The *NAM-B1* transcription factor and the control of grain composition in wheat

Philippa G. M Borrill

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

John Innes Centre

Norwich

September 2014

Abstract

The *NAM-B1* transcription factor increases grain protein content, alters grain micronutrient content and accelerates monocarpic senescence, often without imposing a yield penalty. The aim of this thesis was to understand the mechanisms by which *NAM-B1* influences nutrient remobilisation and monocarpic senescence to cause these effects.

To achieve this I have examined the expression patterns of *NAM-B1* and its homologues during development. I have studied the effects of *NAM-B1* on nutrient transport, photosynthetic capacity and grain filling using a range of molecular biology and physiological techniques. Finally to understand the network of genes which *NAM-B1* regulates I have used chromatin-immunoprecipitation followed by next-generation sequencing (ChIP-seq) to identify downstream targets, and compared these to differentially expressed genes in plants with down-regulated expression of *NAM-B1* homologues (*NAM* RNAi plants).

I have found that *NAM-B1* expression increases after anthesis in both vegetative and reproductive tissues, including the grain. In stem and leaf tissues I identified that *NAM* genes are highly expressed in the vascular bundles, which might be important for nutrient transport. However I did not find evidence for *NAM* genes altering xylem or phloem transport. I found that in *NAM* RNAi plants, grain development was decoupled from flag leaf senescence. In RNAi plants starch synthesis enzymes were less active during the middle of grain filling than in control plants, potentially resulting in the reallocation of photosynthate to the stems as water soluble carbohydrates. Many of the putative NAM-B1 target genes identified by ChIP-seq have functions related to photosynthesis and validation of these candidate genes is ongoing.

In summary I have identified putative NAM-B1 target genes and found that *NAM-B1* may act in a tissue specific manner to regulate monocarpic senescence and grain filling. Furthermore I have highlighted novel functions related to carbohydrate metabolism in stems and the grain.

Acknowledgements

First of all I would like to thank my supervisor Cristobal Uauy for his guidance, his never ending supply of enthusiasm and all his help in developing my scientific skills. I would like to thank my co-supervisor Alison Smith for all her invaluable advice and especially her support during the more difficult times. I am also grateful to Tony Miller who gave excellent advice at my supervisory meetings.

I am very grateful for all the help I have received during my PhD and the support from the members of both the Uauy and Smith labs. Being privy to the expertise and friendship of two such great groups of people has been a delight.

There are many people who have contributed to this project and I especially want to thank Martin Trick for his bioinformatics advice and collaboration, Xana Rebocho for her guidance with RNA *in situ* hybridisation, Alistair McCormick for training me to measure photosynthesis, Brendan Fahy for carrying out the enzyme activity assays and Vasilios Andriotis for his guidance on protein expression.

I was lucky to have great officemates throughout my PhD; Emilie, Nick, Nikolai and Sarah, who have been constant sources of advice, laughter, cake, chocolate and biscuits!

During the entirety of my PhD I have been fortunate to have great friends in Norwich including Hadrien, Nikolai, Stuart, Simon, Albor and especially my housemates Ruth, Athena, Doreen and Jenny, with whom I have shared many fun times – thank you all! I would also like to thank Mike for all his love and support especially during the last few months of thesis mania.

I would like to thank old friends who have stuck with me throughout my PhD: Christine, Louise, Orlando, Roz and Anna – thanks for listening to all the ups and downs and providing a respite from Norwich when it was a bit too much.

Finally, I would like to thank my family for all their support, encouragement and advice throughout this PhD, which at times seemed very long; without them I would not have made it so far.

Table of contents

List of figures

List of tables

Abbreviations

$A\text{-}C_i$	Photosynthetic response to CO_2 concentration
AGPase	ADP-glucose pyrophosphorylase
bp	Base pair
cDNA	Complementary DNA
CER	Controlled environment room
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP followed by next-generation sequencing
CO_2	Carbon dioxide
DAA	Days after anthesis
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMS	Ethyl methanesulfonate
EMSA	Electrophoretic mobility shift assay
F6P	Fructose 6-phosphate
g	Acceleration due to gravity
G6P	Glucose 6-phsophate
GO	Gene ontology
GPC	Grain protein content
His	Histidine
IP	Immunoprecipitation
kb	Kilo base pair
MAPQ	Mapping quality
MAS	Marker assisted selection
Mb	Mega base pair
mRNA	Messenger RNA
N	Nitrogen
NAD+	Nicotinamide adenine dinucleotide (oxidised)
NADH	Nicotinamide adenine dinucleotide (reduced)
NADP+	Nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NIL	Near isogenic line
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PGI	Phosphoglucoisomerase
qPCR	Quantitative PCR
QTL	Quantitative trait loci
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SEM	Standard error of the mean
TF	Transcription factor
UTR	Untranslated region
WSC	Water soluble carbohydrates
WT	Wild type

1 Chapter 1 – General introduction

1.1 The origin of wheat

Approximately 10,000 years ago the "Neolithic Revolution" introduced agriculture for the first time. Cereals, including wheat, were cultivated for their grain and people adopted a settled lifestyle dependent on agriculture rather than on hunting and gathering food. The earliest cultivated wheat was the diploid Einkorn wheat (*Triticum monococcum*, genomes A^mA^m) (Salamini et al. 2002) which was domesticated in southeast Turkey (Heun et al. 1997).

An important next step was the domestication of tetraploid wheats. Tetraploid wild emmer (*Triticum turgidum* ssp. *dicoccoides,* genomes AABB) was formed by the hybridisation of two diploid grasses: the A-genome donor was *Triticum urartu* (genomes A^uA^u) and the B-genome donor was an unknown member of the *Sitopsis* section closely related to *Aegilops speltoides* (genomes SS) (Haider 2013). Emmer wheat (*Triticum turgidum* ssp. *dicoccon,* genomes AABB) was domesticated from wild emmer (Luo M. C. et al. 2007). Domesticated emmer wheat was the most important crop in the Fertile Crescent for thousands of years due to its high yields and adaptability. Domesticated emmer gave rise to modern day pasta wheat cultivars *Triticum durum* (genomes AABB) (Feldman 2001).

The spread of emmer wheat northeast towards the Caspian sea facilitated its hybridisation with *Aegilops tauschii* (genomes DD) (Dvorak et al. 1998) to form hexaploid *Triticum aestivum* (genomes AABBDD) (Kihara 1944). The modern bread wheat cultivars (*Triticum aestivum* ssp. *aestivum*) which arose from this ancestral hexaploid had superior yield and adaptability than their progenitor species.

Selection by farmers led to a clear differentiation between cultivated wheat varieties and their wild ancestors, known as the domestication syndrome. The key trait altered by domestication was reduction of spike-shattering. This enabled the easy harvesting of grain without them being lost by scattering on the ground. The non-shattering trait is encoded by the *brittle rachis (Br)* genes located on the short arms of chromosome 3A and 3B (Nalam et al. 2006). The second important trait altered during domestication was the conversion of hulled forms to free-threshing wheat. This involved the loosening and softening of the glumes surrounding the grain which made separating the grain from the spikelets less labour intensive. The free-threshing forms arose by recessive

mutations at the *tenacious glume* (*Tg*) locus on chromosome 2D, accompanied by modifying effects by the dominant mutation in the *Q* gene on chromosome 5A (Dubcovsky and Dvorak 2007, Jantasuriyarat et al. 2004, Simons et al. 2006). The *Q* gene, which encodes an AP2-like transcription factor, affects a suite of domestication syndrome traits such as rachis fragility, glume shape, plant height and spike emergence time (Simons et al. 2006).

During domestication other traits were also selected including increased seed size, reduced tiller number and a more erect growth habit (Dubcovsky and Dvorak 2007, Purugganan and Fuller 2009). More recently during the Green Revolution semi-dwarf genes from Japanese cultivars were introduced which reduced stem lengths and therefore increased the harvest index (ratio of grain yield to the above ground tissue at maturity (Hay 1995)). The reduced height of the crop also improved the response to nitrogen fertilisation, allowing application of fertiliser to increase yields without the risk of lodging (Hedden 2003). The result of the genetic and agronomic changes introduced during the Green Revolution allowed the production of staple crops to double in forty years (Khush 2001) without a large increase in land area used for farming.

1.2 Global production and consumption

Wheat is a globally important crop and provides approximately 20 % of calories and protein eaten by people around the world (Wheat Initiative 2012). Furthermore wheat is a significant source of micronutrients in the diet. For example in the UK, cereals including wheat provide 44 % and 25 % of daily iron and zinc intake respectively (Shewry 2009).

Despite the great yield increases achieved by the Green Revolution, global food production will need to increase by 70 - 100 % to feed the predicted world population of 9.1 billion people by 2050 (FAO 2009, Tilman et al. 2011). However the current rates of increase of the world's top four crops (maize, rice, wheat and soybean) are far below the rates required to reach this target (Ray et al. 2013). For example global wheat yields increased on average 0.9 % per year from 1961 to 2008, but a yield increase of 2.4 % per year would be required to double wheat production by 2050 (Ray et al. 2013).

1.3 Grain development

Grain development begins with a "double fertilisation" event when one pollen reproductive nucleus fuses with two polar nuclei within the embryo sac to form the triploid endosperm nucleus and the second pollen nucleus fuses with the egg nucleus to form the diploid zygote. After fertilisation the endosperm undergoes mitotic divisions without cell divisions, leading to the formation of one large cell with multiple nuclei. This large cell is subsequently divided into individual cells (cellularisation) and at this stage the cells begin to differentiate (Olsen O. A. 2001). The outer layer of cells differentiate into the aleurone layer and the inner cells into thin-walled cells which will become the starchy endosperm (Shewry et al. 2012). By approximately 2 weeks after anthesis cell division in the endosperm ceases and growth occurs only by cell expansion as storage components (proteins, starch and micronutrients) are deposited. Under UK growth conditions the grain filling period is most active from approximately 14 to 28 days after anthesis and is followed by a phase of grain maturation and desiccation (Shewry et al. 2012).

The mature grain is made up of several distinct tissues including the starchy endosperm, the aleurone, seed coat, pericarp and the embryo (**Figure 1.1**). The starchy endosperm accounts for ~83 % of the dry weight and the aleurone layer for ~6.5 % (Barron et al. 2007). The remainder consists of the embryo (3 %) which is formed by the division and differentiation of the diploid zygote, and the outer layers of maternal tissue (~7.5 %) (Barron et al. 2007), which include the pericarp and seed coat which originate from the ovary wall.

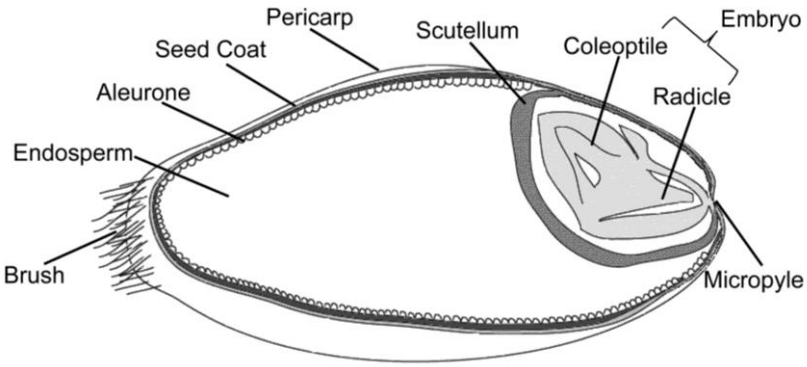


Figure 1.1. Diagram of a longitudinal section through a wheat grain, adapted from (Rathjen et al. 2009).

The pathway by which nutrients enter and fill the grain involves several transport steps. Nutrients are transported to the ear in both the xylem and the phloem, but a block in the xylem between the rachilla and the grain known as the xylem discontinuity (Zee and O'Brien 1970) means that all nutrients must travel the final part of the journey into the grain in the phloem. Upon arrival at the grain crease, nutrients are unloaded from the phloem into specialised maternal nucellar projection cells (**Figure 1.2**) which have extended membrane surfaces to enable the transport across the membrane into the endosperm cavity (Patrick and Offler 2001). Nutrients are taken up by the developing grain at the region of modified aleurone via specialised transporters. For example sucrose is taken up by the sucrose transporter SUT1 (Aoki et al. 2002). Nutrients then move symplasmically into the endosperm via plasmodesmata which are present at high densities (Patrick and Offler 2001).

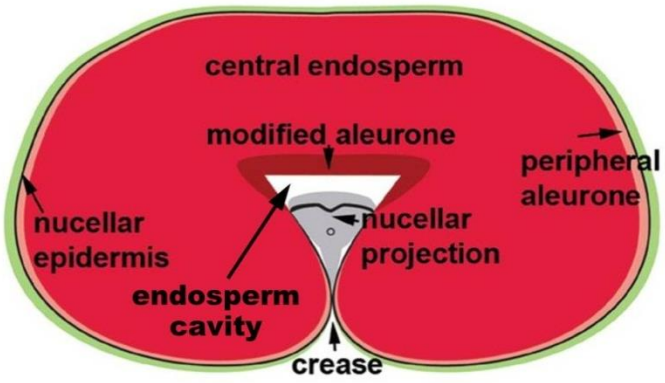


Figure 1.2. Transverse section through a wheat grain, adapted from (Hands et al. 2012).

1.4 Grain filling is intimately linked to monocarpic senescence

In cereal crops a fine balance must be achieved between prolongation of photosynthesis to maximise carbohydrate supply to the grain, and the onset of senescence to ensure efficient remobilisation of nutrients into the grain. A programme of catabolic processes is executed to convert cellular components into exportable nutrients which can be remobilised to the grain during senescence (Cantu et al. 2011, Hörtensteiner and Feller 2002). In wheat, remobilisation provides the majority of the micro- and macro-nutrients to the developing grain and uptake by the roots after anthesis plays a smaller role (Kichey et al. 2007). Monocarpic senescence and nutrient remobilisation are under tight genetic and developmental control, however the genes regulating these processes are poorly characterised at present in wheat. Studies into leaf senescence have revealed many senescence associated genes in both model species such as Arabidopsis (Breeze et al. 2011) and monocots (Gregersen and Holm 2007, Jukanti Aravind K. et al. 2008) with transcription factors co-ordinating this large-scale metabolic re-programming (Uauy et al. 2006b, Wu et al. 2012).

Nutrient remobilisation is not the only hallmark of monocarpic senescence. A suite of other morphological, physiological and molecule changes occur including non-visible senescence such as degradation of DNA, RNA and proteins and a reduction in photosynthetic rate, and visible senescence due to chlorophyll degradation (Wu et al. 2012). This final stage in plant development is influenced by a wide range of genetic and environmental factors including age, phytohormones (Lee et al. 2011, Young et al. 2004), signalling compounds such as reactive oxygen species (Conklin and Barth 2004, Strother 1988), metabolites, particularly sugars (Parrott et al. 2007), and nitrogen compounds (Hörtensteiner and Feller 2002), pathogens (Espinoza et al. 2007), drought (Yang J. C. et al. 2003) and salinity (Zheng et al. 2008).

The process of monocarpic senescence has been extensively reviewed in the past few years (Davies and Gan 2012, Distelfeld et al. 2014, Fischer 2012, Guo and Gan 2014, Thomas H. and Ougham 2014, Wu et al. 2012) therefore below I will discuss briefly how senescence affects the content of carbohydrates, proteins and micronutrients in the grain. These components are discussed in more detail in later chapters as appropriate.

1.4.1 Grain carbohydrates

During monocarpic senescence the photosynthetic apparatus is disassembled and the rate of photosynthesis declines. The breakdown of chloroplasts provides nitrogen and micronutrients which can be remobilised to the grain (see 1.4.2 and 1.4.3). This reduction in photosynthesis affects grain filling because the majority of sucrose used for carbohydrate synthesis in the grain is produced by photosynthesis in the leaf, stem and ear tissues (Araus et al. 1993, Sanchez-Bragado et al. 2014). The timing of senescence is essential to balance the demand for carbohydrates in the grain with the demand for remobilised nutrients from photosynthetic tissues. Delaying senescence can have variable effects on yield (which is largely determined by starch content in the grain), with extra assimilation leading to increased yield in cases where the source tissue (e.g. flag leaf) is limiting and having no effect on yield in cases where the sink (grain) is limiting: this is discussed in 3.1.1.

An additional source of carbohydrates for starch synthesis in the grain is provided by water soluble carbohydrates stored in stem tissues (Schnyder 1993). During periods when photosynthetic rates produce excess photosynthate that cannot be utilised by the grain, the photosynthates may be stored as water soluble carbohydrates (WSC) in stem tissues. WSC are remobilised during the later stages of grain filling and can help to maintain yield under stress conditions, e.g. drought (Blum 1998).

The sucrose which arrives in the ear is converted by a series of enzymatic steps into a glucose donor (ADP-glucose) for use in starch synthesis (see 3.1.1.6). The starch produced constitutes 70-85 % of the final dry weight of the grain (Housley et al. 1981) and therefore is a key determinant of yield.

1.4.2 Grain nitrogen

After starch, protein is the most abundant component of wheat grain, representing 8-15 % of the dry weight of the mature grain (Emes et al. 2003). The majority (50-95%) of nitrogen (N) in wheat grain comes from the remobilisation of N stored in vegetative tissues before anthesis, rather than from uptake after anthesis (Kichey et al. 2007, Palta and Fillery 1995). The above-ground tissues contribute the majority of N for grain filling (leaves 40 %, glumes 23 % and stem 23 %) whereas roots contribute only 16 % of grain N (Simpson et al. 1983).

The genetic control of N remobilisation is linked to the regulation of monocarpic senescence (Gaju et al. 2014, Masclaux et al. 2001, Sinclair and de Wit 1975, Uauy et al. 2006a). At the onset of senescence, protein degradation is initiated and the chloroplasts begin to be dismantled (Masclaux et al. 2001). In particular the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which can account for over 50 % of total soluble protein in a leaf, is an important source of N to the developing grain (Masclaux et al. 2001). The degradation of chloroplasts also influences the rate of photosynthesis: a balance must be struck between maintaining photosynthesis long enough to export enough sucrose for starch synthesis in the grain, and initiating senescence to provide a N source for grain filling. If photosynthesis is extended by delaying senescence this may not only decrease grain N by preventing the remobilisation of N in vegetative tissues, it may also increase the amount of carbohydrate in the grain and thus dilute the N content. The trade-off between photosynthesis and nitrogen remobilisation may explain why there is a negative correlation between grain yield and grain protein content (GPC) (Simmonds 1995, Slafer et al. 1990).

Not only does senescence affect N remobilisation to the grain, but N availability can accelerate or delay senescence itself. For example when the demands of the grain for nitrogen exceed the supply provided by root uptake, senescence is accelerated to provide the deficit (Triboi and Triboi-Blondel 2002). In the opposite case when N is readily available, for example due to a late application of N fertiliser, leaf greenness can be maintained and senescence is delayed (Bancal et al. 2008). The effect of N on senescence is modulated by a range of genetic factors either by direct effects on senescence (e.g. protein catabolic enzymes) and N uptake (e.g. nitrogen reductase), or by altering the source-sink relationship.

There is a debate whether GPC is limited by the source provision of N (either from remobilisation or uptake after anthesis) or by the ability of the grain to synthesise protein (Foulkes et al. 2009). Early studies comparing *in vitro* cultured wheat grain indicated that differences in grain N between high and low GPC genotypes were due to differences in protein biosynthetic ability in the grain (Donovan et al. 1977), however more recent experiments have shown that source may be limiting under field conditions. For example under non-limiting N supply, removal of the upper portion of the ear resulted in an increase in grain N content (Martre et al. 2003) and increasing N fertiliser generally increases GPC (Dupont and Altenbach 2003). It has been shown that under high N or mild N stress, genetic variation in grain N % was mainly influenced by pre-anthesis N accumulation, rather than post-anthesis remobilisation (Gaju et al. 2014). Therefore the N status of the plant throughout its whole lifespan will affect the final GPC.

The economic value of wheat to farmers depends not only on high yields but also on the level of protein content: high protein content grains receive a premium because they are suitable for bread making. Therefore the negative relationship between yield and GPC poses a problem to wheat breeders who aim to increase yields. One way proposed to escape the negative correlation between yield and GPC is to increase the supply of N to the grain by uptake after anthesis. This route would enable grain N to be increased without requiring extra remobilisation: this may be possible because uptake involving nitrogen reductase activity is strongly correlated to grain N and can provide N to grain after flowering (Kichey et al. 2007). This method could be combined with an approach to simultaneously delay senescence (to increase yield) and increase N demand from the grain (to increase GPC) (Triboi and Triboi-Blondel 2002): the extra N required could be supplied by the increased uptake after anthesis. Recently loci which break the relationship between yield and GPC have been identified in a multi-environment trial using three doubled haploid populations (Bogard et al. 2013) and they may prove useful in wheat breeding to increase GPC without decreasing yield.

1.4.3 Grain micronutrients

Although micronutrients such as iron and zinc are a small proportion of the dry matter within wheat grain, wheat is an important dietary source of micronutrients for millions of people worldwide. Micronutrients imported into the grain may be supplied by uptake by the roots or by remobilisation from other tissues, depending on environmental conditions and the individual micronutrient's properties (Garnett and Graham 2005). Individual micronutrients have differing degrees of mobility in the xylem and phloem (Feller et al. 1992, Herren and Feller 1997, Martin 1982) (see 2.1.5) and are transported by diverse families of proteins such as the ZRT-, IRT-like proteins (ZIPs) and yellow stripe like (YSL) transporters (Borrill et al. 2014). (See (Borrill et al. 2014) for a review of iron and zinc transport in wheat).

Micronutrient content in the grain can be strongly influenced by environmental and agronomic factors. For example uptake and remobilisation of zinc during grain filling was increased by N fertilisation (Erenoglu et al. 2011), and enrichment of atmospheric CO_2 reduced mineral content (Loladze 2014), which has implications for the effects of climate change. Micronutrient remobilisation is intimately connected to senescence. Notably about 70 % of iron in mesophyll cells is located in the chloroplasts (Shikanai et al. 2003) and the other transition metals such as copper, manganese and zinc are associated with enzymes involved in energy production (Pottier et al. 2014). Therefore a trade-off exists between micronutrient use in the leaf to sustain photosynthesis and the need to remobilise the components of the photosynthetic apparatus, which parallels the situation for N.

1.5 Genetic control of grain nutrient content

The control of grain nutrient content at the genetic level in wheat is not well characterised. Many QTL have been identified for yield (representing grain carbohydrates), GPC and micronutrient content which are useful for marker assisted selection strategies in wheat breeding. However to understand the control of grain nutrient content at the molecular level it is necessary to identify the genes underlying these QTL. Relatively few genes have been cloned in wheat, compared to other monocots such as rice, due to the complexity of the hexaploid genome and the difficulty, until recently, to develop reliable markers. However the gene *NAM-B1*, which was originally identified as the causal gene under a QTL for GPC (Uauy et al. 2006b), provides an entry point into the molecular control of grain nutrient content.

1.5.1 The origin and cloning of *NAM-B1*

The *NAM-B1* gene originated from a cross between the high GPC *Triticum turgidum* ssp. *dicoccoides* accession FA-15-3 (Avivi 1978) and *T. turgidum* ssp. *durum* cultivar Langdon. A quantitative trait locus (QTL) for GPC was mapped on chromosome 6BS (Joppa et al. 1997) and this locus was associated with increases in GPC of ~14 g kg^{-1} (Chee et al. 2001, Joppa et al. 1997, Mesfin et al. 1999). The QTL was mapped as a single Mendelian locus *Gpc-B1* (Olmos et al. 2003) and was located within an 0.3 cM interval (Distelfeld et al. 2004). The molecular markers *Xuhw89* and *Xucw71* within this region flank a 245 kb physical contig containing *Gpc-B1* (Distelfeld et al. 2006). The gene underlying this QTL was cloned as the *NAM-B1* NAC (*NAM*/*ATAF1*/*CUC2*) transcription factor (Uauy et al. 2006b).

1.5.2 Agronomic effects of *NAM-B1*

The functional allele of *NAM-B1*, which increases GPC, is largely absent from modern durum and bread wheat (Uauy et al. 2006b), although it has been detected in Swedish spring wheat varieties and in a historical collection from 1862 (Asplund et al. 2010, Asplund et al. 2013). The selection for the non-functional allele may be due to a grain size penalty associated with the functional allele (Asplund et al. 2010, Uauy et al. 2006a). The non-functional allele is either due to a complete deletion of the gene or due to a 1-bp insertion in the first exon which disrupts the reading frame; the resulting predicted protein has no similarity to any Genbank protein and has no NAC domain (Uauy et al. 2006b).

The functional *NAM-B1* gene has been incorporated into several breeding programmes in an effort to improve GPC. This includes breeding programmes at the International Maize and Wheat Improvement Centre (CIMMYT) and many commercial companies. Functional *NAM-B1* has been released in commercial cultivars including Lillian (DePauw R. M. et al. 2005), which was the most widely grown wheat cultivar in Canada in 2011 (DePauw R.M. et al. 2011). The introgression of the QTL containing functional *NAM-B1* in certain varieties is associated with a reduction in grain weight (Brevis and Dubcovsky 2010, Brevis et al. 2010), however in other studies a significant yield effect is not detected (Tabbita et al. 2013). Therefore the benefits of a functional *NAM-B1* allele must be assessed on a case-by-case basis for each variety and environment type.

NAM-B1 is not only important for GPC, it also affects micronutrient content. The high protein allele has a positive effect on iron and zinc concentrations of 18 % and 12 % respectively in recombinant substitution lines (Distelfeld et al. 2007). This combination of improved protein, iron and zinc content in wheat grain is an attractive prospect to improve the nutritional value of wheat in several key respects.

NAM-B1 also affects senescence with the functional allele causing an acceleration in flag leaf senescence of approximately 1.5 days (Uauy et al. 2006a). The accelerated senescence may cause a reduction in grain size due to a shorter period of grain filling (Uauy et al. 2006a). Accelerated senescence is linked to accelerated N remobilisation from the leaves (Triboi and Triboi-Blondel 2002) and in the case of *NAM-B1* this accelerated N remobilisation contributes to higher final GPC (Guttieri et al. 2013, Kade et al. 2005, Uauy et al. 2006a, Waters et al. 2009).

1.5.3 Biological function of *NAM-B1*

NAM-B1 is a transcription factor (TF) which we hypothesise regulates many other genes to bring about these changes to grain nutrient content and senescence. *NAM-B1* has two homoeologous copies (*NAM-A1* and *NAM-D1*) located on the group 6 chromosomes, and paralogous copies on the group two chromosomes (*NAM-A2*, *NAM-B2* and *NAM-D2*), although until recently the *NAM-A2* copy was thought to be deleted (Stephen Pearce, UC Davis, personal communication). These homologues of *NAM-B1* appear to be functional from their protein sequences (Uauy et al. 2006b). Down-regulation of all homologues by RNA interference (RNAi) (termed *NAM* RNAi) caused a dramatic 3 week delay in senescence (**Figure 1.3**), a 30 % decrease in GPC, a 38 % decrease in

iron concentration and a 36 % decrease in zinc concentration compared to non-transgenic controls (Uauy et al. 2006b).

1.5.3.1 Effect on nutrient remobilisation

A further study (Waters et al. 2009) investigated whether the reduction in mineral and protein content of the *NAM* RNAi grain was due to decreased remobilisation, lower plant uptake or decreased partitioning to the grain. Although total accumulation of iron and zinc was similar in RNAi and control plants, the RNAi plants showed less remobilisation than the control plants, for example iron was lost from control plants leaves, stems peduncle and rachis from anthesis until maturity, whereas over the same time-span the RNAi plants retained or accumulated iron in these tissues (Waters et al. 2009). However in nutrient limiting conditions RNAi plants were able to remobilise iron and zinc, although to a lesser degree than control plants. These results show that micronutrient differences observed in *NAM* RNAi lines compared to controls are due an altered ability to remobilise nutrients from vegetative tissues and to partition the micronutrients to the grain, rather than a change to net nutrient uptake. These results were supported by a second study which investigated the effects of abiotic stress (heat and drought) on *NAM* RNAi plants. The authors found that although *NAM* genes had no effect on tolerance to heat and drought stress in terms of yield, nutrient remobilisation was reduced in RNAi plants in all environments (Guttieri et al. 2013).

1.5.3.2 Function of homologues

The RNAi construct down-regulates the expression of all homologues of *NAM-B1*, which produces a strong phenotypic effect but does not allow the relative contributions of each homologue to be determined. Mutations in several of the *NAM-B1* homologues have been identified and characterised. A mutation in the *NAM-A1* gene in tetraploid wheat resulted in a significant delay in senescence, and this delay was enhanced by the addition of a mutation in the *NAM-B2* gene, although the mutant *NAM-B2* allele did not cause a statistically significant delay to senescence on its own (Distelfeld et al. 2012). This suggests that both *NAM-A1* and *NAM-B2* play a role in regulating senescence, similar to *NAM-B1*, although *NAM-B2* has a weaker effect than *NAM-A1*. A further study has shown that a mutation in *NAM-D1* also delays senescence, although not as strongly as a mutation in *NAM-A1*, and the double mutant (*nam-A1*/*nam-D1*) has an even larger delay to senescence (Avni et al. 2014). Mutations in *NAM-A1* and *NAM-D1* caused reductions of 1-7 % and 4-10 % in GPC and 8-25 % and 12-19 % in iron content across two years in multiple sites, although significant differences were not found in all

sites/years (Avni et al. 2014). The double mutant had a significant reduction in GPC (10-21 %), iron (29-32 %) and zinc (9 %) (Avni et al. 2014). These results show that *NAM-A1* and *NAM-D1* have a very similar role to *NAM-B1* in controlling nutrient remobilisation and senescence to affect GPC and grain micronutrient content. Due to the high similarity at the protein level (see below) we hypothesise that the other homologues (*NAM-A2*, *NAM-B2*, *NAM-D2*) would also regulate senescence and grain nutrient content, which is supported by the evidence that *NAM-B2* regulates senescence (Distelfeld et al. 2012).

Figure 1.3. *NAM* RNAi (left) and control (right) 50 days after anthesis (DAA), from (Uauy et al. 2006b).

1.5.3.3 NAC transcription factor characteristics

NAM-B1 is a member of the NAC transcription factor family and like many other NAC TFs consists of three exons, of which exons 1 and 2 encode the N-terminal NAC domain which is conserved between NAC TFs. The NAC domain consists of 5 subdomains, in *NAM-B1* these are 98 % to 100 % identical at the protein level to the homologues in wheat, barley, rice and maize (Uauy et al. 2006b). In other NAC TFs it has been shown that the NAC domain mediates dimerization of NAC TFs and contains the DNA binding domain (Duval et al. 2002, Jensen et al. 2010, Olsen A. N. et al.

2004). The transcriptional regulatory domain of NAC proteins is encoded by the third exon. In *NAM-B1* this C-terminal region is predicted to be disordered in structure and may mediate protein-protein interactions in addition to regulating transcription. It is likely that *NAM-B1* forms dimers and NAC TFs have been shown to act as transcriptional regulators in both homodimeric and heterodimeric forms (Jeong et al. 2009).

1.6 Thesis aims

This thesis aims to understand the mechanisms by which the transcription factor *NAM-B1* influences grain nutrient content, nutrient remobilisation and monocarpic senescence in hexaploid wheat (*Triticum aestivum*). To achieve this aim I will use a range of methods to characterise *NAM-B1*. I will explore the expression patterns of *NAM-B1* through the wheat plant in the context of nutrient transport (Chapter 2). I will investigate the role of *NAM-B1* in the regulation of photosynthesis and its effect on source-sink relations (Chapter 3). Finally I will identify candidate target genes of *NAM-B1* which will reveal the network of genes through which it regulates grain nutrient content, nutrient remobilisation and senescence (Chapter 4).

2 Chapter 2 – Expression patterns of *NAM* genes and their role in nutrient transport

2.1 Introduction

2.1.1 *NAC* transcription factors control diverse developmental processes

NAM-B1 is a member of the large NAC domain family which are plant- specific transcriptional regulators. In rice there are 151 non-redundant *NAC* genes and 117 in Arabidopsis (Nuruzzaman et al. 2010). In particular *NAC* transcription factors control diverse processes in plants including stress tolerance and development (Nuruzzaman et al. 2013, Puranik et al. 2012). NAC genes have been shown to regulate development of seeds (Sperotto et al. 2009), embryos (Kunieda et al. 2008), and shoot apical meristems (Duval et al. 2002), in addition to leaf senescence (Balazadeh et al. 2011, Guo and Gan 2006), nutrient remobilisation (Uauy et al. 2006b), xylem formation (Grant et al. 2010) and cell division (Kim Youn-Sung et al. 2006).

2.1.2 *NAM-B1* function is not well conserved with *NAC* genes in other species

In wheat itself there are several closely related homologous copies of *NAM-B1*. Homoeologous copies of *NAM-B1* are located on the A and D genomes of chromosome group 6 (*NAM-A1* and *NAM-D1*) and paralogous copies are located on chromosome group 2 (*NAM-A2*, *NAM-B2* and *NAM-D2*). Therefore in hexaploid wheat there are 6 copies of genes similar to *NAM-B1*. The role of NAM-B1 in nutrient remobilisation and senescence is conserved in *NAM-A1* and *NAM-D1* (Avni et al. 2014) and *NAM-B2* plays a role in senescence (Distelfeld et al. 2012). Therefore it is likely that the paralogous copies *NAM-A2* and *NAM-D2* also regulate nutrient remobilisation and senescence, but this remains to be tested.

Although the function of *NAM-B1* homologues in wheat is likely to be conserved, the conservation of function breaks down as more distant species are studied. Wheat and barley diverged approximately 12 million years ago (MYA) (Chalupska et al. 2008, Marcussen et al. 2014) but the function of *NAM-B1* in both senescence and nutrient remobilisation may have pre-dated the divergence. In barley an orthologous gene to *NAM-B1* (*HvNAM-1*) has been mapped to a 0.7 cM region of chromosome arm 6HS which encompasses the peak of a barley QTL for grain protein content (GPC)

(Distelfeld et al. 2008). Allelic variation of *HvNAM-1* is associated with higher GPC in wild and cultivated barley (Jamar et al. 2010) and NILs for the high GPC allele at this locus show accelerated senescence (Jukanti A. K. and Fischer 2008). Therefore it is likely that *NAM-B1* function is conserved in barley.

In many species NAC transcription factors which play a role in senescence have been identified. However the role of *NAM-B1* in regulating both senescence and nutrient remobilisation may be restricted to the Pooideae. In rice which diverged from wheat approximately 41-47 MYA (Paterson et al. 2004, Salse et al. 2008), the closest orthologue of both *NAM-B1* and *NAM-B2* is *Os07g37920* (**Figure 2.1**), which is colinear with *NAM-B2* (Distelfeld et al. 2012). However rather than affecting monocarpic senescence, *Os07g37920* plays a role in anther dehiscence: reduction of its expression by RNAi prevented pollen from being shed which may be due to a lack of senescence in the tapetum cells. This effect on anther dehiscence is not seen in wheat *nam-a1/nam-b1/nam-b2* triple mutants, therefore the rice and wheat orthologues have distinct biological functions (Distelfeld et al. 2012). In rice several other *NAC* genes are upregulated during or proposed to control senescence including *OsNAP* (*Os03g21060*) (Zhou et al. 2013) and *OsNAC5* (*Os11g018210*) (Sperotto et al. 2009). However these genes are less closely related to *NAM-B1* and although they play roles in senescence, they appear to be acting more downstream of *NAM-B1* or in a distinct parallel regulatory network (Ricachenevsky et al. 2013).

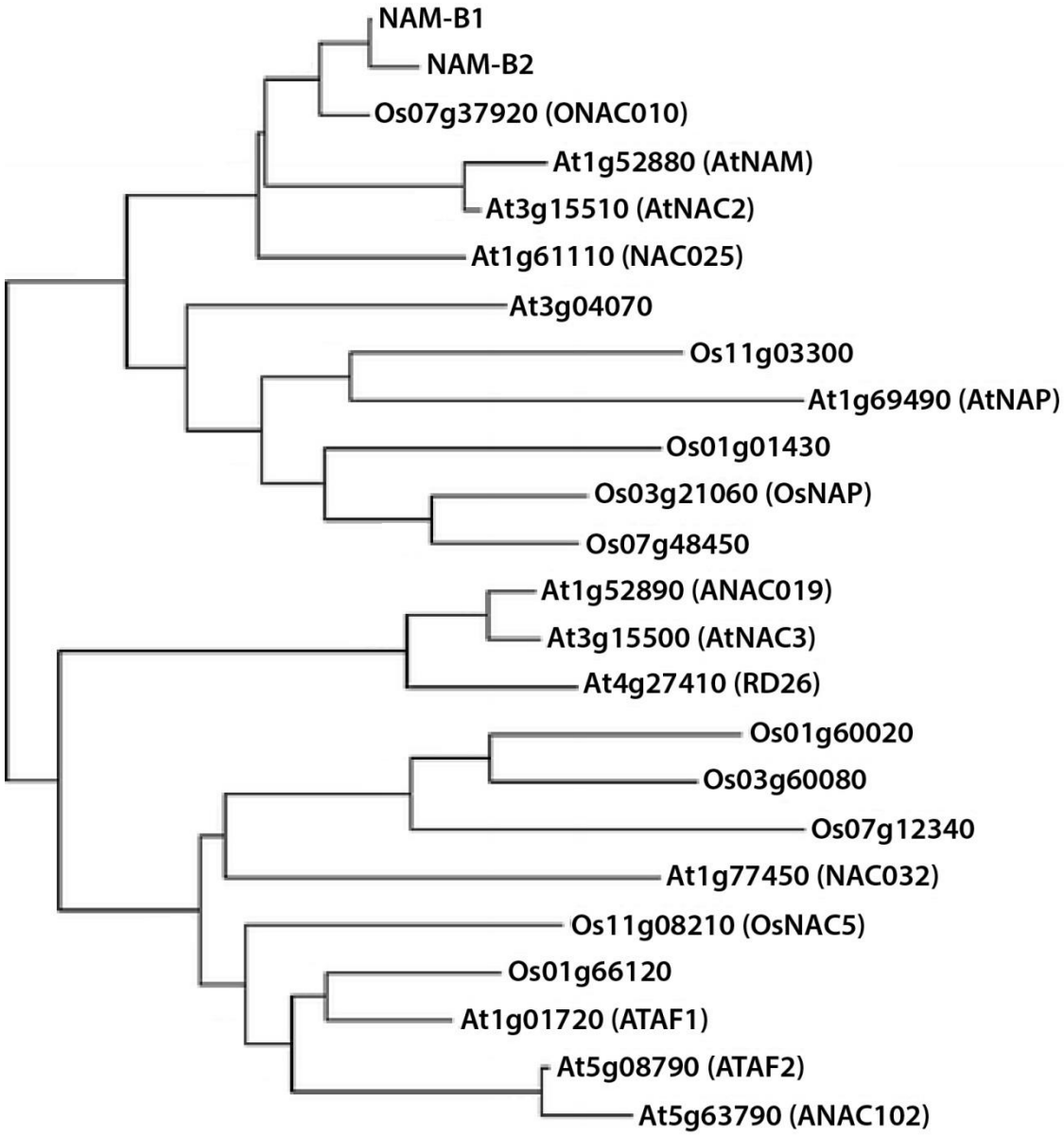


Figure 2.1. Phylogenetic analysis of the NAC proteins closest to NAM-B1 and NAM-B2 from Arabidopsis and rice, adapted from (Distelfeld et al. 2012). The phylogenetic tree was constructed using the amino acid sequences of the NAC domain for each protein sequence.

Since the rice orthologue of *NAM-B1* plays a different role it is unsurprising that the closest orthologues in Arabidopsis to *NAM-B1* have divergent functions. *AtNAM* (no apical meristem) controls shoot apical meristem development (Duval et al. 2002), *AtNAC2* is involved in salt stress responses and lateral root development (He et al. 2005) and *NAC025* affects seed development (Kunieda et al. 2008). Approximately 20 Arabidopsis *NAC* genes are upregulated during senescence (Balazadeh et al. 2008) but many of these are distantly related to *NAM-B1*, therefore their functions and the pathways they control are most likely quite different. For example *AtNAP* has been

reported to play a role in leaf senescence, with a T-DNA insertion mutant showing delayed leaf senescence (Guo and Gan 2006). However *AtNAP* was originally identified to be expressed in floral tissues and to have a role in stamen elongation and anther dehiscence (Sablowski and Meyerowitz 1998) which is similar to the rice orthologue of *NAM-B1 Os07g37920* (Distelfeld et al. 2012). There is directly conflicting evidence about role of *AtNAP* in senescence with transient over-expression reported to accelerate senescence (Guo and Gan 2006) and stable over-expression reported to have no effect on vegetative tissues (Sablowski and Meyerowitz 1998). Therefore even in the model system Arabidopsis much remains to be discovered about the role of NAC transcription factors in the regulation of senescence.

In light of differences between species, to understand the function of *NAM-B1* at the molecular level it is necessary to study this gene in wheat itself. A first step towards understanding the mechanisms by which *NAM-B1* controls nutrient remobilisation, transport and senescence is to characterise where and when it is acting, and any differences in expression between homologues.

2.1.3 *NAM* gene expression localisation in wheat is largely unknown

NAM gene expression was examined in flag leaves in tetraploid wheat (*T. turgidum* (*Tt*)) containing the functional *NAM-B1* gene (Uauy et al. 2006b). *NAM-A1*, *NAM-B1* and *NAM-B2* were found to be expressed after anthesis and expression rose throughout the time-period examined until 23 days after anthesis (DAA) (**Figure 2.2**). It was noted that transcripts were also present in green spikes and peduncles; however a full time-course comparing expression in several tissues was not reported. Expression of *NAM-B2* was also examined in stamens at booting and at flowering, however expression was low compared to flag leaves after anthesis (Distelfeld et al. 2012).

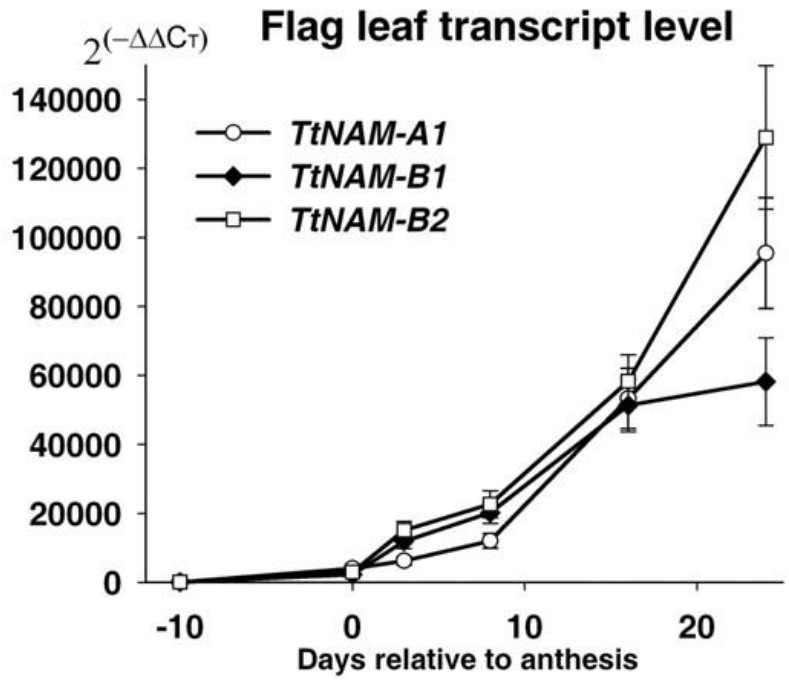


Figure 2.2. Expression of *NAM* genes in tetraploid wheat containing functional *NAM-B1*, from (Uauy et al. 2006b).

2.1.4 Nutrients move to the grain via both the xylem and phloem

NAM-B1 has a significant effect on micronutrient and protein content in the grain and its expression in flag leaves is indicative of the role it plays in nutrient remobilisation from this tissue. However, nutrient content in the grain is not solely determined by remobilisation. Although a large proportion of nutrients from grain filling are from remobilisation, for certain micronutrients uptake from the soil during grain filling is also important (Garnett and Graham 2005). Regardless of whether nutrients come from remobilisation or uptake they must be transported long distances to the grain via the xylem and the phloem. Ultimately all nutrients enter the grain through the phloem due to a block in the xylem between the rachilla and the grain known as the xylem discontinuity (Zee and O'Brien 1970). Multiple experiments have demonstrated that micronutrients are transported by different pathways, including both xylem and phloem transport. Transfer from the xylem to phloem occurs at different points along the pathway including in the roots, stems, peduncle, rachis and at the xylem discontinuity (Pearson et al. 1995).

2.1.5 Xylem and phloem mobility differs between nutrients

One technique that has been widely used to investigate whether nutrients travel via the phloem or the xylem is steam girdling. Heat is used the disrupt the phloem pathway at a

specific point in the plant, for example a fine jet of steam can be directed at a 0.5 cm band of the peduncle. The heat kills the phloem cells and disrupts their transport capabilities, whereas the already dead xylem cells continue to transport water and solutes.

Steam girdling of the phloem has been used to investigate the relative importance of phloem and xylem transport to the grain. All nutrients regardless of whether they originate from the flag leaf or the roots must pass through the peduncle. In this tissue it has been shown that certain micronutrients require phloem transport whereas others do not. For example steam girdling of the peduncle did not alter P and Mg movement to the ear but slightly reduced Ca and nearly stopped all K movement (Martin 1982), indicating that P and Mg can be transported in the xylem through the peduncle, whereas some Ca and nearly all K are transported in the phloem through the peduncle. A separate steam girdling experiment on the peduncle demonstrated that Mn is mainly transported in the xylem (Pearson et al. 1995), and transfer of Zn from xylem to phloem occurs progressively in the peduncle and rachis whereas Mn transfer is hypothesised to occur at the xylem discontinuity.

Other micronutrients have not been studied using steam girdled plants but there is other evidence about phloem and xylem mobility of these elements. However, it is important to note that for many elements the route by which they are transported may be environmentally dependent because the source of the nutrient (e.g. remobilisation or uptake during grain filling) will depend on growing conditions. For example Fe remobilisation is variable, with reports ranging from all grain Fe being supplied by remobilisation in soil grown plants (Garnett and Graham 2005) to no Fe remobilisation being detected in irrigated field-grown spring wheat (Hocking 1994). This difference in the source of Fe for grain filling suggests differences in the transport route used by this element to the grain. Fe can be transported in the xylem in a complex with citrate (Rellán-Álvarez et al. 2010) or transported in the phloem with transfer from xylem to phloem occurring in the root, basal part of the shoot or during remobilisation from the leaves (Borrill et al. 2014). Therefore it is not fully determined which route Fe takes to the grain. Cu has been described as variably phloem mobile (Loneragan 1981) and may be transported in both the xylem and the phloem depending upon conditions. Furthermore Cu is transported to the grain both from post anthesis uptake from the soil and from remobilisation from vegetative tissues (Garnett and Graham 2005) perhaps indicating both xylem and phloem transport.

2.1.6 Micronutrient differences observed in *NAM* RNA interference (RNAi) plants

As discussed above, micronutrients can be tentatively categorised as phloem transported (Ca and K), both xylem and phloem mobile (Cu, Fe and Zn) or xylem transported (Mg, Mn and P). The effect of *NAM-B1* homologues on micronutrient distribution had previously been studied (Waters et al. 2009). Strong differences in grain micronutrient content were found between hexaploid wheat plants in which the expression of *NAM-B1* homoeologues and paralogues was reduced by 50 % by RNA interference (*NAM* RNAi plants) and control plants. We examined this data and observed that in general xylem transported nutrients were reduced in the grain of RNAi lines (e.g. Mg, Mn and P), whereas phloem transported nutrients (e.g. Ca and K) were increased compared to control plants. The most strongly affected were Cu, Fe and Zn which were ~30% lower in RNAi lines. Mg, Mn and P were reduced in RNAi lines to a lesser extent. Contrastingly Ca was slightly higher and K was ~35% higher in RNAi lines.

2.1.7 Aims

We aim to understand more about the molecular function of *NAM-B1*. A key question to address is in which tissues and cell types *NAM-B1* and its homologues (*NAM* genes) are expressed. To answer this we will examine expression in vegetative and reproductive tissues of hexaploid wheat to give a more comprehensive overview of *NAM* gene expression, because previously expression has only been studied in flag leaves in tetraploid wheat. We will characterise the expression patterns of the homoeologous and paralogous copies of *NAM-B1* which might indicate functional specialisation. We will further characterise expression at the tissue level by using RNA *in situ* hybridisation to identify whether *NAM* genes are expressed in specific cell types, which might give clues about which developmental and metabolic processes they control.

A second question we would like to address is the mechanism by which *NAM-B1* affects nutrient transport. As described above, it has been reported that down-regulation of *NAM-B1* homologues alters the grain micronutrient content. We hypothesise that these changes to micronutrient content in RNAi grain are caused by alterations to xylem and phloem transport. To test this hypothesis we will investigate whether *NAM* RNAi plants show alterations in xylem and phloem movement using a tracer study.

To answer these questions we will test the following hypotheses:

2.1.7.1 Hypothesis 1: Patterns of *NAM* gene expression will differ between organs

NAM-B1 affects both nutrient remobilisation and senescence leading to changes in grain nutrient content. We hypothesise that *NAM-B1* and its homologues may play different roles in different organs (e.g. grain and leaves), or homologues may be functionally specialised, therefore we hypothesise they will have a different expression profiles.

2.1.7.2 Hypothesis 2: *NAM* genes will be expressed in specific cell types particularly those related to transport

NAM-B1 affects nutrient remobilisation this may be via an effect on transport. We hypothesise that expression of *NAM* genes will be localised to specific cell types for example the xylem.

2.1.7.3 Hypothesis 3: Xylem and phloem transport to the ear will be affected by reduction of *NAM* gene expression

It has previously been shown that micronutrient content in the grain of *NAM* RNAi plants is altered with xylem-mobile elements generally reduced and phloem-mobile elements unaffected or increased. We hypothesise that these grain micronutrient content differences are due to differences in xylem and phloem transport.

2.2 Methods

2.2.1 Quantitative PCR (qPCR)

2.2.1.1 Plant material for qPCR

Hexaploid wheat plants of the variety UC1041+ GPC (UC1041 with an introgression containing *NAM-B1*(Uauy et al. 2006a)) were germinated in the dark at 4 °C on moist filter paper in petri dishes for 48 hours and incubated for 48 hours in the light at room temperature. They were sown into P40 trays containing peat and sand (85 % Fine Peat, 15% Grit, 2.7 kg m^{-3} Osmocote 3-4 months, Wetting Agent,4 kg m^{-3} Maglime, 1 kg PG Mix) before potting on at 2-3 leaf stage to FP9 pots in cereal mix (40% Medium Grade Peat, 40% Sterilised Soil, 20% Horticultural Grit, 1.3 kg m^{-3} PG Mix 14-16-18 + Te Base Fertiliser, 1 kg m^{-3} Osmocote Mini 16-8-11 2 mg + Te 0.02% B, Wetting Agent, 3 kg m^{-3} Maglime, 300 g m^{-3} Exemptor). All plants were grown in a heated glasshouse ~20 °C with 16 h supplementary light. Samples were harvested between 11 am and noon at anthesis, 7, 20 and 30 days after anthesis (DAA) into liquid nitrogen and stored at -80 °C. At the time of harvest plants were divided into the following tissues: exposed peduncle (top 2 cm of peduncle), enclosed peduncle (2 cm of peduncle just above flag leaf node with flag leaf sheath removed), internode 1 (2 cm of stem just below flag leaf node), flag leaf blade (2-3 cm half way along the blade), flag leaf sheath (2-3 cm half way down the sheath), florets (the palea and lemma of the 1^{st} and 2^{nd} florets from the central 4 spikelets of the spike lengthways) and rachis. A second batch of plants were grown in the same conditions and harvested at the same time-points, but in this case only flag leaf blade and grain (from the central 4 spikelets, selecting only the 1^{st} and 2^{nd} florets) were harvested.

2.2.1.2 RNA extraction from vegetative tissues

Samples from vegetative tissues (exposed peduncle, enclosed peduncle, internode 1, flag leaf blade, flag leaf sheath, florets and rachis) were ground in a liquid-nitrogen cooled mortar and pestle or in an Eppendorf with a micro-pestle depending on the tissue toughness. Approximately 100 mg of ground tissue was added to 1 ml of Tri Reagent (T9424, Sigma) in a fresh Eppendorf and incubated for 5 min at room temperature. 200 µl of chloroform was added and samples were shaken vigorously and incubated for 2 min at room temperature. Samples were centrifuged for 5 min at 10,000 x g and the top aqueous layer (~600 µl) was transferred into a fresh tube containing 500 µl 100% isopropanol. After 3 min incubation at room temperature samples were centrifuged for

10 min at 13,000 g. The supernatant was discarded and the pellet washed in 300 µl 70% ethanol. Samples were centrifuged for 5 min at 10,000 x g and the supernatant was discarded. The pellet was left to dry for 3 min at room temperature and then resuspended in 50 µl nuclease free water (P1193, Promega). The RNA concentration was measured using a NanoDrop 2000 (ThermoScientific) and samples were frozen at -80 °C.

2.2.1.3 RNA extraction from grain

Grain RNA was extracted using a phenol/chloroform method due to the high starch content in the grain which interferes with Tri Reagent RNA extraction. 5-10 grains were ground in liquid nitrogen in a mortar and pestle and ~200 mg were added to 500 µl RNA extraction buffer (0.1 M Tris pH 8.0, 5 mM EDTA pH 8.0, 0.1 M NaCl, 0.5% SDS, 1% 2-meracaptoethanol). 200 µl of Plant RNA Isolation Aid (AM9690, Ambion) was added and the sample was mixed by shaking. After centrifugation for 10 min at 10,000 x g the supernatant was added to 300 µl of 1:1 acidic Phenol (pH4.3) (P4682, Sigma): chloroform. The sample was shaken intermittently for 5 min and then centrifuged for 15 min at 10,000 x g. The supernatant was added to 240 µl isopropanol and 30 µl sodium acetate (3 M, pH 5.2) was added. RNA was precipitated for 15 min at -80 °C. After centrifugation for 30 min at 10,000 x g the supernatant was discarded and the pellet was washed twice with 750 µl 70% ethanol. The supernatant was discarded and the pellet was dried for 3 min at room temperature. The pellet was resuspended in 100 µl nuclease free water (P1193, Promega). The RNA concentration was measured using a NanoDrop 2000 (ThermoScientific) and samples frozen at -80 °C.

2.2.1.4 DNAse treatment

RNA samples were diluted using nuclease free water to 250 ng/µl and treated with RQ1 DNAse (M6101, Promega).

DNase treatment:

- RNA 7 µL
- RQ1 RNase-Free DNase 10x Reaction Buffer 1 µL
- RQ1 RNase-Free DNase 2 µL

1. Incubate samples for 30 min at 37 °C.

2. Add 1 µL of RQ1 DNase Stop Solution to each sample to terminate the reaction.

3. Incubate samples for 10 min at 65 °C to inactivate DNase.

2.2.1.5 cDNA synthesis

Reagents:

- M-MLV Reverse Transcriptase (28025-013, Invitrogen)
- Oligo(dT) Primers (C1101, Promega)
- Random Primers (48190-011, Invitrogen)
- RNasin® Plus RNase Inhibitor (N2611, Promega)
- dNTP mix (U1511, Promega)

Annealing:

1. Put 11 µl DNAse treated RNA into PCR tube
2. Add 1 µl 10 mM dNTPs
3. Add primers (1 µl of 1 in 2 dilution of oligo(dT) and 1 in 20 dilution of random primers)
4. Heat to 65 °C for 5 min then chill on ice

cDNA synthesis:

Add the following to the annealed RNA from previous step:

- 4 µl 5x first strand buffer (Invitrogen)
- 2 µl 0.1 M DTT (Invitrogen)
- 1 µl RNasin (N2611, Promega)
- 1 µl M-MLV RT (28025-013, Invitrogen)

5. Incubate 25 °C for 10 min
6. Incubate 37 °C for 50 min
7. Inactivate reaction at 70 °C for 15 min
8. Store at -20 °C

2.2.1.6 qPCR

qPCR was carried out in 384 well plates (04729749001, Roche) on the LightCycler® 480 (Roche). Reaction components are shown in **Table 4.12** and cycling conditions are shown in **Table 4.13**. The primers used are listed in **Table 2.3**.

Reagent	Volume (µl)
LightCycler® 480 SYBR Green I Master 2x (04707516001, Roche)	5
cDNA (1 in 10 dilution)	2
F primer (2 µM)	1.25
R primer (2 µM)	1.25
H_2O	0.5
TOTAL	**10**

Table 2.1. Standard qPCR composition.

Step	Temperature (°C)	Time (min)
1	95	5:00
2	95	0:10
3	60	0:15
4	72	0:20
5	78	0:01
6	Repeat steps 2-5, 45 times	
7	Cool to 60	
8	Melt curve to 95	5 acquisitions per second

Table 2.2. Standard qPCR cycling conditions.

Primer ID	Gene	5' - 3' sequence (top row = forward primer, bottow row = reverse primer)	Reference
PB57_NAM_A1_F1	*NAM-A1*	GGAGAGGATTCATAAATGCGGA	(Uauy et al.
PB37_NAM_R2		GGCTCCGACCAAACAGTTTC	2006b)
PB36_NAM_B1_F1	*NAM-B1*	TTGTCCACTGCGCCAGC	(Uauy et al.
PB37_NAM_R2		GGCTCCGACCAAACAGTTTC	2006b)
PB38_NAM_D1_F	*NAM-D1*	GCCTGCGGTGGCCTC	(Uauy et al.
PB39_NAM_D1_R		CATGTCTGCATTTATGAATCCTCTTC	2006b)
PB40_NAM_B2_F1	*NAM-B2*	AAAGCGGGTATGCGGTGAC	(Uauy et al.
PB41_NAM_B2_R1		CCTCTCTCAGGTTGGACGAC	2006b)
PB42_NAM_D2_F	*NAM-D2*	GGACAGCCACGAGGACAACT	(Uauy et al.
PB43_NAM_D2_R		GGGCCAGGAACTGTTTGGT	2006b)
PB47_Actin_F	*ACTIN*	ACCTTCAGTTGCCCAGCAAT	(Uauy et al.
PB48_Actin_R		CAGAGTCGAGCACAATACCAGTTG	2006b)
PB67_Ubiquitin_F	*UBIQ*	CCTTCACTTGGTTCTCCGTCT	From Ruth
PB68_Ubiquitin_R		AACGACCAGGACGACAGACACA	Bryant, JIC
PB71_TaGAPDH_F	*GAPDH*	TTAGACTTGCGAAGCCAGCA	From Peter
PB72_TaGAPDH_R		AAATGCCCTTGAGGTTTCCC	Isaac, iDNA
PB69_TaEF1a_F	*EF1A*	TGGTGTCATCAAGCCTGGTATGGT	From Ruth
PB70_TaEF1a_R		ACTCATGGTGCATCTCAACGGACT	Bryant, JIC

Table 2.3. Primers used for homoeologue specific qPCR.

