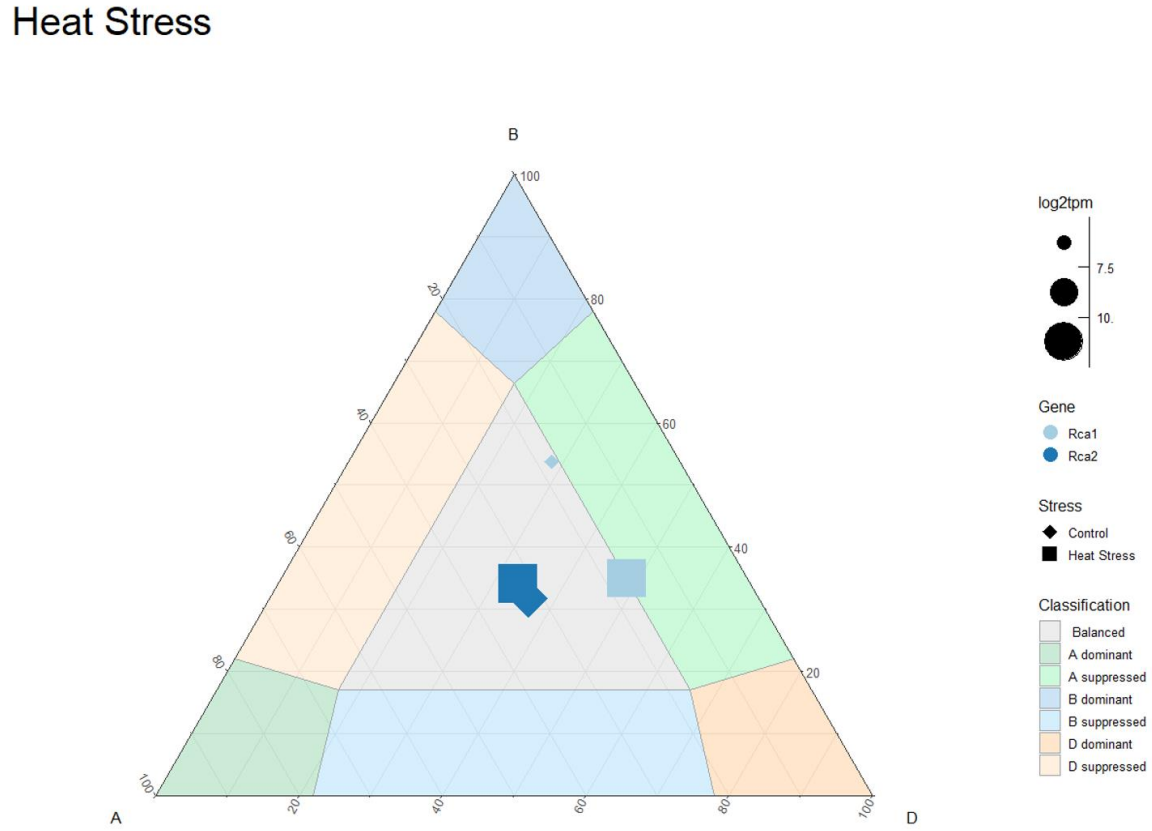


Figure 3.3 Relative expression and expression balance of *Rca1* and *Rca2* in leaves and shoots of hexaploid wheat under control and heat stress conditions. The three axes each correspond to a subgenome indicated by the letter. The position of each point represents the relative contribution of each subgenome specific homoeolog to the overall expression of its respective gene. The size of each symbol is representative of the total expression of each gene triad (Log2 TPM).

3.4 Discussion

Rubisco, the primary carbon-fixing enzyme, can constitute up to 50% of total protein in leaves of C_3 crops such as wheat (Parry *et al.*, 2003; Carmo-Silva *et al.*, 2015), and is a prime target for improving the efficiency of agricultural crop production. Leaves are the primary photosynthetic organs of wheat; however, the importance of photosynthesis in non-foliar tissues is increasingly recognised, with spike tissues shown to contribute up to 39% of grain biomass (Zhang *et al.*, 2020). Here, publicly available gene expression data (Borrill, Ramirez-Gonzalez and Uauy, 2016) was used to explore the relative subgenome contribution to the expression of proteins related to the synthesis and function of Rubisco, termed the Rubiscosome, in leaf and spike tissues of hexaploid wheat.

A total of seventy gene loci were identified across the wheat genome which encode proteins currently known to be essential for Rubisco biogenesis and function. These are well distributed throughout the nuclear genome of hexaploid wheat, with loci on all chromosomes except for the chromosome 3 triplicate (Fig. 3.1). Due to the similarity of the three subgenomes, the three homoeologs corresponding to each gene triad generally occur in a similar location on their respective chromosomes. Homoeologs of *Raf2* and *Cpn60_1* are located within translocated regions (Clavijo *et al.*, 2017), which is why these gene triads span multiple chromosome triplicates. *Cpn20* is encoded by four separate gene triads on chromosomes 2, 5, 6, and 7. The *RbcS* gene family is comprised of 25 loci, seven of which are located in tandem on the chromosome 2A. Chromosomes 2B and 2D each contain 6 copies in tandem, and a further two copies are located on each of the chromosome 5 triplicate. The redundancy of *RbcS* gene copies may potentially be explained as either a gene function protective mechanism, or a subfunctionalisation mechanism, in the ancestral species of the three diploid progenitors(Yamada *et al.*, 2019).

Reference assemblies have struggled to compile the full hexaploid genome due to its large size (~16 Gb) and repetitive sequences (~85%). This 2018 RefSeq1.1 assembly was utilised since this has successfully mapped 14.1 Gb of the wheat genome to the 21 chromosomes, and a further 481 Mb to an 'unassigned chromosome' (Appels *et al.*, 2018). An updated refence genome Triticum_aestivum_4.0 was published in October 2020, which re-localised over 1000 genes from the unassigned chromosome to one of the 21 chromosomes, and uncovered an additional 5799 gene copies across the hexaploid genome (Alonge *et al.*, 2020). The utilisation of this more recent reference genome in future studies may enable more gene loci encoding Rubiscosome proteins to be discovered and resolve the genomic location of the *Raf2* D homoeolog.

For the majority of Rubiscosome proteins, expression was well balanced between each of the three subgenomes in the leaves and spike tissues of hexaploid wheat (Fig. 3.2). In other words, from the perspective of overall Rubiscosome expression, there was no clear dominant subgenome. This is in agreement with previous studies that reported that the expression of over 70% of homoeolog triads are balanced (Ramírez-González., *et al.*, 2018). Interestingly, it appears that the total expression conferred by each of the gene triads is also consistent between the leaves and spike tissues. This highlights that the presence of a functional Rubiscosome is as critical to spike photosynthesis, as it is to leaf photosynthesis.

Whilst the majority of Rubiscosome proteins are expressed in a balanced manner across the three subgenomes, gene loci encoding Rca1, did not display the same balance. Instead, Rca1 features varying degrees of asymmetric expression. A previous report stated that *Rca1* and *Rca2* were most highly expressed by the B subgenome (Carmo-Silva *et al.*, 2015), based on unpublished expressed sequence tags (EST) data. The results herein disagree with this observation, *Rca2* expression remains consistently balanced across tissues and heat stress conditions, and Rca1 displays a dynamic pattern in the relative subgenome expression balance across tissues and across heat stress conditions.

Rca protein content has been shown to increase in wheat under combined heat and drought stress (Perdomo *et al.*, 2017), suggesting that Rca is potentially upregulated at least partially in response to heat. It has been stated that the B and D subgenomes of hexaploid preferentially control the reaction to abiotic conditions (Feldman *et al.*, 2012). The asymmetry observed in the expression of *Rca1*, consistently displays a relatively low contribution from the A subgenome, resulting in the expression balance falling near the A subgenome suppressed category across tissues and across heat stress conditions. The asymmetry observed in the expression of *Rca1* may be linked to the genes response to abiotic stress conditions.

The *Rca1* gene triad in wheat encodes a short β isoform of Rca, while the *Rca2* gene triad produces a short β isoform in addition to a longer α isoform as a result of alternative splicing (Carmo-Silva *et al.*, 2015). Rca1β protein has been shown to feature greater thermostability than the two Rca2 isoforms (Scafaro *et al.*, 2019; Degen, Worrall and Carmo-Silva, 2020). In wheat plants exposed to heat stress (38°C), a 40-fold increase in Rca1β gene expression was observed, with no corresponding increase in the expression of the two Rca2 isoforms(Degen, Orr and Carmo-Silva, 2021). The upregulation of *Rca1* expression has been attributed to heat responsive elements that are present in the promotor regions of all three *Rca1* homoeologs, where in contrast only the *Rca2* A homoeolog contains the identified heat responsive element (Jung *et al.*, 2013; Degen, Orr and Carmo-Silva, 2021). Analysis of expression data from a heat stress study on the heat tolerant variety