2.2.1.7 qPCR data analysis

All melt curves were manually inspected for a single peak at the expected temperature. Crossing point values were called using the Fit Points method on the Lightcycler 480 software. geNorm (Vandesompele et al. 2002) was used to check the stability of reference genes between tissues and time-points and calculate a normalisation factor using all 4 reference genes (*actin*, *ubiquitin*, *GAPDH* and *EF1a*, **Table 2.3**) for each sample. This normalisation factor was used to normalise the *NAM* gene expression levels. All samples were divided by the expression of *NAM-B1* in the flag leaf blade at 0 DAA to enable comparisons between tissues and homologues.

Statistical analysis was carried out in Genstat (v15.2.0.8821) using an unbalanced ANOVA with multiple comparison testing using Fisher's least significant difference test.

2.2.2 RNA *in situ* hybridisation

2.2.2.1 Plant material for RNA *in situ* hybridisation

Hexaploid wheat plants of the variety UC1041+ GPC (Uauy et al. 2006a) were germinated in the dark at 4 °C on moist filter paper in petri dishes for 48 hours and incubated for 48 hours in the light at room temperature. They were sown into P40 trays containing peat and sand. All plants were grown in a heated glasshouse ~20 °C with 16 h supplementary light.

2.2.2.2 Sample harvesting and fixing

Whole tillers were harvested between 2 pm and 3 pm at anthesis and 20 DAA into aluminium foil containing damp tissue to retain moisture until processing in the lab (which was completed immediately following harvest). Plants were divided into the following tissues: exposed peduncle (top 2 cm of peduncle), enclosed peduncle (2 cm of peduncle just above flag leaf node with the flag leaf sheath removed), internode 1 (2 cm of stem just below flag leaf node), tip flag leaf blade (0.5-2.5 cm from the tip of the blade), mid flag leaf blade (2-3 cm half way along the blade), base flag leaf blade (2 cm nearest the leaf sheath). Each tissue was cut with a scalpel into 1-8 mm pieces to create transverse sections and placed immediately into ice cold formaldehyde acetic acid solution (3.7% formaldehyde, 5% acetic acid, 50% ethanol, 0.1% triton-X-100, 0.1% tween-20). Samples were vacuum-infiltrated for 5 min, the vacuum was released and re-applied for a further 5 min. Samples were left in fixative overnight at 4 °C on rollers. Samples were then incubated in 30%, 50% and 70% ethanol for 3 hours each at 4 °C on rollers before storage at 4 °C in fresh 70% ethanol for up to 2 months.

2.2.2.3 Sample wax embedding and sectioning

Samples were transferred into Histosette I tissue processing/embedding cassettes (M490-5, Simport) and run through an ethanol-xylene-wax series (**Table 2.4**) using a Tissue-Tek VIP (Sakura). Samples were then transferred from cassettes to individual wax blocks using the Tissue-Tek TEC (Sakura). This kept the wax melted (60 °C) in the cassettes. Samples were positioned in individual blocks in metal base moulds (Leica) and rapidly cooled on a cold plate. Embedded material was stored at 4 °C. Sections were cut from individual samples to a thickness of 8 µm using a Biocut microtome 2030 (Leica Reichert-Jung) and transferred to polylysine coated slides (631-0107, VWR International GmbH) with a drop of sterile water between the slide and the sample

section. Slides were dried at 40 °C for 24 hours on a slide drying bench (Barnstead electrothermal/Thermo Scientific) and stored at 4 °C.

Step	Solvent	Percentage (%)	Time (hours)	Temperature (°C)
1	Ethanol	70	4	35
2	Ethanol	80	4	35
3	Ethanol	90	4	35
4	Ethanol	100	4	35
5	Ethanol	100	4	35
6	Ethanol	100	4	35
7	Xylene	100	4	35
8	Xylene	100	4	35
9	Xylene	100	4	35
10	Paraffin Wax		4	60
11	Paraffin Wax		4	60
12	Paraffin Wax		4	60
13	Paraffin Wax		4	60

Table 2.4. Programme for wax embedding using the Tissue-Tek VIP.

2.2.2.4 *In situ* probe design and cloning

The NAM probe (**Figure 2.3**) was designed in the third exon of *NAM-B1* (**Figure 2.4**) in order to avoid hybridisation with other NAC transcription factors which shared the conserved NAC domain located in exons 1 and 2. The probe was not specific to *NAM-B1* because the specificity depends on the stringency of the washes used after the hybridisaiton step. Under the conditions we used, any sequence with over 80 % similarity could be bound, which was calculated as follows (McClean 1998):

Percentage similarity of sequences bound by probe = 100 - [(T_m-wash temperature)/1.4]

Where T_m = 81.5 + 16.6(log M [Na+]) + 0.41(% GC content of probe)

We carried out BLAST against the International Wheat Genome Sequencing Consortium chromosome arm survey sequences (IWGSC 2014) to determine which sequences could also be bound by the NAM probe. We identified that the homoeologous and paralogous copies of *NAM-B1* in **Table 2.5** would also be bound by the probe. The homologous sequences are shown as a sequence alignment in Appendix

6.1.1. Therefore the *in situ* hybridisations reflect the expression of *NAM-A1*, *NAM-B1*, *NAM-D1*, *NAM-A2*, *NAM-B2* and *NAM-D2*.

```
>NAM-probe
CGGCGATCAGCAGAGGAACACGGAGTGCGAGGACTCCGTGGAGGACGCGGTCACCGCGT
ACCCGCTCTATGCCACGGCGGGCATGACCGGTGCAGGTGCGCATGGCAGCAACTACGCT
TCACCTTCACTGCTCCATCATCAGGACAGCCATTTCCTGGACGGCCTGTTCACAGCAGA
CGACGCCGGCCTCTCGGCGGGCGCCACCTCGCTGAGCCACCTAGCAGCGGCGGCGAGGG
CAAGCCCGGCTCCGACCAAACAGTTTCTCGCCCCGTCGTCTTCGACCCCGTTCAACTGG
CTCGATGCGTCACCAGTCGGCATCCTCCCACAGGCAAGGAATTTTCCTGGGTTTAACAG
GAGCAGAAACGTCGGCAATATGTCGCTGTCATCGACGGCCGACATGGCTGGCGCAGTGG
ACAACGGTGGAGGCAATGCGGTGAACGCCATGTCTACCTATCTTCCCGTGCAAGA
```

Figure 2.3. *NAM* probe sequence.

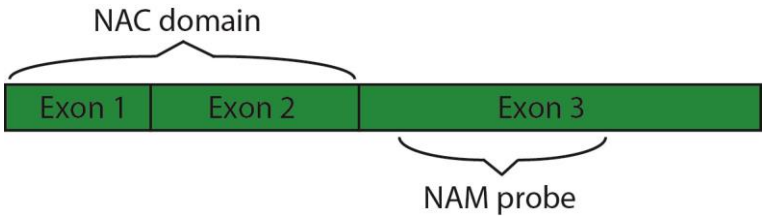


Figure 2.4. Diagram of NAM probe location within the *NAM-B1* coding sequence.

Gene	**IWGSC Contig***	**Length aligned**	**Percentage identity (%)**	**e-value**
NAM-A1	IWGSC_CSS_6AS_scaff_4397602	434	94	0
NAM-B1	N/A[+]	468	100	N/A
NAM-D1	IWGSC_CSS_6DS_scaff_1542870	464	91	0
NAM-A2	IWGSC_CSS_2AS_scaff_5260301	469	86	$1e^{-162}$
NAM-B2	IWGSC_CSS_2BS_scaff_5218890	473	87	$4e^{-168}$
NAM-D2	IWGSC_CSS_2DS_scaff_5388904	473	87	$6e^{-172}$

* International Wheat Genome Sequencing Consortium (IWGSC) contigs available at plants.ensembl.org

[+] *NAM-B1* is not present in the variety Chinese-Spring which has been sequenced.

Table 2.5. BLAST hits to the NAM-probe with an alignment length over 50 bp.

PCR reactions (**Table 4.1**) were carried out using the conditions in **Table 4.2** to amplify the probe sequence (**Figure 2.3**) from cDNA using primers PB91 and PB93 in **Table 4.3**.

Reagent	Volume (µl)
Forward primer (2 µM)	1.5
Reverse primer (2 µM)	1.5
buffer (5x green) (Promega M791B)	3
dNTP (2 mM)	1.5
H_2O	4.5
Homemade taq	0.07
cDNA template (1:10 dilution) from **2.2.1.5 cDNA synthesis,** mixture of tissues 20 DAA	1.5

Table 2.6. PCR reaction for *in situ* probe amplification from cDNA

Step	Temperature	Time (min)
1	95 °C	5:00
2	95 °C	0:30
3	55 °C	0:20
4	72 °C	1:15
5	go to step 2, 39 times	
6	72 °C	10:00
7	10 °C	∞

Table 2.7. PCR cycling conditions for *in situ* probe amplification from cDNA.

Primer	Sequence
PB91_NAM-B1_F1_in_situ	CGGCGATCAGCAGAGGAA
PB93_NAM-B1_R1_in_situ	TCTTGCACGGGAAGATAGGT
M13F	GTAAAACGACGGCCAGT
M13R	CAGGAAACAGCTATGAC

Table 2.8. Primers used to amplify and sequence *in situ* probes

An "A" overhang was added to the purified PCR product by combining:

- 6.5 µl PCR product
- 2 µl PCR 5x PCR buffer (Promega)
- 1 µl 2 mM ATP
- 0.5 µl Homemade taq

The mixture was incubated at 72 °C for 20 min.

The A-tailed PCR product was cloned into pCR™4-TOPO® TA Vector (45-0071, Invitrogen) by incubating the following reagents at room temperature for 10 min:

- 4 µl PCR product with A overhang
- 1 µl salt solution

- 1 µl TOPO vector

The ligation product was subsequently transformed into Mach1™ T1R competent cells (C862003, Life technologies):

1. Thaw 20 µl aliquot of Mach1 cells on ice
2. Add 2 µl TOPO cloning ligation product
3. Incubate on ice for 5-10 min
4. Heat shock at 42 °C for 45 secs in a water bath
5. Put cells on ice
6. Add 250 µl pre-warmed SOC medium
7. Shake at 30 °C for 1 hour
8. Spread 50 µl and 150 µl onto pre-warmed LB plates containing 50 µg/ml of ampicillin
9. Incubate overnight at 37 °C

Colonies were screened by PCR using the primers PB91 with either M13F or M13R (**Table 4.3**) using the reagents listed in **Table 4.1**, except the DNA template used was 1.5 µl of water containing some cells from an individual colony (a pipette tip was used to transfer cells from the plate into the water), the cycling conditions were used as in **Table 4.2**. 3 colonies for each orientation of the insert were selected, grown overnight at 37 °C in LB containing 50 µg/ml kanamycin. Plasmid DNA was prepared using QIAprep Spin Miniprep Kit (27104, Qiagen) and glycerol stocks (500 µl overnight culture with 500 µl 40% glycerol) were stored at -80 °C.

Three plasmids containing the *NAM* probe in each orientation were sequenced. BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used to prepare DNA samples for Sanger sequencing:

1. Measure DNA concentration of the samples to be sequenced.
2. Prepare the following reaction:

- 1 µL Template DNA (~150 ng/µl)
- 1.5µL 5x BigDye sequencing buffer
- 1.75µL Primer (2 µM)
- 1µL BigDye reagent
- 4.75µL ddH_2O

3. Carry out PCR as per **Table 2.9**.

Step	Temperature	Time (min)
1	96 °C	2:00
2	96 °C	0:10
3	55 °C	0:30
4	60 °C	4:00
5	go to step 2, 24 times	
6	15 °C	∞

Table 2.9. Big dye sequencing PCR cycling conditions.

4. Submit samples to TGAC/Eurofins for sequencing.
5. Check template sequence against expected probe to ensure no errors have been introduced during PCR and cloning.

2.2.2.5 *In situ* probe template purification

4x 50 µl PCR reactions were carried per probe using the reaction components in **Table 2.10** under the cycling conditions in **Table 2.11**.

Reagent	Volume (µl)
Forward primer: M13 F (2 µM)	5
Reverse primers: PB91 (for antisense probe), PB93 (for sense probe) (2 µM)	5
buffer (5x green) (Promega M791B)	10
dNTP (2 mM)	5
H_2O	23.5
Homemade taq	0.07
plasmid DNA (NAM_A for antisense probe, NAM_D for sense probe)	1.5

Table 2.10. PCR reaction mix for *in situ* probe template amplification.

Step	Temperature	Time (min)
1	95 °C	5:00
2	95 °C	0:30
3	55 °C	0:20
4	72 °C	1:00
5	go to step 2, 29 times	
6	72 °C	10:00
7	10 °C	∞

Table 2.11. PCR cycling conditions for *in situ* probe template amplification.

PCR products were cut from a 1% agarose gel and purified using QIAquick Gel Extraction Kit (28704, Qiagen). The purified PCR products for each probe combined (total volume 400 µl) and phenol-chloroform extracted:

1. Add 400 µl phenol-chloroform-isoamylalcohol (25:24:1) pH8 (P2069, Sigma)
2. Centrifuge 5 min 10,000 x g
3. Add 400 µl chloroform to the supernatant, vortex
4. Centrifuge 5 min 10,000 x g
5. Precipitate the DNA from the supernatant by adding:
    - 40 µl 3 M sodium acetate pH 5.2
    - 1 ml 100% ethanol
6. Incubate overnight -20 °C or 10 min on dry ice
7. Centrifuge 10 min 10,000 x g
8. Discard supernatant and wash pellet in 1 ml 70% ethanol
9. Centrifuge 5 min 10,000 x g
10. Discard supernatant and dry the pellet (5 min room temperature)
11. Resuspend in 50 µl nuclease free water
12. Check 1 µl on a gel

2.2.2.6 *In situ* probe labelling

1. Mix in the following order:

- 13.5 µl dH_2O
- 2.5 µl 10x T7 transcription buffer
- 1 µl RNAse inhibitor (RNAsin N211-A Promega)
- 2.5 µl 5 mM ATP, GTP, CTP (11277057001, ribonucleoside triphosphate set, Roche)
- 2.5 µl 1 mM DIG-UTP (11093088910, Roche)
- 2 µl purified probe template
- 1 µl T7 polymerase (10881767001, Roche)

2. Incubate 2 h at 37 °C.
3. To stop reaction add:
    - 75 µl 1x MS (autoclaved)
    - 2 µl tRNA (100 mg/ml) (type XXI R-4251 Sigma)
    - 1 µl DNase (27-0514-01, Amersham)

4. Incubate for 30 min at 37 °C.
5. Precipitate with 100µl 3.8 M ammonium acetate (VWR) (autoclaved) and 600µl ice cold ethanol
6. Incubate -20 °C overnight
7. Centrifuge 20 min at 4 °C (13000-15000 pm).
8. Wash pellet with 100-200 µl ice cold 70% ethanol.
9. Centrifuge 20 min at 4 °C (13000-15000 pm), remove supernatant, air dry.
10. Resuspend in 50 µl dH_2O (sterile).
11. Add 50 µl 200 mM carbonate buffer pH 10.2.
12. Hydrolise at 60 °C for 70 min (to break the 468 bp probe down into 150-200 bp fragments to penetrate the tissue)
13. Precipitate by adding:
    - 10 µl 10% acetic acid (keep frozen)
    - 12 µl 3 M sodium acetate pH 4.8
    - 312 µl ethanol (ice cold)
14. -20 °C for 2 hours to overnight
15. Centrifuge 20 min at 4 °C (13000-15000 pm).
16. Wash pellet with 100-200µl ice cold 70% ethanol.
17. Centrifuge 20 min at 4 °C (13000-15000 pm), remove supernatant, air dry.
18. Resuspend in 50µl H_2O.

2.2.2.7 *In situ* slide clearing and preparation

All the solutions in (**Table 2.12**) were prepared using autoclaved distilled water and RNAse free reagents. Slide racks were autoclaved and measuring cylinders were soaked overnight in sodium hydroxide (4 pellets per 1 L cyclinder) to remove RNAses before use. Small quantities of liquids were measured using RNAse free 50 ml plastic centrifuge tubes.

Slides were put into slide racks and passed through the series of solutions in **Table 2.12**.

Solution	Time (min)	Notes	Function
Histoclear (12358637, Fisher)	10	Use glass box	Removes wax
Histoclear	10	Use glass box	Removes wax
100% ethanol	1	Use glass box	
100% ethanol	1		
95% ethanol	1		
85% ethanol, 0.85% saline	1		
50% ethanol, 0.85% saline	1		
30% ethanol, 0.85% saline	1		
0.85% saline	2		
PBS*	2		
0.125 mg/ml pronase (P6911-1G, Sigma) in buffer (5 mM Tris-HCl pH7.5, 5 mM EDTA)	14		Digests cell walls
0.2% glycine in PBS*	3		Inactivates pronase
PBS*	2		
4% paraformaldehyde (100503-916, VWR) in PBS*	10	Use glass box	Fixes and crosslinks
PBS*	2		
PBS*	2		
0.5% acetic anhydride (320102, Sigma) in 0.1 M triethanolamine pH8 (VWR)	10	Add acetic anhydride immediately before use, stir whilst incubating	Creates ionic charge to help probe bind
PBS*	2		
0.85% saline	2		
30% ethanol, 0.85% saline	1		
50% ethanol, 0.85% saline	1		
85% ethanol, 0.85% saline	1		
95% ethanol	1		
100% ethanol	1		
100% ethanol	1 – 120	Put at 4 °C if leaving for more than 10 min	

* PBS is phosphate buffered saline made from 10x stock pH 7. Final concentrations in solutions used are 130 mM NaCl, 7 mM Na_2HPO_4 and 3 mM NaH_2PO_4 .

Table 2.12. Series of solutions to prepare slides for hybridisation.

2.2.2.8 Probe hybridisation

Probe hybridisation was carried out, adapted from Coen lab (JIC) *in situ* protocol.

1. Prepare hybridization buffer (100 µl per slide, always make a bit extra), defrost all reagents at 37 °C, see **Table 2.13**.

Hybridization buffer	**24 slides**	**48 slides**
10x salts (3 M NaCl, 0.1 M Tris-HCl pH 6.8, 0.1 M $NaPO_4$, 50 mM EDTA)	375 µl	687.5 µl
Hi-Di™ Formamide (4311320, Applied Biosystems)	1500 µl	2750 µl
tRNA 100 mg/ml (Type XXI, R4251, Sigma)	37.5 µl	68.75 µl
50x Denhardt's (30915, Sigma)	75 µl	137.5 µl
H_2O	262.5 µl	481.25 µl
50% dextran sulphate (D8906 Sigma)	750 µl	1375 µl
Final volume	**3000 µl**	**5500 µl**

Table 2.13. Hybridisation buffer composition.

2. Vortex the hybridization buffer, spin it down and leave it at room temperature.
3. Mix (for each slide) and keep on ice:
   - 2 µl probe
   - 2 µl H_2O
   - 4 µl Formamide (fridge)

   Or use 4 µl probe and no water.
4. Heat the probe/H_2O/formamide 2 min at 80 °C and cool on ice.
5. Mix the probe with the hybridization buffer (100 µl/slide), vortex, spin down and leave at room temperature.
6. Take the slides out of the rack, allow ethanol to evaporate completely.
7. Use 100 µl of hybridization buffer/probe per slide, cover with a plastic coverslip, (Hybri-slips H0 784-1006A, Sigma), avoid bubbles.
8. Prepare 2x SSC (stock 20X SSC: 3 M NaCl in 0.3 M sodium citrate (pH7.0)), 50% formamide (F9037, Sigma) (50 ml/box=12 slides).
9. Place blue towel at the bottom of each box, soak 2x SSC buffer, cover the towel with parafilm, put the slides on top and seal it with masking tape so that slides cannot dry out.
10. Hybridize overnight floating in a water bath at 50 °C.

2.2.2.9 Washes to remove unhybridised probe

1. Set a water bath at 37 °C.
2. Add 2x SSC to a square petri dish. Soak slides in this solution to pop cover slips off naturally. Do not force this process or tissue will be lost.

3. Add 50 ml of 0.2x SSC to a plastic containing and place the rack so that you keep slides wet after removing the cover slips.
4. When rack is full place in 55 °C water bath for 35 min.
5. Wash twice more in 0.2x SSC for 25 min each at 55 °C.
6. During the washes prepare NTE buffer (0.5 M NaCl, 10 mM Tris-HCl pH7.5, 1 mM EDTA pH8.0)
7. Wash in NTE twice 5 min at 37 °C. (preheat 15 min before).
8. Incubate in NTE containing 20 µg/ml RNase A (R4897, Sigma) for 30 min at 37 °C. (300 mL + 600 µl RNase)

   *****Pre-incubate the NTE-RNase at for 5 min at 37 °C to activate it.
9. Wash in NTE twice for 5 min at room temperature.

2.2.2.10 Antibody staining and imaging

Use an oven at 60-70 °C to dissolve buffer 2.

1. Prepare the following solutions. Each tray for 8 slides takes 30 ml.

Buffer 1: 100 mM Tris-HCl, 150 mM NaCl, 300 ml per box, make 2.5 litres

Buffer 1 + triton: 100 mM Tris-HCl, 150 mM NaCl, 0.3% triton X-100, make 2.5 litres

Buffer 2: 0.5% (w/v) blocking reagent (11 096 176 001, Roche) in buffer 1. Make it fresh, let it dissolve for 1 hour at 60-70 °C, the solution will remain turbid, and leave to cool. 30 ml per tray

Buffer 3: 1% BSA (A7906, Sigma) (w/v), 0.3% (v/v) Triton X-100 in buffer 1 Make fresh. 30 ml per tray plus 100 ml for Buffer 4

Buffer 4: Anti-digoxigenin-AP (11093274910, Roche) 1:3000 in buffer 3. 15 ml per tray of 8 slides. Make fresh shortly before use (100 ml Buffer 3 + 33 µl anti-digoxigenin-AP)

Buffer 5: Mix 60 ml 1 M Tris-HCl 1 M NaCl pH 9.5 and 60 ml 0.5 M $MgCl_2$ with 480 ml dH_2O

Buffer 6: Buffer 5 + NBT 0.15 mg/ml (stock 50 mg/ml, 11383213001, Roche) + BCIP 0.075 mg/ml (stock 50 mg/ml, 11383221001, Roche), 25 ml per plate

2. Transfer all slides to flat trays, 8 slides per tray. Wash steps are all carried out at room temperature on a rocking platform:

- Buffer 1: 5 min
- Buffer 2: 1 h
- Buffer 3: 30 min
- Buffer 4: 1 h 30 min

3. Transfer slides back to racks. Wash steps are all carried out in in separate boxes, on the shaker at room temperature:

- Buffer 1 + Triton: 4 x 25 min
- Buffer 1: 5 min
- Buffer 5: 5 min

4. Transfer slides to clean trays, add 25 ml Buffer 6 per tray, seal places with masking tape and store in the dark overnight.
5. After overnight staining slides were transferred into dH_2O and imaged using a light microscope (Leica DM6000). Multiple sections (5+) from at least 2 independent plants were examined for each tissue, time-point and probe.

2.2.3 Xylem and phloem tracer experiment

2.2.3.1 Plant material

NAM RNAi wheat plants and control plants (null segregants) (Uauy et al. 2006b) were germinated in the dark at 4 °C on moist filter paper in petri dishes for 48 hours and incubated for 48 hours in the light at room temperature. They were then sown into P40 trays containing peat and sand (85% Fine Peat, 15% Grit, 2.7 kg/m³ Osmocote 3-4 months, Wetting Agent,4 kg/m³ Maglime, 1 kg PG Mix) before potting on at 2-3 leaf stage to 1 L pots in cereal mix (40% Medium Grade Peat, 40% Sterilised Soil, 20% Horticultural Grit, 1.3 kg/m³ PG Mix 14-16-18 + Te Base Fertiliser, 1 kg/m³ Osmocote Mini 16-8-11 2 mg + Te 0.02% B, Wetting Agent, 3 kg/m³ Maglime, 300 g/m³ Exemptor). All plants were grown in a heated glasshouse ~20 °C with 16 h supplementary light.

2.2.3.2 Strontium (Sr) and Rubidium (Rb) treatment

Plants were treated with Sr and Rb solution at 4 time-points: 0, 14, 22 and 30 DAA. Between the 10 am and 2 pm on the day of harvest the first tiller of each plant was cut 3 cm from the soil and the cut end was submerged into a large tray of water. The stem was cut underwater 3 cm below the flag leaf node and the cut tiller was transferred to a 50 ml centrifuge, keeping the cut end underwater. Water was removed with a pipette until only ~200 µl remained, enough to immerse the cut end. To the centrifuge tube

either 40 ml of water (control) or Sr and Rb solution (2 mM $SrCl_2$, 2 mM RbCl) was added carefully, avoiding the introduction of bubbles. The top of the tube was sealed with parafilm around the stem to prevent evaporation. These immersed cut tillers were transferred to a controlled environment room with 16 hour light 20 °C, dark 8 hours18 °C, relative humidity 50-70% and incubated for 72 hours to allow the plants to take up the solution. The tillers were then placed in paper bags and dried for 72 h at 60 °C. Tillers were then divided into 4 tissues: ear (cut just above node between rachis and peduncle), flag leaf blade and sheath, peduncle (enclosed and exposed, includes node leading to the ear), node and internode 1 (cut just above the flag leaf node). The individual tissues were weighed and ground in a ball mill or genogrinder using plastic weighing tubes and agate 5 mm balls (5 per tube). Three individual plants (biological replicates) were analysed for RNAi and control plants at each time-point, for Sr and Rb-fed plants and water-fed plants. Samples were sent to UC Davis analytical lab (http://anlab.ucdavis.edu/) for nitric acid/hydrogen peroxide microwave digestion and measurement by Inductively Coupled Plasma Atomic Emission Spectrometry (ICP-AES). Samples were blocked into 3 replicates which were run separately on the ICP-AES. Each block contained peach leaf standards (1547 from the catalogue of uniform samples available at the National Institute of Standards and Technology (NIST), USA), Sr and Rb spiked samples, and $SrCl_2$ and RbCl standards to ensure accurate measurements.

In addition to analysing these samples for their Sr and Rb content to examine xylem and phloem transport respectively, the samples were also analysed for other micronutrients in the same ICP-AES run. These other elements were: calcium (Ca), copper (Cu), iron (Fe), potassium (K), magnesium (Mg), manganese (Mn), phosphorous (P) and zinc (Zn). We examined whether these other elements were altered in RNAi plants compared to control plants, which had been previously reported (Waters et al. 2009). This acted as a control to ensure the RNAi plants were performing as we expected.

2.2.3.3 Statistical analysis

Statistical significance in mineral concentrations and percentages were calculated in Genstat (v15.2.0.8821) using t-tests between RNAi and control plants for each specific time-point and tissue.

2.3 Results

2.3.1 Hypothesis 1: Patterns of *NAM* gene expression will differ between organs

2.3.1.1 *NAM-B1* expression

To establish where *NAM-B1* and its homoeologous and paralogous copies act we determined their expression patterns during a time-course after anthesis using qPCR. Expression in vegetative tissues was examined using homoeologue specific primers normalised to 4 housekeeping genes. *NAM-B1* expression increased after anthesis in the flag leaf blade as previously reported (**Figure 2.5B**) (linear mixed model, days after anthesis effect: $p < 0.001$). Surprisingly *NAM-B1* expression was also seen in all other tissues examined, including the flag leaf sheath, ear (rachis and floret) and stem (internode 1, enclosed and exposed peduncle) and levels of expression were statistically significantly different ($p < 0.001$) which interacted with the days after anthesis ($p = 0.004$). Generally expression levels increased after anthesis with a maximum reached at 20 days after anthesis (DAA) which was maintained at 30 DAA. The highest expression level at 20 DAA was observed in the enclosed peduncle, suggesting that *NAM-B1* might be controlling transport in addition to remobilisation from the flag leaf. A technical explanation for this could be if the reference genes were expressed at a lower level in the enclosed peduncle than in other tissues. However reference genes were expressed at a similar level in the enclosed peduncle to flag leaf blade across all time-points which suggests that the high expression observed in the enclosed peduncle is biologically real. Expression in the internode 1 (the internode immediately below the flag leaf node) continued to increase from 20 to 30 DAA, unlike in all other tissues where the expression level at 20 and 30 DAA was the same.

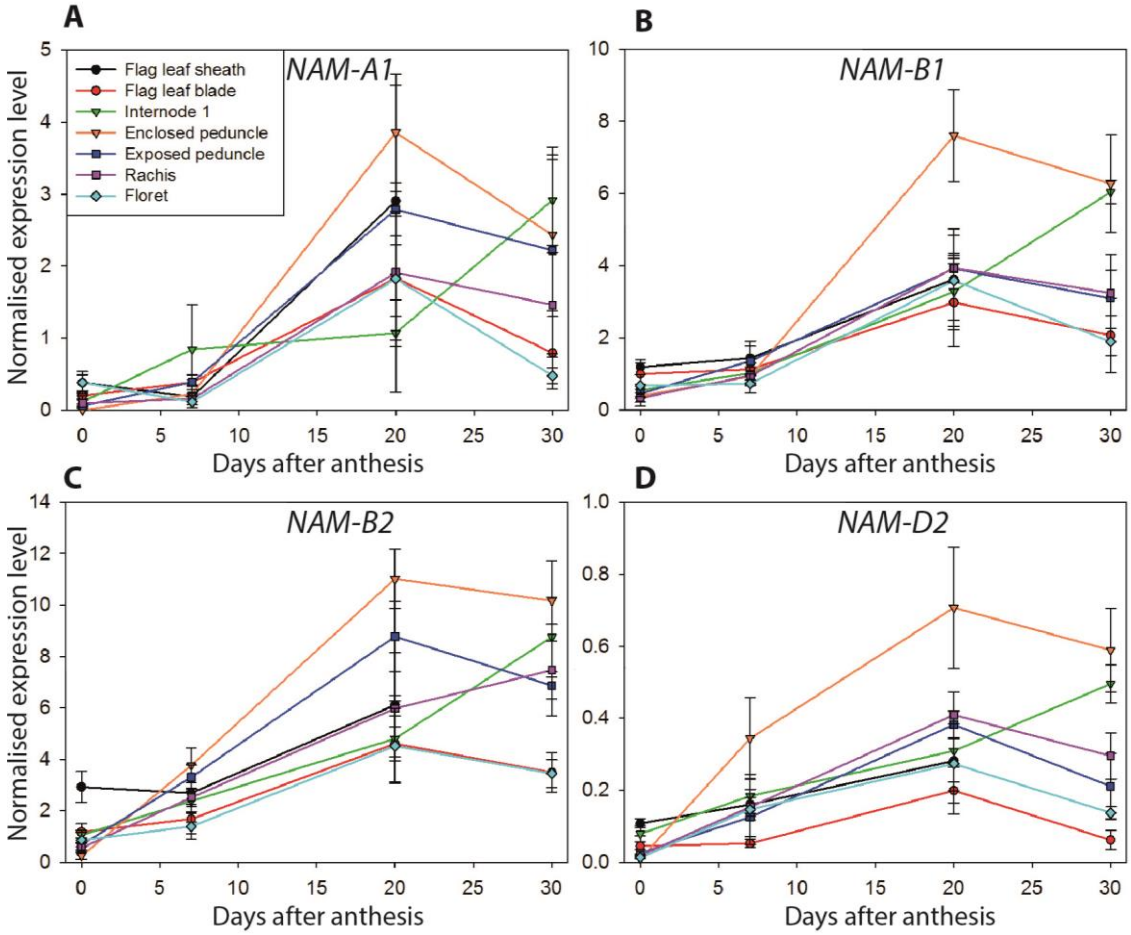


Figure 2.5. Homoeologue specific expression of *NAM* genes after anthesis in vegetative organs. A) *NAM-A1*, B) *NAM-B1*, C) *NAM-B2*, D) *NAM-D2*. n=5, error bars = standard error of the mean (SEM). All samples were divided by the expression of *NAM-B1* in the flag leaf blade at 0 DAA to enable comparisons between tissues and homologues.

2.3.1.2 *NAM-B1* homologue expression

Gene expression of homoeologous and paralogous copies of *NAM-B1* was also examined in vegetative organs after anthesis because differential expression patterns might imply functional specialisation of homologues. In general similar expression patterns were observed in all homologues (*NAM-A1*, *NAM-B1*, *NAM-B2*, *NAM-D2*) between organs (**Figure 2.5**). All homologues showed significant differences in expression over the time-course ($p < 0.001$). Significant differences between expression between tissues was found for *NAM-B1*, *NAM-B2* and *NAM-D2* (all $p < 0.001$) but not for *NAM-A1*. The interaction between time-point and tissue was only significant for *NAM-B1* ($p = 0.004$) and *NAM-B2* ($p = 0.006$). Expression of *NAM-D1* was below detection in all samples (data not shown). Homologues were expressed at significantly different levels in all tissues ($p < 0.001$). *NAM-B2* was generally expressed at higher levels in all organs than all other homologues, with *NAM-A1* and *NAM-B1* having

intermediate expression levels and *NAM-D2* being expressed at low levels in all organs compared to the other homologues.

2.3.1.3 *NAM* gene expression in grains

To understand the full extent of *NAM* gene expression we carried out a further experiment to examine gene expression in the grain. It was unknown whether *NAM-B1* was also expressed in the grain, where it might control nutrient import. Flag leaf blades were harvested to compare between the previous vegetative organs experiment and this (grain) experiment.

The expression pattern observed in the flag leaf blade in the grain experiment (**Figure 2.6B**) is shifted compared to the expression pattern observed in the vegetative organs experiment, with maximum *NAM* gene expression at 7 DAA compared to 20 DAA (**Figure 2.6A**). In the grain experiment it was observed that the plants were senescing more rapidly than in the previous experiment which might explain this shift in gene expression. Nevertheless it is clear that in the grain *NAM-B1* and *NAM-B2* expression levels rise after anthesis, peaking 20 DAA (**Figure 2.6C**) (statistically significant effect of time-point: *NAM-B1*, $p = 0.02$ and *NAM-B2*, $p = 0.017$). The changes in *NAM-A1* expression over the time-course in the grain are non-significant due to large variation within the data, therefore this should be repeated. Using the flag leaf blade as a calibrator, 20 DAA in the grain experiment corresponds to a later time-point in the vegetative organ experiment, suggesting that *NAM* gene expression occurs later in grains than in other organs. Additionally in the grains *NAM-A1*, *NAM-B1* and *NAM-B2* are expressed to a similar maximum level (**Figure 2.6C**), whereas in the vegetative organs gene expression is consistently *NAM-B2 > NAM-B1 > NAM-A1* (**Figure 2.7**). In both grain and vegetative tissues *NAM-D2* is expressed at very low levels. In general the level *NAM-A1, NAM-B1 and NAM-B2* expression in the grain is in the same range as the expression levels found in the flag leaf, showing consistency between these two tissues.

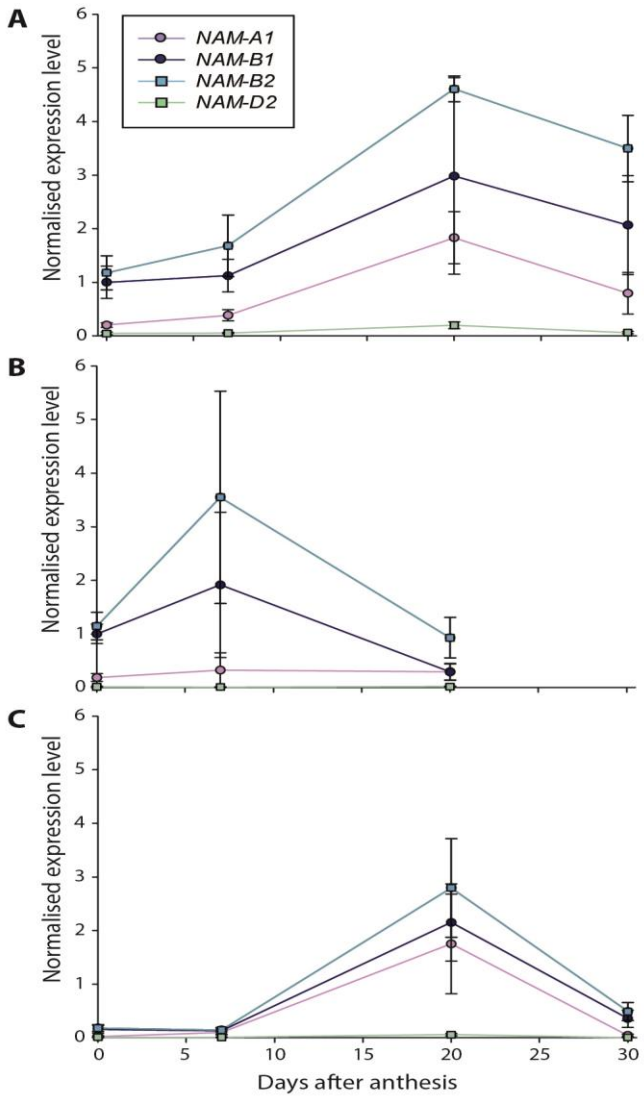


Figure 2.6. Grain and flag leaf blade expression of *NAM* genes after anthesis. A) Flag leaf blade from vegetative organs experiment, B) flag leaf blade from grain experiment, C) grain. n=5, error bars = SEM. All samples were divided by the expression of *NAM-B1* in the flag leaf blade at 0 DAA from the vegetative tissues experiment to enable comparisons between tissues and homologues.

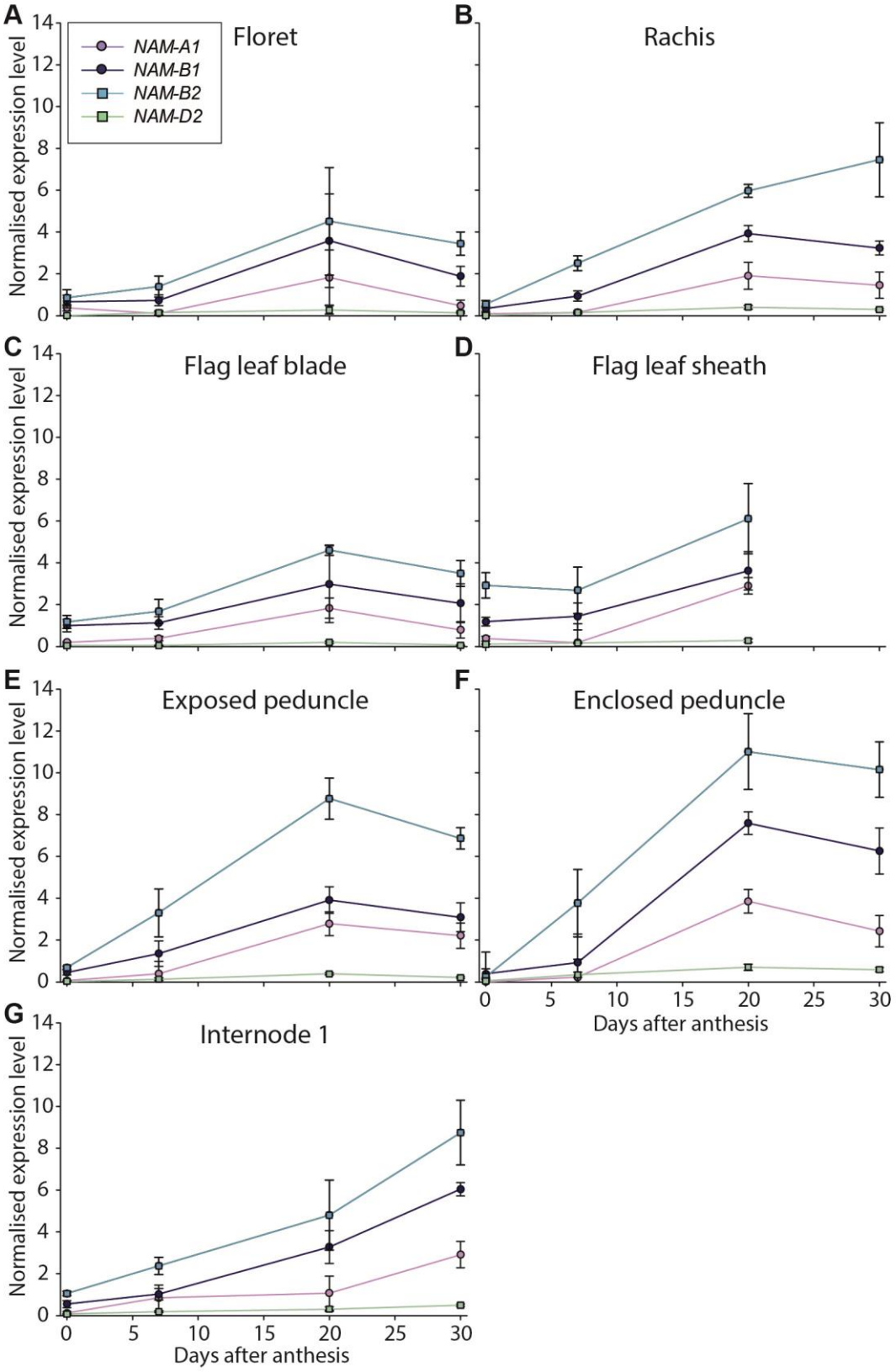


Figure 2.7. Expression of *NAM* genes after anthesis in vegetative organs. A) floret, B) rachis, C) flag leaf blade, D) flag leaf sheath, E) exposed peduncle, F) enclosed peduncle, G) Internode 1. n=5, error bars = SEM. All samples were divided by the expression of *NAM-B1* in the flag leaf blade at 0 DAA to enable comparisons between tissues and homologues.

2.3.2 Hypothesis 2: *NAM* genes will be expressed in specific cell types particularly those related to transport

To explore *NAM* gene expression patterns at a tissue and cellular level RNA *in situ* hybridisations were carried out using a probe which would hybridise with all *NAM* gene mRNA (*NAM-A1*, *NAM-A2*, *NAM-B1*, *NAM-B2*, *NAM-D1* and *NAM-D2*) due to their sequence similarity (> 80 %) within the probe region (see methods section 2.2.2.4 and **Table 2.5**). Two time-points were examined when the maximum changes to gene expression were observed by qPCR: 0 DAA when very little gene expression was observed and 20 DAA when gene expression was high. As mentioned above the 20 DAA time-point may have variable gene expression between batches of plants due to environmental conditions. The plants used for the RNA *in situ* hybridisations showed no visual signs of senescence at 20 DAA and therefore were similar to the plants used for the vegetative tissues qPCR experiment.

A transverse section illustrating the main cell types of interest is shown in **Figure 2.8** and their general functions are described in **Table 2.14**.

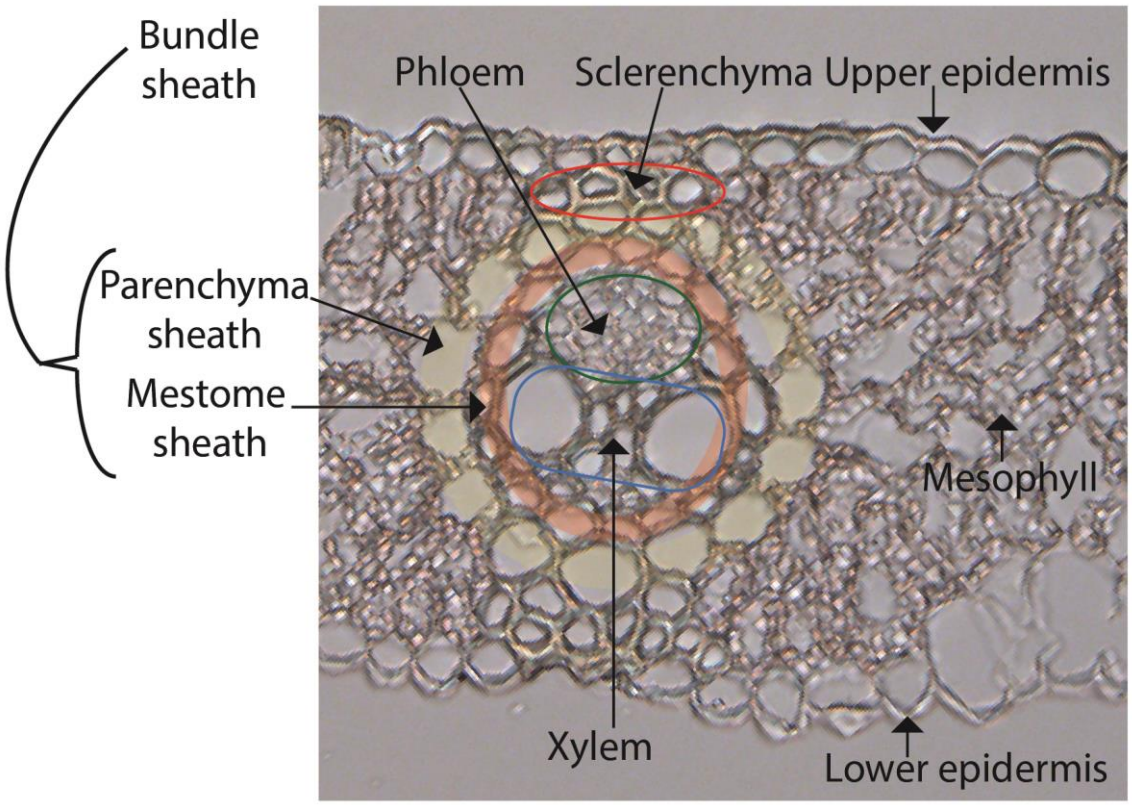


Figure 2.8. Transverse section through wheat flag leaf with cell types annotated.

Tissue	Abbreviation	Function
Epidermis	E	A single layer of cells covering the outside of a tissue, protects against water loss, regulates gas exchange, secretes and absorbs compounds
Parenchyma	Par	Thin walled cells with large central vacuoles which make up the bulk of the soft tissues within a plant and may be important for storage
Chlorenchyma	Chl	Specialised type of parenchyma with chloroplasts for photosynthesis
Sclerenchyma	S	Supporting tissue containing thick secondary cell walls composed of cellulose, hemicellulose and lignin
Outer vascular bundle	OVB	Collection of transport tissues (phloem, xylem and supporting tissues) located towards the outer edge of the stem
Inner vascular bundle	IVB	Collection of transport tissues (phloem, xylem and supporting tissues) located towards the inside of the stem
Mestome sheath	MS	Specialised cells surrounding the xylem and phloem. Hypothesised to control the direction of nutrient transport into/out of the xylem and phloem.
Xylem	X	Transports water and certain nutrients, important for transpiration
Metaxylem	Mx	Specialised type of xylem cell with a wider aperture
Phloem	P	Transports photosynthate and other organic (and some inorganic) nutrients

Table 2.14. Cell types observed in transverse sections through flag leaf and stem tissues used for RNA *in situ* hybridisation.

2.3.2.1 *NAM* expression in the flag leaf blade

Three distinct regions of the flag leaf blade were studied: the tip of the flag leaf, the middle region of the flag leaf and the base of the flag leaf. In all three regions broadly similar staining patterns were observed with strong signal in the vascular bundles particularly the mestome sheath cells and in the sclerenchyma (**Figure 2.9**,**Figure 2.10** and **Figure 2.11**). However the different regions showed distinct patterns of expression with respect to time-point. In all tissues and time-points the sense control probes detected no gene expression (**B** and **D** in **Figure 2.9**, **Figure 2.10** and **Figure 2.11**).

In the tip of the flag leaf blade lower expression was detected at 0 DAA than at 20 DAA (**Figure 2.9A** and **C**) as was shown by qPCR (**Figure 2.6A**). At 0 DAA weak expression was seen in sclerenchyma and strong expression was visible in the lateral mestome sheath cells. At 20 DAA strong expression in mestome sheath was maintained and strong expression was also detected in the xylem. Expression in the sclerenchyma was also more visible at 20 DAA.

Contrastingly in the middle of the flag leaf blade expression at 0 DAA was surprisingly similar or higher than expression at 20 DAA (**Figure 2.10A** and **C**). Similar to the tip of the leaf, strong expression was observed in the xylem, mestome sheath and sclerenchyma. The base of the flag leaf showed a more similar pattern to the tip, with expression increasing from 0 to 20 DAA particularly in the mestome sheath, xylem and sclerenchyma (**Figure 2.11A** and **C**). In all regions of the leaf very little or no gene expression was detected in the mesophyll and epidermal cells.

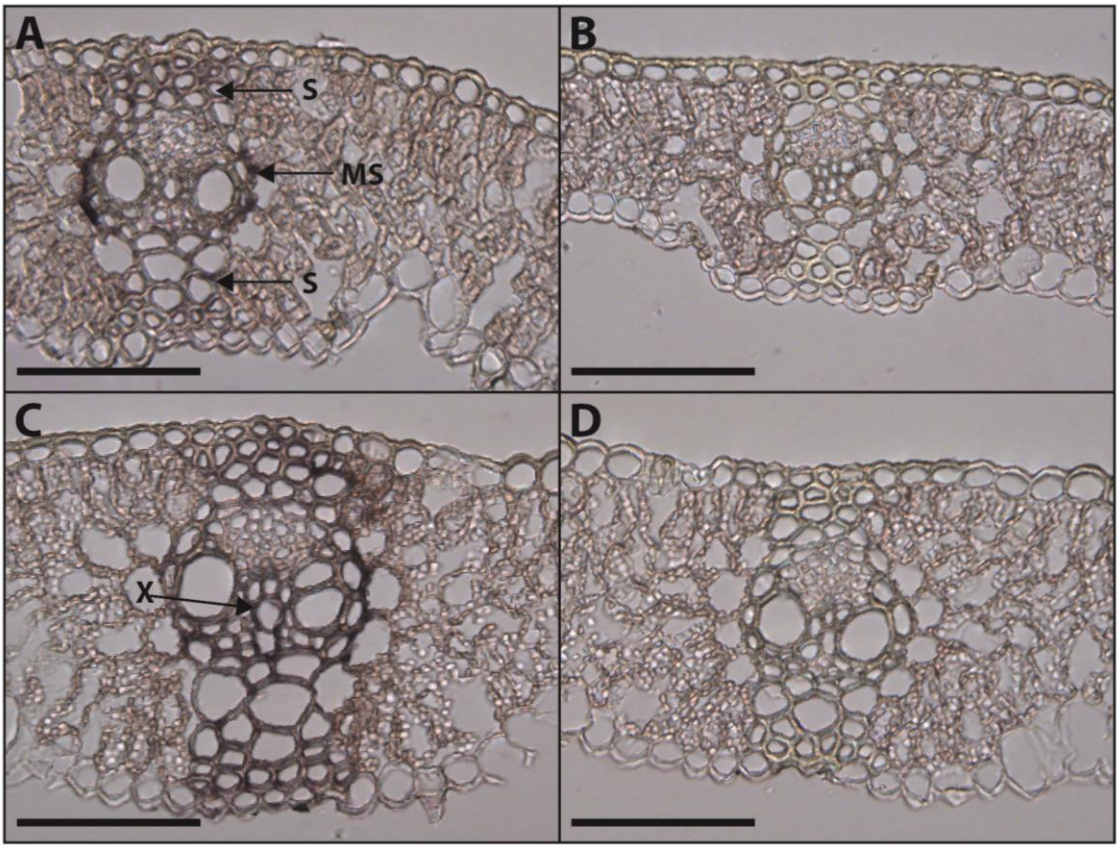


Figure 2.9. Tip of flag leaf blade transverse sections stained by RNA *in situ* hybridisation using antisense (A and C) and sense (B and D) *NAM* probes. 0 DAA (A and B) and 20 DAA (C and D). S = sclerenchyma, MS = mestome sheath, X = xylem. Scale bars 100 µm.

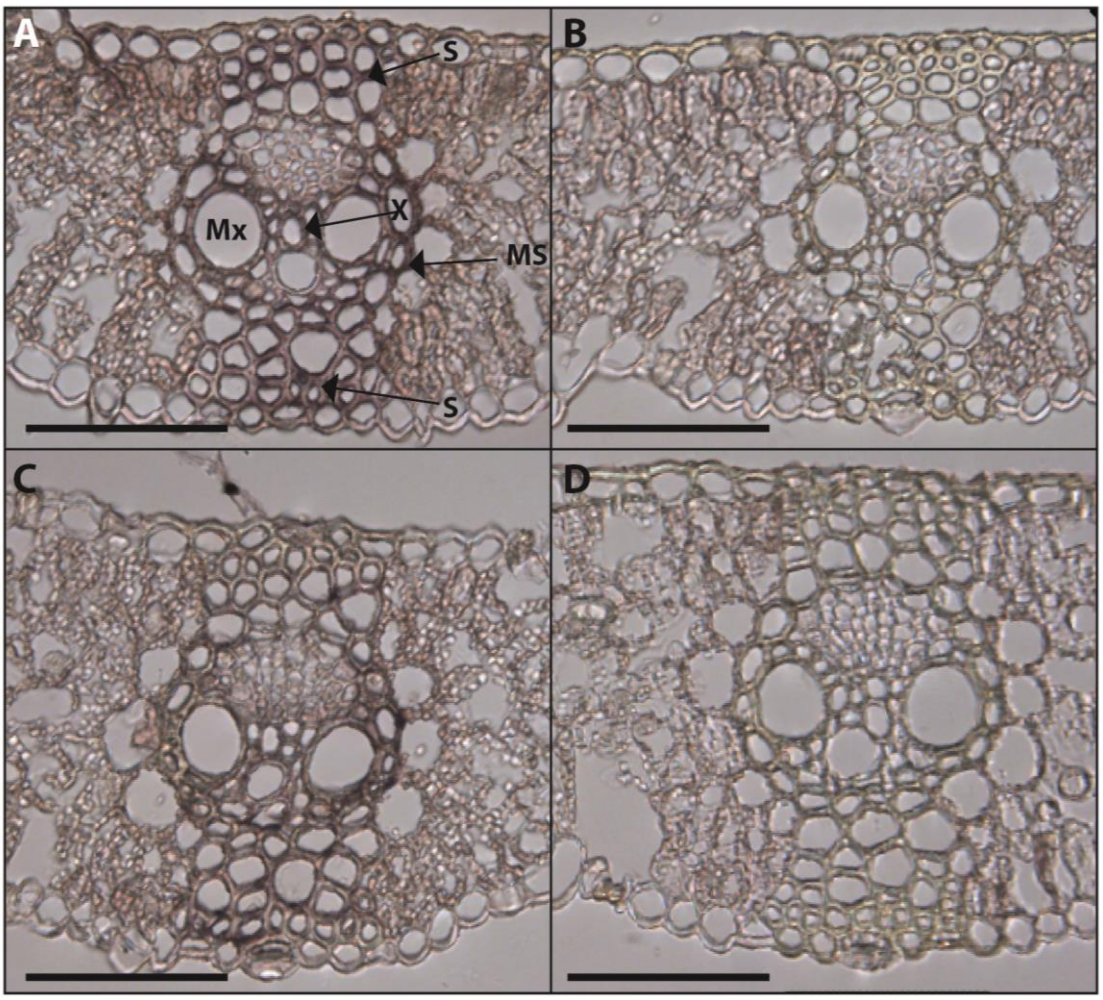


Figure 2.10. Middle of flag leaf blade transverse sections stained by RNA *in situ* hybridisation using antisense (A and C) and sense (B and D) *NAM* probes. 0 DAA (A and B) and 20 DAA (C and D). S = sclerenchyma, MS = mestome sheath, X = xylem, Mx = metaxylem. Scale bars 100 µm.

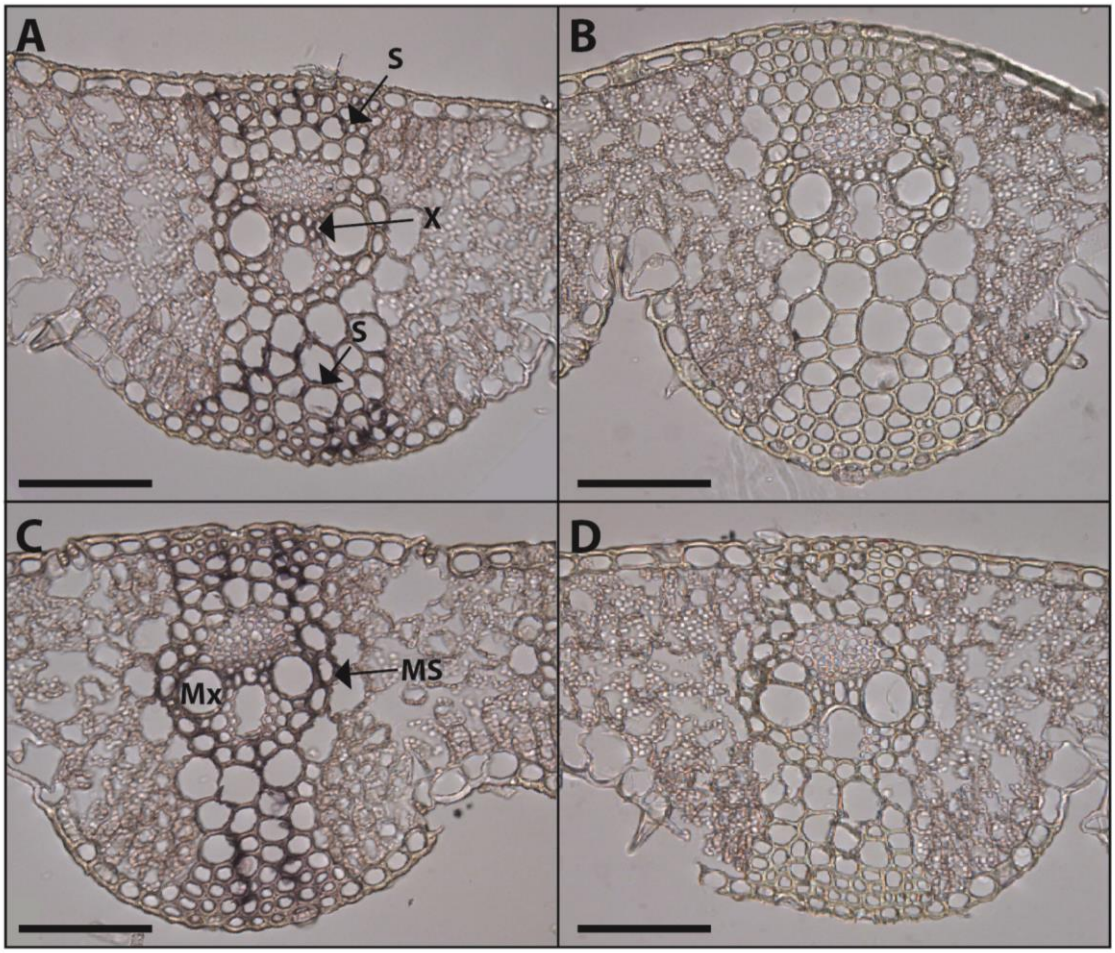


Figure 2.11. Base of flag leaf blade transverse sections stained by RNA *in situ* hybridisation using antisense (A and C) and sense (B and D) *NAM* probes. 0 DAA (A and B) and 20 DAA (C and D). S = sclerenchyma, MS = mestome sheath, X = xylem, Mx = metaxylem. Scale bars 100 µm.