TAM107 of hexaploid wheat (Liu *et al.*, 2015) showed an increase in Rca1 expression in plants exposed to heat stress (40°C) for 6 hours relative to control temperatures 18-22°C as expected (Fig. 3.3). Interestingly, the relative expression asymmetry of Rca1 appears to be dynamic, while expression was upregulated by all three loci under heat stress conditions relative to control, the D subgenome loci displayed a much greater increase than the A and the B subgenome loci, therefore shifting the asymmetry towards D subgenome dominance. However despite the dynamic expression balance observed for Rca1 under control and heat stress conditions, the expression balance remains on the boundary of balanced expression and therefore these findings are consistent with the identification of heat responsive elements upstream of each homoeolog by Degen, Orr and Carmo-Silva., (2021).

For *Rca2* there was no evidence of a shift in subgenome expression under heat stress (Fig. 3.3) suggesting that the presence of the heat responsive element in the *Rca2* A loci promotor sequence (Degen, Orr and Carmo-Silva, 2021) did not effectively promote increased expression of this homoeolog in relation to the *Rca2* B and D homoeologs in wheat TAM107 plants exposed to 40°C for 6 hours The potential role of heat responsive elements in Rca gene expression under stress conditions is an area that warrants further investigation.

To summarise, the results of this study demonstrate that the majority of Rubiscosome proteins are expressed in a balanced manner across the three subgenomes, and that the balanced expression of these proteins is consistent across the leaves and spike tissues. The findings resolve some uncertainty on the contribution of the three subgenomes to the expression of the Rubiscosome, particularly of Rca in hexaploid wheat. Except for the asymmetric expression observed in *Rca1*, there was no dominant subgenome in the overall expression of the remaining Rubiscosome proteins. Therefore, engineering strategies aiming to increase CO_2 fixation by targeting the Rubiscosome must ensure that the target gene is successfully edited across all three subgenomes. The sequences of all three homoeologs should be considered with equal importance in designing constructs to target and alter the genomic sequences of the Rubiscosome genes. Failure to successfully edit all three homeologs of the target gene target may otherwise result in the intended phenotype being partially buffered by the natural phenotype conferred via the non-edited subgenomes of hexaploidy wheat.

4. Conclusion

Food demand is projected to fall behind global food production. Crop yields can be increased by increasing the efficiency of photosynthesis, and one possible strategy to achieve this is through increasing the rate of assimilation of CO_2 by Rubisco. Rubisco biogenesis and activity are complex, requiring interaction with a number of specific auxiliary factors. It has previously been reported that increasing the content of Raf1 in maize yielded a 36% increase in Rubisco and subsequently an increase in biomass (Salesse-Smith *et al.*, 2018). This project aimed to investigate wheat lines that had previously been transformed to overexpress *Raf1*. However, despite validating that the construct was conferring expression, there was no significant increase in Raf1 protein content in the ten independent transgenic lines investigated relative to wild type and azygous plants. It may be possible to increase Rubisco content in wheat via the simultaneous overexpression of *Raf1* and *RbcS*, however in order to maintain Rubisco activation, overexpression of Rca may also be required. This approach would require further increases in the allocation of nitrogen to the Rubiscosome, which already consumes 23% of leaf nitrogen in C_3 crops such as wheat (Evans and Clarke, 2019), meaning that alternative targets may be more promising to sustainably improve Rubisco function and crop yields in wheat.

The hexaploid genome in wheat adds complexity and possible diversity to the nuclear encoded Rubiscosome involved in Rubisco assembly and function. Despite it being previously reported that over 70% of wheat genes are expressed in a balanced manner across the three subgenomes, the expression balance of the nuclear encoded Rubiscosome genes has not been reported in hexaploid wheat. Each of the Rubiscosome proteins except for Rca1 feature a balanced expression. These findings show that all three homoeologs need to be considered in future studies aiming to optimise the expression of the Rubiscosome in wheat.

Acknowledgements

I would like to thank Dr Elizabete Carmo-Silva for all of her support and guidance over the past two years. I would also like to thank Dr Rhiannon Page for all of her help in the lab, and Dr Doug Orr for all of his helpful feedback throughout the project. Finally, I would like to thank the whole Lancaster Photosynthesis Group for being a supportive and welcoming team.

References

Aigner, H. *et al.* (2017) 'Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2', *Science*, 358(6368), pp. 1272–1278. doi: 10.1126/science.aap9221.

Ainsworth, E. A. and Long, S. P. (2005) 'What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2', *New Phytologist*, 165(2), pp. 351–372. doi: https://doi.org/10.1111/j.1469-8137.2004.01224.x.