2.3.2.2 *NAM* expression in the stem tissues

High *NAM* gene expression levels were detected in both enclosed and exposed peduncle and the internode by qPCR at 20 DAA (**Figure 2.7**). Therefore in addition to examining gene expression patterns at the tissue level in the flag leaf, RNA *in situ* hybridisations were also carried out in these stem tissues.

In the enclosed peduncle (the portion of peduncle wrapped inside the flag leaf sheath) *NAM* gene expression was not detected at anthesis, but high expression was observed 20 DAA (**Figure 2.12A** and **C**). *NAM* gene expression was present in all cell types observed including the parenchyma cells where expression was not seen in flag leaves. Particularly high expression was observed in sclerenchyma and in outer vascular bundles including the mestome sheath, which is similar to the patterns observed in the flag leaf. However, surprisingly expression was much lower in inner vascular bundles

than in outer vascular bundles. Additionally strong *NAM* expression was detected in the epidermal cell layer whereas in flag leaves expression was largely absent in the epidermis (e.g. **Figure 2.11C**).

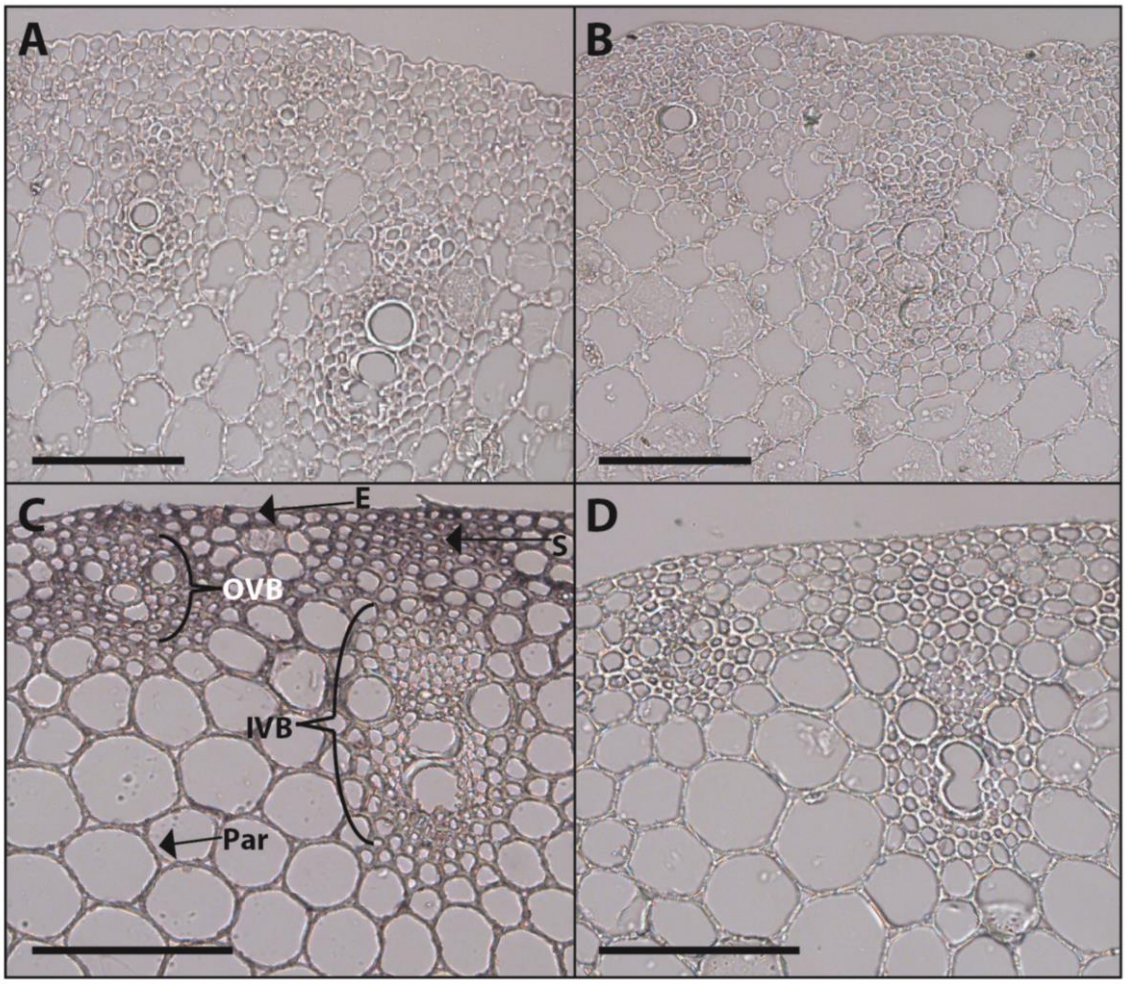


Figure 2.12. Enclosed peduncle transverse sections stained by RNA *in situ* hybridisation using antisense (A and C) and sense (B and D) *NAM* probes. 0 DAA (A and B) and 20 DAA (C and D). S = sclerenchyma, E = epidermis, OVB = outer vascular bundle, IVB = inner vascular bundle, Par = parenchyma. Scale bars 100 µm.

The expression of *NAM* genes in the exposed peduncle (**Figure 2.13A** and **C**) is quite different to that in the enclosed peduncle (**Figure 2.12A** and **C**). In the exposed peduncle *NAM* genes are expressed at 0 DAA in the sclerenchyma, the parenchyma and the mestome sheath, whereas no expression is seen in the enclosed peduncle. At 20 DAA in the exposed peduncle *NAM* expression has increased in these tissues. It is very noticeable that *NAM* genes are not expressed in the chlorenchyma or the phloem at either 0 or 20 DAA in the exposed peduncle.

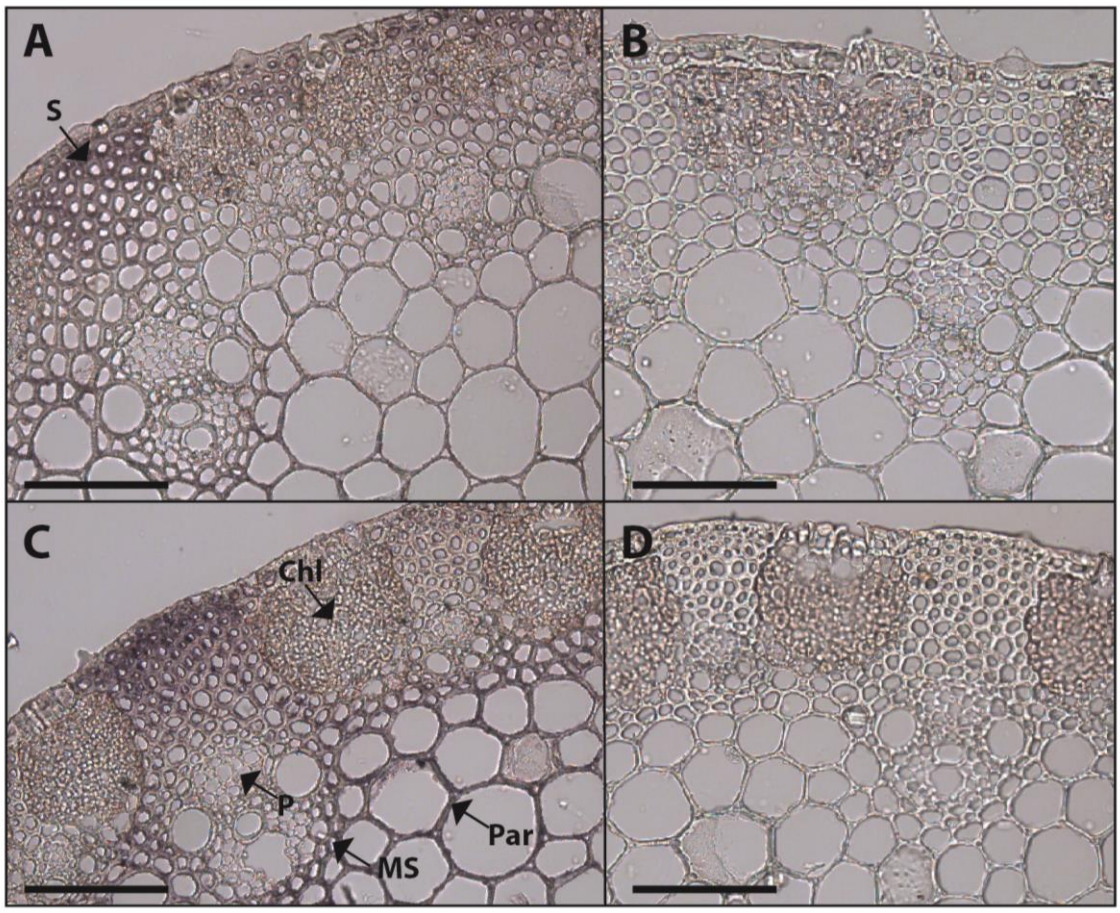


Figure 2.13. Exposed peduncle transverse sections stained by RNA *in situ* hybridisation using antisense (A and C) and sense (B and D) *NAM* probes. 0 DAA (A and B) and 20 DAA (C and D). S = sclerenchyma, Chl = chlorenchyma, MS = mestome sheath, Par = parenchyma, P = phloem. Scale bars 100 µm.

Internode 1 shares some features of *NAM* expression with both regions of the peduncle. In the internode (**Figure 2.14A** and **C**), similar to in the exposed peduncle (**Figure 2.13**A and C) *NAM* expression is absent from the chlorenchyma at both 0 and 20 DAA. In internode 1 at 0 DAA *NAM* expression is restricted to the parenchyma and the mestome sheath cells surrounding the xylem (**Figure 2.14**A). However at 20 DAA in internode 1 *NAM* gene expression is also present in the epidermis and the sclerenchyma as is seen in both peduncle tissues at 20 DAA. In internode 1 at 20 DAA *NAM* gene expression is also present in the xylem in both the outer vascular bundles and to a lower level in the inner vascular bundles. Overall *NAM* gene expression is stronger at 20 DAA in all three stem tissues; however this difference is smallest in the exposed peduncle.

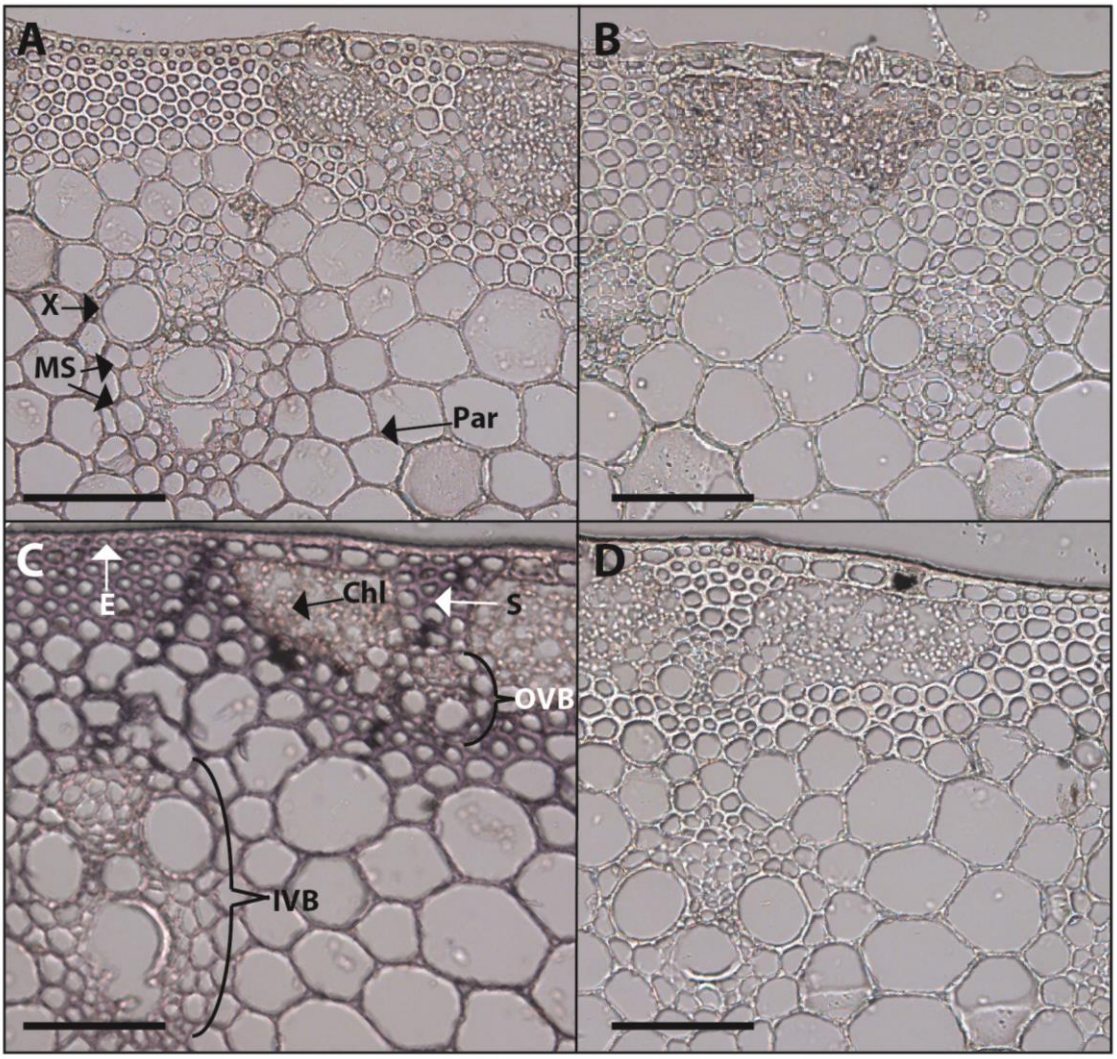


Figure 2.14. Internode 1 transverse sections stained by RNA *in situ* hybridisation using antisense (A and C) and sense (B and D) *NAM* probes. 0 DAA (A and B) and 20 DAA (C and D). E= epidermis, OVB = outer vascular bundle, IVB = inner vascular bundle, S = sclerenchyma, Chl = chlorenchyma, Par = parenchyma, X = xylem, MS = mestome sheath. Scale bars 100 µm.

In summary *NAM* gene expression is cell type specific in both leaf and stem organs, as is shown in **Table 2.15**. In general stronger *NAM* gene expression is detected by RNA *in situ* hybridisation at 20 DAA than at 0 DAA, as was found by qPCR. In all organs at 20 DAA *NAM* genes are strongly expressed in the sclerenchyma and the mestome sheath. In the leaves strong *NAM* gene expression is present in the xylem but absent from the phloem, which is mirrored in the internode 1 and exposed peduncle.

	Epidermis		Parenchyma		Chlorenchyma		Sclerenchyma		Mestome sheath		Xylem		Phloem	
	0	**20**	**0**	**20**	**0**	**20**	**0**	**20**	**0**	**20**	**0**	**20**	**0**	**20**
Tip of flag leaf blade	-	-	-	-	-	-	+	++	+++	+++	+	+++	-	-
Middle of flag leaf blade	-	-	-	-	-	-	+++	++	+++	++	+++	++	-	-
Base of flag leaf blade	-	-	-	-	-	-	+	+++	+	+++	+	+++	-	-
Enclosed peduncle	-	+++	-	++	n/a	n/a	-	+++	-	+++OVB +IVB	-	+++OVB +IVB	-	+++OVB -IVB
Exposed peduncle	+	+	+	+++	-	-	+++	+++	+	+++	-OVB +IVB	+OVB +IVB	-	-
Internode 1	-	+++	+	++	-	-	-	+++	+	+++OVB ++IVB	-	+++OVB +IVB	-	-

Table 2.15. Summary of *NAM* gene expression results from RNA *in situ* hybridisation. 0 = 0 DAA, 20 = 20 DAA, OVB = outer vascular bundle, IVB = inner vascular bundle. Expression level is denoted on the scale: - = no expression, + = weak expression, ++ = medium expression, +++ = strong expression.

2.3.3 Hypothesis 3: Xylem and phloem transport to the ear will be affected by reduction of *NAM* gene expression

It had previously been found that micronutrients accumulate to different levels in *NAM* RNAi plants compared to control plants (Waters et al. 2009). We had observed that in general xylem-mobile elements were decreased in RNAi grain, whereas phloem mobile elements were increased compared to control plants. To test whether xylem and phloem transport were altered in RNAi plants compared to control plants we carried out a tracer study by feeding cut wheat tillers phloem-mobile rubidium (Rb) and xylem-mobile strontium (Sr). We then analysed where in the plant Rb and Sr were accumulated.

The analysis of Rb and Sr distribution was carried out by ICP-AES which also analysed several other micronutrients including calcium (Ca), copper (Cu), iron (Fe), potassium (K), magnesium (Mg), manganese (Mn), phosphorous (P) and zinc (Zn). These elements had previously been analysed in NAM RNAi plants, therefore as a control we decided to check that the distribution of these elements within the RNAi plants we grew were as expected from the previously published results (Waters et al. 2009).

2.3.3.1 Micronutrient accumulation in *NAM* RNAi ears compared to control ears

To confirm that the RNAi plants were behaving as we expected we examined the concentrations of Ca, Cu, Fe, K, Mg, Mn, P and Zn in the ears of RNAi and control plants. We hypothesised that in the RNAi ears the xylem-mobile elements would be reduced whereas the phloem-mobile would be increased, as was previously observed.

Amongst the phloem mobile elements we found that RNAi plants had significantly higher K concentration in their ears than control plants at later time-points (**Figure 2.15**A), for example at 30 DAA RNAi plants had 155 % of the concentration found in control ears. This is similar to the ~170 % difference found previously in grain at 28 DAA (Waters et al. 2009). We found no statistically significant difference in Ca concentration (**Figure 2.15**B), and in the previous report RNAi grain were not significantly different to the control grain at 28 DAA (Waters et al. 2009).

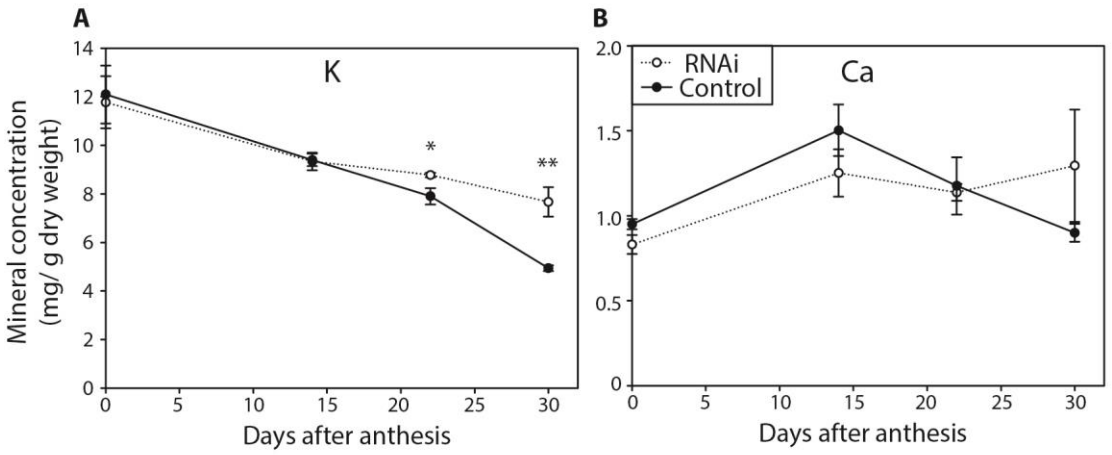


Figure 2.15. Mineral concentrations of phloem-mobile elements in ears of *NAM* RNAi and control plants. A) Potassium, B) Calcium. n=6, error bars = SEM. Asterisks denote significant differences at $p < 0.05$ (*) and $p < 0.01$ (**) as determined by Student t-test between genotypes at each time-point.

Elements which are transported through the plant in both xylem and phloem (Cu, Fe and Zn) (**Figure 2.16**) were accumulated to lower levels in the RNAi plants' ears than in control plants. The values obtained were similar to those previously reported. For example comparing ear values from this data at 30 DAA (**Figure 2.16**) to grain values at 28 DAA (Waters et al. 2009) the proportion of the element found in RNAi plants compared to controls was 78 % and 74 % for Cu, 70 % and 47 % for Fe and 80 % and 64 % for Zn. The overall trend is the same with RNAi plants containing less Cu, Zn and Fe than control plants, and the small differences between these results and the previously reported values may be due to the 2 day difference in time-point, growing conditions or because the values being compared are not for exactly the same tissue because our data did not separate the grains from the glumes and rachis.

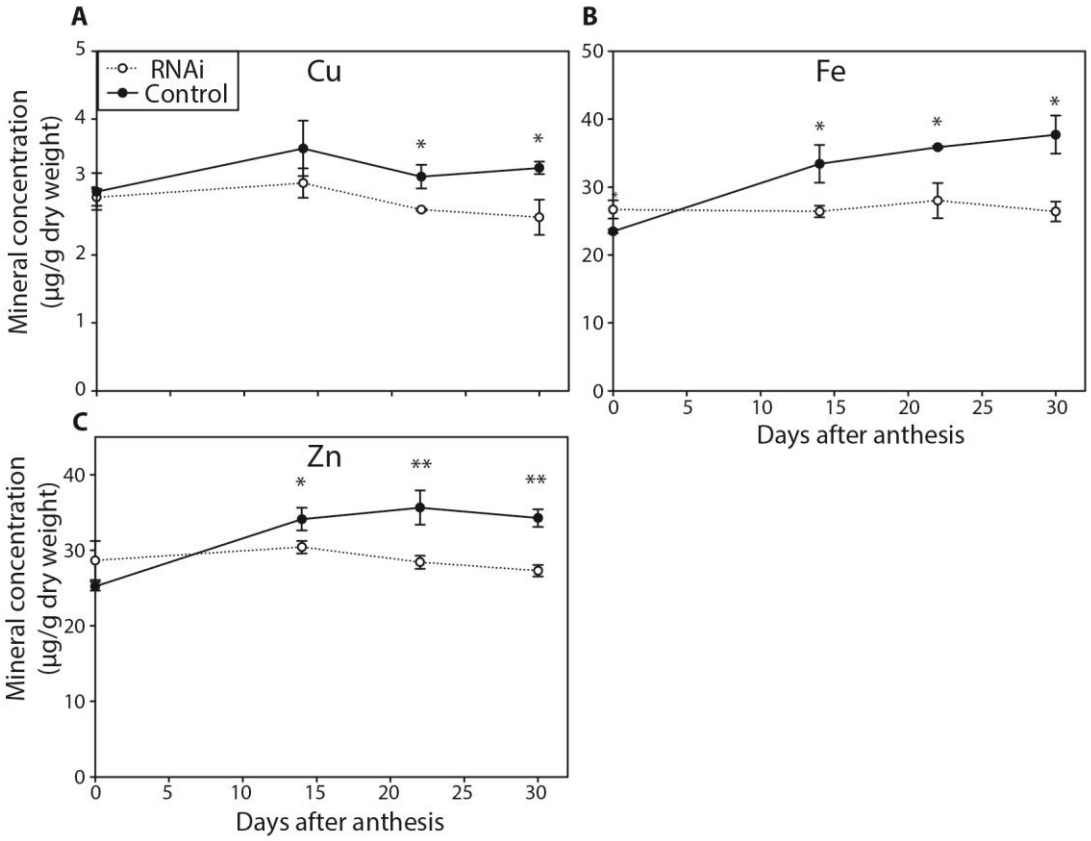


Figure 2.16. Mineral concentrations of xylem/phloem-mobile elements in ears of *NAM* RNAi and control plants. A) Copper, B) Iron, C) Zn. n=6, error bars = SEM. Asterisks denote significant differences at $p < 0.05$ (*) and $p < 0.01$ (**) as determined by Student t-test between genotypes at each time-point.

None of the principally xylem transported elements (Mg, Mn and P) were significantly different between RNAi and control plants' ears (**Figure 2.17**A, B, C) at any time-point. Previously at a similar early time-point (28 DAA) Mg was not found to be significantly different between RNAi and control grain (Waters et al. 2009) and significant differences occurred later at 35 and 42 DAA. However in the previous experiment at 28 DAA RNA grain contained only 66 % of the Mn of control grain and 83 % of the P of control grain (Waters et al. 2009), where at 30 DAA in this experiment no difference was found (**Figure 2.17** B, C). This might be because we did not separate the grain from the rachis and the glumes, whereas the previous experiment only used the grain or because of different growing conditions. The soil type and environment, particularly nitrogen status, can have strong effects on micronutrient uptake, remobilisation and allocation (Cakmak et al. 2010).

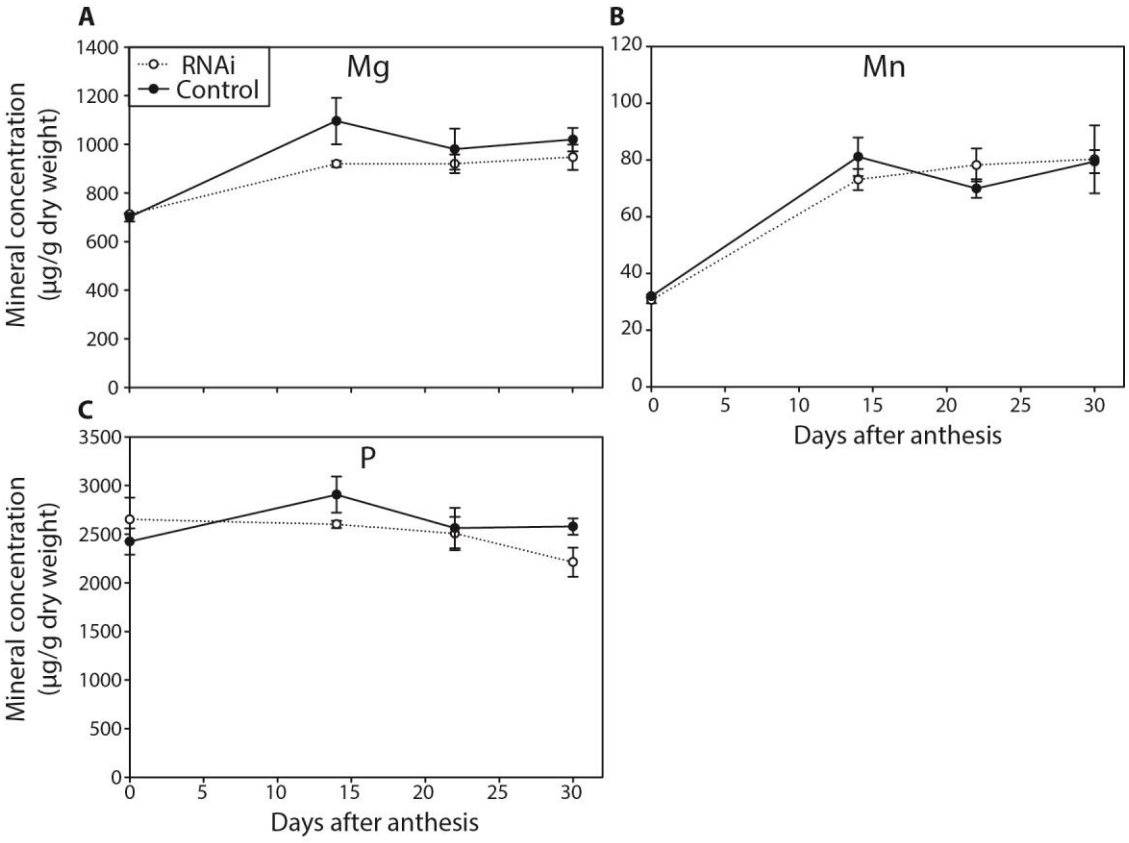


Figure 2.17. Mineral concentrations of xylem-mobile elements in ears of *NAM* RNAi and control plants. A) Magnesium, B) Manganese, C) Phosphorous. n=6, error bars = SEM.

These results broadly agree with the previously reported study, although the experiments are not directly comparable due to different growing conditions and the shorter time-course in this experiment. We found that Zn, Fe and Cu were strongly

reduced in RNAi plants' grain and K was greatly increased. Overall the pattern of micronutrient accumulation in the RNAi grain compared to the control plants was similar to the previously reported distribution (Waters et al. 2009). This confirmed that the RNAi plants were behaving as we expected in terms of micronutrient accumulation. Therefore we went on to test whether the RNAi plants had alterations in their phloem and xylem transporting by using the tracer elements Rb and Sr.

2.3.3.2 Rubidium and strontium transport in *NAM* RNAi plants compared to control plants

To ensure that our detection of rubidium (Rb) and strontium (Sr) was sensitive and accurate we first tested whether a known amount of Rb or Sr, added to the ground tissue, could be recovered. Over 90% of the Rb or Sr spike was recovered (**Figure 2.18**), indicating good accuracy of measurements.

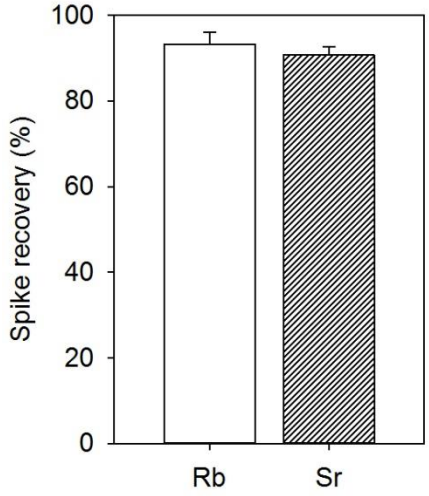


Figure 2.18. Recovery of rubidium (Rb) and strontium (Sr) spike from wheat samples. n= 3, error bars = SEM.

Upon feeding of Rb (phloem mobile) and Sr (xylem mobile) no significant difference was observed in the amount of tracer transported to the ears in *NAM* RNAi and control cut tillers at any time-point after anthesis (**Figure 2.19** and **Figure 2.20**). However, both Rb and Sr are differentially distributed between RNAi and control plants at 30 DAA in vegetative tissues (**Figure 2.21** and **Figure 2.22**), although the differences are not as large as those seen for other minerals. For example the RNAi flag leaves contain 120 % of Rb and 121 % of Sr found in the control leaves, whereas the RNAi flag leaves contain 253 % of the Zn found in control leaves. Both tracers are preferentially accumulated in the leaves of RNAi plants compared to control plants. This is surprising because although Zn, Fe and Cu were found to have their remobilisation from the flag leaf reduced in RNAi lines (Waters et al. 2009), remobilisation of other elements was not affected. It is also noticeable that RNAi plants have a lower proportion of Rb in their peduncles compared to control plants perhaps due to the reduced remobilisation from the flag leaf. At later time-points a lower percentage of Rb and Sr enter the ear (**Figure 2.19** and **Figure 2.20**), instead more accumulates in the peduncle in both RNAi

and control lines (**Figure 2.21** and **Figure 2.22**), perhaps indicating a transport barrier between the peduncle and the ear. Overall these results show that changes to phloem and xylem movement in *NAM* RNAi lines are subtle and cannot explain the large differences in micronutrient content in the grain.

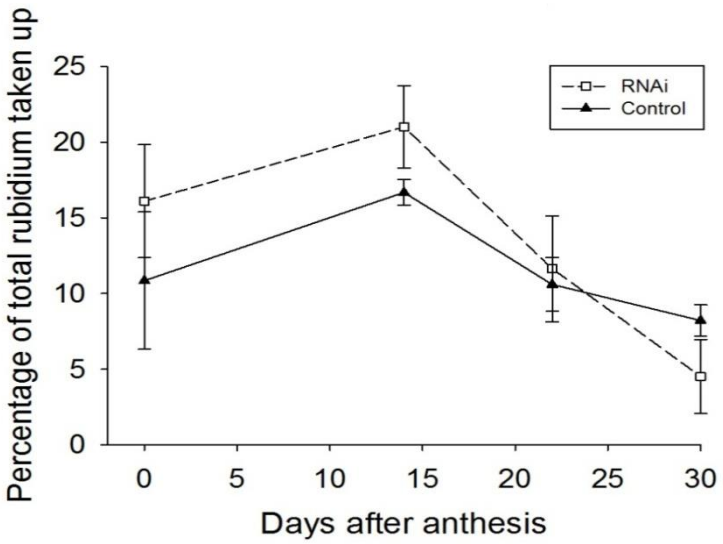


Figure 2.19. The percentage of total rubidium taken up which was transported into the ear. No significant difference was found between RNAi and control at any time-point. n = 3, error bars = SEM.

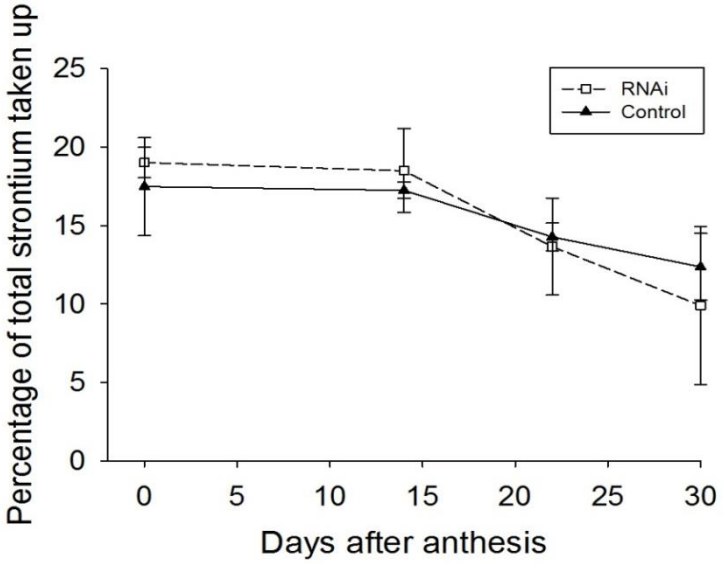


Figure 2.20. The percentage of total strontium taken up which was transported into the ear. No significant difference was found between RNAi and control at any time-point. n = 3, error bars = SEM.

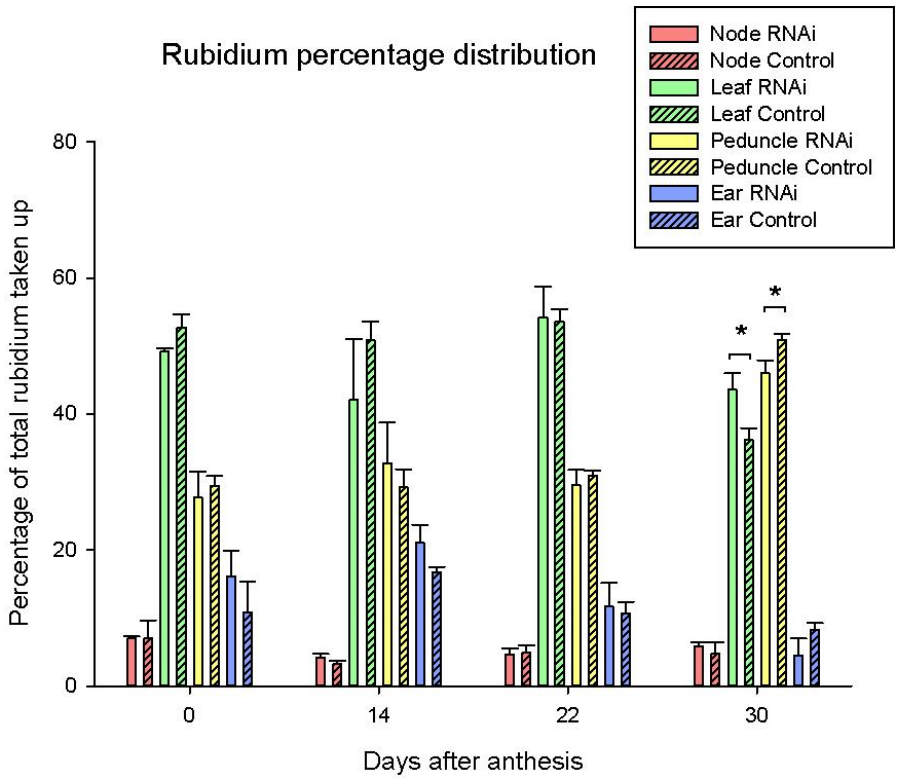


Figure 2.21. The distribution of rubidium taken up by cut wheat tillers. n = 3, error bars = SEM. Asterisk (*) indicates $p < 0.05$.

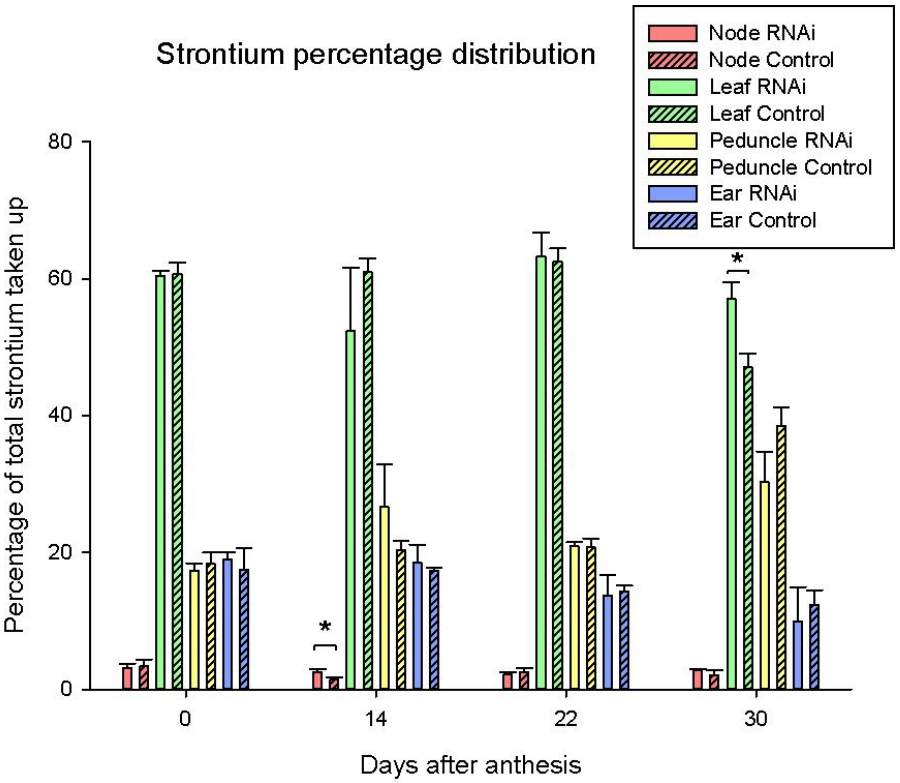


Figure 2.22. The distribution of strontium taken up by cut wheat tillers. n = 3, error bars = SEM. Asterisk (*) indicates $p < 0.05$.

2.4 Discussion

In this chapter we aimed to identify where *NAM* genes are expressed at both the organ and tissue level which would build up a picture of where NAM genes act and might indicate possible functions. We also wanted to test whether *NAM-B1* exerts its effects on nutrient remobilisation and grain nutrient content by altering phloem and xylem transport.

2.4.1 Hypothesis 1: Patterns of *NAM* gene expression will differ between organs

We have found that *NAM-B1* is not only expressed in flag leaves as previously reported (Uauy et al. 2006b) but is also expressed in all other tissues examined including the grain. Expression is temporally regulated, with expression levels rising after anthesis to a maximum at 20 DAA in all tissues (except the internode 1) and the grain where expression reaches a maximum later than 20 DAA. *NAM-B1* therefore may play roles in addition to remobilising nutrients from the flag leaf blade. The expression patterns indicate possible roles in remobilising nutrients from other tissues such as the peduncle and internode, in regulating long distance transport through the peduncle and rachis, and in controlling loading of nutrients into the grain itself. In particular *NAM-B1* may be important in stem tissues where expression is 2-3 fold higher than in the flag leaf blade at 20 DAA.

We have also shown that the homologues of *NAM-B1* are expressed in very similar patterns to *NAM-B1*. The expression of *NAM-B1* homologues at high levels in tissues other than the flag leaf is in conflict with a report where much lower expression of *NAM-A1*, *NAM-D1*, *NAM-D2* and *NAM-B2* was found in the peduncle, rachis, florets/grain and stem tissues compared to in the flag leaf at 14 DAA (Guttieri et al. 2013). This difference in results may be due to the genotypes used. The previous report used a genotype lacking *NAM-B1* expression (hexaploid Bobwhite) which perhaps would affect the expression of the homologues, whereas we used the variety UC1041 with an introgression containing *NAM-B1*. The difference in growing conditions, which we have shown can strongly affect *NAM-B1* expression, may also affect expression of homologues. Technical issues may have also contributed to differences observed: the previous study relied on only one reference gene (actin) whereas our results used four independent reference genes to improve normalisation between tissues. Furthermore the previous study does not report the error bars, making it very hard to compare between

tissues. We believe that our results, that the *NAM* genes are expressed in many tissues, are reliable because they are supported by the independent RNA *in situ* hybridisation results. The phenotype observed in *NAM* RNAi plants also supports this notion: the senescence delay is not restricted to the flag leaf but also is observed in the stem, peduncle, rachis and florets.

The analogous expression patterns of *NAM-B1* and its homologues start to help us understand why these homologues play a role in similar processes. For example single and double *NAM-A1* and *NAM-D1* TILLING mutants have been shown to reduce grain protein content and delay senescence in a similar manner to the *NAM-B1* mutation (Avni et al. 2014). Both *NAM-A1* and *NAM-D1* have phenotypic effects despite their low expression levels compared to *NAM-B1* in our experiment. However the precise expression level of *NAM-A1* compared to *NAM-B1* is different between studies. We generally found that *NAM-A1* was expressed at a similar or lower level to *NAM-B1*, including in the flag leaf blade, but previously it was found that *NAM-A1* expression was higher than *NAM-B1* at 25 DAA in the flag leaf of tetraploid wheat (Uauy et al. 2006b). This might be due to differences in the plant material used (hexaploid compared to tetraploid) or due to differences in growing conditions, which we have shown can alter expressed patterns temporally, and we hypothesise they may also influence the relative expression levels of the homoeologues. *NAM-B2*, which we have found to have the highest level of expression across all tissues, as was previously reported in flag leaves (Uauy et al. 2006b), remains to be characterised in detail, and may have an even stronger effect. Characterisation of TILLING mutants in this gene, as well as *NAM-D2*, would help to delineate the contribution of these homologues. Creating combinations of mutations in multiple homologues might give finer control over the grain nutrient content and senescence profile in wheat.

2.4.2 Hypothesis 2: *NAM* genes will be expressed in specific cell types particularly those related to transport

We have examined *NAM* gene expression more closely at the tissue level, looking at stem and leaf tissues at anthesis and 20 DAA. In the flag leaf blade expression in the tip and base increased from anthesis to 20 DAA, whereas this difference was not clear in the mid flag leaf sections. This was surprising given that a large difference was found by qPCR for the mid region. Subtle changes to growing conditions, however, can cause strong differences in expression pattern. For example flag leaves of plants grown in two separate experiments in the same glasshouse show very different expression profiles

(**Figure 2.6**). The plants used for qPCR and RNA *in situ* hybridisation were grown separately so this may explain the discrepancy between the two techniques, although the visual senescence progression was the same between the two experiments. It is also necessary to note that whilst RNA *in situ* hybridisation can give information about relative expression strength it is only semi-quantitative and perhaps the difference detected by qPCR in this tissue is too subtle to detect by the hybridisation. The localisation of *NAM* gene expression was very similar in all three regions of the leaf, suggesting that *NAM* genes may play a similar role along the length of the leaf, although this might change as senescence progresses (the leaves used in this study showed no visible signs of senescence).

NAM genes were expressed highly at 20 DAA around the vascular bundles, indicating a function in remobilisation and transport of nutrients from the flag leaf. In particular strong expression was observed in the mestome sheath cells which have thickened cell walls containing a suberized lamellae (O'Brien and Kuo 1975). This tissue is hypothesised to keep the flux of water outwards from the xylem separate from the flux of sugars inwards towards the phloem (Canny 1986). Therefore the mestome sheath may play an important role in regulating the direction of movement of nutrients also during remobilisation. Strong *NAM* gene expression was also seen in the xylem, but was absent from the phloem across all tissues, suggesting that *NAM* genes may control xylem specific transport processes. It is of note that strong *NAM* gene expression in flag leaves is located in cells with secondary cell wall thickening (mestome sheath, xylem and sclerenchyma) and perhaps *NAM* genes are important in these cells, for example in regulating transport.

NAM expression in stem tissues shares some common features with expression in the leaves. In particular strong expression is seen at 20 DAA in sclerenchyma and in the xylem of outer vascular bundles. These are also lignified tissues, further strengthening our hypothesis that *NAM* genes may be important in cells containing secondary cell walls. The expression of *NAM* genes occurs after anthesis, by which time secondary cell walls have already formed, therefore perhaps *NAM* plays a role in their maintenance or in nutrient movement through these cells. However expression of *NAM* genes is not detected in the xylem of inner vascular bundles in the exposed peduncle or internode 1, indicating functional specialisation of these different vascular bundle types. In leaves it is known that different types of vascular bundles play different roles, with larger bundles (laterals and midrib) transporting 96% of water and relatively little sugar,

whereas smaller intermediate bundles transport most of the sugar and very little water (Kuo et al. 1974). We hypothesise that similarly in stem tissues the outer and inner vascular bundles may play different roles, which would explain why *NAM* expression is different between the two bundle types. In peduncle and internode 1 expression increased greatly from anthesis to 20 DAA and staining was more widespread than in leaf tissues with epidermal and parenchyma cells showing strong staining, however no staining was observed in chlorenchyma. In the enclosed peduncle *NAM* gene expression is seen in all cell types suggesting a unique function in this tissue where the highest level of expression was detected by qPCR. Therefore the role of this tissue in nutrient remobilisation bears further investigation.

2.4.3 Hypothesis 3: Xylem and phloem transport to the ear will be affected by reduction of *NAM* gene expression

NAM RNAi plants had been reported to have altered grain nutrient concentrations (Waters et al. 2009) which we hypothesized to be due to differences in xylem and phloem transport. In particular we noticed that phloem transported nutrients were generally increased or unaltered in RNAi grain, whereas xylem transported nutrients were reduced. Our *in situ* hybridisations supported this hypothesis because *NAM* genes were expressed strongly around the vascular bundles in the mestome sheath indicating a role in nutrient transport. Strong *NAM* expression was observed in the xylem, suggesting that down-regulation in RNAi lines might reduce xylem transport to the grain, whereas *NAM* gene expression was not detected in the phloem suggesting that phloem transported nutrients should be unaffected. We examined the distribution of elements in flag leaf, flag leaf node, peduncle and ears of RNAi and control plants from anthesis to 30 DAA. We found that Zn, Fe and Cu, which are all xylem transported to some degree were reduced in ears of RNAi lines, supporting the hypothesis that *NAM* genes are necessary for xylem transport. However other xylem transported elements such as Mg, Mn and P were not altered, but this may be due to differences only accumulating at later stages of grain filling. Amongst the phloem transported elements K content was increased in the ears of RNAi plants compared to control plants, but Ca content was not affected at the time-points examined. *NAM* genes were not expressed in the phloem directly, but expression was strong in the mestome sheath suggesting that transfer of nutrients to/from the phloem might be influenced by *NAM*. The difference was only seen in K concentration therefore we hypothesise that the effect might be element specific.

To directly test xylem and phloem transport we fed the phloem mobile Rb and xylem mobile Sr to cut wheat tillers. We did not see any difference in the percentage of Rb or Sr that was transported to the ears in RNAi plants compared to controls. This may be because *NAM* genes do not affect xylem and phloem transport or due to limitations to our experimental design. In our experiment the tillers were only immersed in these solutions for 72 hours which may not have been long enough to produce significant differences between genotypes. Additionally we examined the whole of the ear as if it were a homogenous tissue, whereas there are differences in nutrient accumulation between the rachis, palea and lemma and grain, and differences in distribution between these tissues might have masked any genotype specific differences. Further experiments would be required to confirm whether *NAM* genes play a role in xylem and phloem transport, for example feeding Rb and Sr to plants in pots or in hydroponics and examining distribution over a longer time span. Despite the lack of difference in xylem and phloem transport into the ear between RNAi and control plants, RNAi plants accumulated more Rb and Sr in their flag leaves than control plants, which could suggest a change to both xylem and phloem transport away from the flag leaf. However we favour the hypothesis that in *NAM* RNAi plants the remobilisation of many micronutrients from the flag leaf is altered, but this is controlled specifically for each nutrient. This is supported by some micronutrients e.g. Zn, Fe and Cu having their remobilisation reduced from RNAi flag leaves whereas other micronutrients are unaffected e.g. K (Waters et al. 2009). An interesting observation in both genotypes is that in older plants a lower proportion of both Rb and Sr were transported to the ear. This is indicative of reduced grain filling as the plants mature and over the same time period the peduncles accumulated more Rb and Sr, suggesting a transport control point between the peduncle and the ear.

2.4.4 Summary

In summary these results show that *NAM* genes may play a more diverse role in regulating grain nutrient content and senescence than previously thought. *NAM* genes are not only expressed in flag leaves, but they are also expressed in all tissue types examined. Their role in transport of nutrients from the flag leaf to the grain is supported by strong induction of expression after anthesis in peduncle and internode tissues, particularly in vascular bundles in the mestome sheath and xylem. However we have not found any evidence for overarching control of phloem and xylem transport by *NAM* genes. Instead we hypothesize that *NAM* genes regulate a wide range of downstream

genes which individually control nutrient remobilisation, transport and import into the grain. It is exciting to have found *NAM* gene expression in the grain because this indicates that *NAM* genes may also control import of nutrients in addition to remobilisation from source tissues, and suggests a role for *NAM* genes in source-sink relations.

3 Chapter 3 – The effect of *NAM* genes on photosynthesis, yield and carbohydrate accumulation

3.1 Introduction

3.1.1 Senescence and yield

There is a complex relationship between senescence and yield in monocarpic crops such as wheat. A simple hypothesis would be that delaying senescence would increase yield due to the ability of the plant to photosynthesise for a longer time period. However the impact of senescence on yield is modulated by both environmental and physiological effects which are genotype dependent.

3.1.1.1 Stay-green mutants

Stay-green mutants, in which the leaves remain green for longer than in wild types, have been identified in a wide range of species (Hörtensteiner 2009) including cereals such as rice (Jiang et al. 2007), maize (Young et al. 2004), sorghum (Borrell et al. 2000) and wheat (Spano et al. 2003). Stay-green mutants can be divided into two types: cosmetic and functional. Cosmetic stay-green mutants remain green but do not maintain photosynthesis. In these mutants chlorophyll degradation pathways are impaired, causing the green phenotype, but the rest of the photosynthetic machinery is disassembled during senescence, and protein mobilisation has been shown to proceed as normal (Thomas Howard and Stoddart 1975). A second class of stay-green mutants, which include *nam-B1*, are termed functional stay-green mutants. In these plants initiation of senescence is delayed and photosynthesis continues for longer, which might be expected to result in higher grain yield (Thomas Howard and Howarth 2000).

3.1.1.2 Effect of stay-green in sorghum

The majority of studies trying to understand the effect of stay-green on yield have focused on sorghum, in which the stay-green trait is associated with resistance to terminal drought and leads to an increase in yield under drought conditions (Borrell et al. 2014b). In sorghum stay-green lines have been shown to accumulate more nitrogen in their leaves resulting in a higher specific leaf nitrogen (SLN) before anthesis (Borrell and Hammer 2000). It is hypothesised that the high SLN causes several responses: enhanced radiation use efficiency and higher transpiration efficiency, therefore a higher yield potential can be set (Borrell et al. 2003). After anthesis the higher SLN delays the initiation of senescence and reduces the rate of senescence. Higher N uptake from the

soil after anthesis compensates for the reduced remobilisation (Borrell and Hammer 2000). It has also been shown that the stay-green trait not only influences the N economy of the plants but improves water use, which may contribute to the improved drought tolerance. Under water-limited conditions stay-green lines had a reduced number of tillers and smaller upper leaves, therefore the canopy size at anthesis was reduced (Borrell et al. 2014a, Borrell et al. 2014b). This smaller canopy area meant that water was conserved in the soil until anthesis and more water was taken up during grain filling, leading to a higher yield (Borrell et al. 2014a). These results show that changes to N and water usage in stay-green sorghum lead to a yield increase under water-limited conditions.

The majority of studies conducted in sorghum have investigated the effect of stay-green on yield under the naturally occurring drought conditions. For example using elite sorghum lines in low to medium yielding sites (< 6 t ha^{-1}) stay-green was positively associated with grain yield or was not significant, which reflects the majority of Australia's northern grain belt yields of 2.5 t ha^{-1} (Jordan et al. 2012). At sites with medium to high yields (6 to 11 t ha^{-1}), which generally had higher rainfall or deeper soils with higher water holding capacities, the association was evenly divided between positive and negative associations, although the magnitude of positive associations tended to be higher (Jordan et al. 2012). This indicates that under conditions with more optimal water availability stay-green presents a lower or no yield advantage.

The genetics underlying the stay-green also influences its effect on yield. Four individual QTLs for stay-green have been identified in sorghum (*Stg1-4*) which differ in how strongly senescence is delayed and the precise physiological parameters affected (Harris et al. 2007). These differences in phenotype also are seen in yield effects: for example under well-watered conditions *Stg1* gave a yield advantage under both high and low density planting, whereas *Stg2* and *Stg3* increased yield under high density planting but decreased yield under low density planting. *Stg4* increased yield under high density planting and had no effect at low density planting (Borrell et al. 2014a). Overall at high density planting all four *Stg* NILs produce a higher yield under well-watered conditions and the slight yield penalty under low density planting is likely to be due to reduced interception of light due to a smaller canopy size (Borrell et al. 2014a).

In conclusion, in sorghum the stay-green characteristic is associated with higher yield under the terminal drought conditions which elite sorghum cultivars frequently

encounter. However, under well-watered conditions the effect of stay-green is variable, and depends on interactions with the environment (Jordan et al. 2012) and also the genetics controlling the stay-green phenotype (Borrell et al. 2014a). The genes underlying QTLs for the stay-green trait have not been cloned although candidate genes related to flowering time and vernalisation have been identified (Borrell et al. 2014a). The identification of these genes will allow further interpretation of the many physiological studies carried out in this system.

3.1.1.3 Effect of stay-green in wheat

Fewer studies have been carried out into the effects of stay-green on yield in wheat, although the results broadly agree with the findings in sorghum. To investigate the effect of stress on the stay-green effect on yield, two wheat populations (294 advanced lines and 169 recombinant inbred lines from CIMMYT (International Maize and Wheat Improvement Centre)) were grown under drought, heat and drought combined with heat stress. It was found that under stress conditions stay-green had a positive effect on yield but under full irrigation no significant correlations were observed (Lopes and Reynolds 2012). The effect of environment on the stay-green yield advantage is also seen when comparing the stay-green line SeriM82 to the cultivar Hartog (Christopher et al. 2008). The significant yield advantage of SeriM82 depends on the availability of deep soil moisture, and in its absence the yield advantage and the stay-green trait are lost (Christopher et al. 2008). The requirement for deep soil moisture indicates a similar mechanism to sorghum where water uptake is enhanced after anthesis in stay-green lines (Borrell et al. 2014b). Further parallels can be drawn with the leaf canopy area: in sorghum stay-green lines the reduced leaf canopy area is hypothesised to reduce transpiration and conserve water for use during grain filling (Borrell et al. 2014b), and in wheat a 1BL/1RS translocation is associated with smaller flag leaves, stay-green and a higher yield which may be due to improved water usage (Luo P. G. et al. 2009).

In wheat, similar to sorghum, the effect of stay-green on yield depends upon the environment and genotype. For example in the UK and France under high nitrogen application the duration of senescence was not positively associated with grain yield (Gaju et al. 2014). A second study in field conditions in the UK with a stay-green EMS mutant also found that under high nitrogen conditions yield was reduced, whereas under low nitrogen conditions yield was increased compared to wild type plants (Derkx et al. 2012). It is possible that the yield was increased at low nitrogen conditions due to

increased post anthesis nitrogen uptake as has been observed in stay-green sorghum (Borrell et al. 2000).

An alternative route to studying the effect of senescence on yield is to use fungicide application to delay senescence. In a naturally inoculated trial, in which the most prevalent pathogens were *Septoria tritici* and *Puccinia striiformis*, application of three fungicides (triazol, strobilurin and oxazolidinedione) at flag leaf and ear emergence caused significant increases in green flag leaf area duration (GFLAD) (Dimmock and Gooding 2002). Increases to GFLAD were associated with increases in yield due to an extension of the grain filling period. In the cultivar Hereward in the year 2000 the increase in GFLAD occurred much later relative to grain filling and the yield gains were smaller (Dimmock and Gooding 2002), which shows the importance of temporal coordination of senescence and grain filling. Even in disease-free greenhouse conditions fungicide application can enhance yields by delaying senescence (Berdugo et al. 2012) showing that these effects are due to physiological changes within the plant itself, not due to reduced disease pressure.

3.1.1.4 *NAM* gene effects on yield and senescence

The *NAM-B1* gene increases grain protein content (GPC) which is explained in part by enhanced nitrogen remobilisation which is linked to accelerated senescence (Uauy et al. 2006a, Uauy et al. 2006b). Non-functional alleles of *NAM-B1*, which are found in most modern wheat cultivars (Uauy et al. 2006b), are in effect functional stay-green mutations which delay senescence: several studies have investigated whether this delay in senescence leads to an increase in yield. Six NILs in hexaploid wheat cultivars and three NILs in tetraploid wheat cultivars were tested in three locations in California over three years (Brevis and Dubcovsky 2010). In both hexaploid and tetraploid varieties the non-functional allele was associated with lower GPC and delayed senescence, however no significant difference in yield was detected (Brevis and Dubcovsky 2010, Tabbita et al. 2013). Similarly mutations in *NAM-A1* and *NAM-D1* have been characterised (*nam-A1* and *nam-D1*), and both cause a significant delay in senescence without an associated yield increase (Avni et al. 2014). The double mutant *nam-A1*/*nam-D1* showed an even stronger effect on senescence, with total flag leaf senescence occurring 20-30 days after control lines. However, even with this strong delay of senescence in most environments no significant effect on yield was detected (Avni et al. 2014).

Although the lack of yield increase in *NAM* mutants may be disappointing, to wheat breeders aiming to increase GPC (and by association accelerate senescence), the reported lack of yield penalty makes it an attractive target for marker assisted selection (MAS). *NAM-B1* was introgressed into Indian varieties using MAS in combination with field phenotyping (Kumar et al. 2011). Significant differences in GPC and yield were found between genotypes and environments, and significant genotype-by-environment interactions were present, indicating that the effect of *NAM-B1* needs to be evaluated in each genotype and environment (Kumar et al. 2011), as was reported previously (Brevis and Dubcovsky 2010). The use of phenotyping in combination with MAS enabled the selection of seven lines with significantly higher GPC (14.8 – 17.9 %) which had no yield penalty compared to controls (Kumar et al. 2011). These results suggest that although altering GPC is linked to altering leaf senescence, the negative relationship to yield can be broken by selection.

3.1.1.5 Source-sink relationships

Crop yield is determined by both physiological and environmental factors. Amongst physiological factors a key determinant of yield is the source-sink relationship. This dynamic interaction depends upon the capacity of source tissues (e.g. leaves) to photosynthesise and fix atmospheric CO_2 and on the grain's ability to convert the fixed carbon into grain mass, mainly in the form of starch. The relative importance source or sink capacity, or co-limitation by both, is influenced by both genetic and environmental factors (Serrago et al. 2013, Sweetlove and Hill 2000, Wang Z. et al. 1997).

Photosynthesis depends on light levels, temperature and CO_2 concentration in the atmosphere (Farquhar et al. 1980). At low light levels photosynthetic rate is limited (Smith 1938) and the photosynthetic rate is proportionate to the activity of Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Perchorowicz et al. 1981). At high light levels photosynthetic rate is limited by CO_2 concentration and the capacity of the photosynthetic apparatus, with increased levels of CO_2 permitting higher photosynthetic rates (Smith 1938, Stitt 1991). Temperature also has a strong effect on photosynthesis, for example in wheat seedlings the photosynthetic rate increased three-fold from 0 to ~22 °C and then declined three-fold as the temperature was increased to 40 °C (Jolliffe and Tregunna 1968).

Under conditions when photosynthesis is not limiting, the ability of the sink organs to utilise photosynthate can limit yield. The capacity is determined by physical constraints

such as grain size and the sink strength which depends on the ability of the plant to unload and retrieve photosynthates from the phloem, and convert these photosynthates into storable carbohydrates in the sink (Tuncel and Okita 2013).

3.1.1.6 Starch synthesis in developing grain

Photosynthates are delivered in the form of sucrose via the phloem to the developing grain. Sucrose is unloaded from the phloem in the vascular bundle into the apoplast by specialized transfer cells. Sucrose is then taken up by transfer cells in the aleurone layer and directed into the endosperm cells (Thorne 1985). The enzymatic steps to convert sucrose into starch are outlined in **Figure 3.1**. Sucrose synthase catalyses the conversion of sucrose into fructose and UDP glucose (UDP-Glc) and is generally considered to be the first step towards starch synthesis. Subsequently UDP-Glc is converted into glucose-1-phospate (G-1-P). The conversion of G-1-P into ADP glucose (ADP-Glc) is catalysed by ADP glucose pyrophosphorylase (AGPase). The cytosolic form of AGPase accounts for 85-95 % of total AGPase activity (Denyer et al. 1996, Thorbjornsen et al. 1996) and the activity of AGPase is believed to exercise strong control over the rate of starch synthesis in wheat grains (Tuncel and Okita 2013). Evidence includes experiments in which the over-expression of cytosolic AGPase, particularly forms insensitive to allosteric regulation by inorganic phosphate, led to seed weight increases of 13-25 % in maize (Wang Zhangying et al. 2007), 13-22 % in wheat (Meyer Fletcher D. et al. 2004) and 23% in rice (Smidansky et al. 2003) indicating that AGPase activity limits yield.

Starch comprises two polymers: amylose and amylopectin. Amylose is a linear molecule composed of glucosyl monomers joined together by α-1,4 linkages. Amylopectin has a branched structure containing variable length chains of linear α-1,4 linked glucosyl monomers, with 1,6 linkages at branch points (James et al. 2003). During starch synthesis the ADP-Glc produced by AGPase is the glucosyl donor for the elongation of the α-1,4-glucosidic chains of both amylose and amylopectin. Starch synthase, which catalyses this elongation reaction, is hypothesised to exert strong control over the rate of starch synthesis in wheat grains, in conjuction with AGPase. This is supported by evidence that starch synthase activity is only just sufficient to support the rate of starch synthesis throughout wheat grain filling, and the rate of starch synthesis declines in parallel with starch synthase activity during grain maturation (Hawker and Jenner 1993, Yang J. C. et al. 2004). Furthermore at high temperatures grain weight is decreased mainly due to the reduction of starch synthase activity, because starch synthase is very sensitive to temperature (Keeling et al. 1993).

In wheat there are at least five isoforms of starch synthase. Four isoforms of soluble starch synthases have distinct roles in amylopectin synthesis for example affecting chain lengths (McMaugh et al. 2014). The granule-bound isoform of starch synthase (GBSSI) is encoded by the *Waxy* gene in cereals and is required to make amylose in the grain (Nelson and Rines 1962, Shure et al. 1983), and in wheat a second isoform GBSSII is present in non-storage tissues (Vrinten and Nakamura 2000). Branching and de-branching enzymes re-organise amylopectin structure by adding or cleaving α-1,6 glycosidic bonds to determine the final structure of starch (James et al. 2003).

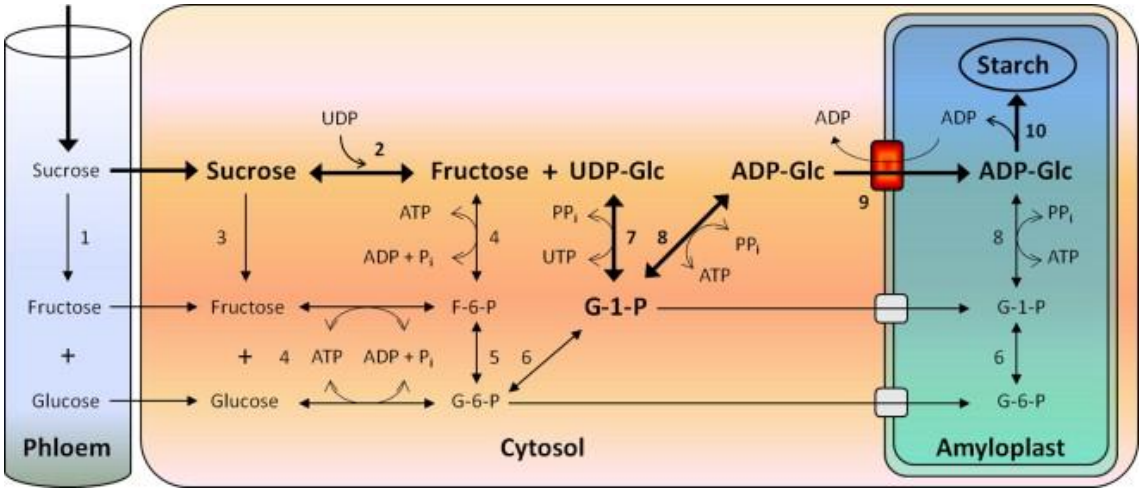


Figure 3.1. Outline of pathway for sucrose unloading and conversion to starch in the endosperm. The sucrose transported from leaves is unloaded at the sinks and converted to ADP-glucose (ADP-Glc) in the cytosol. ADP-Glc is then transported into the amyloplast and utilised by starch synthases for biosynthesis of storage starch. The primary pathway reactions from sucrose to starch in sink tissues are shown in bold arrows. Enzymes catalyzing these reactions are: 1, cell wall invertase; 2, sucrose synthase; 3, cytosolic invertase; 4, hexokinase; 5, phosphoglucoisomerase; 6, phosphoglucomutase; 7, UGPase; 8, AGPase; 9, ADP-Glc transporter; and 10, starch synthases. Figure from (Tuncel and Okita 2013).

3.1.1.7 Water soluble carbohydrates

During grain filling two sources of photosynthate are used: either directly from actively photosynthesising organs or remobilised from temporary storage pools. During the diurnal cycle sucrose is accumulated in the leaves (Jenner C. F. and Rathjen 1972) and supplies carbon to the grain during the night when photosynthesis is not possible. Over the longer term a temporary storage pool of water soluble carbohydrates (WSC) contribute to grain filling. WSC accumulate in the stems and leaf sheaths until approximately 3 weeks after anthesis (Schnyder 1993) and are remobilised during the latter stages of grain filling. This pool of WSC can act as a buffer to store excess

photosynthate at early stages of grain filling when demand is low, and to supplement grain filling when demand is higher than current photosynthesis, for example during senescence or drought stress (Kühbauch and Thome 1989, Yang J. C. et al. 2000).

Most WSC in wheat stem internodes is in the form of fructan (85 %), sucrose constitutes about 10 % of WSC and the rest is in the form of fructose and glucose (Blacklow et al. 1984). Fructans are synthesised in the vacuole by fructosyltransferases which transfer fructose from sucrose onto the growing fructan chain (Vijn and Smeekens 1999). Fructans vary greatly in chain length and have a great structural diversity, but those in the wheat stem mainly contain fewer than 10 hexose units per molecule (Blacklow et al. 1984) and are termed low molecular weight fructans. During remobilisation of stem reserves, fructan exohydrolases release the terminal fructosyl units using water as an acceptor (Valluru and Van den Ende 2008). This released fructose can be converted to sucrose and transported to the developing grain.

3.1.2 Aims and hypotheses

As described above, many studies have examined the relationship between senescence and yield particularly in sorghum. However, relatively few studies have simultaneously characterised sink capacity alongside delayed senescence. Therefore we will carry out an integrated study which examines both the source capacity, by measuring photosynthetic rates, and the sink capacity by examining grain development and enzyme activities involved in starch synthesis. To try to understand the effect of delayed senescence on grain filling we will use *NAM* RNAi lines in which *NAM-B1,* and its homoeologues and paralogues, are down-regulated. These results will shed light onto the effect of *NAM* genes on source-sink relationships. We hypothesise that:

3.1.2.1 Hypothesis 1: The delay in senescence in *NAM* RNAi plants will lead to a yield increase compared to control plants

The delay in senescence in *NAM* RNAi plants is of approximately 3 weeks which we would expect to lead to a large increase in total carbon fixation. We hypothesise that the extra photosynthate may contribute to an increase in yield.

3.1.2.2 Hypothesis 2: Grain maturation in *NAM* RNAi plants is coupled to leaf senescence

Although the flag leaf senescence has been characterised in *NAM* RNAi plants, the grain development has not been well studied. We hypothesise that since *NAM* genes influence whole plant senescence they will also affect the development of the grain.

3.1.2.3 Hypothesis 3: Carbon fixed during the stay-green period in *NAM* RNAi plants will be stored in stem tissues

Previous studies have indicated that delayed senescence due to *NAM* genes generally does not lead to a yield increase (Avni et al. 2014, Brevis and Dubcovsky 2010). If this is also the case in *NAM* RNAi plants we will try to discover the fate of the extra photosynthate. WSC are an important temporary store of carbon during wheat grain filling. We hypothesise that any excess photosynthate fixed during the stay-green period in *NAM* RNAi plants will be stored as WSC in the stems.

3.2 Methods

3.2.1 Plant materials

3.2.1.1 Plants for chlorophyll, yield, ambient photosynthesis measurements, grain moisture and enzyme assays (Batch 1)

Transgenic wheat plants containing a RNA interference (RNAi) construct which down-regulates the expression of *NAM-B1* homologues (*NAM* RNAi) and null-segregants (control) plants were germinated on moist filter paper in petri dishes in the dark at 4 °C for 48 h. They were then incubated for 48 h in the light at room temperature and sown into P40 trays containing peat and sand (see 2.2.1.1). At 2-3 leaf stage they were potted on into 1 L pots containing cereal mix (see 2.2.1.1). Plants were grown in a controlled environment room with 16 h light at 20 °C and 8 h dark at 15 °C with 70 % humidity. Light levels were 300 µmol m^{-2} s^{-1}. Grain from the central third of the ear were dissected from the ears of the primary tiller using tweezers. Only the first and second grain from each floret were used. Grain for moisture content and dry mass were immediately placed in Eppendorf tubes and weighed for fresh weight, while grain for enzyme assays were frozen in liquid nitrogen and stored at -80 °C. Whole ear samples at 10 and 20 days after anthesis DAA were frozen in liquid nitrogen and stored at -80 °C. The required grains were dissected out on dry ice prior to extraction for enzyme assays.

3.2.1.2 Plants for A-C_i Curves (Batch 2) and plants for stem carbohydrates (Batch 3)

Separate batches of plants were used for A-C_i curves and for stem carbohydrates. In both cases *NAM* RNAi and control plants were germinated and potted as described above (3.2.1.1). These plants were grown in a controlled glasshouse with 16 h supplementary lighting (275 µmol m^{-2} s^{-1}) at 20 °C and 8 h dark at 12 °C. For stem carbohydrate measurements the second tillers were harvested at 10, 35 and 77 days after anthesis (DAA). Tillers were cut into the separate regions (exposed peduncle, enclosed peduncle removed from the leaf sheath, internode 1 and internode 2). These regions were further subdivided into 2-5 mm sections to help the grinding process, and the small sections were put into 2 technical replicate Eppendorf tubes, frozen in liquid nitrogen and stored at -80 °C.

3.2.2 Grain yield parameters

3.2.2.1 Grain yield data

The number of tillers per plant was counted and each plant was separated into main tillers (including secondary tillers) and back-tillers. The ears from the main tillers were threshed by hand. The back-tillers were threshed separately. Grain size, number and masses were measured using the MARVIN-universal grain analyser (GTA Sensorik GmbH). Grain mass per plant and tiller number include both main and back-tillers. Thousand grain weight, grain width, length and area only include grain from main tillers because back-tillers were often poorly developed with small grain.

3.2.3 Photosynthesis measurements

3.2.3.1 Chlorophyll measurements

Flag leaf chlorophyll levels were measured using a chlorophyll meter (SPAD502-Plus, Konica Minolta). The meter was clamped onto the primary tiller's flag leaf at 5 equally spaced intervals on each side of the mid-vein, a total of 10 measurements per leaf were recorded and the mean value is reported.

3.2.3.2 A-C_i curves

The photosynthetic response of the primary tiller's flag leaf to CO_2 concentration (A-C_i curve) was measured with a Li6400 portable photosynthesis system (LI-COR Biosciences). The ambient CO_2 concentration in the cuvette (C_a) was reduced from 380 to 0 µmol CO_2 mol^{-1} and then increased from 0 to 50, 100, 200, 380, 600, 800, 1000 and 1200 µmol CO_2 mol^{-1} at a constant photosynthetically active radiation (PAR) of 1000 µmol m^{-2} s^{-1}. The leaf temperature was set at 20 °C. The gas exchange properties including photosynthesis rate and C_i (internal leaf CO_2 concentration) were logged at each C_a once the system had reached a pre-determined stability point (coefficient of variation ≤ 2 %). The cuvette used had a size of 6 cm^2, in cases where the flag leaf was too narrow to cover the whole area, the flag leaf area was measured and the value was inputted into the Li6400 before measurements began.

3.2.3.3 A-C_i curve fitting model

The equation $A = a(1 - e^{(-bCi)} - c)$ was fitted to the A-C_i data using least squares (Lawlor 1987, McCormick et al. 2006), where A = photosynthetic rate, a = J_{max}, b = curvature parameter, c = dark respiration. The slope of the curve approaches zero as a result of substrate limitation (ribulose-1, 5 bisphosphate) and this is equivalent to the CO_2- and light-saturated photosynthetic rate (J_{max}). To calculate the efficiency of

RuBisCO-mediated carboxylation (V_{cmax}) a linear regression was performed on the data between a C_i of 0 and 150 µmol mol^{-1}.

3.2.3.4 Ambient photosynthesis measurements

The rate of photosynthesis of the primary tiller's flag leaf under ambient conditions was measured using a Li6400 portable photosynthesis system (LI-COR Biosciences). The cuvette conditions were C_a at atmospheric level (400 µmol CO_2 mol^{-1}) and a constant PAR of 300 µmol m^{-2} s^{-1}. The leaf temperature was set at 20 °C. The gas exchange properties including photosynthesis rate and C_i (internal leaf CO_2 concentration) were logged once the system had reached a pre-determined stability point (coefficient of variation ≤ 2 %), usually after ~5 min. The chamber area was adjusted for flag leaf size as described above (3.2.3.2).

3.2.3.5 Calculation of glucose fixed during stay-green period

We estimated the amount of extra CO_2 which would be fixed by *NAM* RNAi plants compared to control plants during the stay-green period. We assumed that the only difference between the two genotypes was that the *NAM* RNAi plants had an extra 10 day period of photosynthesis; all photosynthesis before this period, and during the period of senescence were considered identical between genotypes. We assumed that in *NAM* RNAi plants during the extra 10 days of photosynthesis, the rate of photosynthesis was a constant 8.6 µmol CO_2 m^{-2} s^{-1} (this was the average of the 20 and 30 DAA time-points for *NAM* RNAi plants). The average flag leaf area was 0.004025 m^2 (40.25 cm^2), and was not significantly different between genotypes.

During the 10 days with 16 h light (equivalent to 576,000 sec of light): 8.6 µmol CO_2 m^{-2} s^{-1} x 576000 s x 0.004025 m^2 = 19,938 µmol CO_2 were fixed per flag leaf. Assuming that photorespiration decreases the efficiency of CO_2 fixation by 20 %, 7.2 mol CO_2 are required to make 1 mol glucose. Therefore 19,938 µmol CO_2 is equivalent to 2769 µmol glucose. The M_r of glucose is 180 therefore 498 mg glucose equivalent were fixed per flag leaf. The average number of tillers per plant was 4.17, therefore the flag leaves of each *NAM* RNAi plants leaves fixed 2,079 mg glucose more than wild type plants.

3.2.4 Grain development

3.2.4.1 Grain moisture and dry mass

Eight grain from the central portion of the ear of the primary tiller were dissected out and weighed to obtain the fresh weight. The grains were put into Eppendorf tubes and incubated with their lids open in a 60 °C oven for 72 h. Grains were weighed to obtain dry weight. Percentage grain moisture was calculated by:

Grain moisture content = (fresh weight – dry weight) / fresh weight x 100

3.2.4.2 Enzyme assays

3.2.4.2.1 Extracts

The extraction and assay of AGPase activity was carried out by Brendan Fahy in Alison Smith's lab at the JIC. Ten frozen grain from the central part of the ear of the primary tiller were weighed and the required number of grain to obtain an extract with a tissue:extract ratio of ~ 350 mg/ml were homogenized using a 7 mm stainless steel ball in a QIAGEN Retsch ball mill in 0.5 ml extraction buffer (100 mM 3-(N-morpholino)propanesulfonic acid (MOPS) pH 7.2, 5 mM $MgCl_2$, 5 % (v/v) glycerol, 5 mM dithiothreitol (DTT), 10 mg/ml bovine serum albumin (BSA) and 1 % (w/v) polyvinylpolypyrrolidone (PVPP)). After 2 min homogenisation (28 shakes s^{-1}) any additional extraction buffer required to reach the desired concentration was added. Extracts were centrifuged at 10,000 x g for 5 min at 4 °C. 100 µL aliquots of supernatant were snap frozen in liquid nitrogen.

3.2.4.2.2 AGPase assay

AGPase assays were carried out by Brendan Fahy using the method of (Rösti et al. 2006). This assay was adapted for microtitre plate and contained 185 µl of assay mix in each well (100 mM MOPS/NaOH pH 7.2, 2.5 mM $MgCl_2$, 2 mM adenosine diphosphate glucose (ADP glucose), 2 mM nicotinamide adenine dinucleotide (NAD), 1 µl (1 U) glucose-6-phosphate dehydrogenase (10 165 875 001, Roche), 0.6 µl (1 U) phosphoglucomutase (P3397 Sigma))

3.2.4.2.3 Starch synthase assay

The starch synthase assays were carried out by Brendan Fahy and measured the incorporation of ^{14}C from ADP[^{14}C]glucose into glucans. It was carried out according to the Dowex method (Jenner CF et al. 1994).

3.2.5 Stem carbohydrates

3.2.5.1 Fructan

Samples for fructan assay (1 technical replicate from each plant tissue) were ground using 3 mm stainless steel balls in a QIAGEN Retsch ball mill. The stainless steel balls were removed and the samples were weighed to obtain fresh weight. The samples were kept frozen in liquid nitrogen until use.

The fructan assay was carried out using the Megazyme fructan assay (K-Fruc 03/14, Megazyme) according to the manufacturer's instructions, except that all steps were scaled down to fit in 1.2 ml deep-well plates or microtitre plates.

3.2.5.2 Starch assays

Starch was extracted from all stem tissues using the perchloric method (Hargreaves and Aprees 1988). Starch was then enzymatically digested and the glucose within the starch was measured.

3.2.5.2.1 Starch extraction

Samples for starch assay (1 technical replicate from each plant tissue) were ground using 3 mm stainless steel balls in a QIAGEN Retsch ball mill. The stainless steel balls were removed and the samples were weighed to obtain fresh weight. The samples were kept frozen in liquid nitrogen until use.

Ground samples were mixed vigorously with 1 ml 0.7 M perchloric acid. 400 µl aliquots were transferred to two replicate 1.2 ml plates (ThermoScientific) and 300 µl 100 % ethanol was added. Plates were centrifuged for 6 min at 4,000 x g to pellet the starch. The pellets were washed in 600 µl 80 % ethanol. Samples were centrifuged for 6 min at 4,000 x g and supernatants were discarded. Pellets were gelatinised in 150 µl H_2O at 95 °C for 15 min and then cooled on ice. 600 µl 100% ethanol was added, pellets were re-suspended and plates were centrifuged for 6 min at 4,000 x g. The supernatants were discarded, the pellets were air-dried and re-suspended in 300 µl dH_2O.

3.2.5.2.2 Starch digestion

290 µl of 0.2 M sodium acetate pH 4.8 was added to each well. To samples in one plate, 10 µl of a 9:1 (v/v) mixture of amyloglucosidase:α-amylase was added, to the other plate 10 µl water was added. Samples were incubated overnight at 37 °C and then used for glucose measurements.

3.2.5.2.3 Glucose assay

Samples were mixed and centrifuged for 10 min at 4,000 x g. Glucose in the supernatant was assayed enzymatically (Hargreaves and Aprees 1988). All the glucose was converted to glucose 6-phosphate by hexokinase and a blank reading was taken at 340 nm using a plate reader (SPECTRAmax 340PC, Molecular Devices). Glucose 6-phosphate dehydrogenase was added to each sample in the presence of NAD^+. This led to the production of NADH, which was detected by absorbance at 340 nm in a plate reader, at a 1:1 ratio to the original glucose in the sample.

3.2.5.3 Fructose, glucose and sucrose

50 µl of supernatant from the fructan extracts was transferred into 2 replicated 96 well PCR plates and 25 µl of 0.2 M sodium acetate pH 4.8 was added to each well. β-fructosidase (0.625 µl of 10 mg/ml solution) was added to each well of one plate, while the same volume of water was added to the other plate. The plates were incubated at 37 °C overnight. Glucose levels were measured as described above (3.2.5.2.3) and fructose levels were measured by a further enzymatic step. Phosphoglucoisomerase was added to the extracts after the NADH production by glucose 6-phosphate dehydrogenase had reached completion. The production of NADH during this step is due to the conversion of fructose 6-phosphate to glucose 6-phosphate which then reacts with NAD^+ (catalysed by glucose 6-phosphate dehydrogenase) to produce more NADH and a further absorbance change. The level of sucrose was measured by comparing the plates incubated with β-fructosidase to those without – the difference between the two is due to the cleavage of sucrose by this enzyme to release glucose and fructose.

3.3 Results

3.3.1 Hypothesis 1: The delay in senescence in *NAM* RNAi plants will lead to a yield increase compared to control plants

Plants with reduced expression of *NAM-B1* homoeologues and paralogues (*NAM* RNAi) show delayed senescence of approximately 3 weeks (Uauy et al. 2006b) and we observed the same delay in senescence in our glasshouse experiments (**Figure 3.2**). We noticed that the precise timing of the visual senescence in each genotype was environmentally dependent with control plant senescence being visualised between 20-40 days after anthesis (DAA) and *NAM* RNAi plant senescence beginning between 30-50 DAA. This wide variation in timing of senescence presented a challenge for comparisons of different batches of plants. Therefore we used the same batch of plants (batch 1), grown in a controlled environment room, for all the following experiments, except for the A-C_i curves (batch 2) and stem carbohydrates (batch 3).

Figure 3.2. *NAM* RNAi (left) and control (right) 80 days after anthesis (DAA) (batch 3).

3.3.1.1 Chlorophyll levels in flag leaves

As an indication of senescence we measured the relative chlorophyll content of flag leaves (batch 1 plants) after anthesis using a SPAD meter (**Figure 3.3**). At 10 DAA there was no difference in chlorophyll content between *NAM* RNAi and control flag leaves. From 20 DAA onwards the control plants started to senesce rapidly, reaching minimum chlorophyll levels by 35 DAA. However the *NAM* RNAi plants maintained a high level of chlorophyll until 30 DAA when they also started to senesce at a similar rate to the control plants reaching minimum chlorophyll levels at 45 DAA. We define the stay-green period in RNAi plants as the period from 20 to 30 DAA, when chlorophyll levels in RNAi plants remain high, but the control plants are senescing. The stay-green period is defined to end at the point when the RNAi plants begin to senesce (30 DAA onwards).

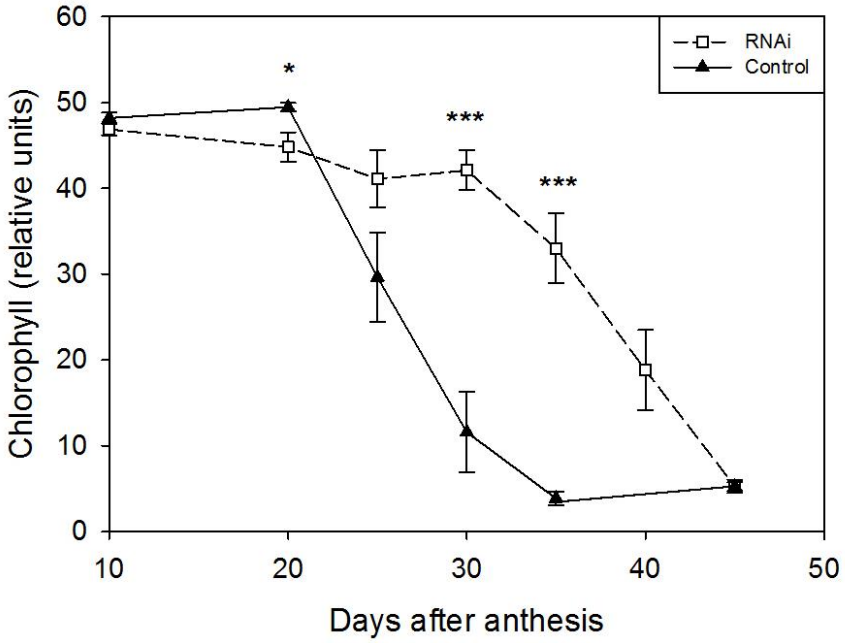


Figure 3.3. Flag leaf chlorophyll content. n = 4-6. Error bars are standard error of the mean (SEM). Asterisks denote significant differences at $p < 0.05$ (*) and $p < 0.001$ (***) as determined by Student t-test between genotypes at each time-point.

3.3.1.2 Photosynthesis of stay-green flag leaves

Although chlorophyll content is an indicator of senescence we wanted to examine the photosynthetic properties of the *NAM* RNAi plants during the period of delayed senescence. It had previously been shown that the maximum quantum efficiency of photosystem II (F_v/F_m) is over 7-fold higher in *NAM* RNAi plants than control plants at the point when control plants are in the middle of senescence (chlorophyll level = ~20 SPAD units) and *NAM* RNAi plants are still green (chlorophyll level = ~47 SPAD units) (Uauy et al. 2006b). The *NAM* RNAi F_v/F_m of 7.7 is close to the range of plants under non-stress conditions (Kitajima and Butler 1975, Maxwell and Johnson 2000) indicating that photosynthesis is functioning efficiently. However F_v/F_m does not directly measure the rate of carbon fixation, which is also affected by other processes such as photorespiration. Therefore we measured A-C_i curves (carbon assimilation rate against internal leaf CO_2 concentration) which give direct information about the biochemical processes happening in photosynthesis.

3.3.1.2.1 Maximum photosynthetic rates

Using batch 2 plants, which had a less rapid rate of senescence due to being grown in different growth conditions (**Figure 3.4**A), we examined in more detail the carbon fixation capacities of *NAM* RNAi and control plants. We found that in control plants during the initial phases of senescence when chlorophyll levels fell from 45 to 25, two key determinants of photosynthetic capacity are dramatically reduced. The maximum rate of RuBP regeneration (J_{max}) falls 4-fold from 20 to 5 µmol m^{-2} s^{-1} (**Figure 3.4**C) and the efficiency of RuBisCO-mediated carboxylation (V_{cmax}) falls from 50 to 12 µmol m^{-2} s^{-1} (**Figure 3.4**E) in control plants. During this same time period in *NAM* RNAi plants J_{max} and V_{cmax} fell much less; from 18 to 13 µmol m^{-2} s^{-1} and 48 to 35 µmol m^{-2} s^{-1} respectively. These parameters are calculated from the A-C_i curves at three time-points (**Figure 3.4**B, D and F). At 23 DAA control and *NAM* RNAi plants show very similar increases in their photosynthesis at higher levels of CO_2. However at 41 DAA the control plants show much less response to elevated CO_2 compared to *NAM* RNAi plants. It was not possible to measure the photosynthetic rate of control plants at the third time-point because the plants were fully senesced therefore not carrying out photosynthesis. At 51 DAA *NAM* RNAi plants maintain a similar ability to respond to CO_2 concentration as at 41 DAA and no change in J_{max} or V_{cmax} was observed.

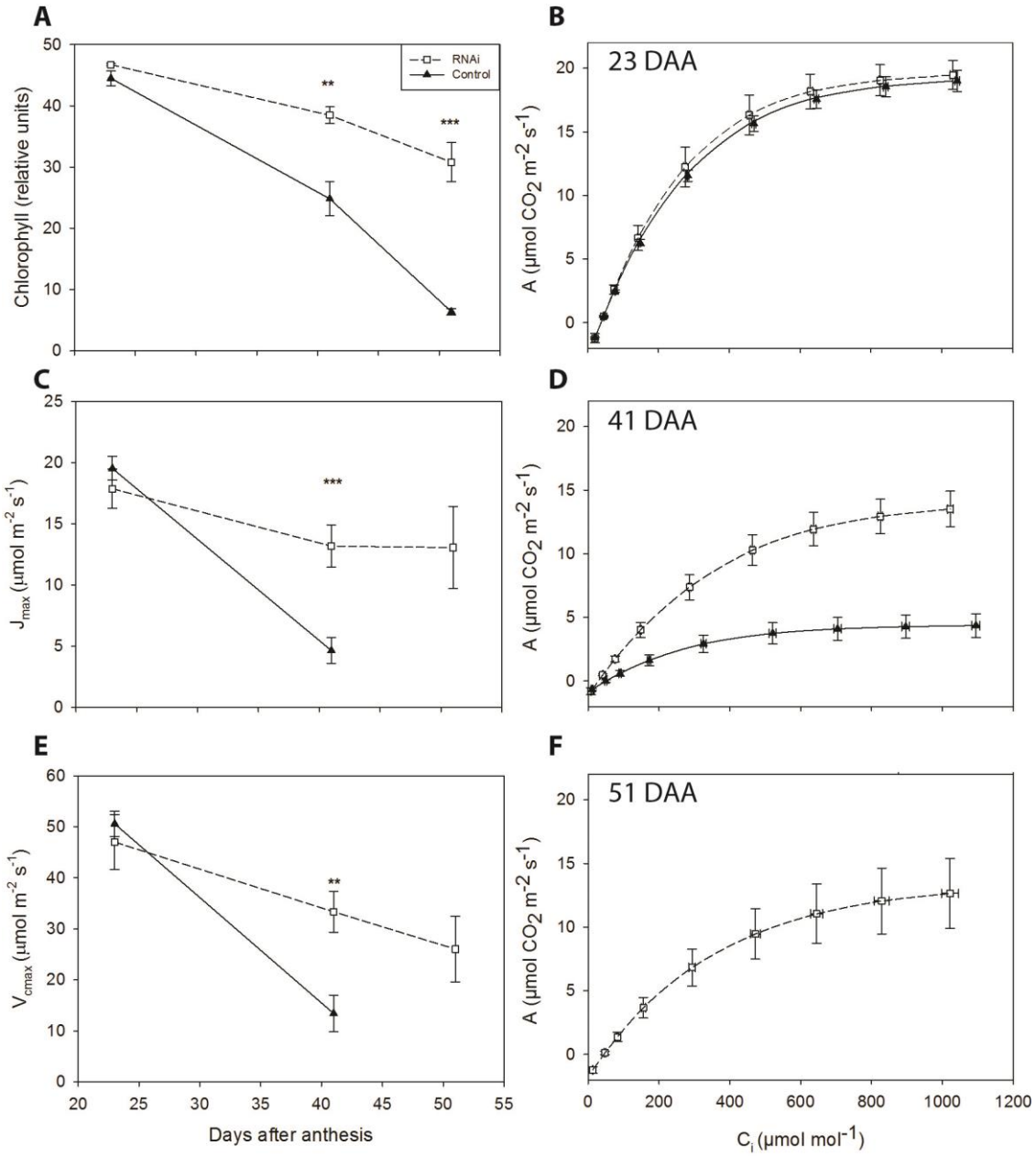


Figure 3.4. Effects of senescence on photosynthetic rate. A) Chlorophyll content, B) A-C_i curve 23 DAA, C) J_{max} (maximum rate of RuBP regeneration), D) A-C_i curve 41 DAA, E) V_{cmax} (maximum rate of RuBisCO-mediated carboxylation), F) A-C_i curve 51 DAA. n = 3-6. Error bars are SEM. Asterisks denote significant differences between genotypes at $p < 0.01$ (**), $p < 0.001$ (***).

3.3.1.2.2 Photosynthesis under ambient conditions

Although A-C_i curves provide information about maximum photosynthetic capacities we also wanted to investigate photosynthetic rates in the conditions in which the plants were growing. Therefore we measured the rate of photosynthesis at ambient CO_2 concentration (400 ppm) and ambient light levels (300 µmol m^{-2} s^{-1}). We used batch 1 plants, the same plants as for the chlorophyll measurements (**Figure 3.3**), to allow comparisons to be made between these data sets. We found that *NAM* RNAi and control flag leaves have the same rate of photosynthesis at 10 and 20 DAA (**Figure 3.5**). From 20 DAA the photosynthetic rate drops dramatically in control plants, whereas photosynthesis is maintained at a high rate until 30 DAA in *NAM* RNAi plants. Photosynthetic rate drops rapidly from 30 to 35 DAA in *NAM* RNAi plants, but by this point photosynthesis is extremely low (or absent) in control plants. *NAM* RNAi plants maintain a low level of photosynthesis from 35 to 40 DAA. These photosynthetic rates mirror changes in the chlorophyll levels measured in flag leaves (**Figure 3.3**) and show that carbon fixation is continuing during the stay-green period in the *NAM* RNAi plants.

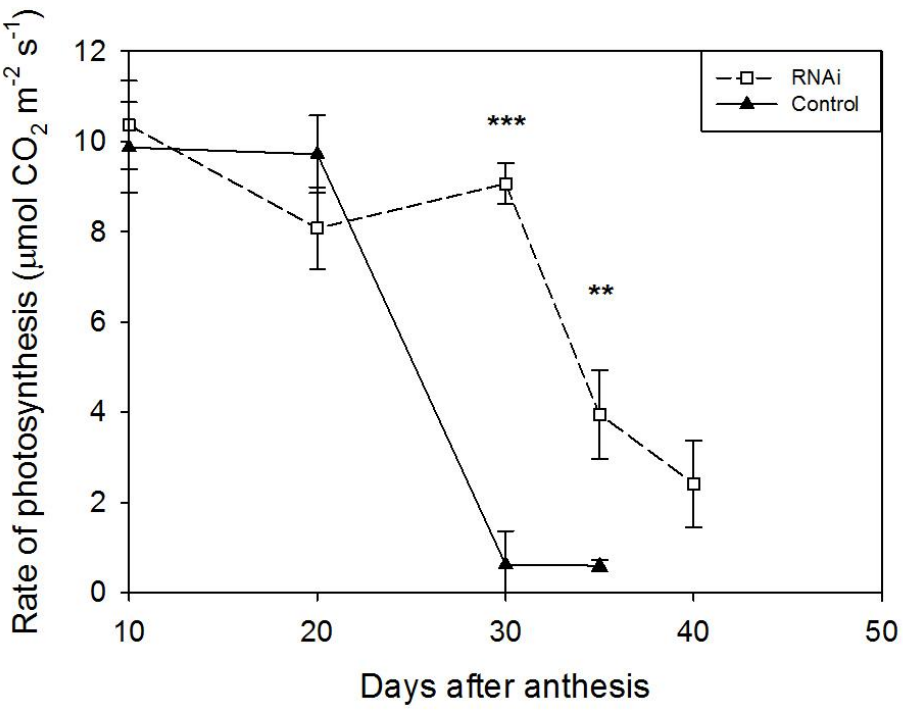


Figure 3.5. Photosynthetic rate at ambient CO_2 concentration (400 ppm). n = 4-6. Error bars are SEM. Asterisks denote significant differences at $p < 0.01$ (**) and $p < 0.001$ (***) as determined by Student t-test between genotypes at each time-point.

We estimated the amount of extra CO_2 which would be fixed by *NAM* RNAi plants compared to control plants during the stay-green period. We assumed that the only difference between the two genotypes was that the *NAM* RNAi plants had an extra 10 day period of photosynthesis. All photosynthesis before this period and during the period of senescence was considered identical between genotypes based on the A-Ci curves and photosynthetic rates at ambient CO_2 concentrations. We assumed that in *NAM* RNAi plants during the extra 10 days of photosynthesis, the rate of photosynthesis was a constant 8.6 µmol CO_2 m^{-2} s^{-1} (this was the average of the 20 and 30 DAA time-points). We measured the average flag leaf area as 0.004025 m^2 (40.25 cm^2). Therefore we estimate that during these extra 10 days ~498 mg glucose was produced by CO_2 fixation per flag leaf. Each plant had on average 4.17 tillers implying that the flag leaves alone in *NAM* RNAi plants produced 2,079 mg glucose more than control plants.

3.3.1.3 Effect of delayed senescence on grain mass

We calculated that the extra 10 days of photosynthesis in *NAM* RNAi plants' flag leaves, compared to control plants, could supply an additional 2,079 mg glucose in total to the ears. However we detected no significant different in grain mass per plant (**Figure 3.6**A) between the *NAM* RNAi plants and the control plants. The thousand grain weight (**Figure 3.6**B) and the number of tillers per plant (**Figure 3.6**C) were the same in both genotypes. The grain area and length were the same in both genotypes (**Figure 3.6**D and F). The *NAM* RNAi plants showed a slightly greater grain width (3.58 mm) compared to control plants (3.49 mm) (**Figure 3.6**E), but this did not translate into increased grain mass or thousand grain weight.

In conclusion, we estimate that the ten day extension of photosynthesis could produce an extra 2 g glucose per plant. However no significant increase in grain mass was observed in RNAi plants compared to controls.

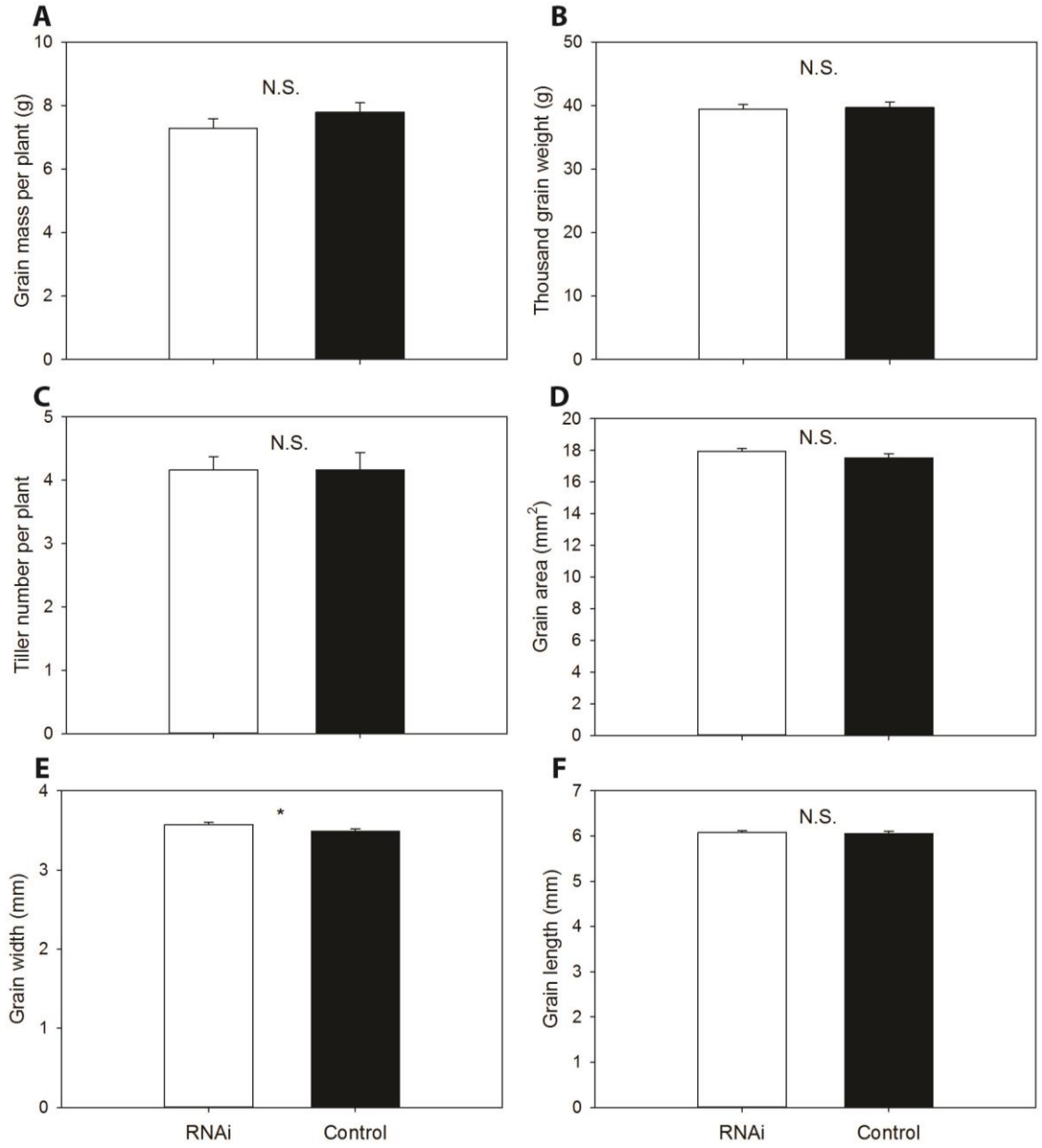


Figure 3.6. Grain yield data at maturity. A) Grain mass per plant, B) Thousand grain weight, C) Tiller number per plant, D) Grain area, E) Grain width, F) Grain length. n = 12. Asterisk denotes significant differences at $p < 0.05$ (*), N.S. denotes non-significant. Genotypes were compared using Student t-tests.

3.3.2 Hypothesis 2: Grain maturation in *NAM* RNAi plants is coupled to leaf senescence

Despite 10 days of extra photosynthesis we found no increase in grain mass in *NAM* RNAi plants. Therefore, we proposed to study the grain development to try to understand why this might be the case. Monocarpic senescence is generally considered to be a co-ordinated whole-plant process. This led to the hypothesis that grain development in *NAM* RNAi lines would be linked to the delayed senescence and therefore would be delayed. However the inability of the plants to convert the extra photosynthate from the stay-green period into grain mass suggests that grain development is not coupled to the delayed leaf senescence.

3.3.2.1 Grain moisture content

To examine the rate of grain development we first measured grain moisture content. Grain moisture content can be used to estimate physiological maturity (maximum dry weight) because this point is usually reached at a moisture content of ~37% (Calderini Daniel F. et al. 2000).

We found that towards the end of grain development the *NAM* RNAi grain have a slightly, but significantly, higher moisture content than the control grain (**Figure 3.7**A). The control grains reach physiological maturity at 40 DAA, whereas the *NAM* RNAi plants reach this point at 43 DAA. These timings correspond well to the attainment of maximum dry weight (**Figure 3.7**B). Although the *NAM* RNAi grains reach physiological maturity 3 days later than the control plants, this is not as large a difference as we had expected from leaf senescence difference of 10 days (the stay-green period). This shows that in *NAM* RNAi plants the delay in leaf senescence does not cause an equivalent change to grain maturation suggesting a disconnect between source and sink tissue.

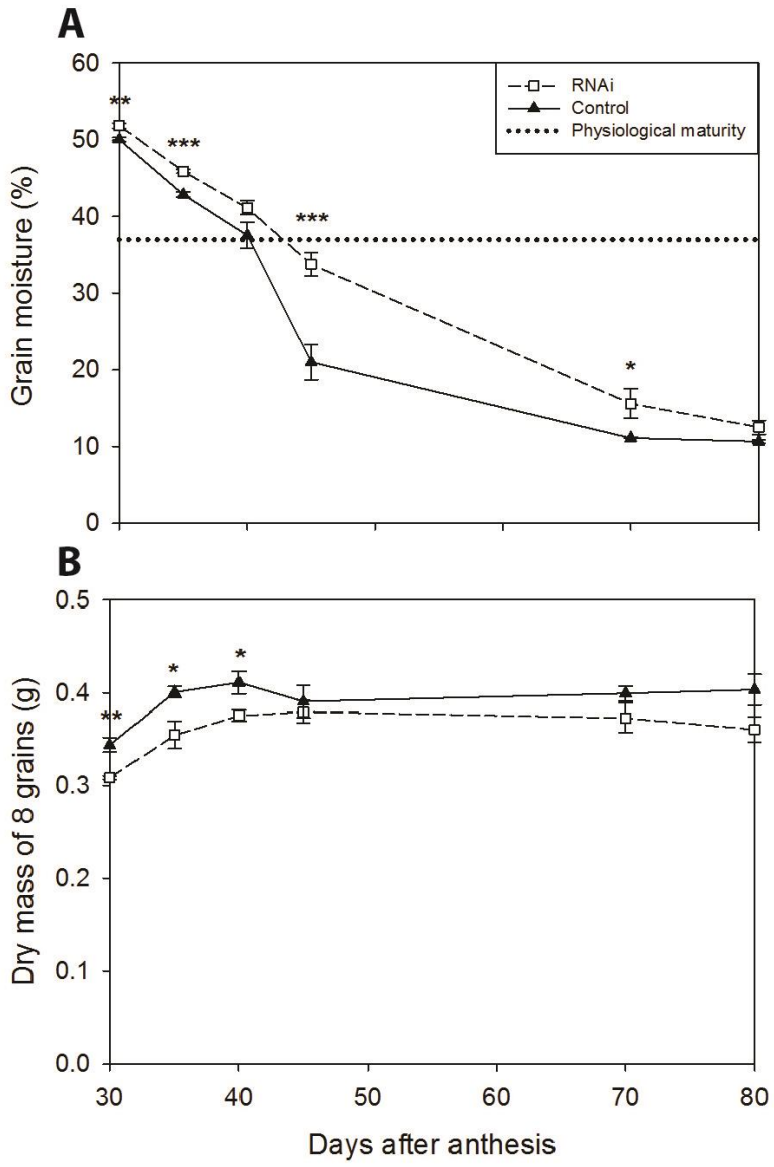


Figure 3.7. Grain moisture content and dry mass. A) Grain moisture content, B) Grain dry mass. n = 5-6. Error bars are SEM. Asterisks denote significant differences at $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***) as determined by Student t-test between genotypes at each time-point.

3.3.2.2 Grain filling enzyme activity

To understand more about grain filling we examined the activity of two key enzymes involved in starch synthesis: ADP-glucose pyrophosphorylase (AGPase) which catalyses the first committed step of starch synthesis and starch synthase. These assays were carried out by Brendan Fahy. We used plants grown in batch 1 which were also used for chlorophyll, grain mass and photosynthesis under ambient conditions to enable comparisons.

3.3.2.2.1 ADP-glucose pyrophosphorylase (AGPase) activity

During the early stage of grain filling at 10 DAA the AGPase activity was the same in *NAM* RNAi and control lines (**Figure 3.8**A and B). By 20 DAA the activity of AGPase on a fresh weight basis was higher in control grain than in *NAM* RNAi grain, which translated into a higher AGPase activity per grain, and grain masses were not different between the two genotypes (**Figure 3.8**C). At 30 DAA the AGPase activity per grain was slightly higher in control than *NAM* RNAi plants, although on a fresh weight basis the activity was the same. By 40 DAA, when both genotypes were near physiological maturity the AGPase activity had fallen to the same level, both on a fresh weight and per grain basis.

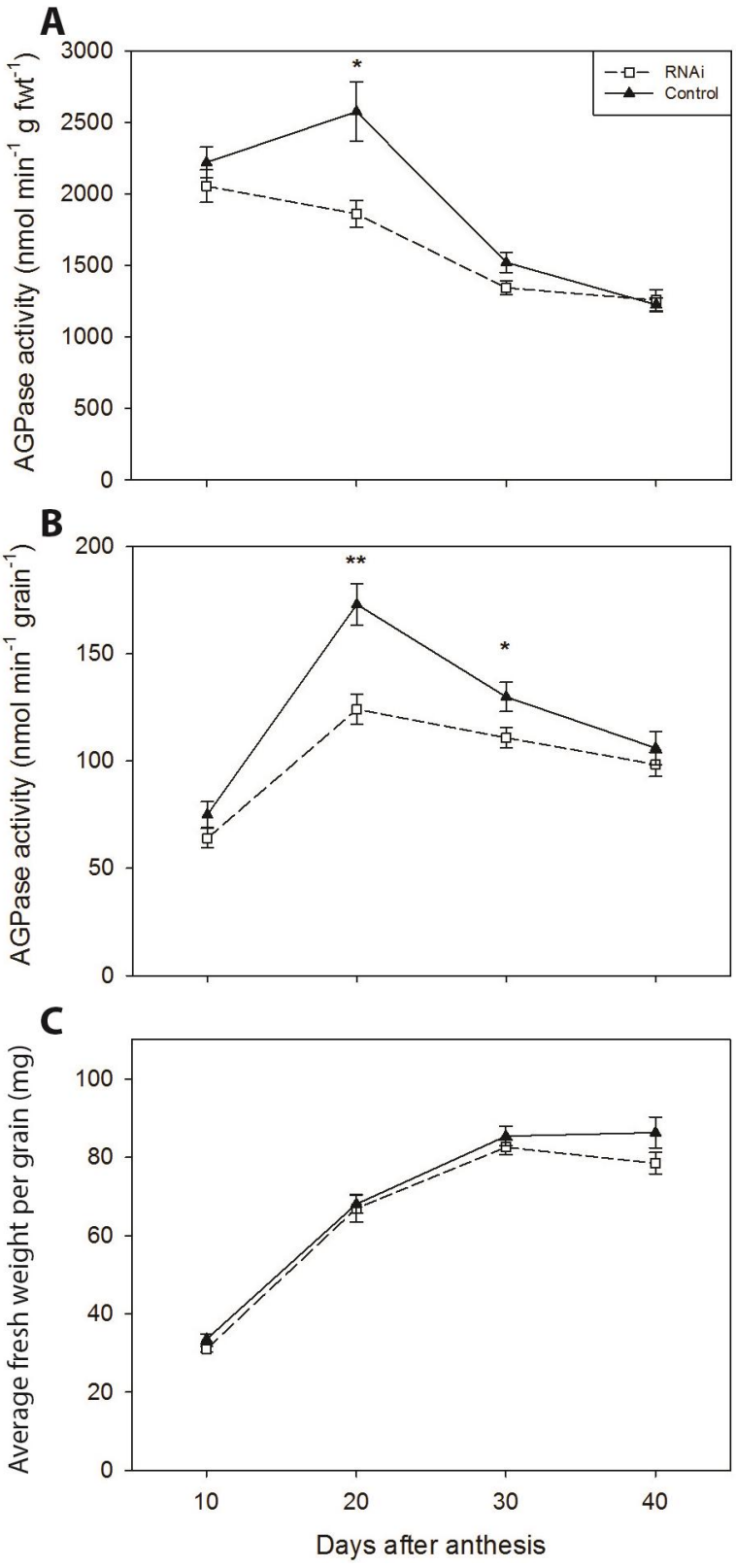


Figure 3.8. AGPase activity in developing grain. A) AGPase activity per gram fresh weight, B) AGPase activity per grain, C) fresh weight per grain used in assay. n = 4-6. Error bars are SEM. Asterisks denote significant differences at $p < 0.05$ (*) and $p < 0.01$ (**) as determined by Student t-test between genotypes at each time-point.

3.3.2.2.2 Starch synthase activity

The activity of starch synthase showed a similar trend to AGPase activity with no difference in activity either on a fresh weight basis or per grain at early and late stages of grain filling (**Figure 3.9**). However, during the middle part of grain filling at 20 DAA the RNAi grain had lower starch synthase activity both per gram fresh weight and per grain.

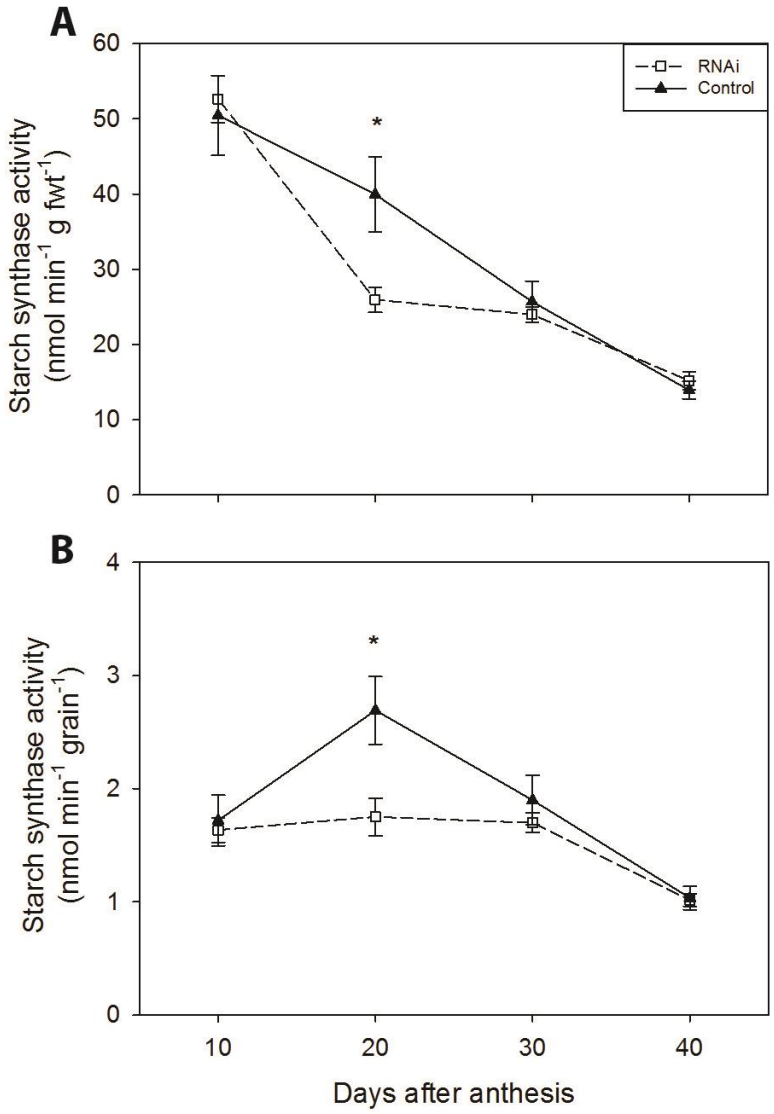


Figure 3.9. Starch synthase activity in developing grain. A) Starch synthase activity per gram fresh weight, B) Starch synthase activity per grain. n = 4-6. Error bars are SEM. Asterisks denote significant differences at $p < 0.05$ (*) as determined by Student t-test between genotypes at each time-point.