Alexandratos, N. (no date) 'World Agriculture towards 2030/2050: the 2012 revision', p. 154.

Alonge, M. *et al.* (2020) 'Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies', *Genetics*, 216(2), pp. 599–608. doi: 10.1534/genetics.120.303501.

Andersson, I. (2008) 'Catalysis and regulation in Rubisco', *Journal of Experimental Botany*, 59(7), pp. 1555–1568. doi: 10.1093/jxb/ern091.

Andralojc, P. J. *et al.* (2018) 'Increasing metabolic potential: C-fixation', *Essays in Biochemistry*. Edited by S. Gutteridge, 62(1), pp. 109–118. doi: 10.1042/EBC20170014.

Appels, R. *et al.* (2018) 'Shifting the limits in wheat research and breeding using a fully annotated reference genome', *Science*, 361(6403), p. eaar7191. doi: 10.1126/science.aar7191.

Birchler, J. A. *et al.* (2010) 'Heterosis', *The Plant Cell*, 22(7), pp. 2105–2112. doi: 10.1105/tpc.110.076133.

Borrill, P. *et al.* (2019) 'Identification of transcription factors regulating senescence in wheat through gene regulatory network modelling', *Plant Physiology*, 180(3), pp. 1740–1755. doi: 10.1104/pp.19.00380.

Borrill, P., Adamski, N. and Uauy, C. (2015) 'Genomics as the key to unlocking the polyploid potential of wheat', *New Phytologist*, 208(4), pp. 1008–1022. doi: https://doi.org/10.1111/nph.13533.

Borrill, P., Ramirez-Gonzalez, R. and Uauy, C. (2016) 'expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform', *Plant Physiology*, 170(4), pp. 2172–2186. doi: 10.1104/pp.15.01667.

Bowes, G., Ogren, W. L. and Hageman, R. H. (1971) 'Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase', *Biochemical and Biophysical Research Communications*, 45(3), pp. 716–722. doi: 10.1016/0006-291X(71)90475-X.

Bracher, A. *et al.* (2017) 'Biogenesis and Metabolic Maintenance of Rubisco', *Annual Review of Plant Biology*, 68, pp. 29–60. doi: 10.1146/annurev-arplant-043015-111633.

Bradford, M. M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Analytical Biochemistry*, 72, pp. 248–254. doi: 10.1006/abio.1976.9999.

Brutnell, T. P. *et al.* (1999) 'BUNDLE SHEATH DEFECTIVE2, a Novel Protein Required for Post-Translational Regulation of the rbcL Gene of Maize', *The Plant Cell*, 11(5), p. 849. doi: 10.1105/tpc.11.5.849.

Carmo-Silva, E. *et al.* (2015) 'Optimizing Rubisco and its regulation for greater resource use efficiency', *Plant, Cell & Environment*, 38(9), pp. 1817–1832. doi: https://doi.org/10.1111/pce.12425.

Christensen, A. H. and Quail, P. H. (1996) 'Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants', *Transgenic Research*, 5(3), pp. 213–218. doi: 10.1007/BF01969712.

Clavijo, B. J. *et al.* (2017) 'An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations', *Genome Research*, 27(5), pp. 885–896. doi: 10.1101/gr.217117.116.

Comai, L. (2005) 'The advantages and disadvantages of being polyploid', *Nature Reviews Genetics*, 6(11), pp. 836–846. doi: 10.1038/nrg1711.

Consortium (IWGSC), T. I. W. G. S. (2014) 'A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome', *Science*, 345(6194). doi: 10.1126/science.1251788.

Cui, Y. *et al.* (2016) 'BioCircos.js: an interactive Circos JavaScript library for biological data visualization on web applications', *Bioinformatics*, 32(11), pp. 1740–1742. doi: 10.1093/bioinformatics/btw041.

Degen, G. E., Orr, D. J. and Carmo-Silva, E. (2021) 'Heat-induced changes in the abundance of wheat Rubisco activase isoforms', *New Phytologist*, 229(3), pp. 1298–1311. doi: https://doi.org/10.1111/nph.16937.

Degen, G. E., Worrall, D. and Carmo-Silva, E. (2020) 'An isoleucine residue acts as a thermal and regulatory switch in wheat Rubisco activase', *The Plant Journal*, 103(2), pp. 742–751. doi: https://doi.org/10.1111/tpj.14766.

Doron, L. *et al.* (2014) 'The BSD2 ortholog in Chlamydomonas reinhardtii is a polysome-associated chaperone that co-migrates on sucrose gradients with the rbcL transcript encoding the Rubisco large subunit', *The Plant Journal: For Cell and Molecular Biology*, 80(2), pp. 345–355. doi: 10.1111/tpj.12638.

Edwards, K., Johnstone, C. and Thompson, C. (1991) 'A simple and rapid method for the preparation of plant genomic DNA for PCR analysis.', *Nucleic Acids Research*, 19(6), p. 1349.

Erb, T. J. and Zarzycki, J. (2018) 'A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme', *Food biotechnology • Plant biotechnology*, 49, pp. 100–107. doi: 10.1016/j.copbio.2017.07.017.

Evans, J. R. and Clarke, V. C. (2019) 'The nitrogen cost of photosynthesis', *Journal of Experimental Botany*, 70(1), pp. 7–15. doi: 10.1093/jxb/ery366.

FAO, I. (2020) *The State of Food Security and Nutrition in the World 2020: Transforming food systems for affordable healthy diets*. Rome, Italy: FAO, IFAD, UNICEF, WFP and WHO (The State of Food Security and Nutrition in the World (SOFI), 2020). doi: 10.4060/ca9692enAlso Available in:Chinese Spanish Arabic French Russian.

Feiz, L. *et al.* (2012) 'Ribulose-1,5-Bis-Phosphate Carboxylase/Oxygenase Accumulation Factor1 Is Required for Holoenzyme Assembly in Maize', *The Plant Cell*, 24(8), pp. 3435–3446. doi: 10.1105/tpc.112.102012.

Feiz, L. *et al.* (2014) 'A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants', *The Plant Journal*, 80(5), pp. 862–869. doi: https://doi.org/10.1111/tpj.12686.

Feldman, M. *et al.* (2012) 'Genomic asymmetry in allopolyploid plants: wheat as a model', *Journal of Experimental Botany*, 63(14), pp. 5045–5059. doi: 10.1093/jxb/ers192.

Feldman, M. and Levy, A. A. (2012) 'Genome Evolution Due to Allopolyploidization in Wheat', *Genetics*, 192(3), pp. 763–774. doi: 10.1534/genetics.112.146316.

ggplot2 - Elegant Graphics for Data Analysis | Hadley Wickham | Springer (no date). Available at: https://www.springer.com/gp/book/9783319242750 (Accessed: 21 February 2021).

Ghannoum, O., Evans, J. and von Caemmerer, S. (2011) 'Chapter 8 Nitrogen and Water Use Efficiency of C4 Plants', in *C4 Photosynthesis and Related CO2 Concentrating Mechanisms*, pp. 129–146. doi: 10.1007/978-90-481-9407-0_8.

Glover, N. M., Redestig, H. and Dessimoz, C. (2016) 'Homoeologs: What Are They and How Do We Infer Them?', *Trends in Plant Science*, 21(7), pp. 609–621. doi: 10.1016/j.tplants.2016.02.005.

Godfray, H. C. J. *et al.* (2010) 'Food Security: The Challenge of Feeding 9 Billion People', *Science*, 327(5967), pp. 812–818. doi: 10.1126/science.1185383.

Gu, Z. *et al.* (2003) 'Role of duplicate genes in genetic robustness against null mutations', *Nature*, 421(6918), pp. 63–66. doi: 10.1038/nature01198.

Hamilton, N. E. and Ferry, M. (2018) 'ggtern: Ternary Diagrams Using ggplot2', *Journal of Statistical Software*, 87(1), pp. 1–17. doi: 10.18637/jss.v087.c03.

Hauser, T. *et al.* (2015) 'Structure and mechanism of the Rubisco-assembly chaperone Raf1', *Nature Structural & Molecular Biology*, 22(9), pp. 720–728. doi: 10.1038/nsmb.3062.

Hayer-Hartl, M. and Hartl, F. U. (2020) 'Chaperone Machineries of Rubisco – The Most Abundant Enzyme', *Trends in Biochemical Sciences*, 45(9), pp. 748–763. doi: 10.1016/j.tibs.2020.05.001.

Hedden, P. (2003) 'The genes of the Green Revolution', *Trends in Genetics*, 19(1), pp. 5–9. doi: 10.1016/S0168-9525(02)00009-4.

Hendrickson, L. *et al.* (2008) 'The effects of Rubisco activase on C4 photosynthesis and metabolism at high temperature', *Journal of Experimental Botany*, 59(7), pp. 1789–1798. doi: 10.1093/jxb/erm373.

Howe, K. L. *et al.* (2020) 'Ensembl Genomes 2020—enabling non-vertebrate genomic research', *Nucleic Acids Research*, 48(D1), pp. D689–D695. doi: 10.1093/nar/gkz890.