In conclusion, the moisture content of *NAM* RNAi grains indicated that they reach physiological maturity 3 days later than the control plants but this is not as large a difference as we had expected from the leaf senescence difference of 10 days. This shows that in *NAM* RNAi plants the delay in leaf senescence does not cause an equivalent change to grain maturation suggesting a disconnect between source and sink tissue. We hypothesise that in *NAM* RNAi grain the capacity to activate the starch synthesis pathway in response to higher carbon availability might be limited, due to decreased enzyme activity, and therefore the grain are unable to use the extra photosynthate provided during the stay-green period.

3.3.3 Hypothesis 3: Carbon fixed during the stay-green period in *NAM* RNAi plants will be stored in stem tissues

By investigating hypothesis 1 and 2 we have shown that although extra CO_2 is fixed by RNAi plants this does not contribute to increasing grain mass per plant. Therefore we hypothesised that fixed carbon which is not entering the grain could be stored in vegetative tissues. In particular stem tissues are known to store water soluble carbohydrates, especially fructans, after anthesis which are then remobilised to the developing grain during the later stages of grain filling (Blacklow et al. 1984). To test this hypothesis *NAM* RNAi and control plants (batch 3) were grown in controlled glasshouse conditions and harvested at 3 time-points to measure carbohydrate accumulation in stem tissues. The time-points harvested were 10 DAA when both genotypes were still fully green, 35 DAA when the *NAM* RNAi plants were still green (flag leaf chlorophyll level = 44.7 ± 0.9 SPAD units) and control plants were in the middle of senescencing (flag leaf chlorophyll level = 21.8 ± 4.6 SPAD units) and 77 DAA when both genotypes were completely senesced (dry with flag leaf chlorophyll level < 10).

We found that at 10 DAA, when both genotypes were fully green, there was no difference in the amount of carbohydrates stored in stems (**Figure 3.10**). At this time-point most of the stem carbohydrate was stored as fructose, although a large amount of fructan and glucose was present in both genotypes. By 35 DAA, when the *NAM* RNAi plants were still green but the control plants had started to senesce, there was still no significant difference in the amount of stem carbohydrate between the genotypes. The amounts of fructose and glucose decreased from 10 DAA to 35 DAA in both genotypes, whereas the amount of sucrose increased. At 77 DAA the *NAM* RNAi plants had 3.9 fold more stem carbohydrates than the control plants, a large proportion of which were stored as fructan. Overall during senescence the RNAi plants maintained a similar amount of stem carbohydrates, whereas the control plants remobilised carbohydrates from the stem.

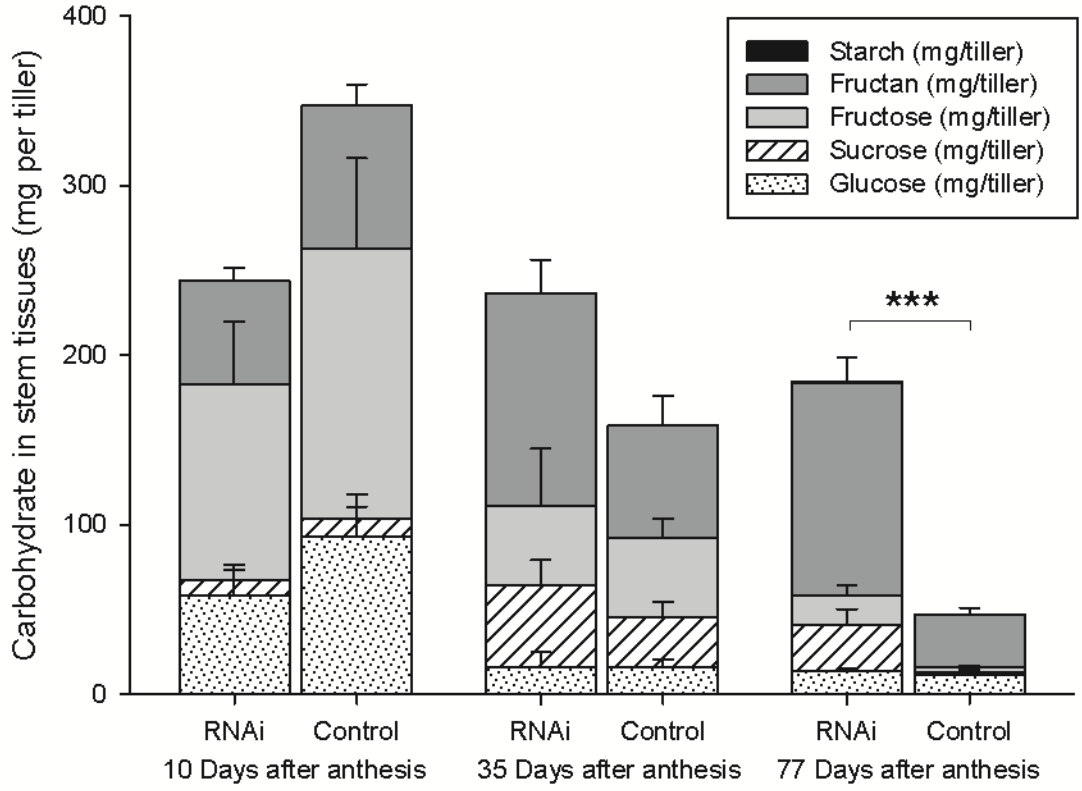


Figure 3.10. Total carbohydrate accumulation in stem tissues. n = 6. Error bars are SEM. Asterisks denote significant differences between total carbohydrates at $p < 0.001$ (***) as determined by Student t-test between genotypes at each time-point.

We investigated where within the stems these carbohydrates were localised, dividing the stems into the exposed peduncle, the enclosed peduncle (wrapped within the flag leaf sheath, but the flag leaf sheath was removed before measurements were taken), internode 1 (below the flag leaf node) and internode 2 (the next internode down). There was a small portion (< 5 cm) of stem (internode 3 and occasionally 4) which was not measured.

We found that a large proportion of carbohydrates including fructan (**Figure 3.11**A), glucose (**Figure 3.11**B), fructose (**Figure 3.11**C) and sucrose (**Figure 3.11**E) were stored in internode 1. Starch, which was only present at very low levels (< 0.15 mg tiller^{-1}), was localised to the peduncle tissues (**Figure 3.11**D) and levels did not differ between genotypes. Glucose (**Figure 3.11**B) was initially present at high levels, particularly in internode 1, but was strongly reduced by 35 DAA and remained present at low levels at 77 DAA. Fructose levels were also initially high (**Figure 3.11**C), but were reduced by 35 DAA in both genotypes. At 77 DAA fructose was nearly absent in the control plant stems but a small but significantly higher amount was present in

internode 2 in *NAM* RNAi plants. Fructan levels decreased steadily in control lines from 10 to 77 DAA whereas in *NAM* RNAi lines fructan levels increased from 10 to 35 DAA, and remained high at 77 DAA in contrast to the control lines. In both genotypes sucrose levels were initially low but increased by 35 DAA, although there was no significant difference in accumulation between genotypes. In control plants sucrose levels declined to nearly zero at 77 DAA, whereas in *NAM* RNAi plants 23 mg tiller^{-1} remained. Overall fructan, fructose and sucrose contributed to there being 137 mg more stem carbohydrates per tiller in *NAM* RNAi compared control plants at 77 DAA.

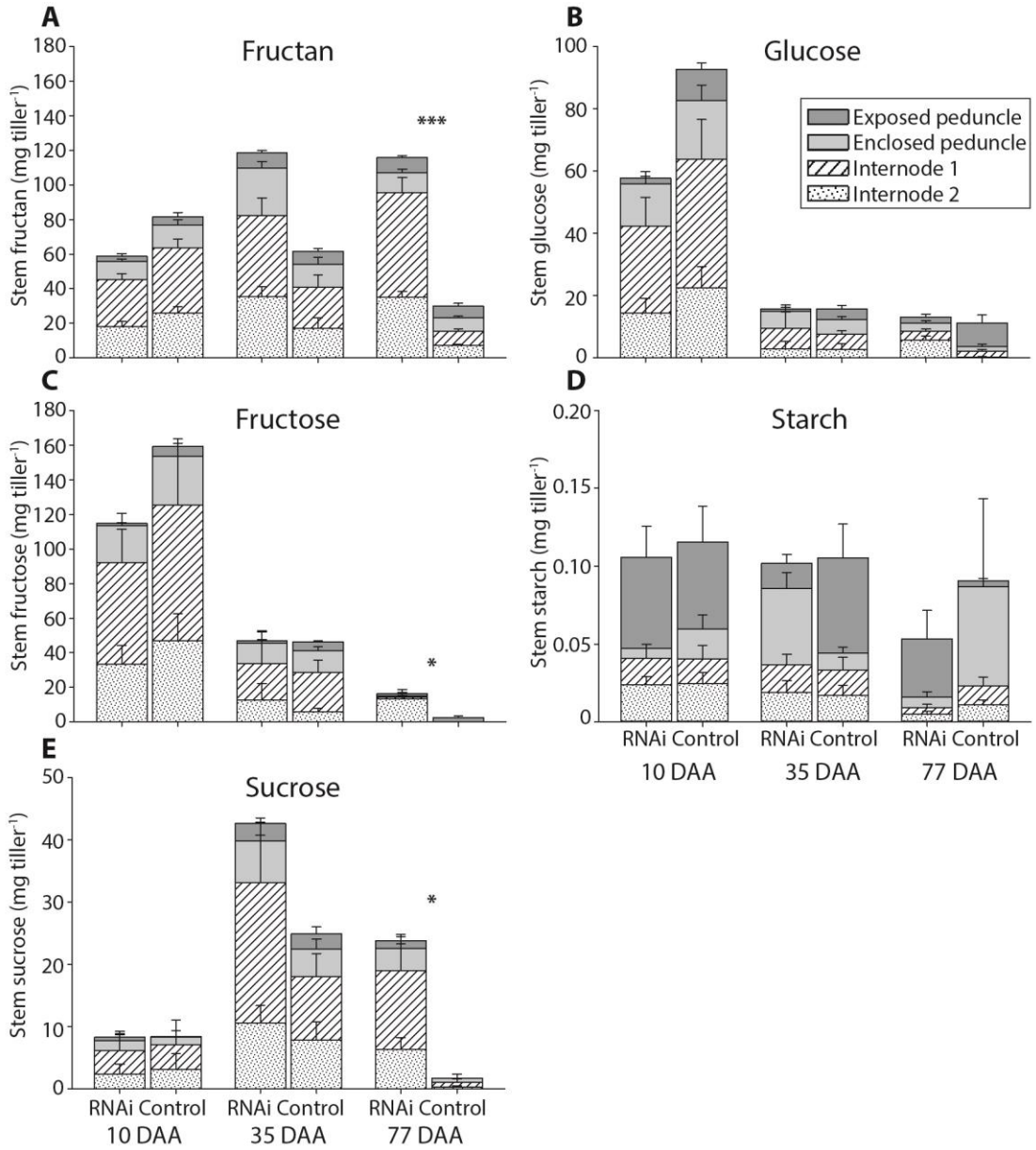


Figure 3.11. Distribution of carbohydrates throughout the stem. A) Fructan, B) Glucose, C) Fructose, D) Starch, E) Sucrose. n = 6. Error bars are SEM. Asterisks denote significant differences at $p < 0.05$ (*) and $p < 0.001$ (***) as determined by Student t-test between genotypes at each time-point.

3.4 Discussion

3.4.1 Hypothesis 1: The delay in senescence in *NAM* RNAi plants will lead to a yield increase compared to control plants

We investigated how photosynthesis changed during the course of senescence in both *NAM* RNAi and control plants. We found that in both genotypes the rate of photosynthesis in flag leaves under ambient environmental levels is correlated with the chlorophyll levels. This confirms that the *NAM* RNAi plants are functional stay-green mutants and that the role of *NAM-B1* is to co-ordinate the onset of senescence rather than directly affect chlorophyll degradation as would be expected in a cosmetic stay-green mutant.

We found that photosynthesis continued at a high rate for approximately 10 days longer in *NAM* RNAi than in control plants. Our direct measurements of photosynthetic rate under ambient conditions enabled the calculation of the amount of extra carbon fixed. Approximately 498 mg of extra glucose were fixed per flag leaf which represents a ~40 % increase in the amount of carbon fixed in total by the flag leaf after anthesis. It has previously been estimated that delaying senescence in *Lolium temulentum* by 2 days would lead to an increase of carbon fixation of 11 % (Thomas Howard and Howarth 2000). Our value of an increase in carbon fixation of 4 % per day of extra photosynthesis is very similar to the 5.5 % per day of extra photosynthesis calculated for *L. temulentum*.

The increase in carbon fixation however does not translate into a grain mass increase in *NAM* RNAi compared to control plants. This result is in line with previous studies using the *NAM* RNAi plants which also found that the delayed senescence did not lead to an increase in yield (Waters et al. 2009). In field conditions non-functional versions of *NAM-A1*, *NAM-B1* and *NAM-D1* have been shown to delay senescence without leading to a yield increased (Avni et al. 2014, Brevis and Dubcovsky 2010). In general the yield advantage associated with delayed senescence is observed in stressed environments, particularly drought conditions (Jordan et al. 2012, Lopes and Reynolds 2012). Therefore it is possible that in the glasshouse where plants were kept well-watered, this advantage of stay-green was lost. However this explanation seems quite unlikely because the *NAM* RNAi plants have previously been grown in pots under heat or drought stress, and even in these conditions they did not have a yield advantage over control plants (Guttieri et al. 2013).

A second mechanism proposed to explain how stay-green sorghum gains its yield advantage is by an enhanced ability to take up N post-anthesis (Borrell and Hammer 2000), and it is possible that improved nitrogen usage in wheat would also lead to a yield increase. Our experiment was carried out in pots, and it is possible that the N became limiting, and therefore enhanced N uptake was not possible late during the growing season, and therefore a yield advantage could not be realised. However, this explanation seems less likely in the light of an experiment carried out at optimal temperatures where pot-grown *NAM* RNAi plants had a yield advantage over control plants at low N, but not at high N (Guttieri et al. 2013). This result for wheat stay-green only being effective at low N is supported by several field studies where the yield advantage of stay-green is absent in high N conditions (Derkx et al. 2012, Gaju et al. 2014), but is present in low N conditions (Derkx et al. 2012).

A third mechanism which may be important for stay-green to improve yield acts before anthesis. Several differences have been observed at anthesis between stay-green and control sorghum lines which are proposed to contributed to the enhanced yield. Stay-green sorghum lines frequently have reduced tillering and smaller leaves, leading to a reduced canopy area at anthesis which is proposed to conserve water for use during grain filling (Borrell et al. 2014a). Higher specific leaf nitrogen at anthesis in stay-green lines has been proposed to allow the plants to set a higher yield potential (Borrell et al. 2003). However in *NAM* RNAi lines it seems unlikely that these mechanisms could be active because we detected no difference in tiller number or flag leaf area between RNAi and control plants and it has been reported that there is no difference in the amount of N in flag leaves in RNAi and control plants at anthesis (Waters et al. 2009). Furthermore *NAM* gene expression has only been detected after anthesis which suggests that the RNAi and control plants should be the same at this time-point.

The translation of extra photosynthesis into a yield advantage requires consideration of the source-sink relationship. Not only must the excess photosynthate be produced and transported to the grain but also the grain must have the capacity to use it to synthesise starch. We had originally hypothesised that because monocarpic senescence is a whole plant process, the grain development would be coupled to the rate of senescence, and therefore would be altered in *NAM* RNAi plants. However, in light of the lack of yield increase, we propose that the grain development might be decoupled from flag leaf senescence, and therefore unable to use the extra photosynthate produced during the stay-green period.

3.4.2 Hypothesis 2: Grain maturation in *NAM* RNAi plants is coupled to leaf senescence

We examined both grain moisture content and AGPase activity as measures of grain development. Around the point of physiological maturity, when grain reach a maximum dry weight, the AGPase activity in both genotypes had fallen to a relatively low level, consistant with the idea that the grain filling period was over. The *NAM* RNAi grain reached physiological maturity ~3 days later than the control plants which is not as significant a delay as would be expected from the 10 day delay in flag leaf senescence (stay-green period). This shows that the grain development is decoupled from flag leaf senescence and suggests that *NAM* genes exert their effects in a tissue-specific manner. However, by the time of the extended photosynthesis in *NAM* RNAi plants (20 to 30 DAA) the AGPase and starch synthase activities in the grain were lower in *NAM* RNAi plants than in control plants. We hypothesise that this lower enzyme activity may have prevented the RNAi plants from accelerating their rate of starch synthesis to take advantage of the extra photosynthate produced during the stay-green period, and therefore the grain yield could not be increased. Alternatively the RNAi plants may not have the capacity to transport the extra photosynthate to the grain. In particular transport between the peduncle and the grain in RNAi plants may be reduced because a large accumulation of micronutrients was observed in the peduncle during nutrient remobilisation (Waters et al. 2009), although we did not detect a build-up of sugars in the exposed peduncle in RNAi plants.

The idea that sink strength may be limiting in RNAi plants is supported by the other data that we have collected. From anthesis until the leaves senesce in control plants, approximately 1.25 g glucose are fixed per flag leaf and 2 g grain are produced per tiller. The balance of 0.75 g glucose is provided by remobilisation of water soluble carbohydrates from stems (~ 0.3 g in control plants) and photosynthesis in tissues such as lower leaves, stems and ears. In particular it has been shown that ear photosynthesis can contribute over 50 % of grain yield (Araus et al. 1993, Maydup et al. 2010, Sanchez-Bragado et al. 2014). In RNAi plants the flag leaf photosynthetic period is extended by 10 days and as result 1.75 g glucose is fixed per flag leaf (0.5 g more than in control plants), but the same 2 g of grain are produced per tiller. In RNAi plants the balance of 0.25 g glucose is provided by photosynthesis in other tissues, and little stem carbohydrate is remobilised. It has been suggested that the presence of stem

carbohydrates at maturity is a sign that plants are unable to use all the photosynthate available and are therefore sink limited (Serrago et al. 2013).

Many studies have investigated the regulation of the source-sink balance in wheat by shading or trimming flag leaves or removing grain. Grain yield has frequently been found to be sink-limited (Calderini D.F. et al. 2006, Jenner CF 1979) across a wide range of genotype by environment interactions (Borrás et al. 2004). For example removal of half the spikelets on two wheat cultivars grown in field conditions did not significantly affect grain mass despite the potential doubling of source capacity available per grain (Slafer and Savin 1994). This situation is similar to the *NAM* RNAi plants which also do not show a yield increase despite a large amount of extra photosynthate (~40 % extra compared to controls).

In the *NAM* RNAi plants the timing of development is altered, therefore some caution must be taken when assuming that extra photosynthate was available to grain filling. Based on our measurements of physiological maturity at ~43 DAA in RNAi plants, the grain were still actively filling during the period of stay-green (20-30 DAA) and therefore might be expected to be capable of using the photosynthate. Our enzyme measurements show that in RNAi grain, despite grain filling still being active, the enzyme activities of key starch synthesis enzymes were lower in RNAi grain than in control grain at 20 DAA. This reduced activity, particularly of starch synthase, may have limited the rate of starch synthesis because the rate of starch synthesis declines in parallel with starch synthase activity during grain maturation (Hawker and Jenner 1993, Yang J. C. et al. 2004). However, ascribing a level of control to an enzyme activity is difficult because the system is interconnected with no one enzyme acting in isolation. One way to quantify the extent to which source and sink reactions control CO_2 flux is using top down metabolic control analysis, which has shown in potato tubers that at least 80 % of control is exerted by the source metabolism (Sweetlove and Hill 2000). The application of these techniques might be one way to quantify the effects of key enzymes on starch synthesis in wheat and determine to what degree the grain capacity limits yield, particularly in *NAM* RNAi plants.

In conclusion the rate of grain development is not co-ordinated with flag leaf senescence and the reduction in enzyme activity in *NAM* RNAi plants gives some insight into the underlying mechanism and the importance of sink capacity. Further

work will be required to understand the control points during wheat grain filling in this germplasm.

3.4.3 Hypothesis 3: Carbon fixed during the stay-green period in *NAM* RNAi plants will be stored in stem tissues

In wheat plants an important long term store of photosynthate is water soluble carbohydrates (WSC), of which the majority (~80%) (Schnyder 1993) are fructans. After anthesis wheat plants normally accumulate WSC in their stems until approximately 20 DAA (Scofield et al. 2009). The WSC is remobilised during the later stages of grain filling to supplement the photosynthate from the leaves, the supply of which starts to decline during senescence (Schnyder 1993).

The extra carbon fixed during the stay-green period in *NAM* RNAi plants was not used in grain filling. We hypothesised that this extra photosynthate would be stored as WSC in the stems. To test this hypothesis we measured the accumulation of WSC: fructan, fructose, sucrose and glucose. We found that over the time-period from 10 DAA to harvest at 77 DAA a different pattern in carbohydrate accumulation was observed between *NAM* RNAi and control plants. In control plants, as expected, WSC were remobilised during grain filling, however *NAM* RNAi plants accumulated WSC during this time period. Therefore by 77 DAA the *NAM* RNAi plants had more fructan, fructose and sucrose in their stems than control plants, with particularly high amounts of fructan remaining in the internode tissues.

It has previously been found that remobilisation of stem WSC starts simultaneously with the decline in photosynthesis of the flag leaf due to senescence and coincides with the period of maximum dry matter accumulation in grains (Kühbauch and Thome 1989, Schnyder 1993). Therefore we hypothesise that since photosynthetic rate in the *NAM* RNAi plants does not decline, the grain continue to be filled using newly assimilated carbon and the WSC is not remobilised from the stems. This is similar to the phenotype observed in the hybrid wheat XN901 which has a 6-day extension to photosynthetic period and accumulates carbohydrates in its stems (Gong et al. 2005). In XN901 only 40 % of ^{14}C fed as CO_2 was exported from the flag leaves within 24 h, whereas the control variety exported 59 %. This delay in transport was also seen at the grain where only 16 % of ^{14}C was deposited in the grain within 24 h in XN901, compared to 52 % in the control variety. This suggests that the ^{14}C was stored within stem tissues, potentially as WSC, and that movement of carbon was reduced in the stay-green variety. However

despite this reduction in grain filling rate in XN901, the grain yield is increased by 15 % compared to a control variety, which is not the case for the *NAM* RNAi lines. This may be because in XN901 the grain filling period was increased by a similar amount to the delay in senescence (6.2 days), therefore the longer grain-filling period compensated for a slower rate of grain filling. However in the *NAM* RNAi plants the time to physiological maturity is only 3 days longer than in control plants, despite the 10 day delay in senescence. Therefore it is possible that a different mechanism may be responsible for the similar phenotype in XN901 and *NAM* RNAi or that environmental conditions have a strong effect on the use of fixed carbon.

Overall the *NAM* RNAi plants had 137 mg more stem carbohydrates per tiller than control plants at maturity and this accounts for 27.5 % of the estimated 498 mg extra photosynthate fixed per flag leaf during the period of stay-green. In addition to storage as WSC in the stems there are several other possibilities for the fate of the extra fixed carbon in *NAM* RNAi plants. One possibility is that the carbon is translocated to another part of the plant and stored there. It has been previously reported that *NAM* RNAi plants produce the same amount of above ground biomass as control plants until harvest (Waters et al. 2009) and that roots had the same biomass at 42 DAA, although later time points were not studied in roots. Therefore it is possible that some of the extra carbon is stored in the roots in *NAM* RNAi plants.

Another alternative explanation for the use of this extra photosynthate could be to maintain the metabolism of the plant. Dark respiration rate (R_d) has been reported to be 22 % of the rate of photosynthesis in flag leaves (Araus et al. 1993) and in ears respiration due to the growth and maturation of grain is 176 % of the rate of photosynthesis of ear tissues. From the A-C_i curves we found that dark respiration rate (y-intercept) did not differ between *NAM* RNAi and control flag leaves, however we did not measure respiration in other tissues. It is likely that the delayed senescence in *NAM* RNAi plants leads to a higher respiratory rate than in control plants due the presence of more green tissues which are still metabolically active. This could be tested by gas-exchange experiments on other tissues, for example the ears and stem tissues. This hypothesised extra respiration may be maintaining metabolism or it might be wasteful respiration engaged in futile cycles to simply use up the extra carbon.

In conclusion we have detected that over one-quarter of the extra photosynthate from delayed senescence in the *NAM* RNAi lines was stored as WSC, particularly fructan in

the stems. However due to the complexity of understanding the whole-plant carbon balance during grain filling we were not able to account for the remainder of the extra fixed carbon. Further work will be necessary to determine the fate of this carbon, for example translocation to the roots or an increased respiration rate.

3.4.4 Summary

In this chapter we have found that the *NAM* RNAi plants continue to photosynthesise for longer than control plants. However this extended photosynthetic period does not increase the grain mass but does increase the WSC in stems at maturity. The increased WSC at maturity in RNAi plants suggests that sink limitation may prevent the use of the excess photosynthate, which is supported by reduced enzyme activity in RNAi grain during grain filling. Although we expected the monocarpic senescence to co-ordinate the rates of development across the whole plant, comparisons between the timing of physiological maturity and senescence indicate that the grain development is decoupled from flag leaf senescence in RNAi plants. There is a complex relationship between the stay-green phenotype and the ability to increase yield: sink strength must be taken into account.

4 Chapter 4 – The identification of direct target genes of NAM-B1

4.1 Introduction

NAM-B1 encodes a NAC transcription factor which is known to regulate monocarpic senescence, nutrient remobilisation and nutrient transport to the developing grain. However the regulation and inter-connections of these processes are not well understood. Therefore using *NAM-B1* as an entry point we aim to dissect the network of genes which control senescence, nutrient remobilisation and nutrient transport. Using a combination of biochemical, molecular biology and next-generation sequencing (NGS) methods we aim to identify direct and downstream targets of *NAM-B1*.

4.1.1 *NAM-B1* is a NAC transcription factor

NAM-B1, like many other NAC transcription factors (TFs) consists of three exons. Exons 1 and 2 encode the N-terminal NAC domain which is conserved between NAC TFs. The NAC domain mediates dimerization of NAC TFs and contains a DNA binding domain. NAC TFs act as transcriptional regulators in both homodimeric and heterodimeric forms (Jeong et al. 2009). The third exon of *NAM-B1* encodes the C-terminal region which is predicted to be disordered in structure and may mediate protein-protein interactions. Like other NAC proteins this C-terminal domain is predicted to have an activation domain which can regulate gene expression.

The structure of NAM-B1 has not been determined, however the highly conserved nature of the NAC domain (Duval et al. 2002) enables some predictions to be made. The structure of an Arabidopsis abscisic acid-responsive NAC (ANAC019) gene revealed that the NAC domain consists of a twisted antiparallel β-sheet sandwiched between two helices (Ernst et al. 2004) and DNA is bound by the curved outer β-strand of the core β-sheet (Welner et al. 2012).

4.1.2 Availability of wheat reference sequences

The availability of sequence resources available for wheat has massively increased over the past five years (Borrill et al. 2014) which has made feasible next-generation sequencing projects such as finding the target genes of NAM-B1. The International Wheat Genome Sequencing Consortium (IWGSC) has coordinated the purification of individual chromosome arms using flow sorting (Safar et al. 2004) followed by shotgun sequencing and assembly into contigs of an average size of 2.5 kb. These genome-specific contigs have been made publicly available through EnsemblPlants (IWGSC

2014). Physical maps of individual chromosome arms are being constructed from BAC libraries (Paux et al. 2008) and eventually a high quality reference sequence supported by physical maps will be available for all chromosomes. In the meantime the contigs have been arranged into a genetic order (IWGSC 2014) using two approaches: POPSEQ (ordering of individual contigs according to the linkage order from a mapping population) (Mascher et al. 2013) and Genome Zipper (relying on synteny between grass species) (Mayer et al. 2011). Until a few months ago the information about the order of the contigs was not published, therefore to carry out ChIP-seq we constructed our own ordered reference sequence.

Many other resources are available which complement the genomic sequences being produced by the IWGSC. Whole genome shotgun (WGS) sequencing followed by synteny-guided *de novo* assembly has produced 5x coverage of the hexaploid wheat genome (Brenchley et al. 2012). Draft sequences for the A and D genome progenitors, *Triticum urartu* (Ling et al. 2013) and *Aegilops tauschii* (Jia et al. 2013), have also been created using a WGS approach followed by *de novo* assembly. Of central importance to understanding gene function is developing reliable gene models associated with the reference genome. A comprehensive set of high quality homoeologue specific gene models is available (Krasileva et al. 2013) and these have been incorporated into the latest release of the IWGSC contigs on EnsemblPlants alongside another set of gene models (IWGSC 2014).

These genomic resources represent a step change in the ability to carry out functional characterisation of wheat genes. We have leveraged these resources to understand more about how *NAM-B1* regulates nutrient remobilisation and senescence in hexaploid wheat.

4.1.3 Existing knowledge about targets of *NAM-B1*

NAM-B1 plays an important role in controlling the grain nutrient content of protein and micronutrients and regulating the rate of terminal senescence (Uauy et al. 2006b). One study has been published which examines the downstream targets of *NAM-B1* using RNA-seq (Cantu et al. 2011). *NAM* RNAi plants with down-regulated expression of *NAM-B1* and its homoeologues were used. The expression levels of genes in flag leaves 12 days after anthesis (DAA) were compared between WT and RNAi plants. A large scale change in gene expression was detected between the two genotypes. In particular transporters, hormone regulated genes and transcription factors showed altered

expression levels. These results give the first molecular insight into the pathways which *NAM-B1* regulates. A significant limitation of this study was the poor availability of a reference sequence which meant the authors had to construct their own reference sequence using 454 sequencing of the transcriptome and this reference sequence was not separated into homoeologues. Therefore it will be valuable to re-analyse this data using the comprehensive reference sequences now available.

4.1.4 Chromatin immuno-precipitation followed by next-generation sequencing (ChIP-seq) to identify direct targets

The targets of *NAM-B1* identified by differential gene expression analysis include not only genes directly regulated by NAM-B1 but also genes which are further downstream in the regulatory cascade. In order to identify direct targets of NAM-B1 ChIP-seq can be utilised. ChIP-seq identifies the precise regions of DNA to which NAM-B1 binds, nearby genes are potential regulatory targets of NAM-B1.

4.1.4.1 Outline of the ChIP-seq method

Tissue samples to be used in ChIP-seq are cross-linked using formaldehyde to make strong connections between the transcription factor (TF) and the DNA region it is binding *in vivo* (**Figure 4.1**). The cross-linked DNA-TF complex is then sonicated which breaks the chromatin into shorter pieces. An antibody to the TF of interest is then used to specifically capture the TF along with the region of DNA to which it is bound. De-cross-linking removes the tight bond between TF and DNA, and the protein is subsequently washed away. The DNA remaining, which was specifically bound by the TF, is then sequenced.

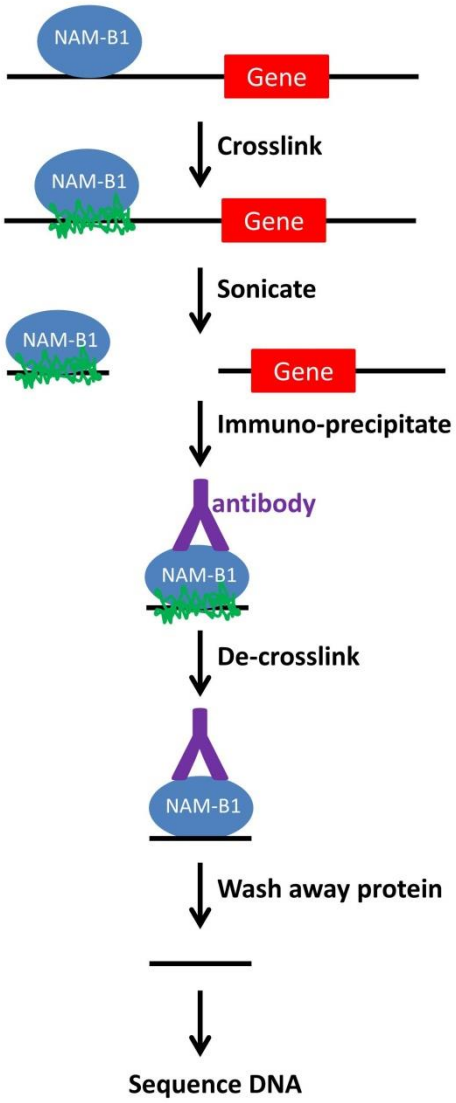


Figure 4.1. Schematic representation of the steps of sample preparation for ChIP-seq.

The DNA is sequenced and then aligned to a reference genome. After alignment it is possible to compare between samples containing the tagged transgene and control samples which did not contain the transgene. Looking at the distribution of the reads across the reference genome it is possible to detect regions where the TF is binding because these regions have more reads in the TF sample than in the control sample. The binding regions are termed 'peaks' (**Figure 4.2**) and a statistical model determines which peaks are statistically significant compared to the control which should represent background noise.

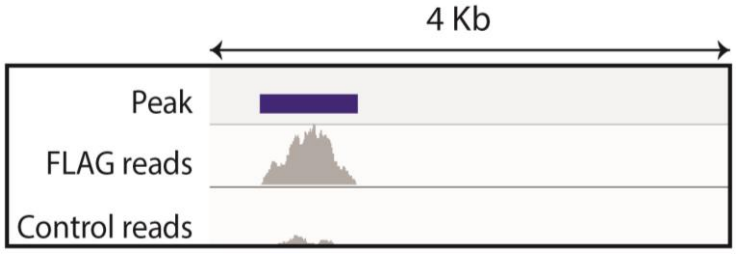


Figure 4.2. Example of NAM-B1 binding region.

4.1.4.2 Uses of ChIP-seq to identify targets of NAC transcription factors and genes important in senescence

ChIP has been successfully applied to identify downstream targets of transcription factors in several plant species including Arabidopsis (Kaufmann et al. 2010), rice (Zhu Jia-Ying et al. 2012), maize (Bolduc et al. 2012) and soybean (Shamimuzzaman and Vodkin 2013). Many different classes of transcription factor have proved amenable to this technique including AP2 (Yant et al. 2010), class 1 homeobox (Bolduc et al. 2012), WRKY (Roccaro and Somssich 2011) and NAC (Shamimuzzaman and Vodkin 2013) transcription factors.

Although biological processes as diverse as stem cell fate (Bolduc et al. 2012), floral development (Pajoro et al. 2014, Yant et al. 2010) and drought stress (Ricardi et al. 2014) have been studied, very few ChIP-seq experiments have been carried out on senescence-associated transcription factors. Instead most senescence-related genomics experiments have focused on differential gene expression analysis to elucidate the networks of genes regulating senescence. ChIP-seq has been used to show that histone marks alter during the course of senescence (Brusslan et al. 2012), however the association between epigenetic regulation and transcription factor regulation of genes during senescence is not well understood. The identification of direct targets of NAM-B1 will give insight to the regulation of nutrient remobilisation and senescence in wheat and will be one of the first examples of the application of ChIP-seq to understand the genes regulating these processes.

4.1.5 Aims and hypotheses

We aim to understand which genes NAM-B1 directly regulates which will be a first step towards understanding the complex network of genes which control senescence, nutrient remobilisation and transport. To achieve this aim we will carry out ChIP-seq and compare the target genes identified to RNA-seq data (Cantu et al. 2011), which we will re-analyse using our own reference sequence.

We will generate our own ordered reference sequence and transgenic plants to use for the specific immuno-precipitation of NAM-B1. After the creation of these resources we will test a number of hypotheses, both about the parameters affecting the ChIP-seq data analysis and about the target genes of NAM-B1. These will be:

4.1.5.1 Hypothesis 1: The stringency of read mapping will affect the number of binding sites identified by ChIP-seq

We expect that if read mapping is made less stringent, more reads will be mapped and more peaks will be detected due to increased read depth. However, these extra peaks may be false positives caused by incorrect mapping, therefore it will be important to optimise the ratio of number of reads mapped to the number of peaks detected.

4.1.5.2 Hypothesis 2: True NAM-B1 binding sites will be detected in several replicates, the significance of binding sites will be correlated between samples and binding sites may be different between leaf and peduncle samples

Replication of binding sites in several samples would confirm that these sites are reproducible and more likely to be biologically relevant targets of NAM-B1. We would expect regions with stronger binding in one sample, to also have stronger binding in other samples, if the strength of binding is biologically meaningful. Thirdly NAM-B1 may play different roles in different tissues, therefore we hypothesise that it might have different target genes in leaf and peduncle tissues.

4.1.5.3 Hypothesis 3: NAM-B1 binding sites will be unique within the genome to ensure specific binding of target genes

The majority (>80 %) of the wheat genome consists of repetitive elements. For NAM-B1 to specifically target its downstream genes we would expect it to bind to non-repetitive (i.e. unique) sequences.

4.1.5.4 Hypothesis 4: NAM-B1 will bind promoter regions to regulate gene expression

To regulate transcription TFs usually bind to promoter regions, therefore we expect that the majority NAM-B1 binding sites will be in promoter regions.

4.1.5.5 Hypothesis 5: NAM-B1 binding will be mediated through a specific DNA sequence motif

Transcription factors bind DNA in a sequence-specific manner. Therefore we hypothesise that NAM-B1 will bind to a specific sequence motif, which we will be able to detect by analysing the sequences of DNA identified by ChIP-seq.

4.1.5.6 Hypothesis 6: NAM-B1 target genes will have common functions related to nutrient remobilisation and senescence

Since NAM-B1 is known to control nutrient remobilisation and senescence we expect its target genes to have functions related to these processes.

4.1.5.7 Hypothesis 7: Differential gene expression analysis will identify genes downstream of NAM-B1 involved in nutrient remobilisation and senescence

Differential gene expression analysis is another approach to identify downstream targets of NAM-B1: we expect these target genes to be related to nutrient remobilisation and senescence.

4.1.5.8 Hypothesis 8: Differentially expressed genes will contain shared DNA motifs in their promoters which facilitate co-expression

Often co-expressed genes involved in similar biological functions will have shared DNA motifs in their promoter regions which are bound by a particular TF. These differentially expressed genes may contain a NAM-B1 binding motif, which can be compared to the motifs identified by ChIP-seq.

4.1.5.9 Hypothesis 9: Direct targets of NAM-B1 will have their promoter sequences bound by NAM-B1 and will be differentially expressed in *NAM* RNAi plants

We hypothesise that when NAM-B1 binds to the promoter region of a target gene that it will induce a change in gene expression which can be detected by RNA-seq. This change in gene expression will have downstream consequences on nutrient remobilisation and senescence.

4.1.5.10 Hypothesis 10: NAM-B1 directly binds the DNA regions (peaks) identified by ChIP-seq

We will carry out an *in vitro* biochemical assay to validate the binding of NAM-B1 to the regions of DNA identified by ChIP-seq.

By examining these hypotheses we will assess whether ChIP-seq, in combination with RNA-seq, identifies downstream targets of NAM-B1 and how this furthers our understanding of the regulation of nutrient remobilisation and senescence.

4.2 Methods

4.2.1 Building pseudomolecules to represent the wheat genome

4.2.1.1 Pseudomolecules v2 (previously created by Martin Trick)

De novo leaf transcriptome assemblies (ABD assemblies) had previously been created by Andrea Harper (University of York) from diploid leaf transcriptomes using the Trinity assembler (Grabherr et al. 2011). Since the B genome diploid was quite dissimilar to the B genome in hexaploid wheat, the B genome assemblies were specifically adjusted to be more similar in their sequence composition to hexaploid wheat using a *de novo* transcriptome assembly from the tetraploid *T. dicoccoides*. The ABD assemblies were BLASTed against the *Brachypodium distachyon* (Brachypodium) genome v2 and the best bit scoring hit with an e-value $\leq 1E^{-20}$ was retained. This anchored the ABD assemblies to the Brachypodium chromosomes and gave a physical position for each gene. A Chinese Spring x Paragon mapping population was transcriptome-sequenced and SNP markers in the ABD assemblies were identified. The linkage group orientation and order of these markers was determined from the mapping population. This linkage map data was used to re-arrange the ABD assemblies (previously in Brachypodium-like order) into a wheat-like order, which was replicated across the A, B and D genome. The same order had to be used for the A, B and D genome because the transcriptome sequencing of the mapping population was not genome specific, therefore all three homoeologues were collapsed together. Any ABD assemblies which were non-polymorphic in the Chinese Spring x Paragon population were interpolated between the nearest genes according to the Brachypodium physical order. This created wheat pseudomolecules v2 (pseudo_v2).

4.2.1.2 Converting ABD assemblies into IWGSC (International Wheat Genome Sequencing Consortium) contigs to create pseudomolecules v3

The ABD assemblies were extracted from the pseudo_v2 and BLASTed against the chromosome arm flow-sorted IWGSC contigs, selecting the best bit-scoring hit with an e-value ≤ 1E-30. Assemblies from pseudo_v2 which were mapped onto the correct chromosome (i.e. chromosome position on pseudo_v2 matched the chromosome of the best BLAST hit from IWGSC) were retained. The IWGSC contigs for these assemblies were concatenated to form wheat pseudomolecules v3 (pseudo_v3). This converted the transcriptome sequence based pseudo_v2 to the genomic sequence based pseudo_v3.

4.2.1.3 Incorporating IWGSC annotated mRNAs to make pseudomolecules v3.3

From a file containing annotation of the IWGSC contigs (Sarah Ayling, TGAC) 65,103 mRNA features were extracted. The names of contigs containing these mRNA were compared with the contigs already present in pseudo_v3. There were 13,329 contigs specific to pseudo_v3; 23,594 in both the pseudo_v3 and the annotation file and 41,508 mRNAs only present in the annotation file. The large proportion of mRNAs missing in the pseudo_v3 was not surprising because mRNA present in pseudo_v3 had to be expressed in the leaf and polymorphic in Chinese Spring x Paragon.

The 41,508 new mRNAs from the annotation file were BLASTed against pseudo_v3 (a match required an e-value $\leq 1E^{-100}$ and a high-scoring segment pair within an annotated coding sequence). Of the 41,508 new mRNAs 6,845 had no matches to pseudo_v3, 34,483 matched to the pseudo_v3 with less than 100 % identity so were likely homoeologues of genes already present, and 90 mRNAs matched with 100 % identity within the coding sequence and were discarded to avoid duplication of identical genes. Contigs containing the mRNAs with no matches to pseudo_v3 or matches less than 100 % identity were concatenated according to which chromosome they came from in an arbitrary order. These concatenated contigs were added to the ordered contigs in pseudo_v3 to form pseudomolecules v3.3 consisting of an arbitrarily ordered (unordered) portion and ordered portion (from pseudo_v3, according to the Chinese Spring x Paragon genetic map).

4.2.1.4 Annotation of the pseudomolecules v3.3 to make a non-redundant set of mRNA sequences

The pseudomolecules v3.3 were annotated by extracting the mRNA positions from the annotation file provided by Sarah Ayling (TGAC) and converting the mRNA position to the position on the pseudomolecules v3.3. For each position on the pseudomolecules if two mRNAs were present (e.g. splice isoforms) the longest transcript was selected. This produced annotation for the pseudomolecules v3.3 consisting of 75,419 cDNAs: 24,608 (A), 29,275 (B), 21,446 (D) which were formatted as a GFF file and made into a fasta file to use as a reference sequence for RNA-seq.

4.2.1.5 GO (gene ontology) term assignation to mRNAs in the pseudomolecules v3.3

All the mRNA sequences from the pseudomolecules v3.3 were translated to proteins and annotated using the software Blast2GO (v2.7.1) run using Java (v1.7.0_51). Specifically BLASTX (v2.2.28+) was used to search for all wheat mRNAs in the UniProtKB/Swiss-Prot database (v2013-12). The mRNA sequences and the BLAST output were loaded into Blast2GO and annotated using the "annotate" feature. GO terms were assigned to mRNA sequences using the Blast2GO mapping feature. InterProScan from Blast2GO was used to annotate the protein domains in peptide sequences encoded by the mRNA, using the InterPro database.

4.2.1.6 Rice homologues of mRNAs in the pseudomolecules v3.3

To get more information about putative gene functions, BLAST+ (v2.2.28) was used to identify the best match to rice for each mRNA in the pseudomolecules v3.3. The database searched was the rice cDNA annotation (v7.0) and the best hit (ranked by lowest e-value and then highest bit score) was retained (e-value $\leq 4E^{-5}$).

4.2.2 Construct creation for transgenic wheat for use in ChIP-seq

4.2.2.1 Overall cloning and expression strategy

We decided to express the full length *NAM-B1* coding sequence, with an N-terminal tag under its native promoter in transgenic wheat plants. The N-terminal tag was either 3xFLAG or 3xHA: both of these peptide tags have high quality commercially available antibodies to enable specific immuno-precipitation of the NAM-B1 protein from the transgenic plants. We used the native promoter to try to maintain similar expression localisation and levels to the native protein. We did not use an overexpression construct because we were concerned that overexpression of NAM-B1 would induce senescence. An outline of the steps to make the construct is given in **Figure 4.3** (details in following sections). The construct was then sent to the National Institute for Agricultural Botany (NIAB) for transformation into the hexaploid wheat variety Fielder (genotype *NAM-A1/nam-B1*/*NAM-D1*/*NAM-A2*/*NAM-B2/NAM-D2*).

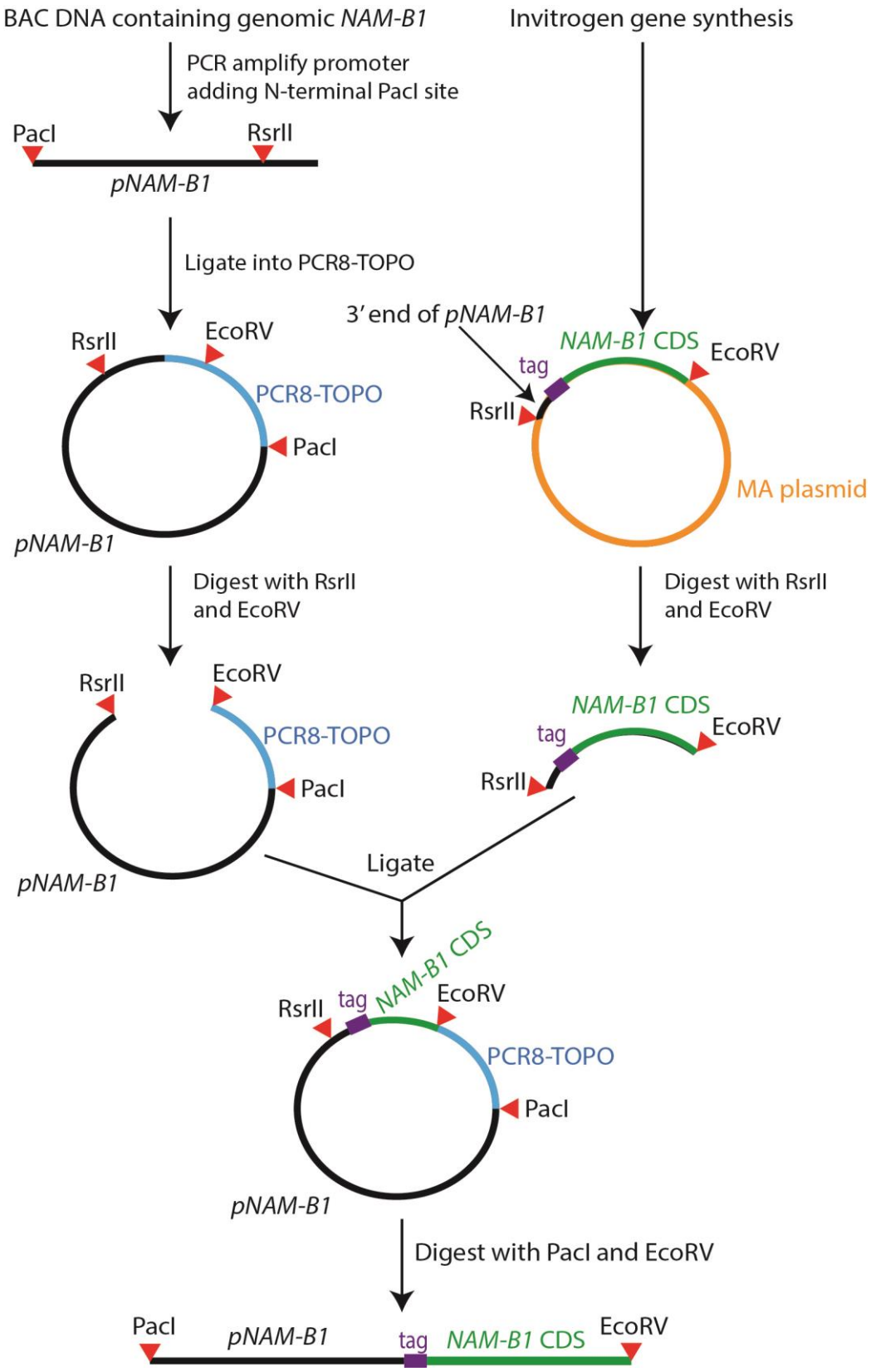


Figure 4.3. Cloning strategy to produce the construct containing the *NAM-B1* promoter (*pNAM-B1*) fused to a tag (either 3x HA or 3x FLAG) in frame with the *NAM-B1* coding sequence (*NAM-B1* CDS). Restriction enzyme sites are shown as red triangles. Plasmids used were PCR8-TOPO (Invitrogen) and MA plasmid (containing a tag-*NAM-B1* synthesised by Invitrogen).

4.2.2.2 *NAM-B1* promoter sequence cloning

A BAC clone from a durum wheat (*Triticum turgidum* ssp. *durum*, AABB) wheat library was obtained from J. Dubcovsky, UC Davis which contained the *NAM-B1* genomic region (Cenci et al. 2003). The Langdon cultivar of durum used to make this library contained *NAM-B1* due to an introgression of a 30 cM segment of chromosome 6BS from *T. turgidum* ssp. *dicoccoides*. Plasmid DNA was purified using the QIAprep Spin Miniprep Kit (27104, Qiagen) and used as a template to amplify the ~8.2 kb promoter region of *NAM-B1*. The PCR reaction mix is listed in **Table 4.1**, cycling conditions in **Table 4.2** and primers used in **Table 4.3**. The amplified product sequence is in the Appendix 6.1.2.

Reagent	**Volume (µl)**
Forward primer (PB1_pGPC_F2) (2 µM)	5
Reverse primer (PB1_pGPC_R2) (2 µM)	5
5x Phusion GC buffer (M0530S, NEB)	4
dNTP (2 mM)	2
BAC DNA template (30 ng/µl)	0.25
Dimethyl sulfoxide (DMSO) 100 %	0.6
H_2O	2.95
Phusion DNA polymerase (M0530S, NEB)	0.2

Table 4.1. PCR reaction for amplification of *NAM-B1* promoter sequence from BAC DNA using Phusion Taq.

Step	**Temperature (°C)**	**Time (min)**
1	98	0:30
2	98	0:10
3	63	0:20
4	72	4:00
5	go to step 2 34x	
6	72	10:00
7	4	∞

Table 4.2. PCR cycling conditions for Phusion taq amplification of *NAM-B1* promoter sequence.

Primer	Sequence	Notes
PB1_pGPC_F2	TTTTTTTTAATTAATACTTTTACACCCGGTCACG	To amplify promoter sequence and add 6xT-PacI site at the 5' end
PB1_pGPC_R2	GTTGGGAGGAAGGAGAGGAG	To amplify promoter sequence
PB28_seqGPC_R11	CCCTAGCCCATCAGAGTTCA	For sequencing
PB29_seqGPC_R12	CATAGGGCTTTCGTGTCTGG	For sequencing
PB30_seqGPC_R13	ATGAACCCAACCAGTTCGTC	For sequencing
PB31_seqGPC_R14	GCTAACAAACAGGTGCCACA	For sequencing

Table 4.3. Primers used to clone and sequence *NAM-B1* promoter sequence.

The ~8.2 kb promoter sequence was purified from the gel using QIAquick Gel Extraction Kit (28704, Qiagen).

An "A" overhang was added to the purified PCR product by combining:

- 5 µl PCR product
- 1 µl PCR 10x PCR buffer (NEB)
- 1 µl 2 mM ATP
- 1 µl Homemade Taq
- 2 µl H_2O

The mixture was incubated at 72 °C for 20 min.

The A-tailed PCR product was cloned into pCR™8-TOPO® TA Vector (K2500-20, Invitrogen) by incubating the following reagents at room temperature overnight:

- 4 µl PCR product with A overhang
- 1 µl salt solution
- 1 µl TOPO vector

The ligation product was subsequently transformed into DH5α competent cells (18265-017, Invitrogen):

10. Thaw 50 µl aliquot of DH5α cells on ice
11. Add 5 µl TOPO cloning ligation product
12. Incubate on ice for 30 min
13. Heat shock at 42 °C for 30 sec in a water bath
14. Put cells on ice for 2 min
15. Add 950 µl pre-warmed SOC medium
16. Shake at 37 °C for 1 h
17. Spread 50 µl and 150 µl onto pre-warmed LB plates containing 100 µg/ml of spectinomycin
18. Incubate overnight at 37 °C

Five colonies were selected and grown overnight in LB with spectinomycin (100 µg/µl) and plasmid DNA was prepared using the QIAprep Spin Miniprep Kit (27104, Qiagen). The presence of the insert was confirmed by restriction digests with EcoRI at 37 °C for 1 h (**Table 4.4**). Glycerol stocks (500 µl overnight culture with 500 µl 40 % glycerol) were stored at -80 °C.

	Volume (µl)
Plasmid DNA (~100 ng/µl)	3
Roche Buffer H	2.5
EcoRI (10703737001, Roche)	1
H_2O	18.5

Table 4.4. Restriction digest with EcoRI.

Colonies with the expected insert were sequenced using primers in **Table 4.3**. The BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used to prepare DNA samples for Sanger sequencing (2.2.2.4).

4.2.2.3 N-terminal tag – *NAM-B1* gene synthesis

To produce the constructs as quickly and accurately as possible it was decided to synthesise the coding sequence of *NAM-B1* with N terminal tags fused in frame (**Figure 4.4**). The tags chosen were 3xHA and 3x FLAG which have commercially available antibodies for specific immuno-precipitation. A portion of the *NAM-B1* promoter sequence and the 5'UTR were also incorporated into the construct to facilitate downstream cloning, along with an EcoRV restriction enzyme site after the stop codon at the 3' end of the construct. The constructs were supplied in the vector pMA by Geneart (Invitrogen).

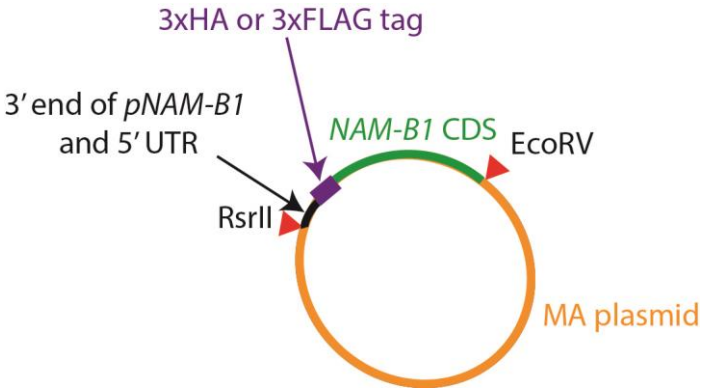


Figure 4.4. Gene synthesised including partial *NAM-B1* promoter (*pNAM-B1*), 5'UTR, tag and coding sequence.

4.2.2.4 Digestion and ligation to produce final construct

The plasmid containing the *NAM-B1* promoter and the plasmids containing the tagged *NAM-B1* were digested with RsrII and EcoRV-HF at 37 °C overnight, reaction mix in **Table 4.5**.

Reagent	**Volume (µl)**
DNA (450 ng/µl)	5
10x buffer 4 (NEB)	5
RsrII (5,000 U/ml) (NEB)	1
EcoRV-HF (20,000 U/ml) (NEB)	1
H_2O	38

Table 4.5. Restriction digest with RsrII and EcoRV.

The bands containing the required fragments of the plasmid were gel purified with the QIAquick Gel Extraction Kit (28704, Qiagen) and ligated together overnight at 37 °C using T4 DNA ligase, reaction mix in **Table 4.6**.

Reagent	Volume (µl)
2x buffer	5
Restriction digested *NAM-B1* promoter (30 ng/µl)	2
Restriction digested tagged *NAM-B1* (25 ng/µl)	1.6
T4 DNA ligase (3 Weiss U/µl) (from pGEM-T, A3600, Promega)	1
H_2O	0.4

Table 4.6. Ligation reaction.

Ligation products were transformed into DH5α (4.2.2.2) and 6 colonies were grown overnight in LB with 100 µg/ml spectinomycin. Glycerol stocks were made (4.2.2.2) and plasmids were extracted using QIAprep Spin Miniprep Kit (27104, Qiagen). Inserts were checked by restriction digest overnight at 37 °C with BamHI and HindIII separately, using the reaction mix in **Table 4.7**.

	Volume (µl)
Plasmid DNA (~500 ng/µl)	2
10x Buffer B	2.5
BamHI or HindIII (NEB)	0.5
H_2O	20

Table 4.7. Restriction digest with BamHI or HindIII.

Plasmids with the expected restriction patterns were sequenced from ~3.5 kb upstream of the NAM-B1 start site until end of the insert using the primers in **Table 4.8**. The BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used to prepare DNA samples for Sanger sequencing (2.2.2.4).

Primer	Sequence
PB1_seqGPC_F1	GCCAGTAGTTGTTGCAGTTTG
PB1_seqGPC_F2	CCATCATCAGGACAGCCATT
PB1_seqGPC_F3	GACCGCAAGTACCCCAAC
PB1_seqGPC_F4	GCAGCTCCGACTCATCTTC
PB1_seqGPC_F5	ATCTGACACGCGGTACGAG
PB1_seqGPC_F6	GGTTTTGCGAGCGTTTTATG
PB1_seqGPC_F7	TTTTTCAAATCCGATGAACTTTT
PB1_seqGPC_F8	CTAGCAGCCAGACAAACATAATC
PB1_seqGPC_F10	CGAGAAACACAATCGGTGCT
PB27_seqGPC_F11	GCTAGGAGTATTTTCGAAGTACAAA
PB28_seqGPC_R11	CCCTAGCCCATCAGAGTTCA
PB29_seqGPC_R12	CATAGGGCTTTCGTGTCTGG
PB30_seqGPC_R13	ATGAACCCAACCAGTTCGTC
PB31_seqGPC_R14	GCTAACAAACAGGTGCCACA
PB49_seqpGPC_R	TGTCTACAGCAGTGCGCTTT

Table 4.8. Primers to sequence final constructs for *NAM-B1* transgenics.