Huang, S. *et al.* (2002) 'Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat', *Proceedings of the National Academy of Sciences*, 99(12), pp. 8133–8138. doi: 10.1073/pnas.072223799.

Jarvis, P. (2008) 'Targeting of nucleus-encoded proteins to chloroplasts in plants', *New Phytologist*, 179(2), pp. 257–285. doi: https://doi.org/10.1111/j.1469-8137.2008.02452.x.

Jarvis, P. and Robinson, C. (2004) 'Mechanisms of Protein Import and Routing in Chloroplasts', *Current Biology*, 14(24), pp. R1064–R1077. doi: 10.1016/j.cub.2004.11.049.

Jarvis, P. and Soll, J. (2002) 'Toc, tic, and chloroplast protein import', *Biochimica Et Biophysica Acta*, 1590(1–3), pp. 177–189. doi: 10.1016/s0167-4889(02)00176-3.

Jung, H.-S. *et al.* (2013) 'Subset of heat-shock transcription factors required for the early response of Arabidopsis to excess light', *Proceedings of the National Academy of Sciences*, 110(35), pp. 14474–14479. doi: 10.1073/pnas.1311632110.

Krasileva, K. V. *et al.* (2013) 'Separating homeologs by phasing in the tetraploid wheat transcriptome', *Genome Biology*, 14(6), p. R66. doi: 10.1186/gb-2013-14-6-r66.

Krasileva, K. V. *et al.* (2017) 'Uncovering hidden variation in polyploid wheat', *Proceedings of the National Academy of Sciences*, 114(6), pp. E913–E921. doi: 10.1073/pnas.1619268114.

Lin, M. T. *et al.* (2020) 'Small subunits can determine enzyme kinetics of tobacco Rubisco expressed in Escherichia coli', *Nature Plants*, 6(10), pp. 1289–1299. doi: 10.1038/s41477-020-00761-5.

Liu, Z. *et al.* (2015a) 'Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (Triticum aestivum L.)', *BMC Plant Biology*, 15(1), p. 152. doi: 10.1186/s12870-015-0511-8.

Liu, Z. *et al.* (2015b) 'Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (Triticum aestivum L.)', *BMC Plant Biology*, 15(1), pp. 1–20. doi: 10.1186/s12870-015-0511-8.

Lobo, A. K. M. *et al.* (2019) 'Overexpression of ca1pase Decreases Rubisco Abundance and Grain Yield in Wheat', *Plant Physiology*, 181(2), p. 471. doi: 10.1104/pp.19.00693.

Long, S. P. *et al.* (2006) 'Can improvement in photosynthesis increase crop yields?', *Plant, Cell & Environment*, 29(3), pp. 315–330. doi: https://doi.org/10.1111/j.1365-3040.2005.01493.x.

Long, S. P., Marshall-Colon, A. and Zhu, X.-G. (2015a) 'Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential', *Cell*, 161(1), pp. 56–66. doi: 10.1016/j.cell.2015.03.019.

Long, S. P., Marshall-Colon, A. and Zhu, X.-G. (2015b) 'Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential', *Cell*, 161(1), pp. 56–66. doi: 10.1016/j.cell.2015.03.019.

Long, S. P. and Ort, D. R. (2010) 'More than taking the heat: crops and global change', *Current Opinion in Plant Biology*, 13(3), pp. 241–248. doi: 10.1016/j.pbi.2010.04.008.

Lorimer, G. H., Badger, M. R. and Andrews, T. J. (1976) 'The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications', *Biochemistry*, 15(3), pp. 529–536. doi: 10.1021/bi00648a012.

Madlung, A. (2013) 'Polyploidy and its effect on evolutionary success: old questions revisited with new tools', *Heredity*, 110(2), pp. 99–104. doi: 10.1038/hdy.2012.79.

'Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation — IPCC' (no date). Available at: https://www.ipcc.ch/report/managing-the-risks-of-extreme-events-and-disasters-to-advance-climate-change-adaptation/ (Accessed: 21 February 2021).

Martín, A. C. *et al.* (2018) 'Genome-Wide Transcription During Early Wheat Meiosis Is Independent of Synapsis, Ploidy Level, and the Ph1 Locus', *Frontiers in Plant Science*, 9, p. 1791. doi: 10.3389/fpls.2018.01791.

Martinez-Perez, E., Shaw, P. and Moore, G. (2001) 'The Ph1 locus is needed to ensure specific somatic and meiotic centromere association', *Nature*, 411(6834), pp. 204–207. doi: 10.1038/35075597.