4.2.3 Plant materials

4.2.3.1 T_0 transgenics for gene and protein expression analysis and preliminary ChIP-seq experiment

The plasmids containing FLAG or HA-tagged *NAM-B1* expressed under the *NAM-B1* native promoter were sent to the Crop Transformation Service at the National Institute for Agricultural Botany (NIAB). Transformation into the spring wheat variety Fielder was carried out using a patented technology (PureIntro; WO 95/06722) from Japan Tobacco.

T_0 transformed plants were returned at the 3-4 leaf stage with copy number information. Control plants were also provided. These were regenerated from immature wheat embryos which were not inoculated with agrobacterium. These plants were potted on into 1 L pots containing cereal mix (2.2.1.1) and grown in a controlled environment room (CER) with 16 h light at 20 °C and 8 h dark at 15 °C at 70 % humidity. The distal two-thirds of the main tiller flag leaf was harvested at 20 days after anthesis (DAA). The flag leaves were immediately divided lengthways into separate samples for mRNA and protein expression analysis which were snap frozen in liquid nitrogen and stored at -80 °C. The distal two-thirds of the second tiller's flag leaf was harvested 20 DAA and snap frozen in liquid nitrogen and stored at -80 °C for use in a preliminary ChIP-seq experiment. All tillers of T_0 plants were bagged and T_1 seeds were harvested.

4.2.3.2 T_1 transgenics for gene and protein expression analysis and replicated ChIP-seq experiment

The T_1 seeds were germinated on filter paper in the dark for 48 hs at 4 °C followed by 48 hs at room temperature. Seeds were sown into peat and sand mix in p40 trays and transplanted into FP9 pots containing cereal mix after 2-3 weeks. These T_1 plants were grown in the CER under the same conditions as the previous generation (4.2.3.1). The main tiller flag leaf blade and peduncle (flag leaf sheath removed) were harvested at 20 DAA. The middle 2 cm of the flag leaf blade was snap frozen in liquid nitrogen for gene expression analysis by qPCR. The rest of the flag leaf blade and peduncle were put into separate 50 ml centrifuge tubes containing 20 ml ice-cold fixative (1 % formaldehyde and 1x phosphate-buffered saline (PBS)) and kept on ice for a maximum of 1 h whilst other samples were harvested. The fixative was vacuum infiltrated into the samples by applying a vacuum for 3 min, releasing the vacuum and re-applying for 7 min, ensuring all tissue pieces were fully immersed. This step cross-links proteins to the DNA and

fixed NAM-B1 protein to the DNA it was interacting with *in planta*. Subsequently 1.5 ml of ice-cold 2 M glycine was added to each sample, and tubes were shaken to mix. Samples were vacuum infiltrated for a further 5 min to prevent further cross-linking of proteins to the DNA. Samples were then washed in dH_2O, patted dry with paper towel and frozen at -80 °C.

4.2.4 Characterisation of transgenic plants

4.2.4.1 RNA extraction, DNAse treatment and cDNA synthesis

These steps were carried out as described in 2.2.1.

4.2.4.2 Quantitative PCR (qPCR) primer design and testing

Primers were designed with the forward primer binding to the FLAG or HA tag sequence of the transgene to provide specificity, and the reverse primer within the first exon (which might also amplify homoeologous copies of *NAM* genes). All product sizes were 81-176 bp long to allow use in qPCR. Primers were tested using standard PCR conditions (**Table 4.10** and **Table 4.11**).

Primer	**Sequence**	**Notes**
PB58_HA-F1	CCCTATGACGTCCCGGACTA	Binds HA-tag
PB59_HA-F2	ATGCGGGCTATCCCTATGAC	Binds HA-tag
PB60_HA-F3	GGTTGATTAACATCTTTTACCCATACG	Binds HA-tag
PB61_FLAG-F1	TGACATCGACTACAAGGATGACG	Binds FLAG-tag
PB62_FLAG-F2	GACTACAAAGACCATGACGGTGA	Binds FLAG -tag
PB63_FLAG-F3	TCATGACATCGACTACAAGGATGA	Binds FLAG -tag
PB64_NAM-B1_R1	TGATACCGCGTTGCTTTTTG	Binds NAM-B1 exon 1
PB65_NAM-B1_R2	GGTGATACCGCGTTGCTTTT	Binds NAM-B1 exon 1
PB66_NAM-B1_R3	CTGATGGTGATACCGCGTTG	Binds NAM-B1 exon 1

Table 4.9. Primers for qPCR of HA or FLAG-tagged *NAM-B1*.

Reagent	Volume (µl)
Forward primer (2 µM)	1.5
Reverse primer (2 µM)	1.5
5x buffer (M791B, Promega)	3
dNTP (2 mM)	1.5
DNA template (15 ng/µl)	3
H_2O	4.5
DNA polymerase	0.07

Table 4.10. Standard PCR reaction mix composition.

Step	Temperature (°C)	Time (min)
1	95	5:00
2	95	0:30
3	60	0:15
4	72	0:20
5	go to step 2 39x	
6	72	10:00
7	10	∞

Table 4.11. Standard PCR cycling conditions.

Three pairs of primers for each construct produced the expected band size and these were tested for primer efficiency using serial dilutions cDNA (4 concentrations, each 10 fold diluted). The standard qPCR reaction mix (**Table 4.12**) was used with standard qPCR cycling conditions (**Table 4.13**).

	Volume (µl)
LightCycler® 480 SYBR Green I Master 2x (04707516001, Roche)	5
cDNA (1 in 10 dilution)	2
F primer (2 µM)	1.25
R primer (2 µM)	1.25
H_2O	0.5
TOTAL	**10**

Table 4.12. Standard qPCR composition.

Step	Temperature (°C)	Time (min)
1	95	5:00
2	95	0:10
3	60	0:15
4	72	0:20
5	78	0:01
6	Repeat steps 2-5, 45 times	
7	Cool to 60	
8	Melt curve to 95	5 acquisitions per second

Table 4.13. Standard qPCR cycling conditions.

Melt curves were checked for a single peak, which indicated the primers specifically amplified only one product. Primer efficiencies were calculated by:

$$\text{Efficiency} = -1 + 10^{(-1/\text{gradient})}$$

The gradient is obtained by plotting crossing threshold (Ct) value from qPCR (initial point when the fluorescence from the PCR amplification is detected) on the y-axis, against the log of the concentration of DNA on the x-axis. The primer pairs with the efficiency closest to 100 % were selected to analyse the transgenic plants for expression levels (80-99 % for PB58 + PB66 which amplify the HA construct, and 61-88 % PB63 + PB66 which amplify the FLAG construct).

4.2.4.3 Transgenic construct mRNA expression analysis by qPCR

Carried out as in 2.2.1.6 and 2.2.1.7 using 4 reference genes: *actin*, *ubiquitin*, *TaGAPDH* and *TaEF1a* (**Table 2.3**).

4.2.4.4 Preparation of extracts for protein expression analysis

The samples (flag leaf blade or peduncle) were ground in a liquid nitrogen-cooled mortar and pestle to a fine powder. 2.5 ml of ice cold extraction buffer (100 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.1 % (v/v) Triton X-100, 1 % (w/v) polyvinylpolypyrrolidone (PVPP) , 5 mM ethylenediaminetetraacetic acid (EDTA), 10 % glycerol, 2 mM dithiothreitol (DTT), protease inhibitor cocktail (P9599, Sigma, diluted 100x)) was added with a small spoon of acid-washed sand. Samples were further ground for 2 min in buffer. The samples were then filtered through 2 layers of miracloth into 15 ml centrifuge tubes and squeezed to ensure all extract went through the cloth. The filtered sample was transferred to Eppendorf tubes and centrifuged at 10,000 x g for 30 min at 4 °C. The supernatant was kept on ice whilst a small portion was used for a Bradford assay.

4.2.4.5 Bradford assay

1. Get a cuvette for each sample and 7 for the standard curve
2. To sample cuvettes add 790 µl water and 10 µl protein
3. To standard curve cuvettes add 800 µl water
4. To standard curve cuvettes:
    i. 0 µl BSA = blank
    ii. Remove 2.5 µl water, add 2.5 µl BSA = 2.5 µg
    iii. Remove 5 µl water, add 5 µl BSA = 5 µg
    iv. Remove 7.5 µl water, add 7.5 µl BSA = 7.5 µg
    v. Remove 10 µl water, add 10 µl BSA = 10 µg
    vi. Remove 15 µl water, add 15 µl BSA = 15 µg
    vii. Remove 20 µl water, add 20 µl BSA = 20 µg
5. Add 200 µl Bradford reagent to all cuvettes (standard and samples)
6. Mix well by holding parafilm over top and inverting several times
7. Incubate 5 min room temperature
8. Blank at OD 595 nm
9. Read all standards and samples at 595 nm in the order that the Bradford's reagent was added
10. Using the standard curve calculate the amount of protein in each sample, thus now much to load to have equal quantities of sample.

4.2.4.6 Protein gels

Sodium dodecylsulphate (SDS) polyacrylamide gels were prepared as per **Table 4.14**. Gels were run in tris-glycine running buffer (125 mM Tris-HCl pH 8.3, 960 mM glycine, 5 % SDS) at 150 V for 1 -1.5 h. Gels were then used for immunoblotting or stained with InstantBlue™ (New England Biolabs) to visualise bands.

Component	Stacking Gel (ml)	Separating Gel (ml)
H_2O	6.8	13.2
30 % acrylamide mix	1.7	16
1.0 M Tris (pH 6.8)	1.25	-
1.5 M Tris (pH 8.8)	-	10
10 % SDS	0.1	0.4
10 % ammonium persulfate	0.1	0.4
TEMED (161-0801, Bio-rad)	0.01	0.16

Table 4.14. SDS polyacrylamide gel composition.

4.2.4.7 Immunoblotting to detect tagged-NAM-B1 protein using anti-HA or anti-FLAG antibodies

Proteins were transferred from SDS polyacrylamide gels to nitrocellulose membrane using Trans-Blot SD transfer apparatus (Bio-Rad) according to the manufacturer's instructions. Blotting was carried out at 4 °C at 95 mA for ~1 h or 35 mA overnight. Membrane were blocked overnight at 4 °C in 1x tris-buffered saline (TBS, 0.05 M Tris, 0.2 M NaCl, pH 7.4) with 5 % dried milk powder. They were incubated with the primary antibody (monoclonal mouse anti-FLAG M2 antibody (F1804, Sigma) or rat anti-HA (11867423001, Roche)) diluted 1:5,000 in 1 % milk in 1x TBS at room temperature for 2 h. The membrane was washed and incubated with secondary antibody (goat anti-mouse conjugated to horse radish peroxidase, sc-2005, Santa-Cruz) diluted 1:10,000 in 1 % milk in 1x TBS for 1 h at room temperature. Membranes were developed using Pierce ECL Western Blotting Substrate (32106, Pierce) according to the manufacturer's instructions and imaged using the chemiluminescence detection system ImageQuant LAS 500 (GE Healthcare).

4.2.5 Sample preparation for ChIP-seq

The T_1 samples were already cross-linked when they were harvested (4.2.3.2), however the T_0 samples were frozen without cross-linking. Therefore the T_0 samples were removed from the freezer and cross-linked by immersion into fixative as described in (4.2.3.2) before proceeding to nuclei extraction (4.2.5.1).

For T_0 samples, the highest expressing 3 lines were selected for *HA-NAM-B1* and for *FLAG-NAM-B1*. The 2nd tiller's flag leaf was harvested as described above (4.2.3.1) and the 3 independent lines were pooled for each construct.

For T_1 samples only the three highest expressing *FLAG-NAM-B1* transgenic lines were used (events CU29.1, CU31.4 and CU31.6). 3 plants from each transgenic line were combined to make a single biological replicate, and 2 biological replicate from each line were processed. The flag leaf and peduncle tissues from these biological replicates were kept separately to allow comparisons between tissues. In parallel 3 control lines were sampled in exactly the same way to give equal numbers of *FLAG-NAM-B1* and control samples. Therefore a total of 24 samples were sequenced (3 independent lines x 2 biological replicates x 2 tissues x 2 genotypes (*FLAG-NAM-B1* and control)).

4.2.5.1 Nuclei extraction

Cross-linked samples were ground in pre-cooled pestle and mortars in liquid nitrogen to a fine powder. HB buffer was added (**Table 4.15**), 5 ml per sample to the mortar and allowed to melt before grinding for 1-2 min. The samples were filtered through one layer of miracloth into pre-chilled 50 ml centrifuge tubes. The mortars were rinsed with a further 2 ml of HB buffer which was used to rinse the miracloth. The filtrate was transferred into 15 ml centrifuge tubes carefully which already contained a 2 ml cushion of Percoll solution (**Table 4.15**). The tubes were centrifuged at 2,500 x g for 10 min at 4 °C. The supernatant above the HB/Percoll interface was removed and discarded. The green layer at the interface between the HB/Percoll was collected which contains nuclei and chloroplasts. The nuclei were resuspended in 5 ml of nuclei wash buffer (**Table 4.15**) and centrifuged at 3,000 x g for 5 min at 4 °C. The supernatant was discarded and nuclei were washed a further 4 times in 1 ml of Triton wash buffer to remove the chloroplasts (**Table 4.15**). Nuclei were resuspended in 0.5 ml freezing buffer (**Table 4.15**) at -80 °C.

Buffer name	**Composition**
HB Buffer	10 mM Tris pH 7.5, 10 mM NaCl, 10 mM $MgCl_2$, 250 mM sucrose, 1.0 M hexylene glycol, 20 % glycerol, 2 mM DTT, 1x protease inhibitors (Protease inhibitor cocktail, Sigma)
Percoll solution	81 % Percoll, 10 mM Tris pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 250 mM sucrose, 0.5 M hexylene glycol, 2 mM DTT, 1x protease inhibitors (Protease inhibitor cocktail, Sigma)
Nuclei wash buffer	10 mM Tris pH 8.0, 25 mM KCl , 5 mM $MgCl_2$, 1 mM DTT, 1x protease inhibitors (Protease inhibitor cocktail, Sigma)
Trition wash buffer	10 mM Tris pH 8.0, 10 mM $MgCl_2$, 250 mM sucrose, 1 % Triton X-100, 1 mM DTT, 1x protease inhibitors (Protease inhibitor cocktail, Sigma)
Freezing buffer	10 mM Tris pH 8.0, 10 mM $MgCl_2$, 250 mM sucrose, 1 mM DTT, 1x protease inhibitors (Protease inhibitor cocktail, Sigma), 50 % glycerol
Nuclei lysis buffer	50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1 % SDS, 1x protease inhibitors (Protease inhibitor cocktail, Sigma)
ChIP dilution buffer	16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA pH 8.0, 167 mM NaCl, 1.1 % Triton X-100, 0.1 % NP-40 (IGEPAL CA630)
Low salt washing buffer	20 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 150 mM NaCl, 1 % Triton X-100
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0
TBS	50 mM Tris, 200 mM NaCl, pH 7.4

Table 4.15. Buffer compositions for ChIP.

4.2.5.2 Sonication

Samples were defrosted on ice and centrifuged at 3,000 x g for 5 min at 4 °C. The supernatant was discarded and nuclei were resuspended in 300 µl nuclei lysis buffer (**Table 4.15**). A small portion (30 µl) of each sample was saved as the input fraction and frozen at -80 °C.

The remaining 270 µl of each sample was sonicated in a cooled water bath using a Bioruptor Plus sonicator (Diagenode) on medium power for 5 min (30 sec on, 30 sec off x 5). The tubes were mixed briefly to remove any bubbles and the sonication was repeated twice using the same conditions. 6 samples including control and transgenic samples were sonicated simultaneously.

Efficacy of sonication was checked by de-crosslinking a small portion of the sample and running on a gel to check for the correct sized smear of DNA (100-500 bp):

1. To 10 µl of sonicated chromatin add 1 µl 10 % chelex resin
2. Heat for 10 min at 95 °C
3. Cool to room temperature, add 2 µl 1:10 dilution of RNAse A (158922, Qiagen) and incubate 30 min 37 °C
4. Add 1 µl proteinase K (158918, Qiagen) and incubate at 55-65 °C for 30 min
5. Heat for 10 min at 95 °C to inactivate proteinase K
6. Centrifuge for 5 min 10,000 x g supernatant
7. Run the supernatant on a 1.5 % agarose gel at 70 V for ~45 min

4.2.5.3 Immunoprecipitation

Samples were diluted ten-fold with ChIP Dilution Buffer (**Table 4.15**). To keep samples in a small enough volume for eppendorf tubes each chromatin sample was divided into two technical replicates. Samples were centrifuged at 10,000 x g for 10 min. The supernatant was transferred to fresh Eppendorf tubes. 20 µl was kept as IP input protein for Western blot and stored at -80 °C.The ten-fold dilution is essential to bring the SDS concentration from the nuclei lysis buffer to 0.1 % because higher concentrations may denature antibodies.

EZview Red ANTI-FLAG M2 (F2426, Sigma) or ANTI-HA (F6779, Sigma) Affinity Gel beads were carefully mixed until completely and uniformly suspended. 720 µl of the beads

were transferred to a clean Eppendorf tube on ice using a wide orifice pipette to allow unrestricted flow of the bead suspension. Beads were washed 3 times:

1. Add 500 µl of ChIP dilution buffer to the tube, vortex, and centrifuge for 30 sec at 8,200 x g.
2. Carefully remove the supernatant with pipette and set tube with the bead pellet on ice. Repeat steps 1 and 2.
3. For third wash split the beads into 24 tubes, each with 30 µl of original beads (ie. 15 µl actual beads, 15 µl buffer). Centrifuge and remove supernatant carefully.

Each individual diluted sample was added to a tube of washed beads and incubated on a rotator overnight at 4 °C.

4.2.5.4 Bead washing

After the immunoprecipitation incubation over night the beads were washed to remove unbound proteins:

1. Add 0.5 ml of low salt washing buffer (**Table 4.15**) to each tube and combine the 2 technical replicates into one new Eppendorf tube. (Add buffer on ice, pipette up and down and incubate for 1 min)
2. Centrifuge 30 sec at 6,000 x g, remove supernatant, keep beads
3. Wash with 1 ml low salt washing buffer
4. Wash twice with TE buffer (**Table 4.15**)
5. Wash once with TBS (**Table 4.15**)

4.2.5.5 Elution and reverse cross-linking of chromatin

Protein-DNA complexes were eluted from the beads and the cross-linking was reversed:

1. For 12 samples: make 2 tubes with each: 54 µl of 5 µg/µl HA (I2149, Sigma) or 5 µg/µl 3xFLAG peptide (F4799, Sigma) and 1800 µl TBS (final concentration of the peptide 150 ng/µl).
2. Put 150 µl of 150 ng/µl 3xFLAG or HA solution onto each sample of beads.
3. Incubate the samples and controls on a rotator for 30 min at 4 °C.
4. Centrifuge the samples for 30 sec at 8,200 x g.
5. Transfer the supernatants to fresh eppendorf tubes. Be careful not to transfer any resin.
6. Repeat elution again and pool supernatants (300 µl per sample). Keep 20 µl for Western blot (IP'ed protein), store at -80 °C.

7. Add 1 µl of RNaseA (10 mg/ml, Qiagen) to all samples (including total input chromatin from freezer) and incubate for 15-20 min at 37°C.
8. De-crosslink by adding 18 µl 5M NaCl (final concentration 0.3M) and incubate overnight at 65°C.
9. Add 6 µl 0.5M EDTA, 12 µl 1M Tris-HCl pH 6.5, 1 µl proteinase K (20 mg/ml) (Sigma P2308) and incubate 1 h at 45 °C to digest proteins.
10. Phenol-chloroform to remove proteins, then EtOH precipitation in the presence of glycogen: add 1 volume (340 µl) of Phenol-chloroform-isoamyl alcohol (P2069, Sigma), vortex, centrifuge 5 min 10,000 x g.
11. Transfer the supernatant to a new tube and add 1 volume (300 µl) of chloroform; vortex, centrifuge 5 min 10,000 x g.
12. Transfer supernatant (~ 300 µl) to a new tube and add 2 µl of glycogen (20 mg/ml), 1/10 volume sodium acetate (~30 µl, 3 M pH 5.2) and 3 volumes of 100 % ethanol (~900 µl). Incubate -80°C for at least 1 h or -20 °C overnight.
13. Centrifuge 30 min at 4 ºC 13,000 x g
14. Wash pellet with 70 % ethanol and centrifuge 5 min at 4 ºC at 13,000 x g.
15. Air dry the pellet (~ 3 min) and resuspend in 50 µl dH_2O. Allow the samples to sit at room temperature for 30 min and then pipette up and down using filter tips.
16. Store the resuspended chromatin at -20 °C.

4.2.5.6 Library construction and sequencing

Chromatin samples were submitted to The Genome Analysis Centre (TGAC), for library construction and sequencing. For the preliminary samples from T_0 the 4 samples (FLAG, HA and 2 control samples) were sequenced in one run of an Illumina Mi-Seq with 250 bp paired- end reads. For the replicated T_1 samples 12 samples were sequenced per lane on an Illumina Hi-Seq with 100 bp paired-end reads. In all cases one Illumina TruSeq ChIP library was constructed per sample and the libraries were barcoded.

4.2.6 ChIP-seq data analysis

4.2.6.1 Read mapping

Reads were mapped to the pseudomolecules v3.3 using the read aligner bwa (v0.7.5) (Li Heng and Durbin 2009). The algorithm used was bwa-aln which can be used to precisely manipulate the number of mismatches permitted between each read and the reference sequence. The number of mismatches (n) allowed was optimised as described in the results (4.3.2.1.1). Mapped reads were subsequently filtered using samtools (v0.1.19) (Li H. et al. 2009) to produce separate files containing all reads, reads in proper pairs, reads with a mapping quality (MAPQ) over 10 and reads with MAPQ over 30.

4.2.6.2 NAM-B1 binding site (peak) detection

After read mapping Macs2 (v2.0.10) (Zhang Yong et al. 2008) was used to detect regions where NAM-B1 was binding. The binding is detected by comparing the distribution of reads between FLAG or HA- tagged NAM-B1 samples and control un-tagged samples. The resulting samples were named according to **Table 4.16**. Regions which have significantly more reads in the FLAG or HA tagged samples than in the control regions are called as peaks, i.e. putative NAM-B1 binding sites. Macs2 takes into account background levels of variation of sequencing across the genome and the total number of reads per sample. The algorithm used in Macs2 was 'callpeak' and the parameters used were: paired end read files (-f BAMPE), the effective genome size of the pseudomolecules v3.3 was calculated to be $4.21e^{+8}$ (-g $4.21e^{+8}$) (the effective genome size is the mappable region of the genome, therefore removes regions with high repeat content) and the shiftsize (i.e. distance expected between reads) was set as 100 bp (--shiftsize 100). Peak calling was conducted for each sample at the 4 different stringencies of read mapping (all reads, reads in proper pairs, reads in proper pairs with MAPQ > 10 and reads in proper paires with MAPQ > 30). Only significant peaks with q-value < 0.05 were recorded and subsequently analysed. The q-value is the p-value adjusted for multiple testing. Macs2 gives the significance values as the -log10(qvalue) to allow the scale to be more interpretable and this was used in many of the following analyses. The q-value = 0.05 equates to –log10(qvalue) = 1.3, and all –log10(qvalues) larger than 1.3 are statistically significant.

Sample name	Tissue	FLAG-tagged transgenic event used	Control line used	Replicate
L1A	Flag leaf (L)	CU29.1 (event 1)	CU28CON	A
L1B	Flag leaf (L)	CU29.1 (event 1)	CU28CON	B
L2A	Flag leaf (L)	CU31.4 (event 2)	CU30CON	A
L2B	Flag leaf (L)	CU31.4 (event 2)	CU30CON	B
L3A	Flag leaf (L)	CU31.6 (event 3)	CU31CON	A
L3B	Flag leaf (L)	CU31.6 (event 3)	CU31CON	B
P1A	Peduncle (P)	CU29.1 (event 1)	CU28CON	A
P1B	Peduncle (P)	CU29.1 (event 1)	CU28CON	B
P2A	Peduncle (P)	CU31.4 (event 2)	CU30CON	A
P2B	Peduncle (P)	CU31.4 (event 2)	CU30CON	B
P3A	Peduncle (P)	CU31.6 (event 3)	CU31CON	A
P3B	Peduncle (P)	CU31.6 (event 3)	CU31CON	B

Table 4.16. Key to sample names.

4.2.6.3 Visualisation of binding sites

The Integrative Genomics Viewer (IGV) (Robinson James T. et al. 2011, Thorvaldsdóttir et al. 2013) was used to visualise mapped reads and peak regions detected by Macs2 on the pseudomolecules v3.3 alongside gene-model annotations.

4.2.6.4 Repetitiveness of peaks

The repetitiveness of peak regions was investigated by taking all the peak region sequences for each individual sample and BLASTing them against the IWGSC chromosome arm survey sequences. The number of hits with over 99 % identity were recorded for each peak region for each sample and plotted as graphs using R.

This task was accomplished by writing custom perl scripts and R scripts:

1) **bed_to_fasta.pl** - Perl script to convert bed files to fasta files using the reference genome (Appendix 6.2.1.1).

For each separate ChIP-seq sample this script uses the co-ordinates of the peak regions to extract the relevant DNA sequence from the reference genome (the pseudomolecules v3.3). This relies on the program samtools (v0.1.19). The script then splits the sequences for each peak region into separate files (using the script split_fasta.pl see below). This process is then repeated for all the other ChIP-seq samples.

2) **split_fasta.pl** - Perl script to split a fasta file containing multiple sequences into one fasta file per sequence (written by John Nash). (Appendix 6.2.1.2)

3) **blast_promoter_regions_to_chroms_lsf.pl -** Perl script to launch computation jobs which BLAST each peak against all the IWGSC chromosome arm survey sequences, one chromosome at a time. (Appendix 6.2.1.3)

Within one sample this script takes each peak sequence in turn (which was extracted by steps 1) and 2)). It writes temporary files which instruct the Unix cluster to BLAST the sequence against each chromosome of the wheat genome. This script then does the same thing for all the other samples to find out how many BLAST hits there are over 99 % identity for each peak region for each sample.

4) **calc_promoter_region_hits_to_chroms.pl** - Perl script to summarise the number of hits to the IWGSC chromosome arm sequences per peak region (as determined by BLAST) (Appendix 6.2.1.4)

Within one sample this script takes each peak region in turn and creates a table listing how many BLAST hits were found per chromosome for that peak region. The script then adds up the number of BLAST hits for that peak region in total across the whole genome and appends it to a table specific to that sample. This is repeated for all peaks within that sample to create a table listing all peak regions and the number of BLAST hits. This is then repeated for other samples to create a separate table of results for each sample.

5) **R_graphs_blast_promoters.pl -** Perl script to launch an R script to make graphs of the number of hits to the IWGSC chromosome arm survey sequences for each sample (Appendix 6.2.1.5)

For each sample this script creates another script in the statistical programming language R which will plot graphs of the number of BLAST hits (see **6)** below). After the graphs have been made they are renamed after the sample that they represent.

6) **loop_plot_no_blast_hits.r** – R script to plot the number of blast hits for a particular sample as a box plot and two histograms with specific axes (see appendix (Appendix 6.2.1.6)

This script which runs in the statistical programming programme R (v3.0.2) makes a pdf and then prints graphs containing information about the number of BLAST hits for the particular sample being examined.

Similarity between A, B and D homoeologues of peaks were calculated using the best BLAST hit from **blast_promoter_regions_to_chroms_lsf.pl** for each homoeologous chromosome and the data was plotted using R.

4.2.6.5 Shared peaks between samples

Peak regions which are shared between samples were detected by using bedops (v2.3.0) (Neph et al. 2012) using the bedmap mode which retains regions which overlap. Samples were compared in a pairwise manner and any peaks which overlapped at least 10 % were considered to be shared between samples. It was observed that five of the peduncle samples shared many peaks with each other so to further characterise this bedops was used to identify peak regions which were shared between all five peduncle samples.

4.2.6.6 Correlation between sample significance value

For the T_1 replicated ChIP-seq the significance of peaks was compared between samples to establish whether the ranking of peaks was the same between samples. For the leaf samples only L1A and L1B (see key in **Table 4.16**) had a high number of shared peaks. For these two samples the log transformed significance values of the shared peaks were plotted against each other as a scatter plot in excel. The Pearson's correlation was calculated between these two leaf samples. The same correlation analysis was carried out pairwise between all peduncle samples using R.

4.2.6.7 Distribution of genomic features across the pseudomolecule

To determine which regions of the genome NAM-B1 binds (e.g. promoter sequence), genomic features were annotated onto the pseudomolecules v3.3. A gff file which contained the locations of mRNAs had already been created (4.2.1.4). This was used as an input to bedtools (v2.17.0) using the 'flank' option which selected a defined number of base pairs around the mRNAs. The genomic features which were defined were:

a. 5' untranslated regions (UTRs) (0 to -150 bp upstream of start codon)
b. Proximal promoter regions (-150 to -2,000 bp upstream of start codon)
c. Distal promoter regions (-2,000 to -5,000 bp upstream of start codon)
d. 3' UTRs (0 to 200 bp downstream of stop codon)
e. Exons (from the mRNA file)
f. Introns (by subtracting the co-ordinates of exons from the entire mRNA's co-ordinates)

g. Intergenic regions (all other regions)

4.2.6.8 Peak location with respect to genomic features

To assign the peaks to specific genomic features bedtools (v2.17.0) was used. The 'intersect' option determined whether each peak overlapped each type of feature. If peak overlapped with more than one type of feature it was weighted accordingly, i.e. if a peak overlapped with 2 features, only half a peak would be counted for each feature when summing up all peaks to create the graphs of distribution. This avoided biasing results by counting some peaks more than once.

4.2.6.9 Distance from peaks to transcription start site

Bedtools (v2.17.0) was used to calculate the distance from each peak to the nearest mRNA. First the central 1-2 bp of the peak were determined by the option 'slop'. The central position of each peak was then compared to the mRNA positions using bedtools option 'closest'. This identified the closest mRNA to each peak, the distance to that mRNA and whether the peak was 5' or 3' of the mRNA. Peaks which were determined to be within genes (i.e. the distance = 0 bp) were further analysed. These peaks within genes had their co-ordinates compared with the co-ordinates of the overlapping mRNA. How far along the mRNA the peak was located was calculated as a percentage of the mRNA length. These percentages were then converted to base pair distances by multiplying by the average length of mRNAs on the pseudomolecules (1779 bp).

It was necessary to combine the information about peaks which surround mRNAs and peaks within mRNAs before plotting graphs. Peaks which were 5' of the mRNAs had their raw distances to the mRNAs used for graph plotting, but as negative values. Peaks which were within mRNAs had already been adjusted to the average mRNA length (1779 bp), therefore for these peaks their distance through the average mRNA was plotted (as positive values). Peaks which were 3' of mRNAs had their distances to the mRNAs adjusted by adding on the average mRNA length (1779 bp), and were subsequently plotted as positive values. Therefore a peak 100 bp 5' of the transcription start site of the mRNAs would be plotted at -100 bp, a peak 10 % along an mRNA from its start would be plotted at 177.9 bp (10 % x 1779) and a peak 100 bp downstream of an mRNA would be plotted at 1879 bp (100 + 1779).

4.2.6.10 Motif identification

To identify motifs enriched within peak regions the sequences of the peak regions were extracted using bed_to_fasta.pl (4.2.6.4). The motif finding programme Multiple Em for Motif Elicitation (MEME) (v4.9.1) (Bailey Timothy L. et al. 2009) was used to identify common motifs within these sequences. The options specified were 'zero or one occurrence (motif) per sequence', 'revcomp' (motifs were allowed on the both strands) and the programme should search until 5 motifs were identified.

4.2.6.11 GO enrichment of genes identified by ChIP-seq

To determine whether genes identified by ChIP-seq had particular biological functions, enrichment for any GO terms was examined using the software Blast2GO. The nearest gene model to each peak was identified using bedtools, and we verified that the gene model and the peak were present on the same contig from the IWGSC. We then selected only genes which were within 1 kb of a significant peak because binding events which are regulatory are generally close to the gene body.

These genes were analysed using a Fisher's exact test to examine whether the genes regulated by NAM-B1 were specifically over- or under-represented for specific GO terms, compared to the entire pseudomolecule-reference set of mRNAs. The significance cut-off used was a false discovery rate < 0.05. Any mRNAs which had several peaks nearby were only counted once for the enrichment test.

4.2.7 RNA-seq data analysis

4.2.7.1 Experimental details

A previous study sequenced RNA extracted from the flag leaves of *NAM* RNAi and wild type plants at 12 DAA (Cantu et al. 2011). Four replicates were sequenced from RNAi plants and 3 replicates from wild type plants. We re-analysed this data using our new pseudomolecules v3.3 reference.

4.2.7.2 Read mapping

85 bp single-end reads were mapped using bwa (v0.7.5) as previously described (4.2.6.1) except the reference sequence used was the mRNA sequences extracted from the pseudomolecules v3.3 (cDNA_v3.3.fasta) rather than the genomic sequence, and single-end reads were used.

4.2.7.3 Differential expression analysis

Samtools (v0.1.19) was used to count the number of reads aligned to each mRNA, which was used as an input to the differential expression analysis software. Two differential expression analysis methods were used: edgeR (Robinson M. D. et al. 2010) and DESeq2 (Love et al. 2014), to identify differentially expressed genes in wild type plants compared to RNAi plants. Both methods are available as R/Bioconductor packages and assume that the count data has a negative binomial distribution. However edgeR and DESeq2 estimate the variance for each gene (which depends on the dispersion, a key parameter in the final model) using different methods. Both methods rely on information from other genes but use different methods to incorporate this information to adjust the individual genes' dispersions. This results in different dispersion estimates between the two software packages, and some different genes being identified as differentially expressed (although many genes should be detected by both software packages).

4.2.7.4 Gene ontology (GO) enrichment

GO enrichment analysis was carried out as for ChIP-seq target genes (4.2.6.11) using a Fisher's exact test in Blast2GO. Many GO terms which were detected to be over- or under-represented in the differentially expressed genes (436 for upregulated genes, 0 for downregulated genes). To make understand this data two approaches were taken to reduce the complexity.

Firstly the over or under-represented GO terms were analysed with REVIGO which summarises and visualises GO terms (Supek et al. 2011). REVIGO collapses GO terms with similar functions. The amount of similarity to condense the GO terms must be user-specified. In this case the settings used were 'medium similarity' and the algorithm to measure similarity was 'SimRel'. The REVIGO interface provides R scripts for the user's data. These scripts were used as a basis to create graphs in R (v3.0.2).

Secondly GO terms related to biological processes were examined in more detail. The over- or under-representation of level 3 GO terms (which describe broad categories of biological processes) were examined in bar graphs.

4.2.7.5 Motif identification in the promoter regions of differentially expressed genes

The 1000 bp upstream of differentially expressed mRNAs (i.e. putative promoter sequences) were extracted from the pseudomolecules using bedtools (v2.17.0) using the 'flank' option. The extracted sequences were examined for motifs by MEME (see 0).

4.2.8 Comparison of ChIP-seq and RNA-seq target genes

4.2.8.1 Genes identified by both ChIP-seq and RNA-seq

384 genes were detected as differentially expressed by both DESeq2 and edgeR (FDR < 0.1). The only ChIP-seq sample which had any overlapping genes to the differentially expressed genes was L1B proper pairs MAPQ > 30. Out of the 1128 significant peaks only 447 were within 1 kb of the nearest gene on the same contig. Out of these 447 genes which are directly bound by NAM-B1 only 4 were differentially expressed. We used a Chi-squared test with Yates' correction to see if there was a significant overlap between the RNA-seq and ChIP-seq identified genes.

4.2.8.2 ChIP-seq motifs found in differentially expressed contigs from RNA-seq (MAST)

It might be expected that the motif sequences identified to be bound by NAM-B1 would be found in genes regulated by NAM-B1, as identified by RNA-seq. Therefore the MAST tool of the MEME suite (v4.9.1) (Bailey Timothy L. et al. 2009) was used to identify whether each motif was present in contigs containing differentially expressed genes.

4.2.9 ChIP-seq target validation

4.2.9.1 Cloning NAM-B1 for expression in *E. coli*

The full length coding sequence (CDS) of *NAM-B1* was amplified by two rounds of PCR (**Table 4.17**) using standard PCR composition (**Table 4.10**) and conditions (**Table 4.11**) except an extension time of 2 min 30 sec was used.

Step	Forward primer	Reverse primer	Template DNA	Notes
1	PB107	PB109	cDNA from UC1041+GPC, mixture of vegetative tissues (see 2.2.1)	Amplify B genome specific region containing *NAM-B1* CDS (1316 bp)
2	PB112	PB114	Amplified product from step 1	Amplify only the *NAM-B1* CDS (i.e. just start to stop codon) (1218 bp)

Table 4.17. Two rounds of PCR to amplify *NAM-B1* coding sequence (avoiding homoeologues).

The gel purified product was cloned into the PCR8 vector, transformed into DH5α (see 4.2.2.2) and colonies were screened for the insert by PCR using standard composition (**Table 4.10**) and conditions (**Table 4.11**). An extension time of 2 min 30 sec was used and the primers used were M13F and PB114 (**Table 4.18**). The DNA template used was water which had a pipette tip immersed into it which had touched the relevant colony on the plate of bacteria. Colonies with the expected insert were sequenced, as described previously for the *NAM-B1* promoter (4.2.2.2). The sequences of primers used to amplify the CDS and for sequencing are listed in (**Table 4.18**).

Primer	Sequence	Notes
PB107	TGCCCGTATGTGGTGTTCAT	1st step of PCR amplification – homoeologue specific
PB109	CTCTGGTGGGATCATCTGGT	1st step of PCR amplification – homoeologue specific
PB112	ATGGGCAGCTCCGACTCATCT	2nd step of PCR amplification – from start to stop codon
PB114	TCAGGGATTCCAGTTCACG	2nd step of PCR amplification – from start to stop codon
M13F	GTAAAACGACGGCCAGT	Sequencing
M13R	CAGGAAACAGCTATGAC	Sequencing
T7F	TAATACGACTCACTATAGGG	Sequencing
T7R	GCTAGTTATTGCTCAGCGG	Sequencing

Table 4.18. Primers used to amplify and sequence *NAM-B1* coding sequence.

One synonymous mutation was detected in the third exon of the *NAM-B1* CDS in the selected clone. This mutation changed the codon TCG to TCT, which did not result in an amino acid substitution (serine was maintained). The clone was carried forwards because a synonymous mutation would have no effect on the final protein produced.

The PCR8 vector containing *NAM-B1* CDS was used as an insert for an LR reaction to transfer the *NAM-B1* CDS into a vector suitable for expression in *E. coli*. The destination vectors pDEST17 (11803-012, Invitrogen) and pETG-10A (A. Geerlof, EMBL) were used which both had N-terminal his-tags in the backbone to simplify protein purification. The LR reaction was carried out:

1. Add 1 µl insert vector, 1 µl destination vector and 2 µl TE (10 mM Tris, 1 mM EDTA, pH 8) to an Eppendorf tube
2. Add 1 µl LR-Clonase II (11791-100, Invitrogen) and mix well by pipetting
3. Incubate overnight at room temperature
4. Add 0.5 µl 2 µg/µl proteinase K solution (11791-100 Invitrogen) and incubate for 10 min at 37 °C
5. Use 3 µl to transform bacteria

Transformation was carried out as previously described except the antibiotic used was carbenicillin (100 µg/ml) (4.2.2.2). Colonies were screened by PCR as described above using the T7F and PB114 primers (**Table 4.18**). Colonies with the expected insert were sequenced using the T7F and T7R primers (**Table 4.18**).

4.2.9.2 Expression of NAM-B1 in *E. coli*

The two vectors containing *NAM-B1* CDS (pETG-10A and pDEST17) were transformed into differential bacterial strains for optimisation of protein expression. The strains used were BL21-CodonPlus (Agilent), Rosetta-gami 2 (Merck Millipore) and SoluBL21 (AMS Biotechnology). The same transformation protocol was used as for DH5α *E. coli* (4.2.2.2).

The transformed colonies were grown overnight at 37 °C in 10 ml LB with carbenicillin (100 µg/ml). The next day 0.1 ml of the overnight culture was put into 10 ml of fresh LB containing 100 µg/ml carbenicillin and grown until the absorbance at 600 nm reached 0.5. At this point expression of NAM-B1 was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG), generally at the concentration of 1 mM IPTG, but 0.5 mM IPTG was also tested. Growth conditions of the bacteria after

induction were varied to optimise the amount and solubility of protein. The effect of length of expression was tested from 3 h to overnight and temperature was tested from 18 – 37 °C (the temperatures of the cultures were brought down to the relevant temperature before induction by incubation on ice for 10 min).

4.2.9.3 Analysis of NAM-B1 expression by protein gel and Western blot

After NAM-B1 expression had been induced and the cultures incubated in the relevant conditions the cultures were centrifuged at 4,000 x g for 15 min at 4 °C. The pellets were resuspended in bacterial extraction buffer pH 7.5 (50 mM Tris/HCl pH 7.5, 500 mM NaCl, 5 % glycerol, 20 mM Imidazole, 0.1 % Triton X-100, 2 mM DTT, ¼ cOmplete protease inhibitor tablet per 10 ml (11873580001, Roche)). In general for 10 ml of culture was resuspended in 1 ml bacterial extraction buffer. For large scale protein expression 2 L overnight culture was resuspended in 50 ml bacterial extraction buffer.

The bacterial crude extraction was separated into insoluble and soluble proteins by sonication using either a 24 multi-probe or a larger single probe depending on sample size (Sonics, Vibra-Cell). For the large probe 40 % amplitude was used, with 20 sec on and 60 sec off, repeated 8 times. For the small probe amplitude 40 % was used, with with 1 sec on and 3 sec off repeated 20 times. Extracts were kept on ice during sonication. After sonication the samples were centrifuged for 20 min at 10,000 x g at 4 °C. The supernatant was kept as the soluble fraction and the pellet was resuspended as the insoluble fraction in the same amount of bacterial extraction buffer as the original extract.

Soluble and insoluble fractions were run on SDS-polyacrylamide gels (4.2.4.6). Total protein expression was visualised with instant blue (Expedeon). NAM-B1 expression was detected by immunoblot using a tetra-his antibody (34670, Qiagen) as previously described (4.2.4.7).

4.2.9.4 Purification of NAM-B1 from *E. coli*

When suitable conditions for expression had been decided a large scale (2 L) culture was induced to express NAM-B1. The soluble extract was purified by high performance liquid chromatography (HPLC) using an AKTA FLPC (GEHealthcare/Amersham). A 1 ml HisTrap FF column (17-5255-01, GeHealthcare) was washed with 5 ml buffer A (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.1 % Triton X-100). The sample was filtered through a 0.2 µM syringe filter and loaded onto the column. The column was washed with 20 ml buffer A. The column was then eluted by a gradient of

buffer B (buffer A + 500 mM imidazole) which went from 0 % buffer B to 100 % buffer B over 30 ml (the remaining percentage was buffer A). 1 ml fractions were kept throughout the run. The fractions were examined by protein gel and immunoblotting as described above.

4.2.9.5 Probe annealing

Probes to use in electrophoretic mobility shift assays (EMSA) were made by ordering single stranded complementary oligonucleotides (Sigma) and annealing the oligonucleotides. An unlabelled reverse oligonucleotide was annealed to an unlabelled forward oligonucleotide to make an unlabelled probe. To make the biotin labelled probe, a biotin labelled forward oligonucleotide was ordered (Sigma) and annealed to the unlabelled reverse oligonucleotide. To create annealed oligonucleotides:

1. Add 1 µl 100 µM forward oligonucleotide and 1 µl 100 µM reverse oligonucleotide
2. Add 98 µl of buffer (10 mM Tris, 1 mM EDTA, 50 mM NaCl pH 8.0)
3. Incubate in a PCR machine:
4. 5 min at 95 °C
5. 70 x 1 min starting from 95 °C, decreasing 1 °C per cycle
6. Store the annealed oligonucleotides at -20 °C

The oligonucleotides used are listed in **Table 4.19**. These probes were the motifs identified by ChIP-seq for sample L1B proper pairs MAPQ > 30. All probes were 25 bp.

Oligonucleotide (primer) ID	Sequence	Motif ID
PB123_498F	TTGCGCTCGCCTCCCCCTCCAGACC	Peak_498
PB124_498R	GGTCTGGAGGGGGAGGCGAGCGCAA	
PB125_483F	CACCCTCACCTGCGACGCCGAGCTG	Peak_483
PB126_483R	CAGCTCGGCGTCGCAGGTGAGGGTG	
PB128_biotin_904F	GAGCCAGCATCGTCCTCACCTCTCC	Peak_904
PB120_904R	GGAGAGGTGAGGACGATGCTGGCTC	
PB117_94F	GCCGTCTCCTCCTCCTTCCCCTGTG	Peak_94
PB118_94R	CACAGGGGAAGGAGGAGGAGACGGC	
PB245_11F	GGCCGCCACCATGGCCTCCCTCACG	Peak_11
PB247_11R	CGTGAGGGAGGCCATGGTGGCGGCC	
PB251_11F	GGCCGTATATATATATATATATTCACG	Peak_11_mutated
PB253_11R	CGTGAATATATATATATATATACGGCC	
PB254_508F	GCCCCTCCACGTCGCCCACGCTCGC	Peak_508
PB256_508R	GCGAGCGTGGGCGACGTGGAGGGGC	
PB260_508F	GCCCCATCAGGTAGCTCAGACTCGC	Peak_508_mutated
PB262_508R	GCGAGTCTGAGCTACCTGATGGGGC	
PB263_641F	CCACTCCACATCTTCCTCCCGTGCA	Peak_641
PB265_641R	TGCACGGGAGGAAGATGTGGAGTGG	
PB269_641F	CCACTGGAAGGCTCGTAGTAGTGCA	Peak_641_mutated
PB271_641R	TGCACTACTACGAGCCTTCCAGTGG	
PB281_randomF	TGTTCTTGCTTAATGAGTTGCCGGA	Random_1
PB283_randomR	TCCGGCAACTCATTAAGCAAGAACA	
PB284_randomF	ATCTGTGGACTGTCTGGTCCATGCA	Random_2
PB286_randomR	TGCATGGACCAGACAGTCCACAGAT	

Table 4.19. Oligonucleotides for EMSA. Motif sequence (if present) in red.

4.2.9.6 Electrophoretic mobility shift assay (EMSA)

EMSA was carried out using the LightShift Chemiluminescent, EMSA Kit (20148, Thermo Scientific). Assays were carried out as per the manufacturer's instructions and the detailed protocol is described below. All buffers were provided with the kit.

Make the gel and prepare the binding reactions:

1. Make native polyacrylamide gels (5 %) by combining the reagents in **Table 4.20** and pouring into 8 x 8 x 0.1 cm protein gel plates (BioRad) used only for EMSA.
2. Pre-run gel at 100 V for 60 min at 4 °C.
3. Thaw all binding reaction components on ice.
4. On ice prepare binding reactions by adding the components in the order listed in the **Table 4.21**. Do not vortex reaction tubes. Incubate reactions at room temperature for 20 min between adding the protein and biotin labelled DNA to ensure unlabelled DNA can be bound preferentially by the protein if present.

5. Incubate binding reactions for a further 20 min at room temperature
6. Add 5 µL of 5x loading buffer to each 20 µL binding reaction, pipetting up and down several times to mix.

Component	2 gels	1 gel
19:1 acrylamide mix	5 ml	2.5 ml
10X TBE gel buffer	1 ml	0.5 ml
water	14 ml	7 ml
25 % APS	120 µl	60 µl
TEMED	50 µl	25 µl

Table 4.20. Components to make 5 % native polyacrylamide gels.

Component	Final Amount	Labelled probe only (µl)	Labelled probe and NAM-B1 (µl)	Labelled probe, unlabelled probe and NAM-B1 (µl)
water	-	To total volume 20	To total volume 20	To total volume 20
10x binding buffer	1x	2	2	2
1 µg/µl Poly(dI·dC)	50 ng/µl	1-4	1-4	1-4
50 % glycerol	2.50 %	(1)	(1)	(1)
100 mM MgCl2	5 mM	(1)	(1)	(1)
1 % NP-40	0.05 %	(1)	(1)	(1)
1 M KCl		(1)	(1)	(1)
200 mM EDTA		(1)	(1)	(1)
Unlabelled Target DNA (1 µM = 1 pmol/µl) or 50 µM	4 pmol	-	-	4
Protein (NAM-B1)	varied	-	2	2
Biotin-Target DNA (dilute 1 µM stock 100x to get 10 nM = 10 fmol/µl)	20 fmol	2	2	2
Total Volume		20	20	20

Table 4.21. Binding reaction components for EMSA.

Run the binding reactions on the gel

1. Switch off current to the polyacrylamide gel.
2. Load 20 µL of each sample.
3. Run gel at 100 V until the bromophenol blue reaches ~2/3 to 3/4 down the gel.

Transfer binding reactions to a nylon membrane

1. Transfer the gel to a pre-soaked nylon membrane using the Trans-Blot SD transfer apparatus (Bio-Rad) running at 380 mA (~100V) for 30 min at 4 °C in 0.5 x TBE.
2. Remove the membrane from the transfer equipment and place it on a paper towel for 5 sec to allow the buffer to absorb to the membrane
3. UV-crosslink the membrane 45-60 sec using the auto crosslink function of the UV Stratalinker 2400 (Stratagene). Settings: 120 mJ/cm2 with 254 nm bulbs.

Washing and detection of biotin-labelled DNA

1. Warm the blocking buffer and the 4x wash buffer to 37-50 °C in a water bath until all particulate matter is dissolved. All incubation steps are with gentle shaking unless otherwise specified.
2. Incubate the membrane in 20 ml blocking buffer for 15 min
3. Incubate the membrane with 20 ml blocking buffer containing 66.7 µl stabilised streptavidin-horseradish peroxidase conjugate (1:300 dilution) for 15 min
4. Rinse the membrane with 20 ml 1x wash buffer in a new container
5. Wash the membrane 4 x 5 min in 20 mL of 1x wash solution.
6. In a new container incubate the membrane in 30 ml substrate equilibration buffer
7. Blot the membrane on paper towel to remove excess buffer and place in a new container.
8. Add 8 ml of 1:1 luminol/enhancer solution : stable peroxide solution. Incubate for 5 min without shaking.
9. Blot the membrane and wrap it in cling film.
10. Detect chemiluminescence using the ImageQuant LAS 500 (GE Healthcare)

4.3 Results

4.3.1 Materials generated

4.3.1.1 Pseudomolecules represent the gene-rich portion of the wheat genome

4.3.1.1.1 Pseudomolecules gene content and size

We selected all IWGSC contigs over 200 bp in length (10,880,661 contigs). From these we created a set of wheat pseudomolecules by selecting the contigs which contained a gene model (Krasileva et al. 2013) or a *de novo* transcript assembly (unpublished, Martin Trick) (77,795 contigs). These contigs were ordered with reference to a Chinese Spring x Paragon mapping population. For each chromosome in addition to an ordered pseudomolecule, an unordered pseudomolecule was also created because many contigs did not contain markers (40,872 contigs could not be ordered, 36,923 were ordered (**Table 4.22**)). The N50s of contigs which contained genes and were therefore included in the pseudomolecules were 20-23 fold higher than the N50 of all IWGSC contigs (**Table 4.22**). The mean length of chromosomes were similar between the ordered and unordered pseudomolecules (**Table 4.23**) and for each chromosome group (e.g. 1A) the unordered and ordered pseudomolecules were similar in length (**Figure 4.5**). Overall the gene-rich pseudomolecules were 659 Mb in length which is ~4 % of the expected 17 Gb wheat genome. The pseudomolecules contain 75,419 gene models of which 42 % are on ordered chromosomes and 58 % are on unordered chromosomes across the genome (**Figure 4.6**). These are quite evenly distributed between the A, B and D genomes, although the D genome has fewer genes (**Figure 4.7**) as has been previously observed (IWGSC 2014).

	Number of contigs	**N50 of contigs**
IWGSC contigs >200 bp	10,880,661	326 bp
Ordered pseudomolecules	36,923	7,591 bp
Unordered pseudomolecules	40,872	6,638 bp

Table 4.22. Contigs in the pseudomolecules.

	Mean length of chromosomes	**Total number of gene models**
Ordered pseudomolecules	15,833,956 bp	31,640
Unordered pseudomolecules	15,578,619 bp	43,779

Table 4.23. Length and gene content of pseudomolecules.

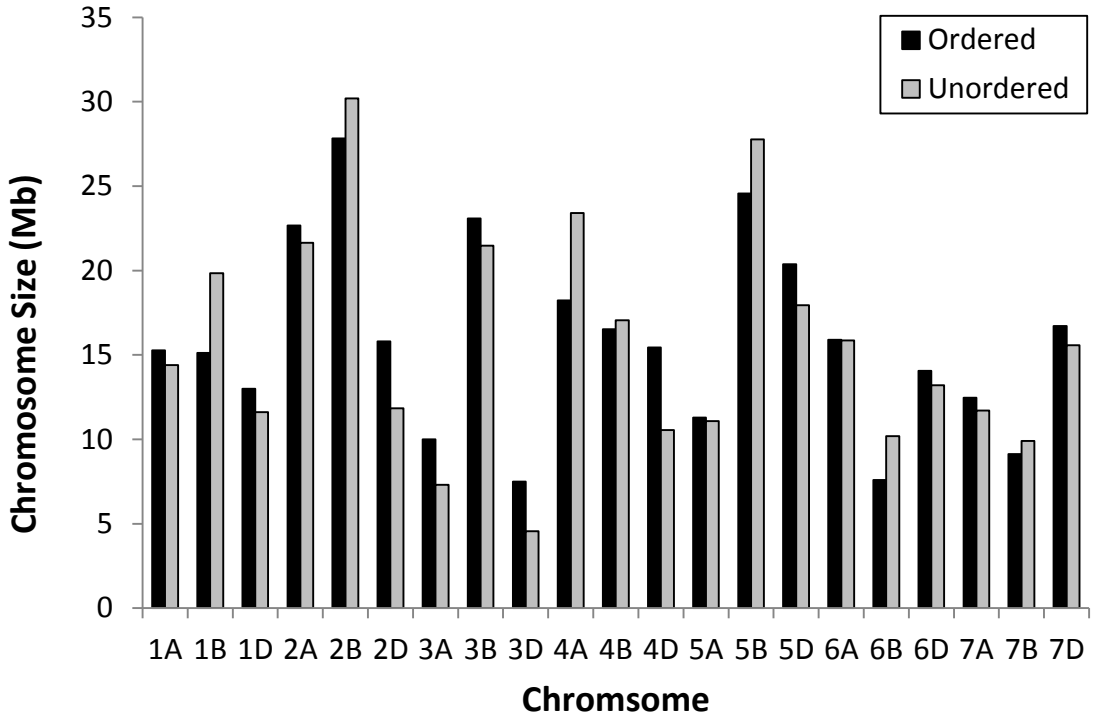


Figure 4.5. Lengths of individual chromosomes in the pseudomolecule.

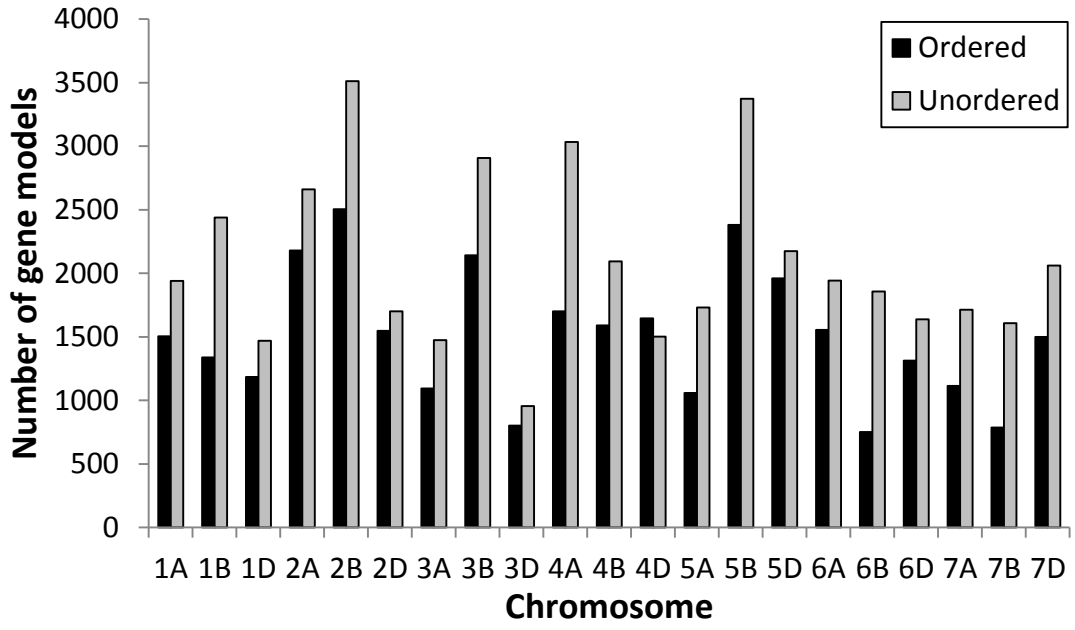


Figure 4.6. Gene distribution across chromosomes on the pseudomolecules.

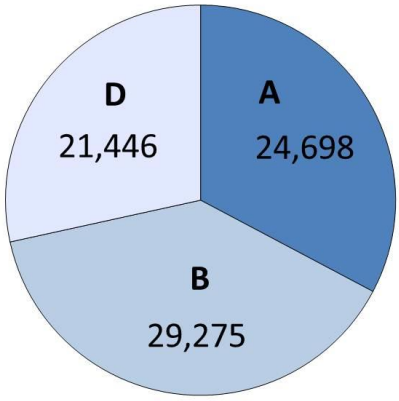


Figure 4.7. Distribution of genes across homoeologous pseudomolecules.

4.3.1.1.2 Gene ontology (GO) annotation of the mRNAs

The software Blast2GO (v2.7.2) was used to annotate the mRNAs within the pseudomolecules. This software uses BLAST to predict the function of genes by comparison with other species. GO terms with information about biological processes, cellular compartment and molecular function are assigned to each mRNA.

Over 45,000 mRNAs were assigned GO terms (**Figure 4.8**), however 35 % of sequences did not have a BLAST hit in the database therefore were not used in subsequent GO enrichment analyses.

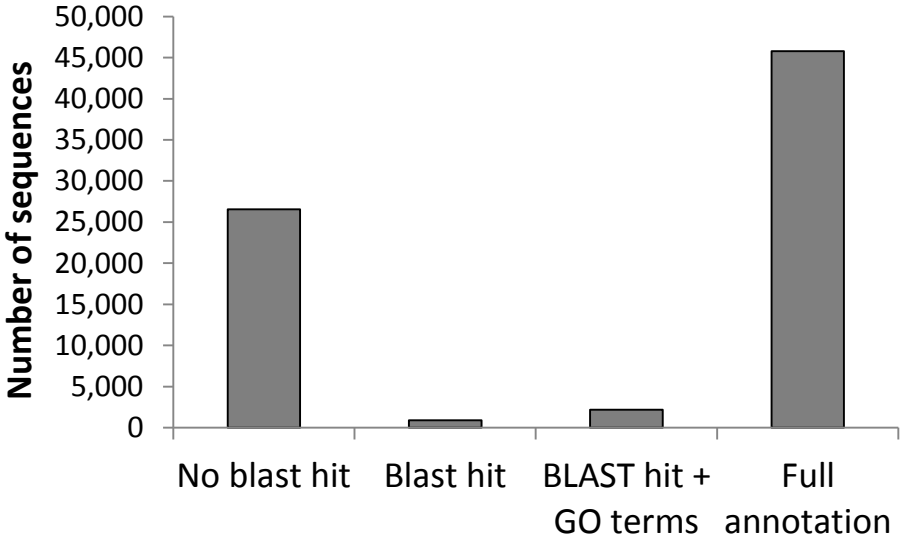


Figure 4.8. Annotation of mRNAs in pseudomolecules v3.3.

4.3.1.1.3 Annotation of mRNAs using rice cDNAs

All mRNA sequences were compared to the rice genome cDNAs (v7.0) using BLAST (v2.2.28). In total 36,659 mRNAs had a hit with a percentage identity >70 % and the majority of these matches were highly similar (31,108 mRNAs were over 80 % similar to the rice cDNA).

4.3.1.2 Characterisation of T_0 transgenic plants

T_0 transgenic plants were provided by NIAB containing the HA or FLAG-tagged *NAM-B1* constructs (**Figure 4.9**). To characterise these plants copy number had been determined by qPCR at NIAB and we developed primers to examine expression level of the constructs by quantitative PCR (qPCR).

Figure 4.9. HA and FLAG-tagged *NAM-B1* constructs.

4.3.1.2.1 Copy number analysis of T_0 plants

Copy number was determined by qPCR by Emma Wallington (NIAB). Several lines for both HA and FLAG-tagged NAM-B1 had a single insertion of the transgene, however the majority of lines had multiple insertions (**Figure 4.10**).

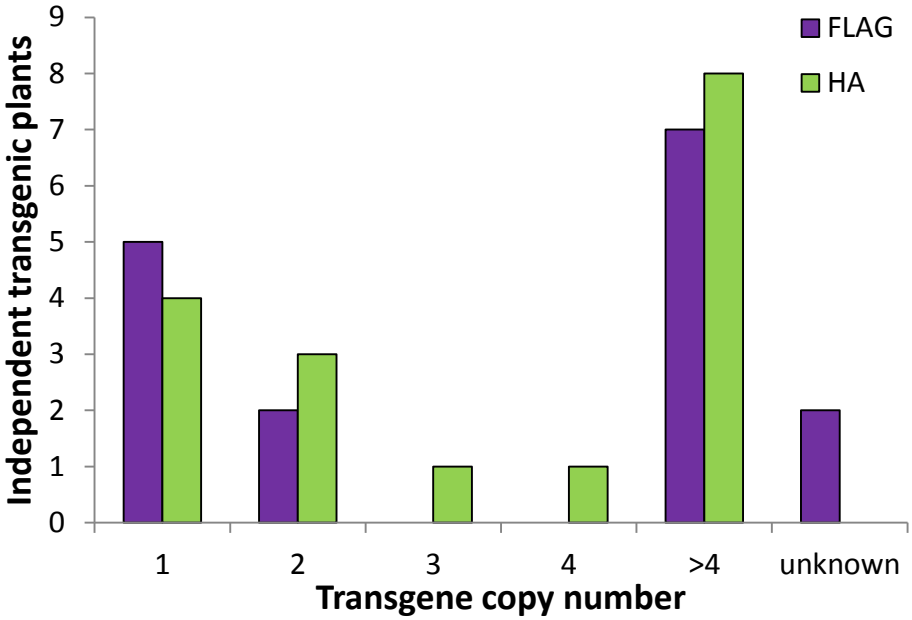


Figure 4.10. Copy number analysis of transgenic plants containing tagged-*NAM-B1*.

4.3.1.2.2 RNA expression of T_0 plants

Quantative PCR (qPCR) was carried out to determine which transgenic lines showed expression levels of the FLAG or HA-tagged *NAM-B1* most similar to those from the endogenous copy. Three of the lines expressing *FLAG-NAM-B1* showed a similar expression level to the endogenous *NAM-B1* expression (highlighted in green, **Figure 4.11**). However the *HA-NAM-B1* construct was expressed at low levels in all lines, with a maximum expression level of 2 % of the endogenous *NAM-B1* (**Figure 4.12**).

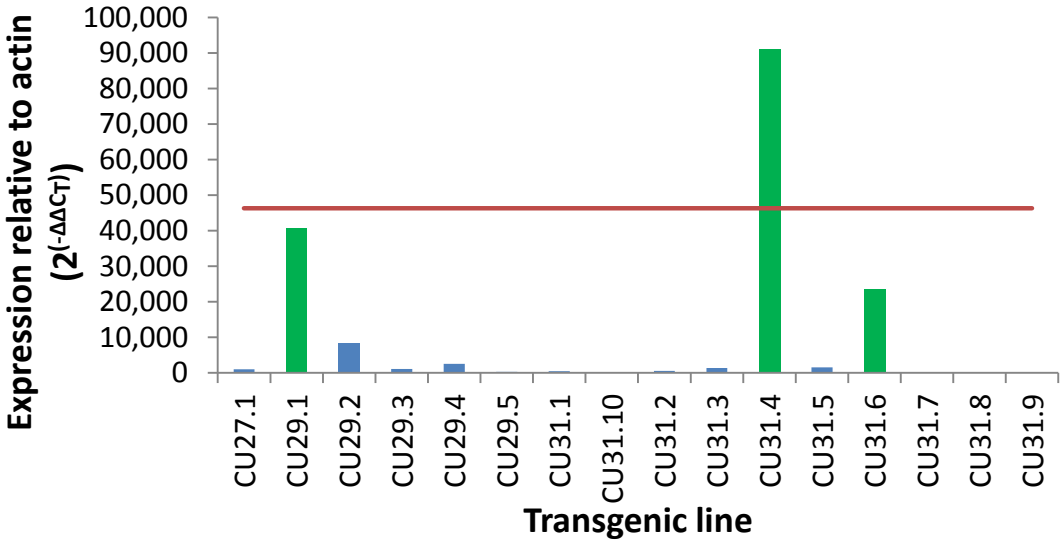


Figure 4.11. Expression level of *FLAG-NAM-B1* relative to actin expression. The red horizontal line is the level of expression expected from the native promoter. The green individuals were used for the preliminary ChIP-seq experiment and continued into the next generation.

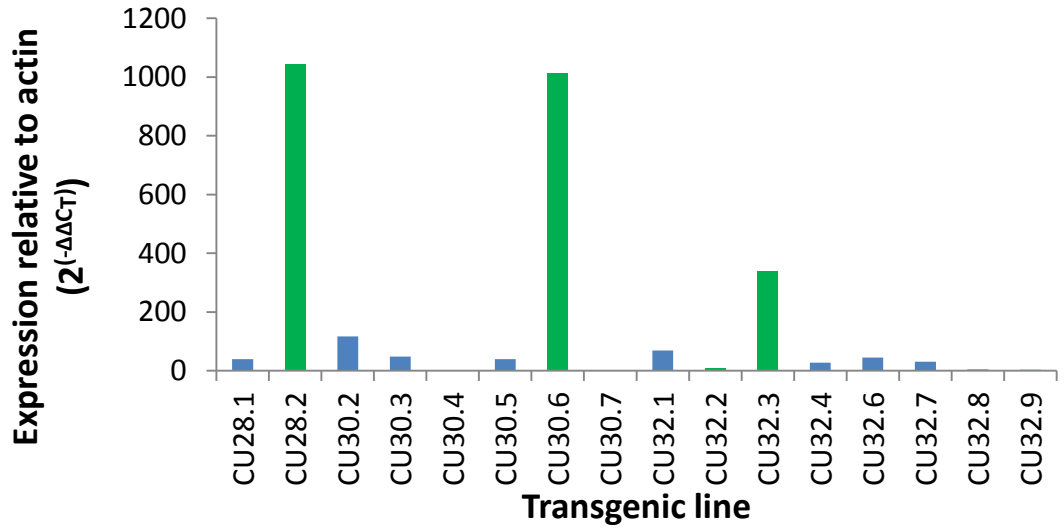


Figure 4.12. Expression level of *HA-NAM-B1* relative to actin expression. The green individuals were used for the preliminary ChIP-seq experiment.

4.3.1.3 Characterisation of T_1 generation transgenics

The seeds of the T_0 transgenic plants were harvested and the three lines expressing *FLAG-NAM-B1* at the most similar level to native gene expression were sown to be used in the replicated ChIP-seq experiment.

4.3.1.3.1 RNA expression levels in T_1

We examined the expression levels of *FLAG-NAM-B1* in the T_1 plants to ensure that the constructs had not been silenced (**Figure 4.13**). The graph shows the plants we selected to use in ChIP-seq. These plants had expression levels similar to the native *NAM-B1* expression in the variety UC1041+GPC (the transformed variety had a null allele for *NAM-B1*).

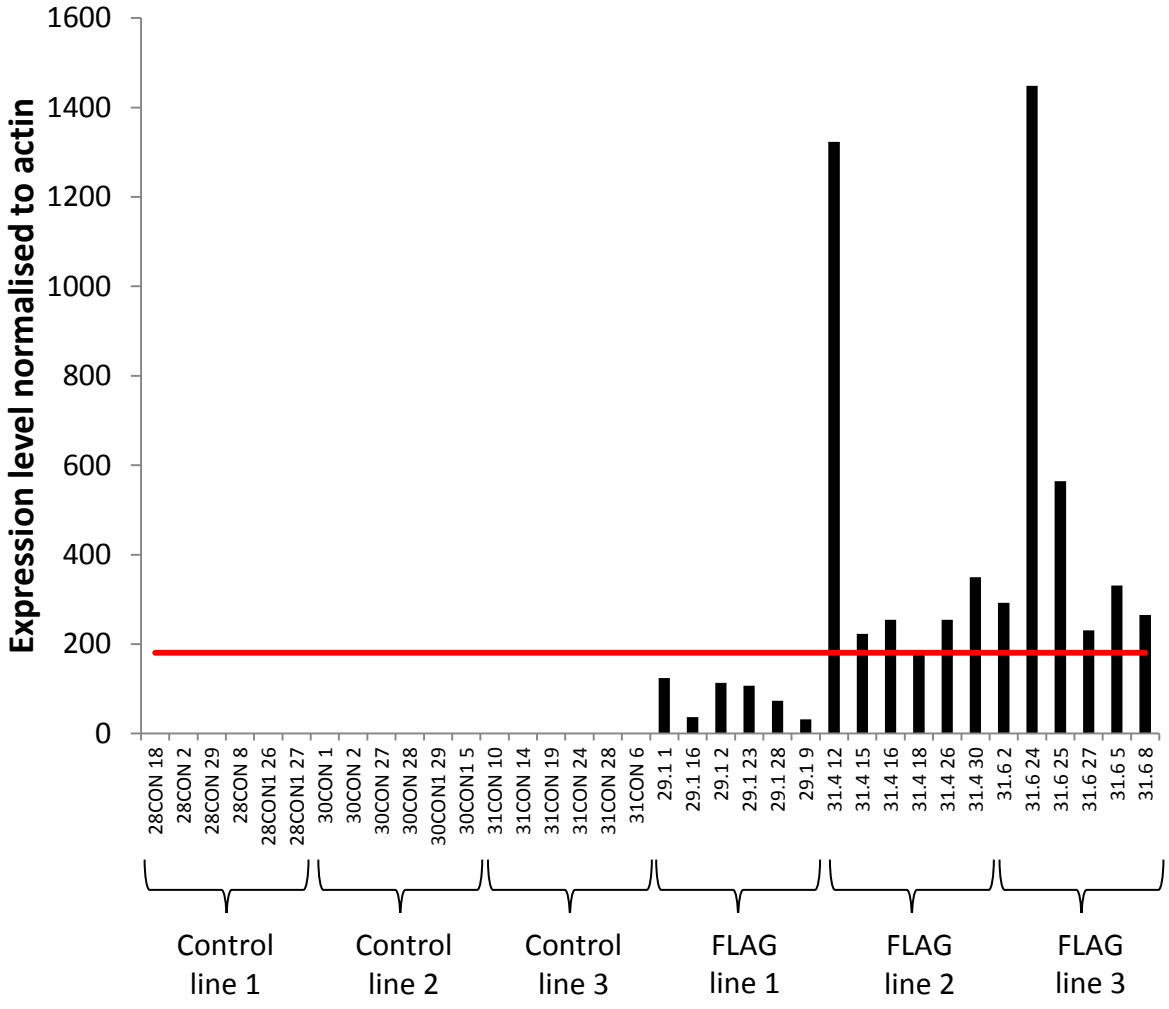


Figure 4.13. Expression levels of FLAG-tagged *NAM-B1* in T_1 normalised to actin. The expression level of native *NAM-B1* in the variety UC1041+ GPC is indicated by the red line.

4.3.1.3.2 Protein expression in T_1

We also examined the protein expression in these transgenic plants using a small portion of the flag leaf and peduncle sample produced by ChIP. However in many cases the protein was not detectable in Western blots in either the flag leaf or the peduncle. For example in **Figure 4.14** a strong band is seen in the positive control (lane 1), and bands corresponding to the heavy and light chains of IgG (50 and 25 kDa) can be seen in the eluted fraction from the immunoprecipitation (IP) (lanes a in FLAG2A and CON2A). This is due to cross reactivity between the secondary antibody and the anti-flag antibody used in the IP. In no case was the FLAG-NAM-B1 detected in this eluted fraction. However in one peduncle sample (out of 6) a band at the expected size of FLAG-NAM-B1 (~48 kDa) was detected in the purified nuclei (red arrow). FLAG-NAM-B1 was not detected in any leaf sample (data not shown).

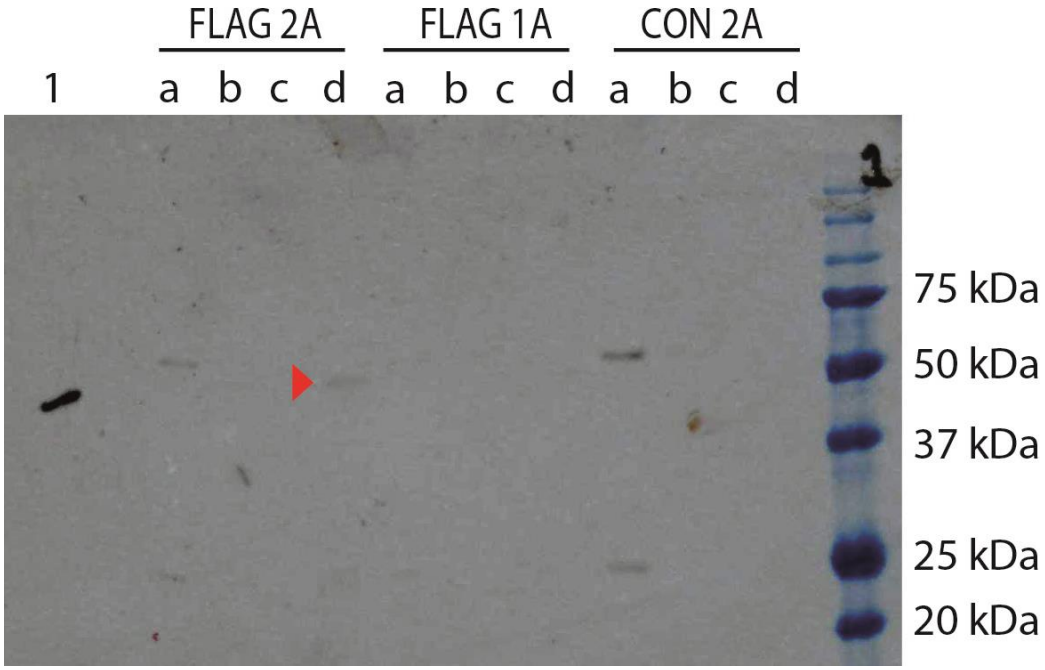


Figure 4.14. Western blot on peduncle samples for two FLAG-tagged samples and a control sample. Lane 1 is FLAG-BAP positive control (49 kDa). a = eluted fraction from immunoprecipitation, b = unbound fraction from immunoprecipitation, c = input total protein extraction, d = purified nuclei. Samples used were FLAG transgenic lines 1 and 2, and control line 2, biological replicate A. The red arrow indicates a band the size of FLAG-NAM-B1.

4.3.2 Hypothesis 1: The stringency of read mapping will affect the number of binding sites identified by ChIP-seq

4.3.2.1 Small scale ChIP-seq with T_0 plants

In this experiment one FLAG tagged sample and one HA tagged sample were used, with a control sample for each tag type, using plants from the T_0 generation. The aim of this experiment was to determine which transgenic lines to use to carry out a fully replicated experiment with T_1 plants and to set up bioinformatics methods for the data analysis. For each construct 3 lines showed much higher expression levels than all the others (**Figure 4.11** and **Figure 4.12**), therefore the flag leaves from the 2nd tiller of these lines were used, pooled per construct.

4.3.2.1.1 Read mapping parameter optimisation

For each of the samples 4-6 million pairs of reads were sequenced (**Figure 4.15**). We found that as the number of mismatches allowed during alignment with bwa-aln increased, the percentage of reads mapped increased from an average of 4.7 % mapping with 2 mismatches allowed (98.8 % identity) to 10.4 % mapping with 5 mismatches allowed (97.1 % identity) up to 20.4 % mapping with 8 mismatches allowed (95.3 % identity) (**Figure 4.16**). Regardless of the mismatches allowed, in all cases many of the reads mapped were not in unique positions. Therefore we selected reads which were in proper pairs and mapped with high mapping quality (MAPQ > 30) to ensure unique mapping locations. Again we found that as mismatch number increased, the percentage of total reads mapped increased but to a much lower level: even with 8 mismatches permitted on average only 7.6 % of reads mapped (**Figure 4.17**). We did not want to permit too many mismatches because we wanted to map reads to the correct homoeologue. As the number of mismatches increased the proportion of high quality mapped reads (MAPQ > 30) decreased (**Figure 4.18**) which indicated that increasing the number of mismatches allows lower quality alignments, therefore filtering for mapping quality is an important step, especially if using relatively high numbers of mismatches.

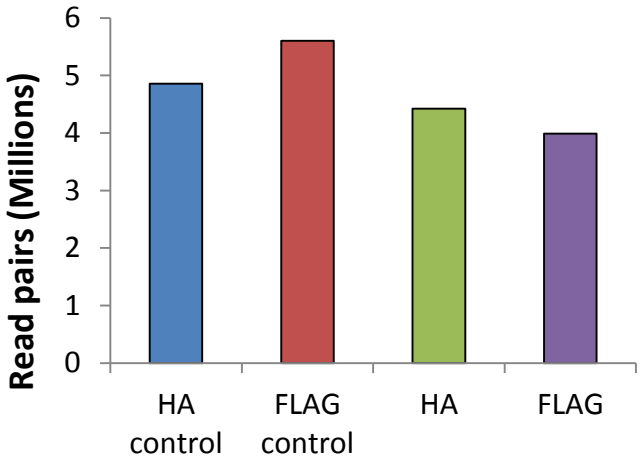


Figure 4.15. Number of read pairs per sample.

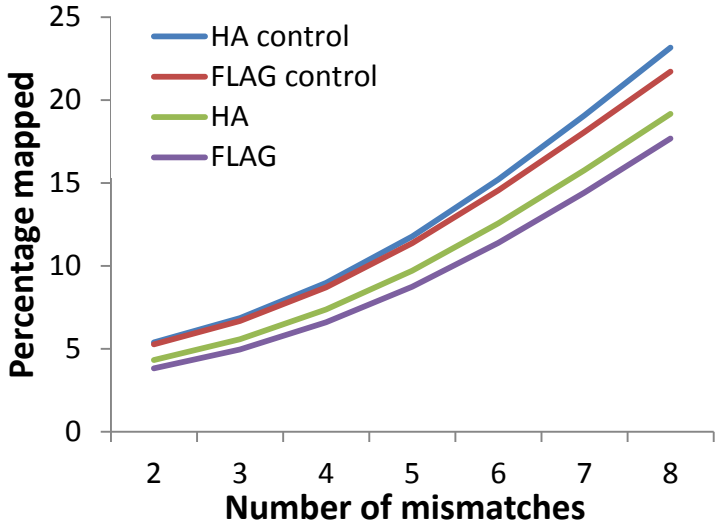


Figure 4.16. Percentage of reads mapped with increasing numbers of mismatches permitted during alignment.

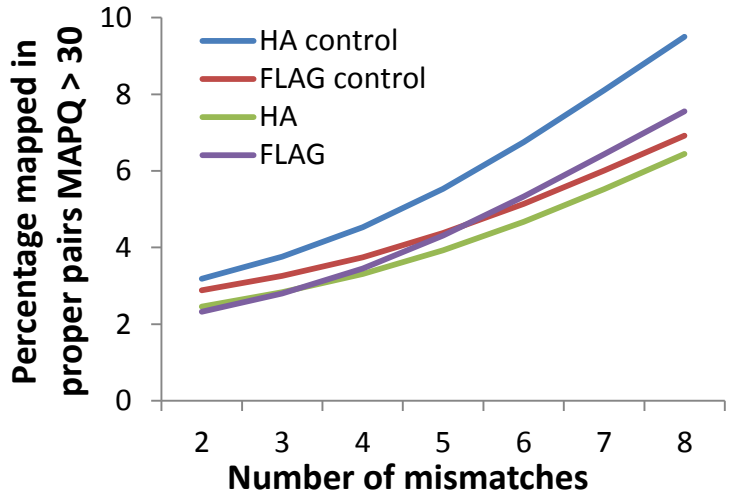


Figure 4.17. Percentage of total reads which are properly paired with MAPQ > 30 with increasing numbers of mismatches permitted during alignment.

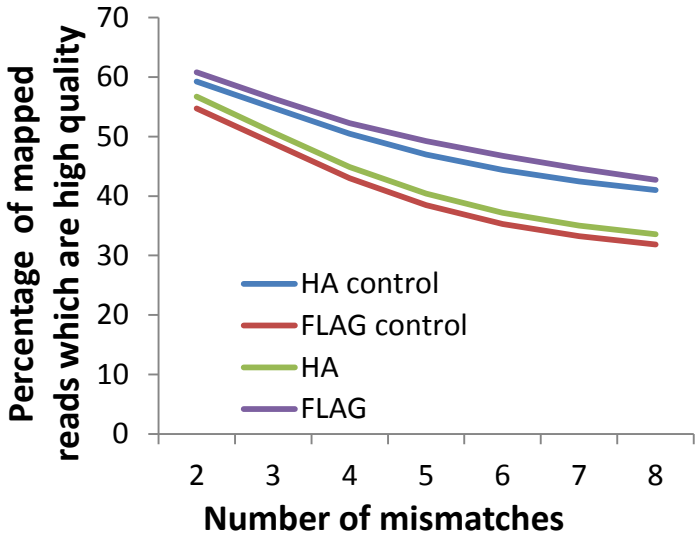


Figure 4.18. Percentage of mapped reads which are in proper pairs and have MAPQ > 30.

4.3.2.1.2 Mapping quality affects the number and type of peaks identified

We investigated whether the quality of the mapping and the number of reads mapped would affect the peaks called by Macs2.0, i.e. the number of NAM-B1 binding sites detected. We decided to compare: 2 mismatches, 5 mismatches and 8 mismatches, using either all mapped reads or only reads with a high mapping quality (MAPQ > 30).

We found that in general with less stringent read mapping more peaks were called by Macs2.0 (**Figure 4.19**). Surprisingly the HA tagged sample had far more peaks than the

FLAG construct when using all the mapped reads. However when only high quality mapped reads were selected (proper pairs with MAPQ >30) the numbers of peaks were similar between FLAG and HA samples (**Figure 4.20**) and when allowing fewer mismatches the FLAG sample had more peaks called. It was also noticeable that the FLAG sample had more peaks which were highly significant ($p < 0.001$) than the HA sample (**Figure 4.21**). This was not simply because the FLAG sample had more peaks called in total, in fact a higher proportion of peaks were highly significant (**Figure 4.22**) than in the HA sample. We decided to carry out the replicated experiment with the FLAG tagged transgenic lines because the FLAG tagged sample had expression levels more similar to the native gene. The FLAG sample also has more peaks detected with stringent read mapping (e.g. proper pairs with MAPQ > 30 with 5 mismatches (97 % identity)), and these peaks are more highly significant than in the HA sample.

These results indicate that mapping quality does affect the number and type of putative binding sites bound by NAM-B1.

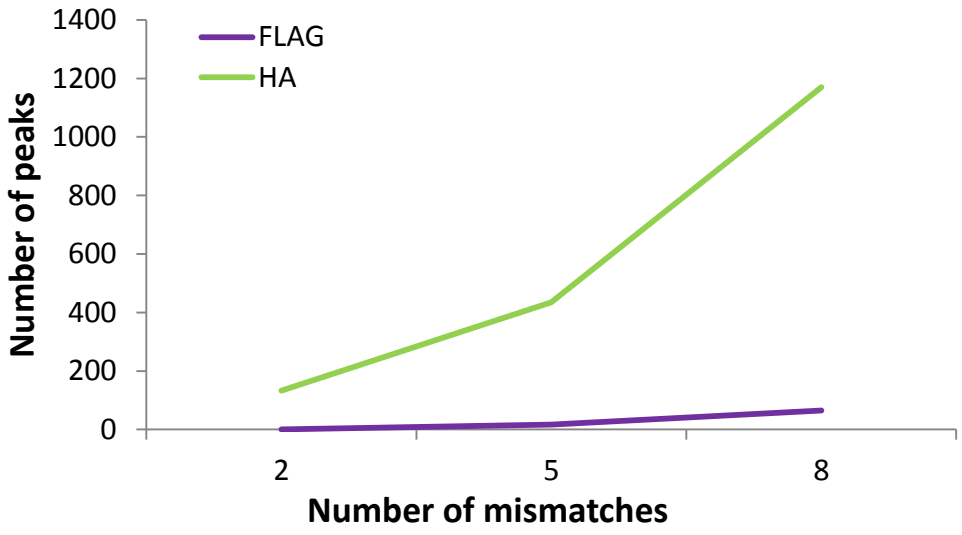


Figure 4.19. Number of peaks detected using all mapped reads.

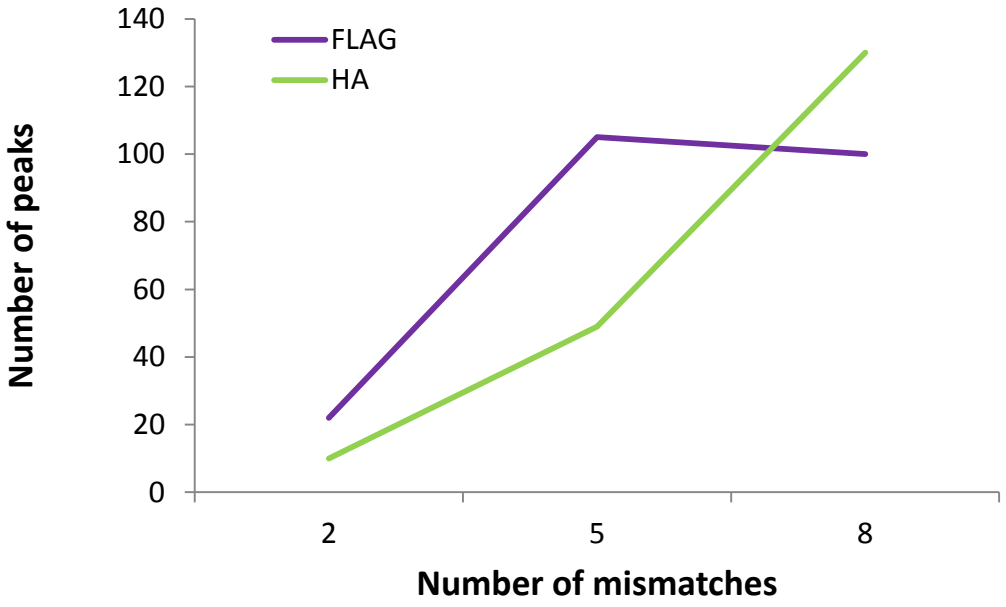


Figure 4.20. Number of peaks detected using only properly paired reads with MAPQ > 30.

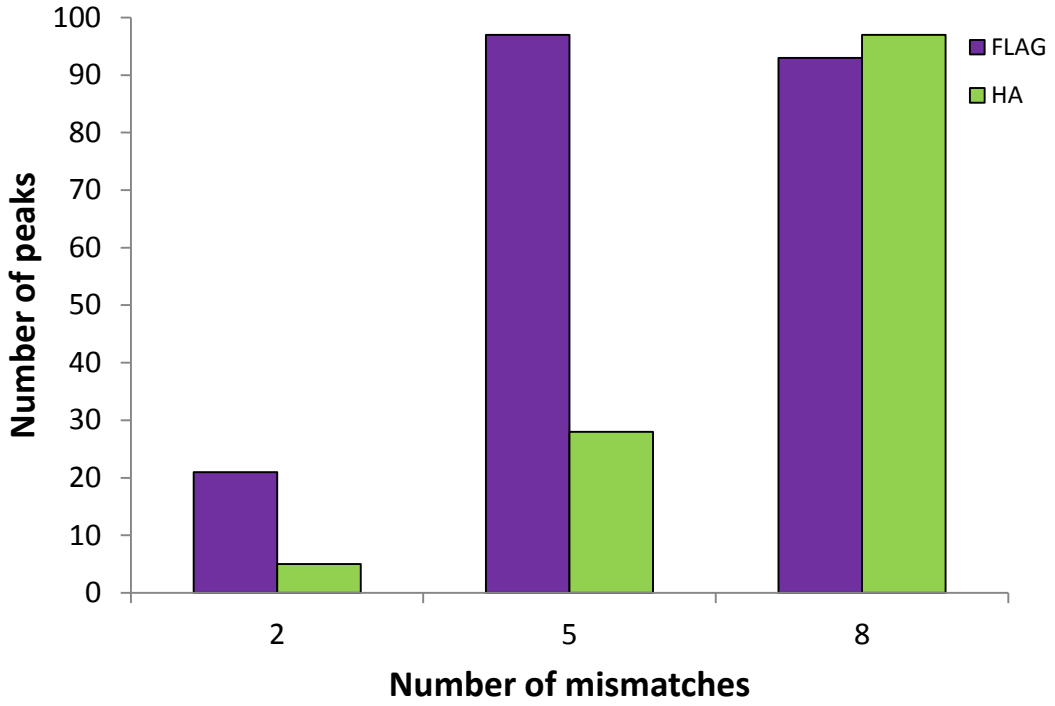


Figure 4.21. Highly significant peaks (p <0.001) using properly paired reads with a MAPQ > 30.

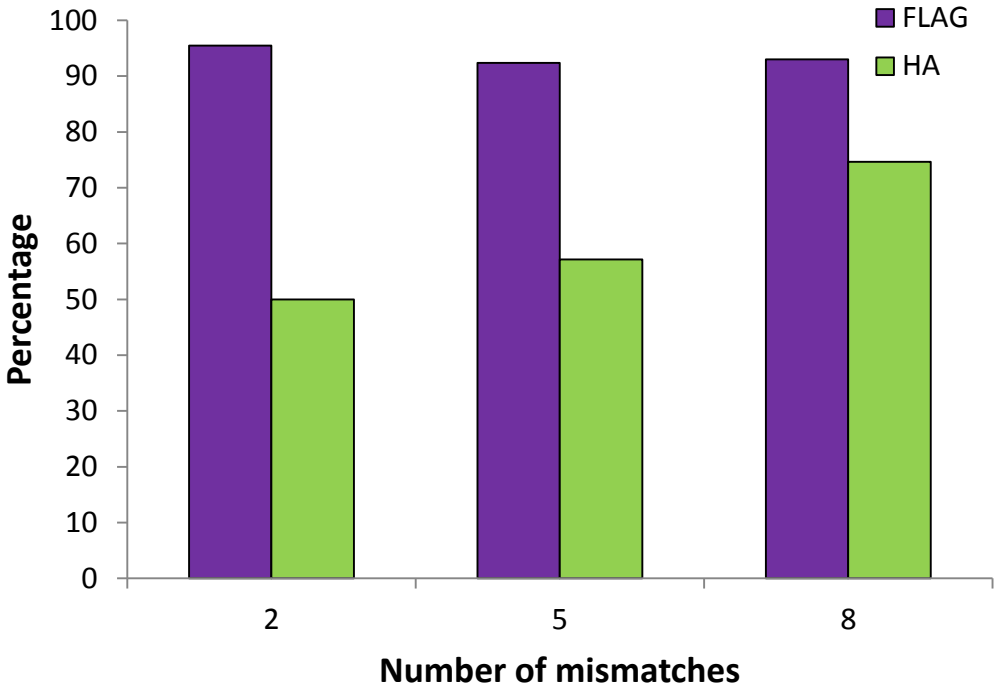


Figure 4.22. Percentage of highly significant peaks (p< 0.001) out of all significant peaks (p<0.05). Peaks were determined using reads mapped in proper pairs MAPQ > 30).

4.3.2.2 Replicated ChIP-seq using T_1 plants

The small-scale ChIP-seq with T_0 samples had shown that the number and type of peaks identified depends on the mapping stringency. However because we used shorter reads for the T_1 plant samples (100 bp paired for T_1 vs. 250 bp paired end for T_0) we wanted to check whether the same effects would be seen with these shorter reads. Additionally we wanted to select the optimal read filtering conditions to identify peaks.

In all graphs below the samples are labelled according to this following code: L = leaf, P = peduncle, number = transgenic line, A/B = biological replicate. Also see **Table 4.16**.

4.3.2.2.1 Read mapping and filtering

On average 11.6 M paired reads were sequenced per sample with quite similar numbers of reads per sample (**Figure 4.23**). From the preliminary experiment we found that is necessary to have stringent read mapping parameters therefore we decided to allow 2 mismatches per read (n=2) because these were shorter reads than used in the preliminary experiment. This ensures 98 % identity between the reads and the reference sequence and should help map reads to the correct homoeologue whilst allowing for single nucleotide polymorphisms between the reference variety (Chinese Spring) and our experimental variety (Fielder). We found that again selecting for properly paired reads and high mapping quality dramatically reduces the proportion of reads mapped from an average of 17.8 % to 4.3 % (**Figure 4.24**).

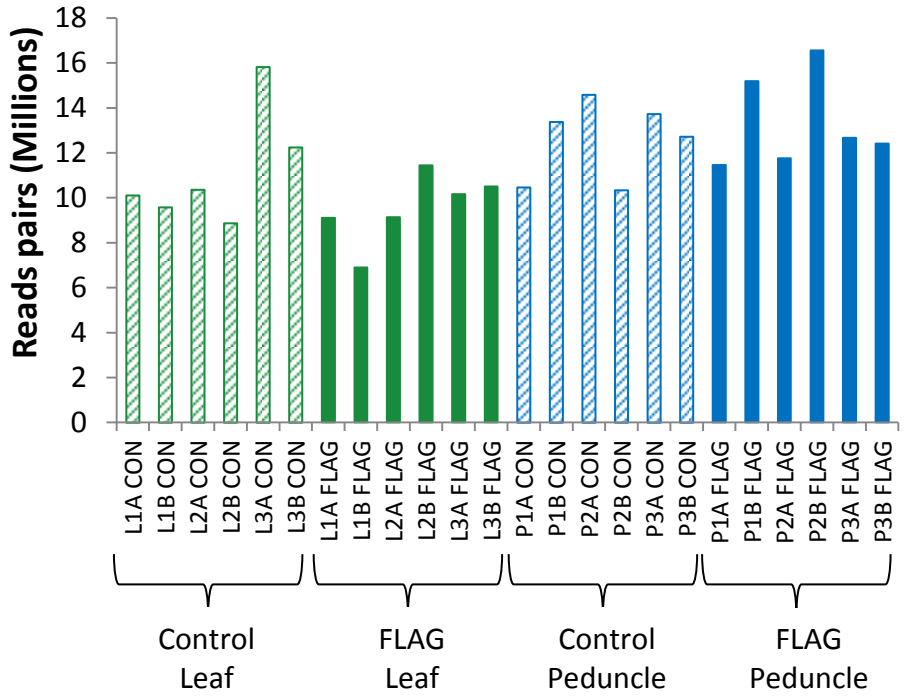


Figure 4.23. Total reads sequenced per sample for replicated ChIP-seq experiment.

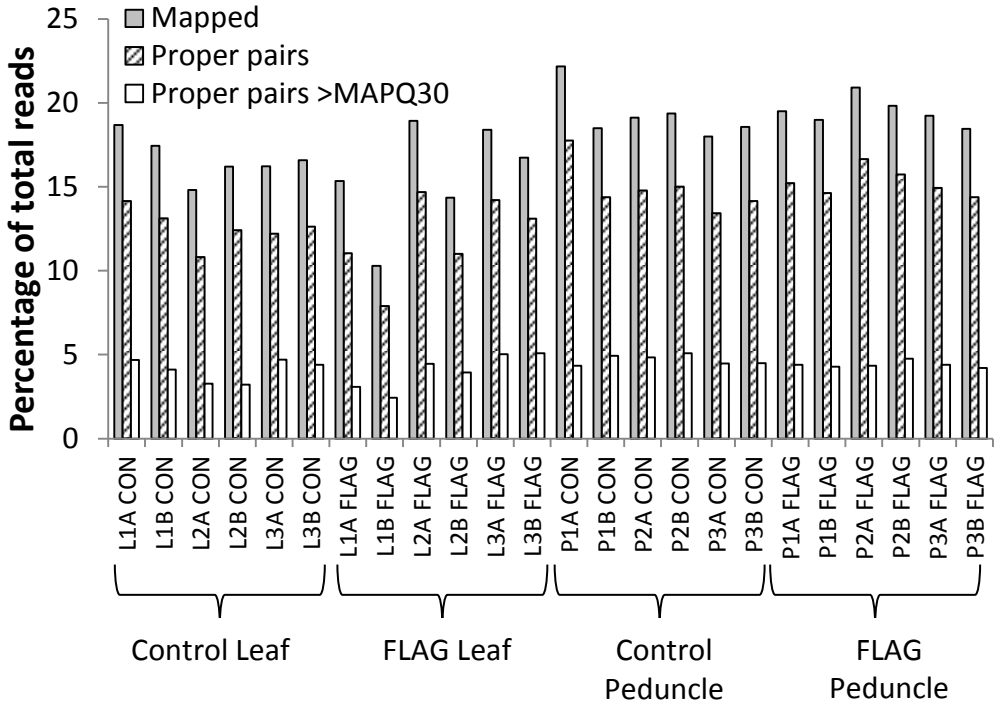


Figure 4.24. Comparison between the percentage of reads which are mapped in proper pairs and in proper pairs at a high mapping quality (MAPQ > 30). Reads were mapped using bwa-aln allowing 2 mismatches.

4.3.2.2.2 Comparison of peaks identified using different read filtering stringencies

In general far more peaks were identified using all mapped reads than were identified using reads filtered to be in proper pairs and with a MAPQ > 30. When reads were filtered to include only proper pairs before peak calling, nearly all the peaks identified using all mapped reads were identified. However when reads were filtered to only allow proper pairs with MAPQ >30 far fewer peaks were detected in all samples. L1B still had 1128 peaks although only approximately half of these were also detected using all reads. The overlap of peaks between samples or between mapping conditions was calculated using the programme Bedops which requires 10 % of the peaks to be overlapping.

In conclusion the stringency of read mapping significantly affects the peaks identified, with fewer peaks identified in more stringent conditions.

4.3.3 Hypothesis 2: True NAM-B1 binding sites will be detected in several replicates, the significance of binding sites will be correlated between samples and binding sites may be different between leaf and peduncle samples

4.3.3.1 Overlapping peaks between samples

We wanted to determine whether the same peaks were identified in more than one sample because we hypothesised that if peaks are replicated they are higher confidence binding sites of NAM-B1.

Using only properly paired reads (**Figure 4.25**) to call peaks, 5 out of 6 peduncle samples share over 50 % of their peaks (P2A, P3A, P1B, P2B and P3B). Out of these P3A and P3B, which are biological replicates of the same transgenic line, have 1010 shared peaks out of 1555 and 1470 peaks respectively. L1A and L1B share 330 peaks out of 1253 and 4107 respectively.

We also filtered the reads to select high quality mapped reads (either MAPQ > 10 or MAPQ > 30) and then carried out the same analysis to establish whether many peaks were shared between samples when more stringent mapping conditions were used.

Using MAPQ > 10 most samples had very few peak regions shared except L1A and L1B which shared 221 peaks out of 980 and 5430 respectively (**Figure 4.26**). Using MAPQ > 30 most samples had 0 overlapping peaks and the maximum number of shared peaks was 23 for 3AP and 3BP out of 28 and 32 respectively (graph not shown).

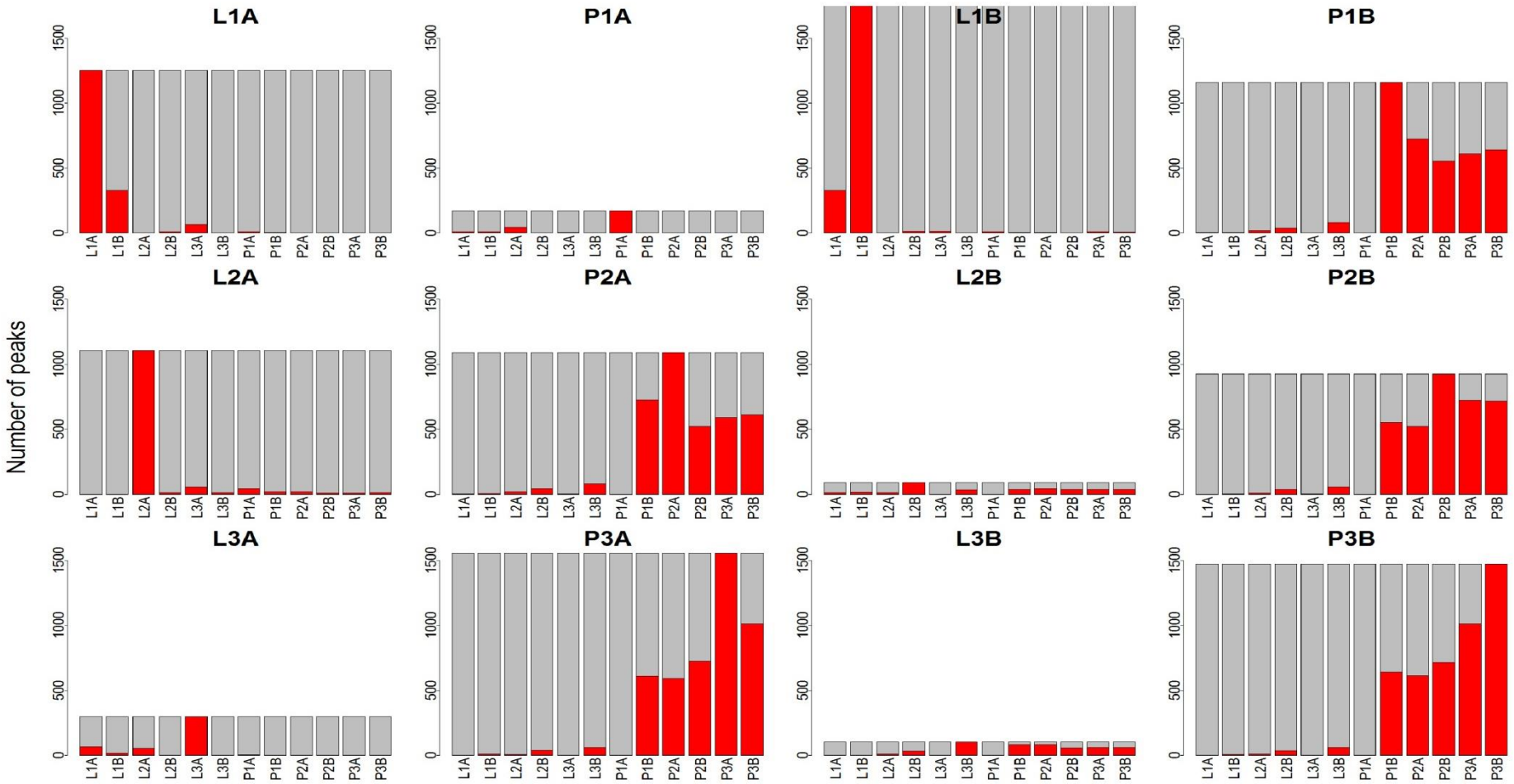


Figure 4.25. Peaks which overlap between samples using properly paired reads as the input for Macs2. Grey bars indicate total number of peaks, red indicate the overlap between samples. P – peduncle, L = leaf, 1/2/3 = transgenic line used, A/B biological replicate.

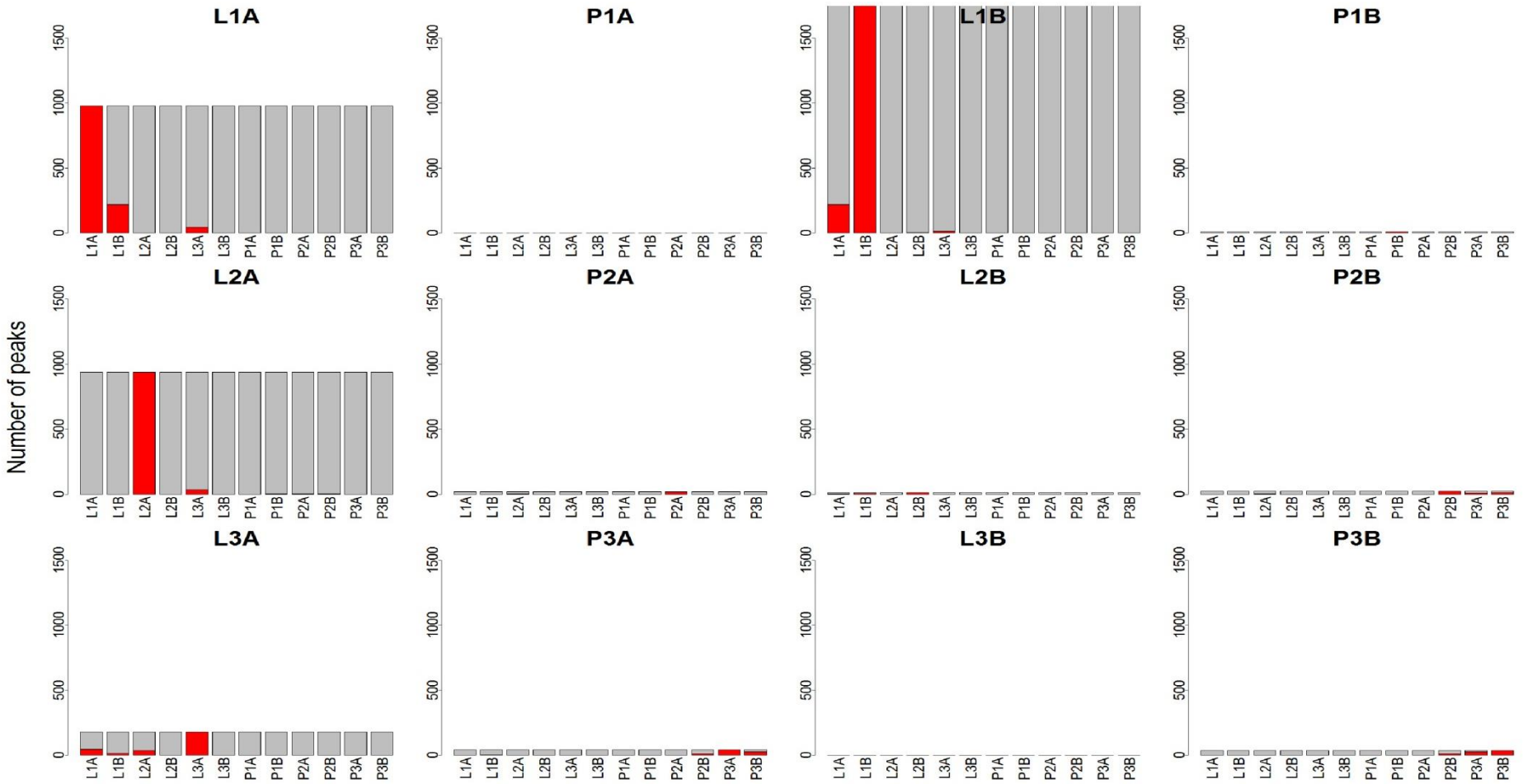


Figure 4.26. Peaks which overlap between samples using properly paired reads with MAPQ > 10 as the input for Macs2. Grey bars indicate total number of peaks, red indicate the overlap between samples. P – peduncle, L = leaf, 1/2/3 = transgenic line used, A/B biological replicate.

4.3.3.2 Correlations of peak significance values between leaf samples

For the samples which had overlapping peaks we were interested to find out if there was a correlation between the significance values between samples. It was noticeable that irrespective of the read filtering (proper pairs, MAPQ >10 or MAPQ >30) many peaks overlapped between L1A and L1B (**Table 4.15**).

	Filtering of reads:		
	Proper pairs	**Proper pairs MAPQ > 10**	**Proper pairs MAPQ > 30**
Peaks in L1A	1253	980	61
Peaks in L1B	4107	5430	1128
Overlapping peaks between L1A and L1B	330	221	8
Percentage of L1A peaks which overlap with L1B	26.3 %	22.6 %	13.1 %

Table 4.24. Peaks overlapping between L1A and L1B.

Using properly paired reads across all significant peaks there is not a strong correlation between L1A and L1B ($R^2 = 0.102$) (**Figure 4.27**A), however there is a stronger correlation amongst the 20 most significant peaks ($R^2 = 0.336$) (**Figure 4.27**B). Similarly when using the properly paired reads with MAPQ > 10, the top 20 peaks are more strongly correlated than the rest of the peaks ($R^2 = 0.280$ compared to $R^2 = 0.09$) (**Figure 4.28**), however all of these correlations are weak. Using only properly paired reads with MAPQ > 30 there is a very strong correlation (**Figure 4.29**) between the two samples ($R^2 = 0.94$), but there are only 8 data points. Two of the peaks identified are found in all three read filtering conditions (red points), and many peaks are identified in 2 conditions (green and yellow). It is clear that the peaks for L1B have much more significant values (higher $-\log_{10}$qvalue) than the peaks in sample L1A, ant this is true regardless of read mapping filtering.

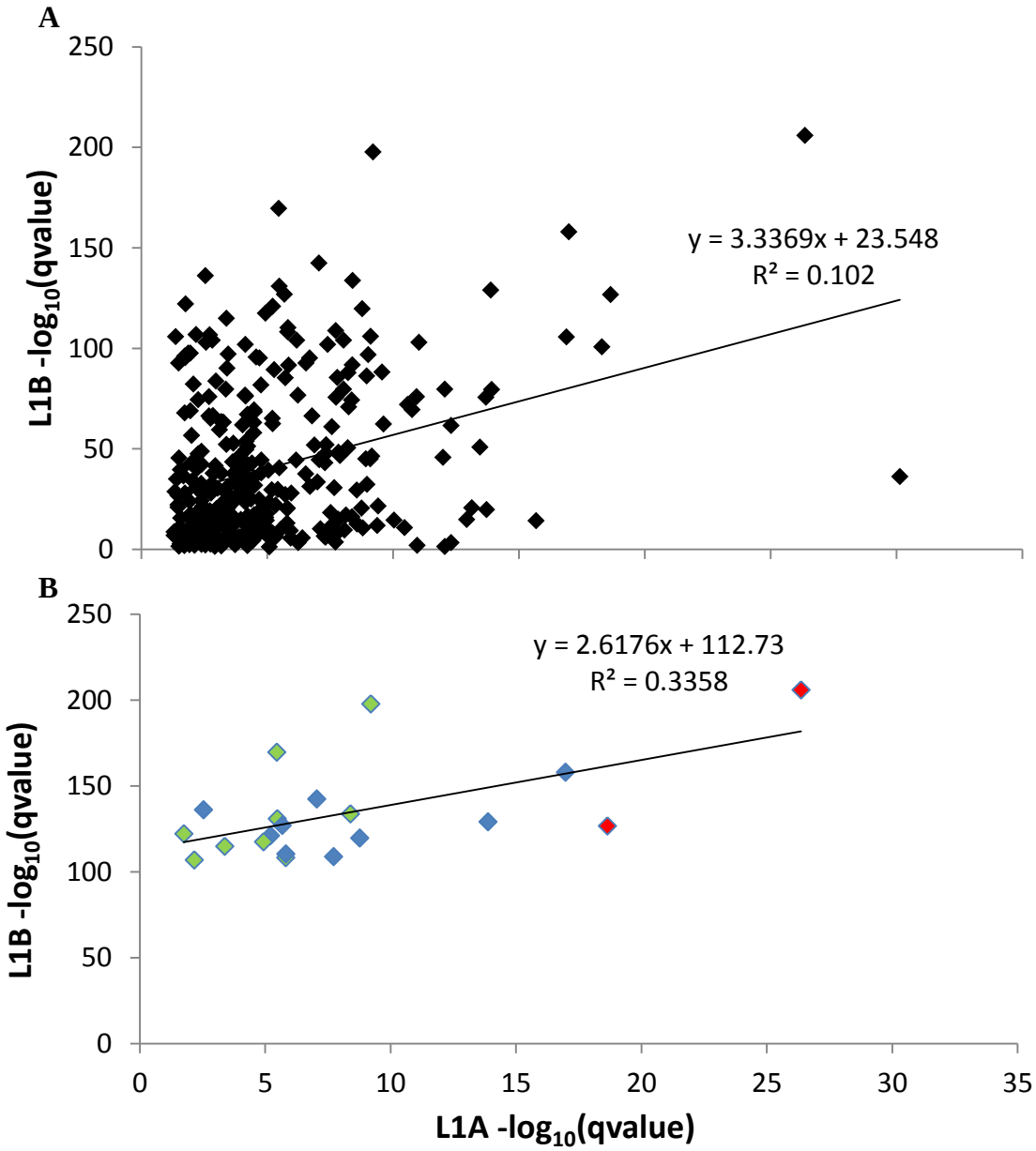


Figure 4.27. Correlation between peaks found using all properly paired reads in 2 independent biological replicates from the same transgenic line in leaf tissue. A) all peaks and B) top 20 most significant peaks. Red filled points are found in the top 20 using proper pairs, proper pairs MAPQ >10 and proper pairs MAPQ > 30, green filled points are found in the top 20 using proper pairs and proper pairs MAPQ >10, blue points are unique to the properly paired samples.

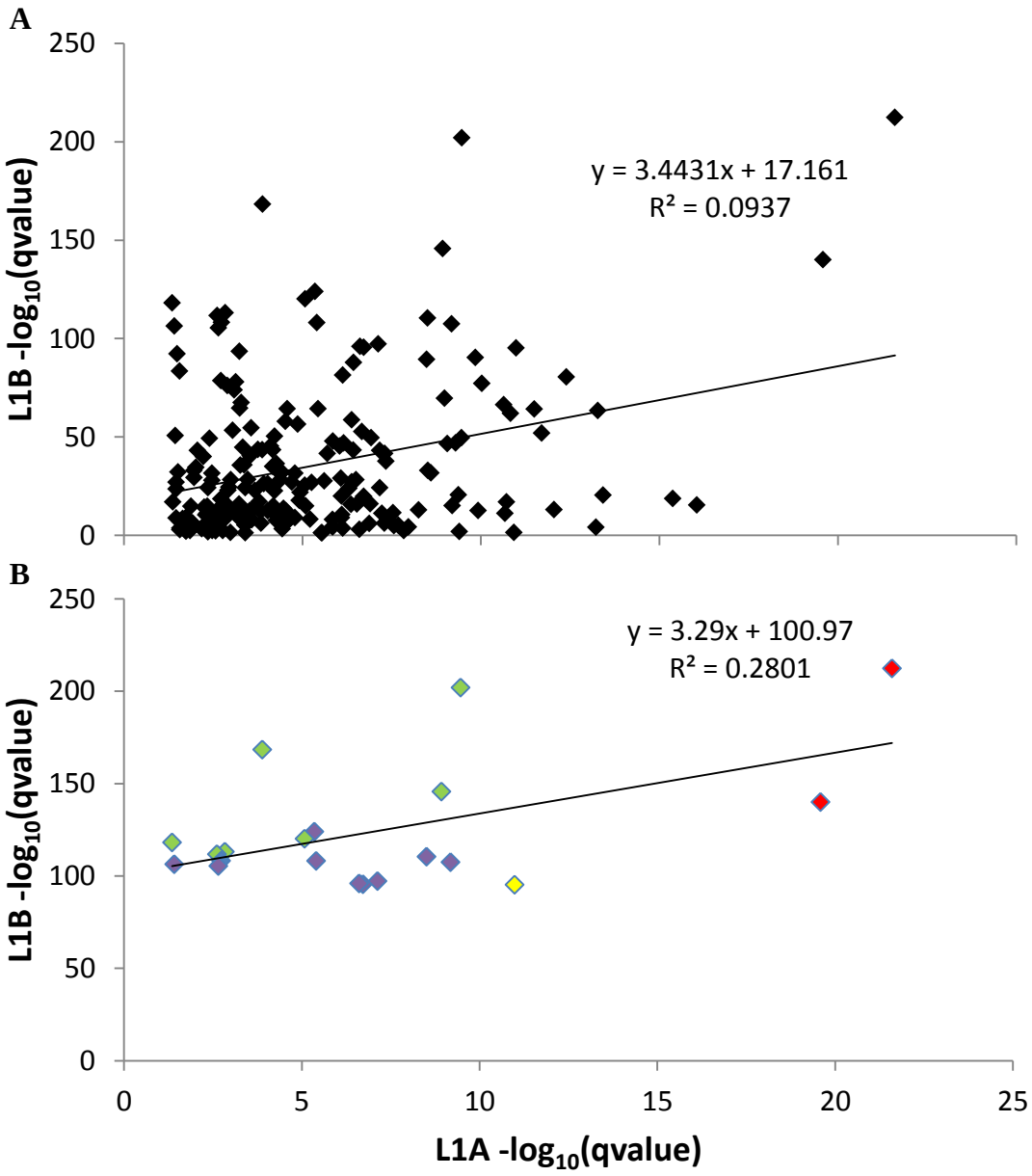


Figure 4.28. Correlation between peaks found using all properly paired reads MAPQ > 10 in 2 independent biological replicates from the same transgenic line in leaf tissue. A) all peaks and B) top 20 most significant peaks. Red filled points are found in the top 20 using proper pairs, proper pairs MAPQ > 10 and proper pairs MAPQ > 30, green filled points are found in the top 20 using proper pairs and proper pairs MAPQ > 10, yellow filled points are found in the top 20 in proper pairs MAPQ > 10 and proper pairs MAPQ > 30, and purple filled points are unique to the proper pairs MAPQ > 10 samples.