Moore, R. C. and Purugganan, M. D. (2005) 'The evolutionary dynamics of plant duplicate genes', *Current Opinion in Plant Biology*, 8(2), pp. 122–128. doi: 10.1016/j.pbi.2004.12.001.

Morita, K. *et al.* (2016) 'Identification and expression analysis of non-photosynthetic Rubisco small subunit, OsRbcS1-like genes in plants', *Plant Gene*, 8, pp. 26–31. doi: 10.1016/j.plgene.2016.09.004.

Otto, S. P. and Whitton, J. (2000) 'Polyploid incidence and evolution', *Annual Review of Genetics*, 34, pp. 401–437. doi: 10.1146/annurev.genet.34.1.401.

Paolacci, A. R. *et al.* (2009) 'Identification and validation of reference genes for quantitative RT-PCR normalization in wheat', *BMC Molecular Biology*, 10(1), p. 11. doi: 10.1186/1471-2199-10-11.

Parry, M. A. J. *et al.* (2003) 'Manipulation of Rubisco: the amount, activity, function and regulation', *Journal of Experimental Botany*, 54(386), pp. 1321–1333. doi: 10.1093/jxb/erg141.

Parry, M. A. J. *et al.* (2008) 'Rubisco regulation: a role for inhibitors', *Journal of Experimental Botany*, 59(7), pp. 1569–1580. doi: 10.1093/jxb/ern084.

Perdomo, J. A. *et al.* (2017) 'Rubisco and Rubisco Activase Play an Important Role in the Biochemical Limitations of Photosynthesis in Rice, Wheat, and Maize under High Temperature and Water Deficit', *Frontiers in Plant Science*, 8. doi: 10.3389/fpls.2017.00490.

Pfaffl, M. W. (2001) 'A new mathematical model for relative quantification in real-time RT-PCR', *Nucleic Acids Research*, 29(9), p. e45. doi: 10.1093/nar/29.9.e45.

Phillips, R. and Milo, R. (2009) 'A feeling for the numbers in biology', *Proceedings of the National Academy of Sciences*, 106(51), p. 21465. doi: 10.1073/pnas.0907732106.

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., Ex, F. van, *et al.* (2018) 'The transcriptional landscape of polyploid wheat', *Science*, 361(6403). doi: 10.1126/science.aar6089.

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., Van Ex, F., *et al.* (2018) 'The transcriptional landscape of polyploid wheat', *Science*, 361(6403). doi: 10.1126/science.aar6089.

Ray, D. K. *et al.* (2012) 'Recent patterns of crop yield growth and stagnation', *Nature Communications*, 3(1), p. 1293. doi: 10.1038/ncomms2296.

Ray, D. K. *et al.* (2013) 'Yield Trends Are Insufficient to Double Global Crop Production by 2050', *PLOS ONE*, 8(6), p. e66428. doi: 10.1371/journal.pone.0066428.

Renny-Byfield, S. and Wendel, J. F. (2014) 'Doubling down on genomes: polyploidy and crop plants', *American Journal of Botany*, 101(10), pp. 1711–1725. doi: 10.3732/ajb.1400119.

Robinson, S. P. and Portis, A. R. (1988) 'Release of the nocturnal inhibitor, carboxyarabinitol- 1-phosphate, from ribulose bisphosphate carboxylase/oxygenase by rubisco activase', *FEBS LETTERS*, 233(2), p. 4.

Salesse, C. *et al.* (2017) 'The Rubisco Chaperone BSD2 May Regulate Chloroplast Coverage in Maize Bundle Sheath Cells', *Plant Physiology*, 175(4), pp. 1624–1633. doi: 10.1104/pp.17.01346.

Salesse-Smith, C. E. *et al.* (2018) 'Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize', *Nature Plants*, 4(10), pp. 802–810. doi: 10.1038/s41477-018-0252-4.

Salman-Minkov, A., Sabath, N. and Mayrose, I. (2016) 'Whole-genome duplication as a key factor in crop domestication', *Nature Plants*, 2(8), pp. 1–4. doi: 10.1038/nplants.2016.115.

Saschenbrecker, S. *et al.* (2007) 'Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco', *Cell*, 129(6), pp. 1189–1200. doi: 10.1016/j.cell.2007.04.025.

Scafaro, A. P. *et al.* (2019) 'A Conserved Sequence from Heat-Adapted Species Improves Rubisco Activase Thermostability in Wheat', *Plant Physiology*, 181(1), pp. 43–54. doi: 10.1104/pp.19.00425.

Simkin, A. J., López-Calcagno, P. E. and Raines, C. A. (2019) 'Feeding the world: improving photosynthetic efficiency for sustainable crop production', *Journal of Experimental Botany*, 70(4), pp. 1119–1140. doi: 10.1093/jxb/ery445.

Sparks, C. A. and Jones, H. D. (2014) 'Genetic transformation of wheat via particle bombardment', *Methods in Molecular Biology (Clifton, N.J.)*, 1099, pp. 201–218. doi: 10.1007/978-1-62703-715-0_17.

Spreitzer, R. J. (2003) 'Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase', *Archives of Biochemistry and Biophysics*, 414(2), pp. 141–149. doi: 10.1016/s0003-9861(03)00171-1.

Spreitzer, R. J. and Salvucci, M. E. (2002) 'RUBISCO: Structure, Regulatory Interactions, and Possibilities for a Better Enzyme', *Annual Review of Plant Biology*, 53(1), pp. 449–475. doi: 10.1146/annurev.arplant.53.100301.135233.

Stern, D. B., Hanson, M. R. and Barkan, A. (2004) 'Genetics and genomics of chloroplast biogenesis: maize as a model system', *Trends in Plant Science*, 9(6), pp. 293–301. doi: 10.1016/j.tplants.2004.04.001.

Studer, R. A. *et al.* (2014) 'Stability-activity tradeoffs constrain the adaptive evolution of RubisCO', *Proceedings of the National Academy of Sciences*, 111(6), p. 2223. doi: 10.1073/pnas.1310811111.

Suzuki, Y. and Makino, A. (2012) 'Availability of Rubisco Small Subunit Up-Regulates the Transcript Levels of Large Subunit for Stoichiometric Assembly of Its Holoenzyme in Rice', *Plant Physiology*, 160(1), p. 533. doi: 10.1104/pp.112.201459.

Tilman, D. *et al.* (2011) 'Global food demand and the sustainable intensification of agriculture', *Proceedings of the National Academy of Sciences*, 108(50), p. 20260. doi: 10.1073/pnas.1116437108.

Vitlin Gruber, A. and Feiz, L. (2018) 'Rubisco Assembly in the Chloroplast', *Frontiers in Molecular Biosciences*, 5. doi: 10.3389/fmolb.2018.00024.

Watson, A. *et al.* (2018) 'Speed breeding is a powerful tool to accelerate crop research and breeding', *Nature Plants*, 4(1), pp. 23–29. doi: 10.1038/s41477-017-0083-8.

Way, D. A. and Yamori, W. (2014) 'Thermal acclimation of photosynthesis: on the importance of adjusting our definitions and accounting for thermal acclimation of respiration', *Photosynthesis Research*, 119(1), pp. 89–100. doi: 10.1007/s11120-013-9873-7.

Wickham, H. *et al.* (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4(43), p. 1686. doi: 10.21105/joss.01686.

Wilson, A. T. and Calvin, M. (1955) 'The Photosynthetic Cycle. CO2 Dependent Transients', *Journal of the American Chemical Society*, 77(22), pp. 5948–5957. doi: 10.1021/ja01627a050.

Wostrikoff, K. and Stern, D. (2007) 'Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts', *Proceedings of the National Academy of Sciences*, 104(15), p. 6466. doi: 10.1073/pnas.0610586104.

Yamada, K. *et al.* (2019) 'Duplication history and molecular evolution of the rbcS multigene family in angiosperms', *Journal of Experimental Botany*, 70(21), pp. 6127–6139. doi: 10.1093/jxb/erz363.

Zhang, M. *et al.* (2020) 'The contribution of spike photosynthesis to wheat yield needs to be considered in process-based crop models', *Field Crops Research*, 257, p. 107931. doi: 10.1016/j.fcr.2020.107931.

Zhao, C. *et al.* (2017) 'Temperature increase reduces global yields of major crops in four independent estimates', *Proceedings of the National Academy of Sciences*, 114(35), p. 9326. doi: 10.1073/pnas.1701762114.

Zhao, Q. and Liu, C. (2018) 'Chloroplast Chaperonin: An Intricate Protein Folding Machine for Photosynthesis', *Frontiers in Molecular Biosciences*, 4. doi: 10.3389/fmolb.2017.00098.

Ziska, L. H. *et al.* (2012) 'Food security and climate change: on the potential to adapt global crop production by active selection to rising atmospheric carbon dioxide', *Proceedings of the Royal Society B: Biological Sciences*, 279(1745), pp. 4097–4105. doi: 10.1098/rspb.2012.1005.