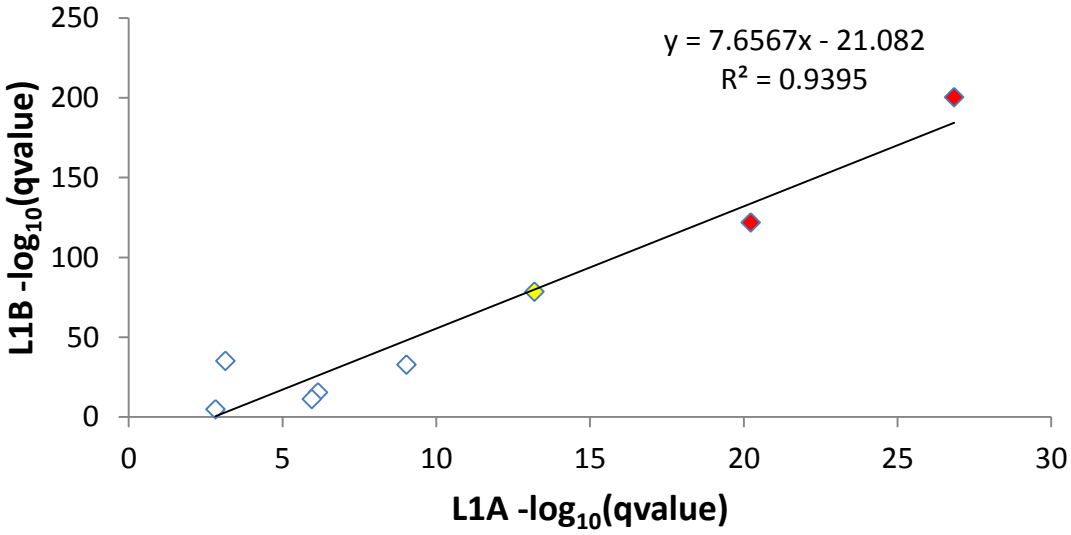


Figure 4.29. Correlation between peaks found using all properly paired reads MAPQ > 30 in 2 independent biological replicates from the same transgenic line in leaf tissue. Red filled points are found in the top 20 using proper pairs, proper pairs MAPQ > 10 and proper pairs MAPQ > 30, yellow filled points are found in the top 20 in proper pairs MAPQ > 10 and proper pairs MAPQ > 30, and white filled points are unique to the proper pairs MAPQ > 30 samples.

4.3.3.3 Correlations of peak significance values between peduncle samples

Amongst the six peduncle samples we noticed that five shared over 50 % of their peaks when the reads used for peak calling were in proper pairs

An example region of a pseudomolecule is shown in **Figure 4.30** and although a statistically significant peak is called in all five samples, in certain cases the enrichment of reads in the FLAG sample is much clearer than in others (e.g. P2B compared to P2A). To try to characterise these shared regions further we examined whether the significance value was correlated for each region between samples (**Figure 4.31**). Between most samples there was quite low correlation between the significance values of shared peaks, however between P3A and P3B there is a higher correlation (R^2 = 0.38).

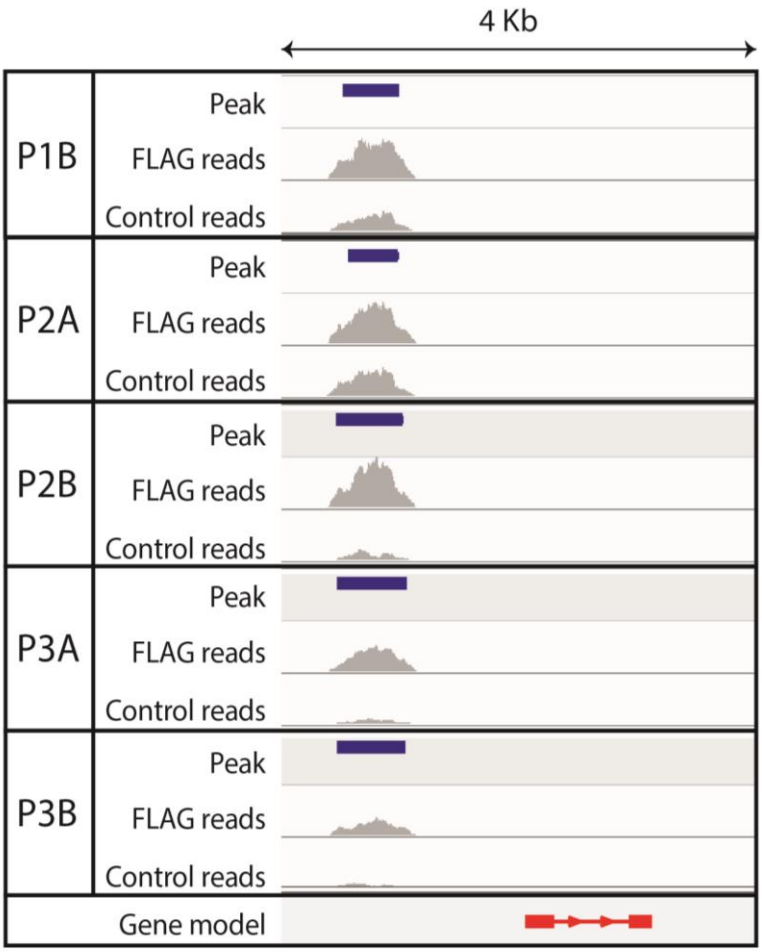


Figure 4.30. Data from five independent peduncle samples with overlapping peaks on the unordered region of chromosome 4D of the pseudomolecule. Grey regions indicate read pile ups. Blue regions indicate binding sites identified by MACS2. Gene model shown in red

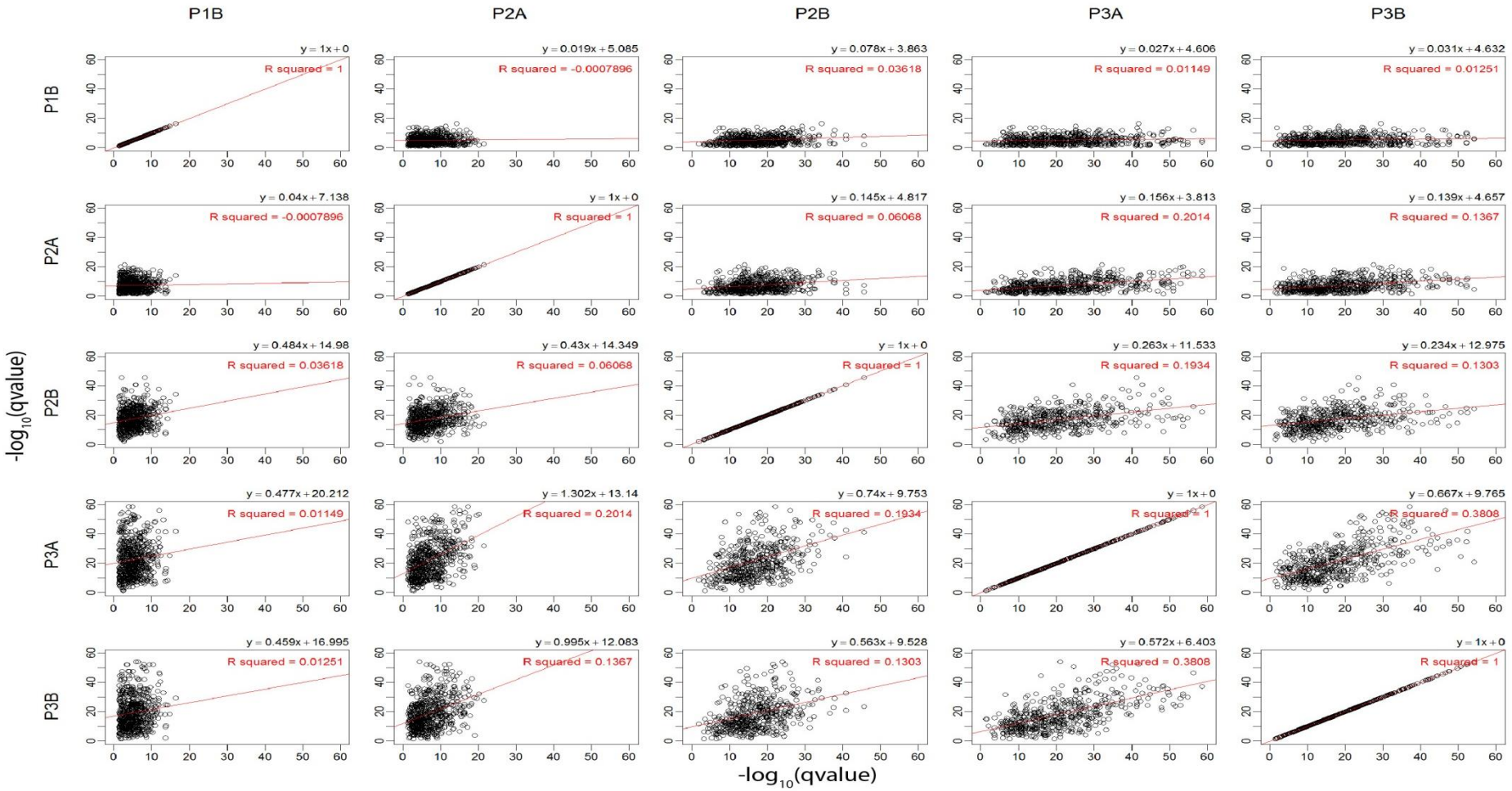


Figure 4.31. Correlation of significance value between peduncle samples for overlapping peaks. .

4.3.3.4 Peaks identified are different between leaf and peduncle samples

Although overlapping peaks were identified using both relaxed and more stringent mapping conditions (**Figure 4.25** and **Figure 4.26**), the majority of the overlapping peaks are shared between samples from the same tissue. Very few peaks were shared between both leaf and peduncle samples. The maximum number of peaks shared between tissues are the 82 common to L3B and P1B, however using the same filtering of reads (proper pairs) over 1000 peaks are shared by two peduncle samples (P3A and P3B). This suggests that NAM-B1 may have different target genes in the two tissue types.

In conclusion, within peduncle samples many regions are replicated between samples which may be biologically relevant regions. However in leaf samples less replication was observed and one sample (L1B) was very different from the others. In general there was a poor correlation between the significance of peaks between samples, although most significant peaks (e.g. top 20) found in each sample were often correlated between samples.

4.3.4 Hypothesis 3: NAM-B1 binding sites will be unique within the genome to ensure specific binding of target genes

To ensure specific binding of NAM-B1 to promoter elements near its target genes we hypothesised that the bound regions would be unique within the wheat genome. To test this hypothesis we examined how many copies there were of each peak sequence within the wheat genome.

4.3.4.1 Peaks identified are often highly repeated throughout the genome

We wrote a set of custom Perl scripts (Appendix 6.2.1) to calculate the number hits found by BLAST at 99 % identity for each peak region within the whole wheat genome (IWGSC chromosome arm survey sequences, including regions which were not included in the pseudomolecules because they do not contain genes). We used the peaks identified using properly paired reads as an input. We found that for all peduncle samples the majority of peaks identified had multiple matches (50-100) to the IWGSC contigs at over 99 % identity (**Figure 4.32**B, D, F, H, J and L). This implies that the peak regions are highly repetitive. Several of the leaf samples also had highly repetitive peaks (**Figure 4.32**A, G and K).

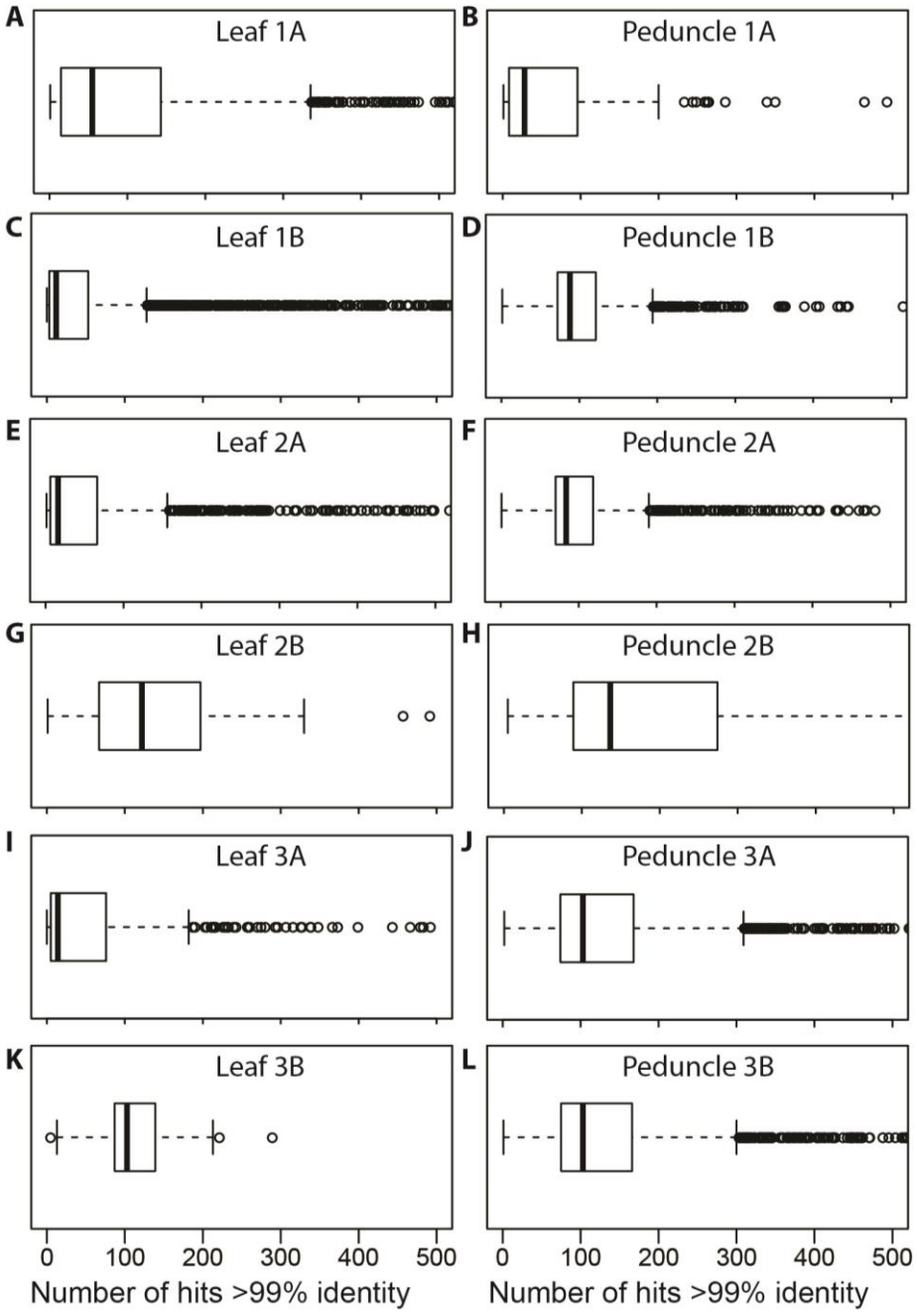


Figure 4.32. Number of BLAST matches found in the IWGSC contigs with over 99 % similarity to the peaks in all samples. A) Leaf 1A, B) Peduncle 1A, C) Leaf 1B, D) Peduncle 1B, E) Leaf 2A, F) Peduncle 2A, G) Leaf 2B, H) Peduncle 2B, I) Leaf 3A, J) Peduncle 3A, K) Leaf 3B, L) Peduncle 3B.

Contrastingly three of the leaf samples showed fewer repetitive peak regions (**Figure 4.32**C, E and I). On closer examination of leaf samples most of the samples had relatively few peaks (if any) which had 0-20 BLAST hits (**Figure 4.33**A, C, D, E and F). However one sample stood out: L1B (Leaf sample from transgenic line 1, biological replicate B). This sample had a different distribution from other samples with >600 of its peaks having only 1 BLAST hit (**Figure 4.33**B), and distribution of peaks from 0-20

BLAST hits was skewed towards having fewer BLAST hits. This indicated that the putative NAM-B1 binding sites identified in sample L1B were low copy number regions, even when using not very stringently mapped reads such as the properly paired reads used for this analysis.

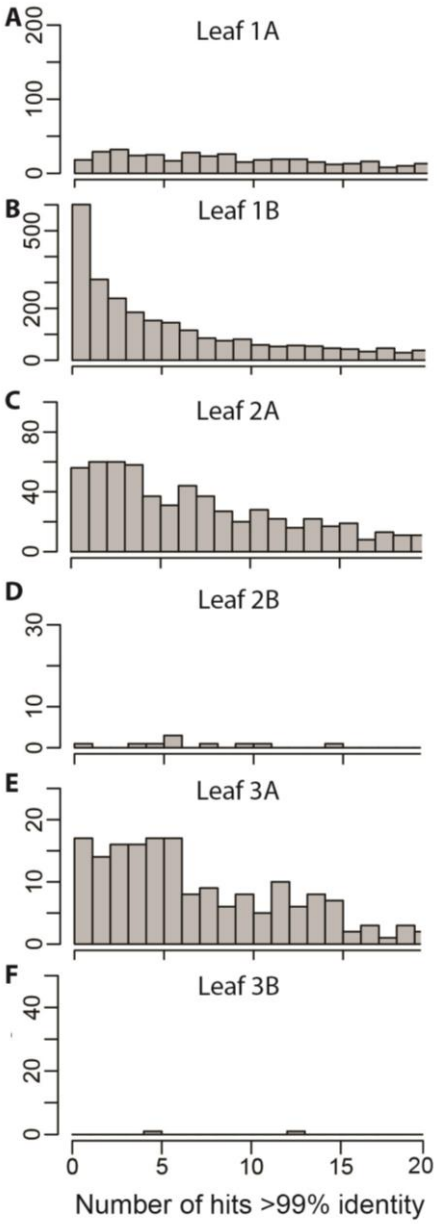


Figure 4.33. Number of BLAST matches found in the IWGSC contigs with over 99 % similarity to the peaks in leaf samples. A) Leaf 1A, B) Leaf 1B, C) Leaf 2A, D) Leaf 2B, E) Leaf 3A, F) Leaf 3B. NB. y-axes scales are different between samples.

4.3.4.2 Similarity between homoeologues decreases outside peak regions

To try to understand more about the repetitive nature of the peak regions identified in all samples except L1B, we selected peaks which were identified in 2 or more independent biological replicates for peduncle samples using properly paired reads. We calculated the percentage identity between the A, B and D homoeologous copies of these peak regions. Surprisingly we found that most of these regions were extremely similar between A, B and D genome (Figure **4.34**A) with a mean identity of 95.0 % and a median identity of 97.8 %. This was unexpected because although coding regions are recognised to be well conserved between genomes (~97 %) we expected promoter regions (where we expect NAM-B1 to bind) to be less conserved. We also noticed that the exact peak regions bound were more similar between A, B and D genomes than the surrounding regions (Figure **4.34**B, C and D). This may indicate either that NAM-B1 binds to repetitive regions and therefore is not specific in its binding. Alternatively the specific binding to regions which are highly similar within promoter sequences of homoeologues, may be of biological relevance to regulate homoeologous genes.

In conclusion peaks in the two tissues identified showed different properties, with the leaf sample L1B having many unique binding sites, whereas the peduncle samples had high repetitive binding sites.

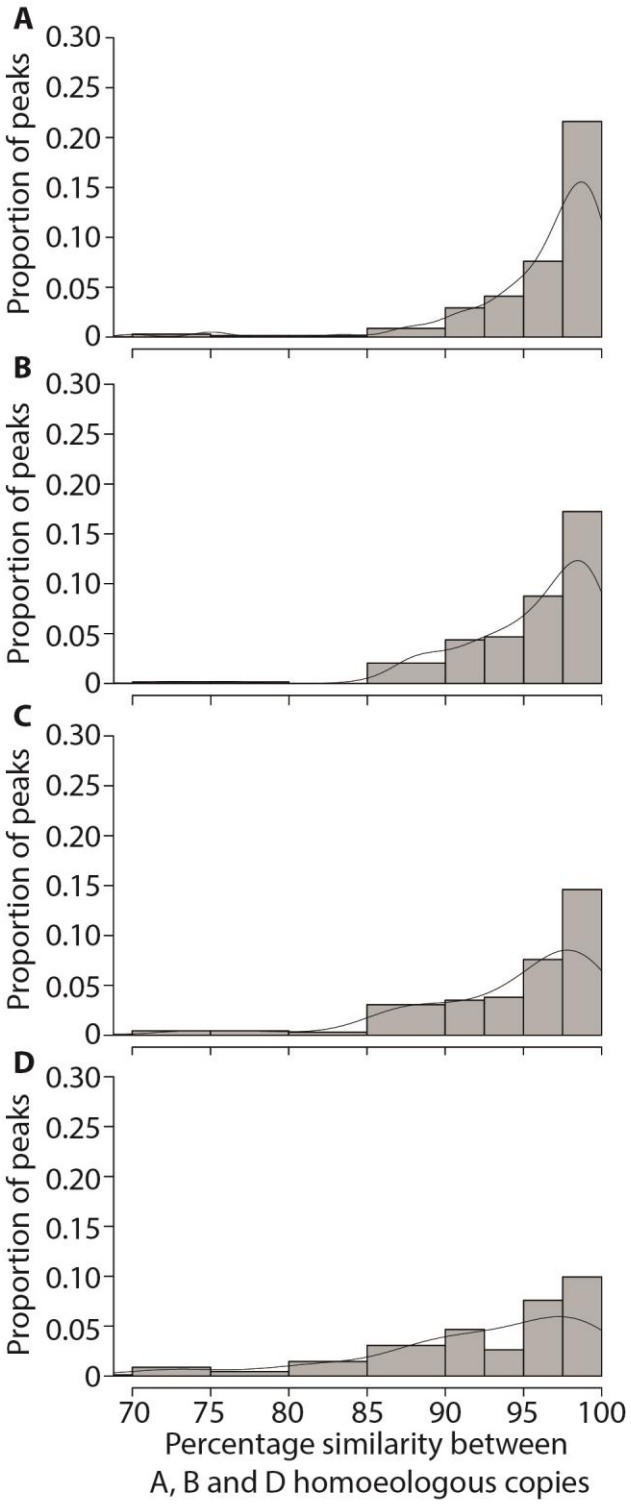


Figure 4.34. Percentage similarity between peak regions and their homoeologous sequences. A = peak region only, B = peak region plus 50 bp of surrounding sequence, C = peak region plus 100 bp of surrounding sequence, D = peak region plus 200 bp of surrouding sequence.

4.3.5 Selecting peaks to further characterise

We analysed all of the samples using different read filtering stringencies, examined overlapping peaks between samples and examined the repetitiveness of peak regions. Using this information we used several criteria to determine which peaks to examine further. We were interested in samples with low copy number peaks and we were interested peaks detected in several independent biological replicates especially where correlations between peak significance were detected. None of our samples fitted all our criteria of interest therefore we decided to further characterise several lists of peaks to try to understand the biologically relevant binding sites of NAM-B1. The true nature of the binding sites would be further validated by comparison to differential expression analysis (4.3.11) and biochemical assays (4.3.12). Five lists of peaks were selected for further characterisation:

- Overlapping peaks between L1A and L1B proper pairs MAPQ > 10
    - These peaks are identified independently in 2 biological replicates, more peaks are identified than using MAPQ > 30 which is very stringent and only leaves 8 peaks.
- Overlapping peaks between L1A and L1B proper pairs
    - These peaks are identified independently in 2 biological replicates and these include even more peaks than using proper pairs MAPQ > 10
- Peaks from L1B proper pairs MAPQ > 30
    - This sample appears to be quite different from all the other samples because it has far more significant peaks using proper pairs MAPQ > 30 (1128 vs 61 in the next highest) and these peaks are in non-repetitive regions (**Figure 4.33**).
- Overlapping peaks between P1B, P2A, P2B, P3A, P3B proper pairs
    - Five out of the six peduncle samples share over 50 % of their peaks, therefore the regions which are present in all five replicates will be examined.

- Overlapping peaks between P3A and P3B proper pairs

P3A and P3B share even more regions between them than the five peduncle regions do (1010 vs. 330) and there is a correlation between the significance value of the peaks between these two samples (**Figure 4.31**) which lends further confidence in the biological relevance of these regions.

4.3.6 Hypothesis 4: NAM-B1 will bind promoter regions to regulate gene expression

Transcription factors are generally expected to bind in the promoter region 5' of the transcription start site of target genes. In this location the transcription factor will interact with RNA polymerase II to initiate or repress transcription of the target genes. Therefore we hypothesised that NAM-B1 binding will be found in promoter regions of target genes. We tested this in two complementary manners: identifying which types of sequence peaks were located in (e.g. promoter, exon, intron) and by calculating the distance from peaks to the transcription start site of target genes. First we added more annotation to our pseudomolecules to describe the location of genomic features (e.g. promoters).

4.3.6.1 Distribution of genomic features across the pseudomolecule

We annotated the pseudomolecules to have various genomic features: distal promoters, proximal promoters, 5'UTRs, exons, introns and 3'UTRs. The locations of these features were determined by comparison with other species because little is known about these regions in wheat since genomic sequence has only recently become available. For example for the 5'UTR it has been reported that eukaryotic 5'UTRs range from 100-200 bp (Pesole et al. 2001). In Arabidopsis the 5'UTR average length is 131 bp (Kim Byung-Hoon et al. 2007) and similar lengths are seen in Brachypodium (Mochida et al. 2013). Therefore we took the 150 bp immediately upstream of translation start site (TSS) to be the 5'UTR. Average eukaryotic 3'UTRs are 200 bp (Pesole et al. 2001) therefore we annotated 200 bp downstream of final exon as the 3'UTR. It is difficult to determine the length of promoter regions because they can be variable between and within species. For example in Arabidopsis the median length of promoters is ~1,672 bp for stress related genes and 1,113 bp for non-stress related genes (Kristiansson et al. 2009). To capture the majority of promoter events we created two categories of promoter regions: distal promoter regions (-2000 to -5000 bp) and proximal promoter regions (-150 to -2000 bp), relative to the TSS. From our BLAST between the gene models and the pseudomolecules we had identified exons and introns.

All remaining regions of the pseudomolecules were categorised as intergenic regions. The distribution of these features is represented in **Figure 4.35**A. Approximately 10 % of the pseudomolecules consists of exons, with a similar proportion of introns. A very small proportion of the pseudomolecules (2 %) represents UTRs. Over 50 % of the pseudomolecules are defined as promoter regions (21 % proximal and 34 % distal) which is quite surprising, however the contigs which make up the pseudomolecules were selected to be gene-rich and our definition of the promoter sequence (from -150 to 5000 bp) is very wide, to ensure all potentially regulatory interactions are captured. A very low proportion of the pseudomolecules (20 %) is intergenic, which is explained by not only our large promoter regions but also the incomplete nature of the pseudomolecules which only represent ~4 % of the total wheat genome and is specifically enriched for gene-containing regions.

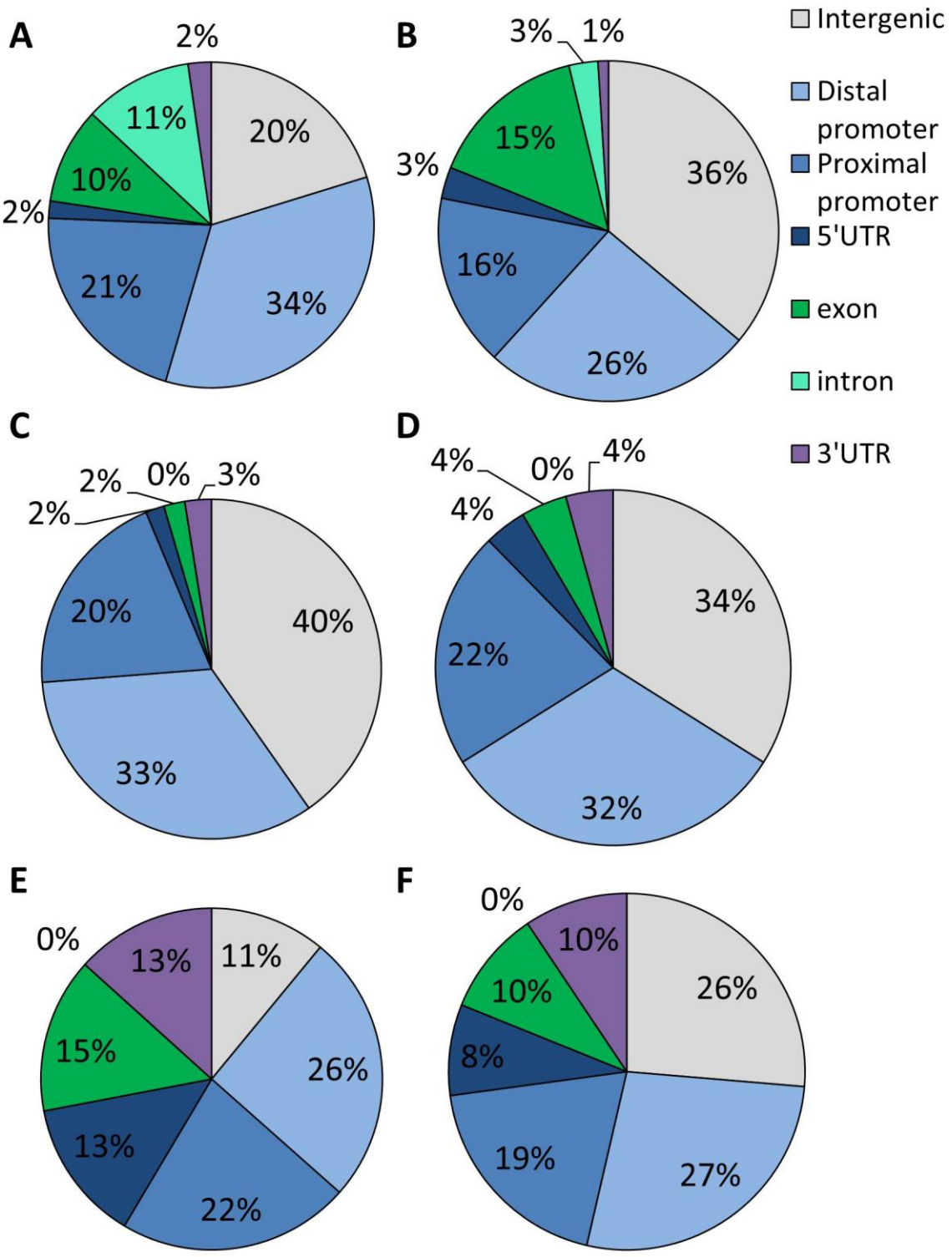


Figure 4.35. Distribution of peaks across genomic features. A) Overall distribution of features across the pseudomolecules, B) peaks in L1B proper pairs MAPQ > 30, C) peaks in L1A and L1B using proper pairs MAPQ > 10, D) peaks in L1A and L1B using proper pairs, E) peaks in P1B, P2A, P2B, P3A and P3B using proper pairs, F) peaks in P3A and P3B using proper pairs.

4.3.6.2 Distribution of peaks across genomic features

In the L1B proper pairs MAPQ > 30 sample the distribution of peaks (**Figure 4.35**B) is quite similar to the distribution of features across the entire pseudomolecule. However L1B has an increased proportion of peaks within the intergenic regions and a reduction in binding in promoter regions. Surprisingly many peaks are found in exons (15 %) but binding is not frequently found in introns.

In the peaks identified in L1A and L1B whether using all properly paired reads or only properly paired reads MAPQ > 10, (**Figure 4.35**C and D), a similar proportion of peaks are found in promoter regions to the distribution of promoter regions across the pseudomolecules (~20 % in proximal promoter and ~33 % in the distal promoter). However very few peaks are identified in exon and intronic regions and more peaks are located in intergenic regions.

For the five overlapping peduncle samples the distribution of peak binding looks quite different from the overall genome feature distribution (**Figure 4.35**E). In particular a large proportion of binding occurs in the 5' UTR (13 % of peaks), compared to the 2 % of the pseudomolecules which consists of 5'UTRs. Similarly the 3'UTR is much over-represented (13 % vs 2 %). Additionally many peaks are found in the exons (15 %) and very few in the intergenic regions. This distribution of peaks in the UTRs suggests a regulatory role for the peaks identified. A similar distribution of peaks is observed in the P3A and P3B sample (**Figure 4.35**F) with strong enrichment of peaks within UTRs. In all samples very few peaks are identified in introns.

4.3.6.3 Distribution of peaks relative to transcription start sites

Transcription factors are known to bind to the upstream sequences near to transcription start sites (TSS) to regulate transcription of genes. Therefore to determine whether NAM-B1 binds near to TSS and therefore is regulating genes, we calculated the distance from each peak to the TSS. Firstly as a control we examined the distribution of random peaks relative to TSS (**Figure 4.36**A). With randomly distributed peaks there is a strong trend in the location of peaks, with most peaks present in the regions surrounding genes (red box on graph) from -1,500 bp to +5,000 bp. This enrichment of random peaks around genes is due to the lengths of contigs which were used to make the pseudomolecules: the median size of contigs was 7,591 bp in ordered chromosomes and 6,638 bp in unordered contigs (**Table 4.22**), which are very close in size to the enriched region found using random peaks (~6,500 bp).

Despite this underlying trend due to contig size, the peaks identified by ChIP-seq are not randomly distributed (**Figure 4.36**B-F). In all samples examined there are far fewer peaks in the gene body region (red boxes) than would be expected from a random distribution. In L1A and L1B (**Figure 4.36**C and D), regardless of whether only reads MAPQ > 10 were used, the gene body has few peaks but the enrichment of peaks in the promoter and 3' downstream regions is not very different from the randomly distributed peaks.

For the sample with peaks which overlap between 5 peduncle samples a very strong enrichment of peaks is found in the proximal promoter and 5'UTR with up to 40 peaks per 200 bp bin (**Figure 4.36**E). This mirrors the distribution of peaks across genomic features (**Figure 4.35**E). Another very strong enrichment is observed in this sample in the 3' UTR. For the peaks which overlap between P3A and P3B (**Figure 4.36**F) a similar pattern is observed to the other peduncle samples: a very strong enrichment is seen in the proximal promoter, 5'UTR and 3'UTR.

A different pattern is observed in the sample L1B proper pairs MAPQ > 30 (**Figure 4.36**B) where the largest enrichment of binding is located within the first half of the gene body, as is reflected in **Figure 4.35**B. Strong enrichment is also observed in the proximal promoter and 5'UTR.

In conclusion, in the peduncle samples the putative NAM-B1 binding sites are concentrated in the proximal promoter and 5'UTR, as might be expected for transcription factor binding sites. The L1B proper pairs MAPQ > 30 samples has many binding sites in the 5' end of the first exon in addition to enrichment within the proximal promoter. The L1A and L1B samples show enrichment in the proximal promoter and also in the region 3' of the gene.

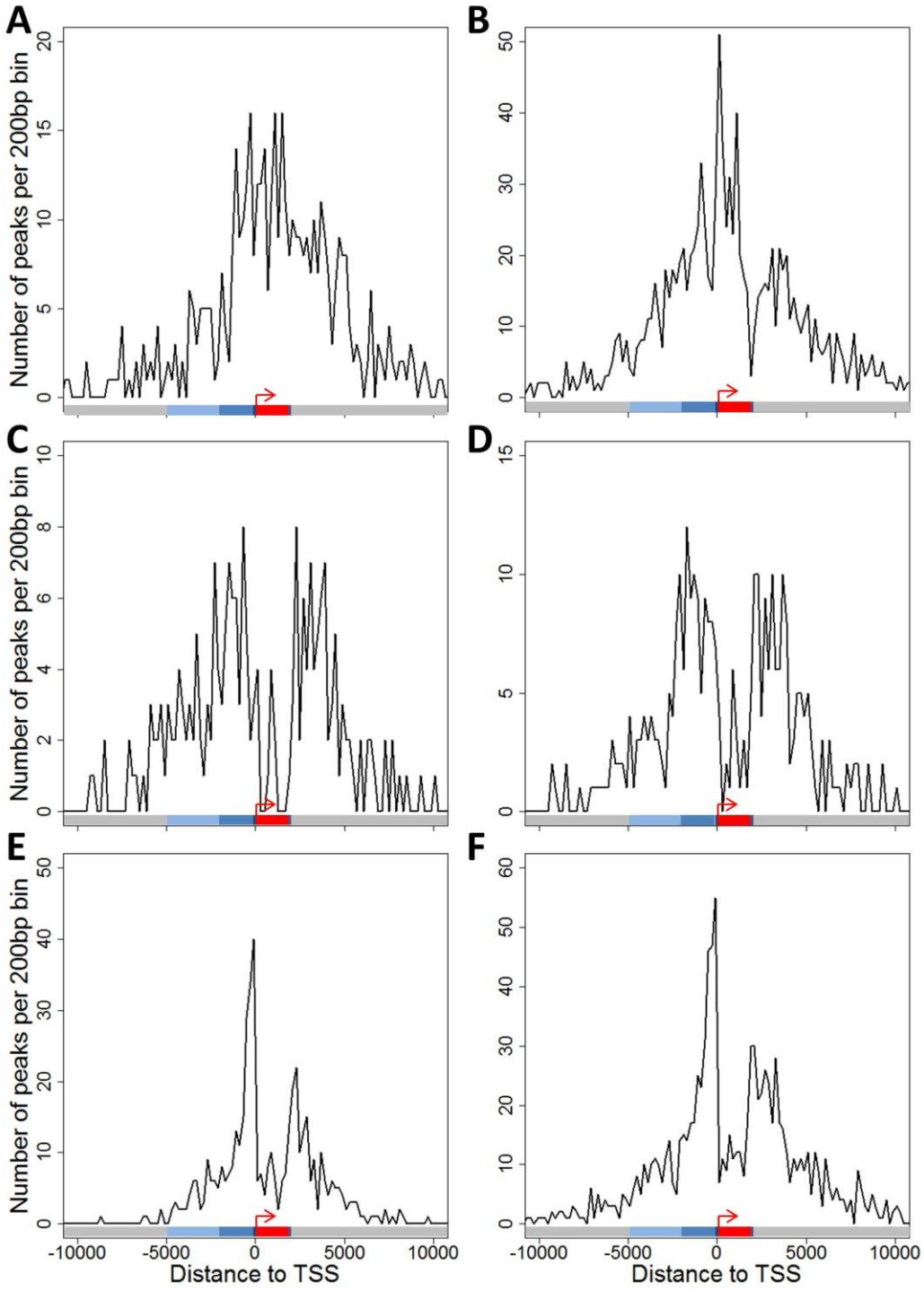


Figure 4.36. Peak location relative to transcription start site (TSS). A) Randomly distributed peaks across the pseudomolecules, B) peaks in L1B proper pairs MAPQ > 30, C) peaks in L1A and L1B using proper pairs MAPQ > 10, D) peaks in L1A and L1B using proper pairs, E) peaks in P1B, P2A, P2B, P3A and P3B using proper pairs, F) peaks in P3A and P3B using proper pairs. Colour code on x-axis: grey = intergenic region, light blue = distal promoter, mid-blue = proximal promoter, black = UTR, red = gene body (including exons and introns).

4.3.7 Hypothesis 5: NAM-B1 binding will be mediated through a specific DNA sequence motif

Transcription factors bind to DNA in a sequence-specific manner which allows them to target particular genes. A consensus DNA sequence which is bound by the Arabidopsis ANAC019 and other Arabidopsis NAC TFs has been identified: (CGT[GA])(Welner et al. 2012). We were interested whether this sequence, or a different motif, might be identified within the peak regions bound by NAM-B1.

4.3.7.1 Motifs present in peaks

We used MEME (Bailey Timothy L. et al. 2009) to identify the top 5 sequence motifs enriched within peak regions, and within random DNA sequences with the same sizes as the peaks found in L1B proper pairs MAPQ > 30.

In all samples examined MEME discovered sequence motifs (**Figure 4.37**). In the L1B proper pairs MAPQ > 30 sample two similar motifs were identified: motif 1 and 5 (**Figure 4.37**B) which were present in 83 % and 38 % of peaks respectively (referred to as CxxC motifs). However, motifs similar to these CxxC motifs were also identified using random peaks, although in a lower proportion of peaks (17 % for both motif 2 and 5 in **Figure 4.37**A). In the L1A and L1B overlap samples (**Figure 4.37**C and D) the motifs found were not present in high proportions of peaks (6-30 %) with the most common motif (motif 2, **Figure 4.37**C) resembling the CxxC motif identified at a low frequency in random peaks. In the sample with peaks which overlap between five peduncle samples the motifs identified had very strong e-values and were often present in a high proportion of peaks. Motif 2 which was present in all peaks (**Figure 4.37**E) and bears little resemblance to any motif identified in random peaks. Similar A-rich motifs were identified in the P3A and P3B overlap sample (**Figure 4.37**F), and were again present in a high proportion of peaks (80 % and 89 % for motifs 1 and 3 respectively). We only found a motif resembling the NAC consensus binding motif (CGTA) in one the L1A and L1B using proper pairs (**Figure 4.37**D, motif 5) and this was only present in 11 % of peak regions.

In conclusion we have identified several sequence motifs which may represent NAM-B1 binding sites, however the most common of these do not resemble the known NAC consensus binding motif.

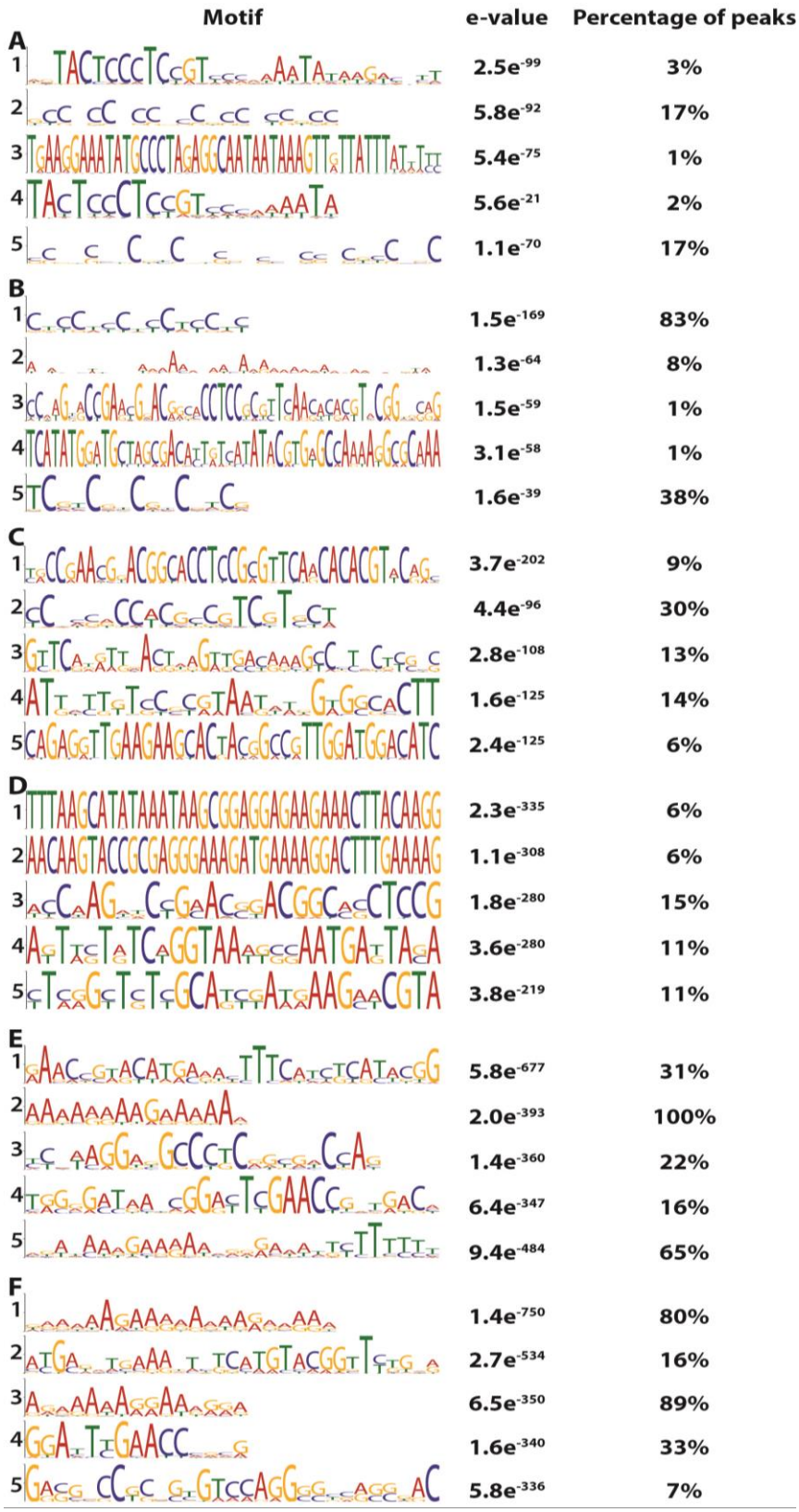


Figure 4.37. Motifs identified in peak regions by MEME. A) Randomly distributed peaks across the pseudomolecules (equivalent sizes to peaks in L1B proper pairs MAPQ > 30), B) peaks in L1B proper pairs MAPQ > 30, C) peaks in L1A and L1B using proper pairs MAPQ > 10, D) peaks in L1A and L1B using proper pairs, E) peaks in P1B, P2A, P2B, P3A and P3B using proper pairs, F) peaks in P3A and P3B using proper pairs. 5 motifs were identified for each set of peaks (1-5).

4.3.8 Hypothesis 6: NAM-B1 target genes will have common functions related to nutrient remobilisation and senescence

NAM-B1 controls nutrient remobilisation and senescence, therefore we expect the direct targets identified by ChIP-seq to have functions related to these processes. We examined whether target genes of NAM-B1 were enriched for particular GO terms and studied what the predicted functions of the genes near the most significant peaks were. For this analysis we focused on the genes identified in sample L1B proper pairs MAPQ > 30 and the genes identified by five of the peduncle samples.

4.3.8.1 GO enrichment of ChIP-seq target genes in L1B proper pairs MAPQ > 30

For L1B proper pairs MAPQ > 30 out of the 1128 peaks identified, 1018 mRNAs were in the same contig as the peak. We only used the 447 peaks which were within 1 kb of the mRNA because these are more likely to be regulatory binding events due to their proximity to the mRNAs. Only 309 out of the 447 peaks had GO term annotation and these were analysed for enrichment in GO terms compared to the whole pseudomolecule.

In leaves in sample L1B proper pairs MAPQ > 30, we found that the target genes of NAM-B1 were enriched for several GO terms, which are mainly related to secondary metabolism (**Figure 4.38**). In particular GO terms related to phenylpropanoid metabolism are enriched including the molecular function "heme-binding" which may reflect the heme-containing monooxygenases involved in phenylpropanoid biosynthesis (Meyer K et al. 1996). The enrichment of very-long chain fatty acid metabolism related genes may reflect the macromolecular breakdown and nutrient translocation from leaves to the grain during senescence (Gregersen and Holm 2007).

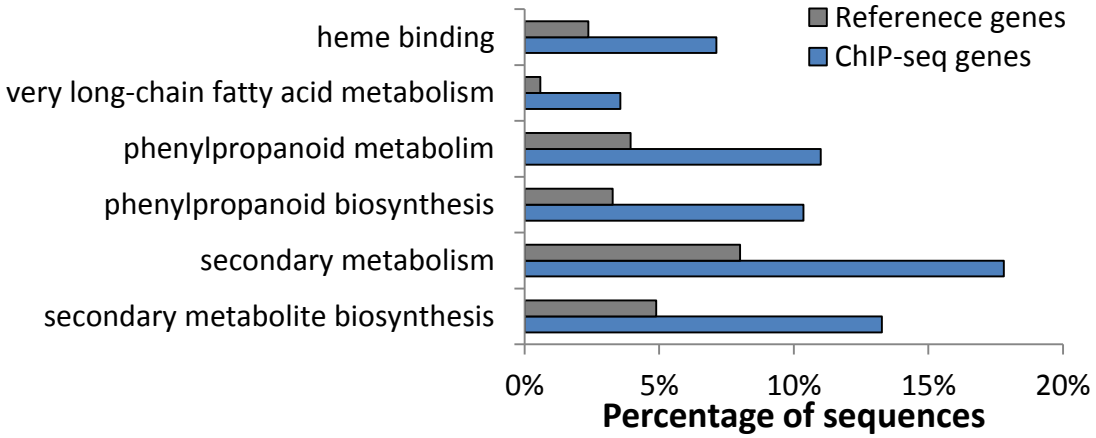


Figure 4.38. GO term enrichment of NAM-B1 target genes from sample L1B proper pairs MAPQ > 30.

4.3.8.2 Top 15 genes from L1B proper pairs MAPQ > 30

To identify the most significantly bound genes, as above, we selected genes within 1 kb of the peak which were in the same contig. We further filtered candidate target genes by requiring that the peak sequence was not a repetitive element to ensure the correct target gene was identified, we did this by allowing a maximum of 3 hits to the IWGSC contigs at >99 % similarity to allow for the A, B and D homoeologous copies. We then only kept genes which had a significant rice BLAST hit. This narrowed the list down to only 175 target genes of which the most significant 15 genes ranked on q-value are shown in **Table 4.25**. Many enzymes are found in the most significant target genes including transferases and hydrolases. Genes involved in transporting molecules across membranes, intracellular protein trafficking, protein degradation and signalling are also present amongst NAM-B1 target genes.

Peak ID	-LOG$_{10}$ (qvalue)	Distance from peak to mrna (bp)	mRNA ID	Rice BLAST	RICE BLAST % ID
648	10.2	0	mrna019927	receptor-like protein kinase 5 precursor, putative, expressed	82.3
665	8.4	-723	mrna102949	haloacid dehalogenase-like hydrolase family protein, putative, expressed	86.0
245	6.6	0	mrna027900	methyltransferase domain containing protein, expressed	81.9
415	6.6	-441	mrna093410	S-formylglutathione hydrolase, putative, expressed	87.0
480	6.6	0	mrna025081	U-box domain containing protein, expressed	85.3
663	6.6	0	mrna023882	COBRA-like protein precursor, putative, expressed	82.8
725	6.6	0	mrna069598	glycerol-3-phosphate acyltransferase, putative, expressed	78.5
737	6.6	0	mrna130148	bifunctional monodehydroascorbate reductase and carbonic anhydrasenectarin-3 precursor, putative, expressed	81.2
1057	6.6	0	mrna031476	transferase family protein, putative, expressed	73.4
45	4.9	0	mrna064514	expressed protein	83.4
171	4.9	0	mrna113118	DUF581 domain containing protein, expressed	89.2
308	4.9	0	mrna087796	amino acid transporter family protein, putative, expressed	81.2
331	4.9	933	mrna086102	aminotransferase, putative, expressed	86.8
342	4.9	0	mrna052756	sec23/Sec24 trunk domain containing protein, expressed	84.2
489	4.9	-898	mrna071751	homeobox domain containing protein, expressed	81.2

Table 4.25. The top 15 NAM-B1 target genes identified in sample L1B proper pairs MAPQ > 30. Only genes within 1 kb of the peak and located on the same contig were included. A distance of 0 bp from the gene denotes that the peak is within the gene body (exons or introns). A negative distance denotes a peak 5' of the gene, a positive distance denotes a peak 3' of the gene. All genes have only 1 hit to IWGSC contigs at 99 % similarity.

4.3.8.3 GO term enrichment amongst the genes identified by the five peduncle samples

For the 5 peduncle samples out of the 465 peaks identified, 380 mRNAs were in the same contig as the peak and 318 peaks were within 1 kb of the mRNA. Only 239 out of the 318 peaks had GO term annotation and these were analysed for enrichment in GO terms.

For the target genes identified by the five peduncle samples 110 GO terms were enriched compared to the pseudomolecule-reference. This higher amount of enrichment compared to the leaf sample raises the possibility that these samples are more biologically representative. We used level 4 biological process GO terms to give an overview of possible function, having condensed the most specific GO terms together. The level 4 GO terms which are over-represented in the peduncle targets are photosynthesis, generation of precursor metabolites and energy and oxidation-reduction process (**Figure 4.39**).

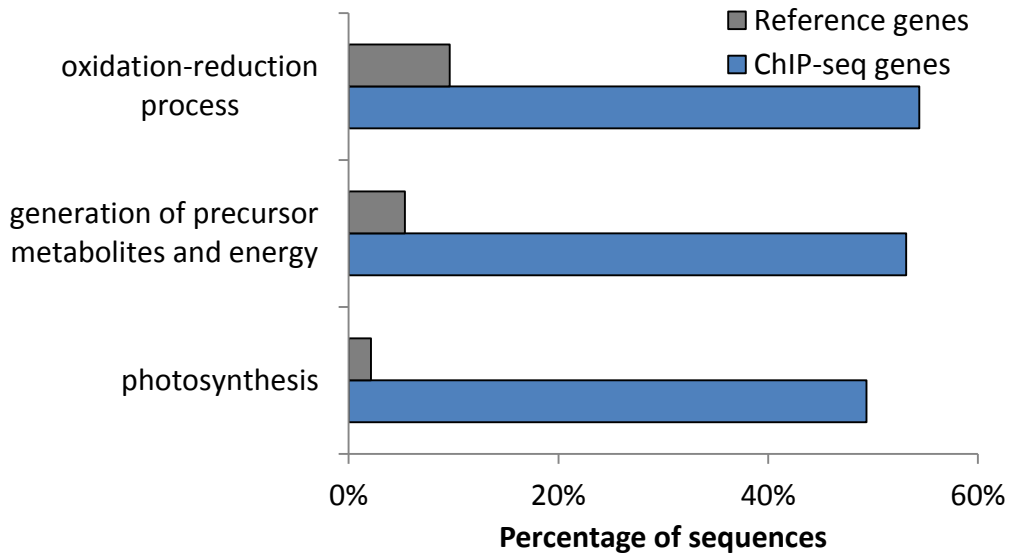


Figure 4.39. GO term enrichment of NAM-B1 target genes from regions overlapping in five peduncle samples.

4.3.8.4 Top genes from the genes identified by the five peduncle samples

For the peduncle samples out of the 465 peaks identified to be shared in five samples, 380 mRNAs were in the same contig as the peak and 318 peaks were within 1 kb of the mRNA. Only 276 mRNAs had BLAST hit to rice. Due to the poor correlations in significance value between peduncle samples for each individual peak (4.3.3.3) it was impossible to select the top ranking genes from the 5 peduncle samples. However nearly all the genes identified are related to chloroplast function. In particular there were 65 NADPH-dependent oxidoreductases, 38 DNA-directed RNA polymerase, 46 chloroplast ribosomal proteins and 70 photosystem related proteins (including cytochromes and apocytochrome f precursors).

In conclusion the peduncle samples had many functions related to photosynthesis, as would be expected from the biological function of *NAM-B1*. A wide range of target genes is bound in L1B proper pairs MAPQ >30 with some functions related to nutrient remobilisation such as transporters, although other target genes have less of an obvious link to known *NAM-B1* functions.

4.3.9 Hypothesis 7: Differential gene expression analysis will identify genes downstream of *NAM-B1* involved in nutrient remobilisation and senescence

RNA-seq is a powerful technology which can be used to analyse gene expression levels. In particular we hypothesise that using differential gene expression analysis comparing wild type and *NAM* RNAi plants (with reduced expression of *NAM-B1* homologues) we will identify genes regulated by *NAM-B1*. However unlike in ChIP-seq these genes may be indirect targets of *NAM-B1* and *NAM-B1* may regulate their expression via intermediate genes.

4.3.9.1 Read mapping

RNA was extracted from the flag leaves of *NAM* RNAi and wild type (WT) plants at 12 DAA (Cantu et al. 2011). Four replicates were sequenced from RNAi plants and 3 replicates from WT plants. Reads were mapped using bwa-aln allowing 1 mismatch to the gene models present on the pseudomolecule. This meant mapped reads were over 98.8 % similar to the reference sequence because reads were 85 bp single end reads (84/85 bp = 98.8 %). An average of 29.9 M reads were sequenced per sample with a mean of 4.2 % of reads mapping with 1 mismatch (**Figure 4.40**). Surprisingly RNAi replicate 2 had many more reads mapping despite the total read number being slightly

lower than average (23.5M reads). Filtering for reads with a higher mapping quality (MAPQ > 10 or > 30) reduced the proportion of reads which mapped. We decided to use all reads with MAPQ > 10 in the differential gene expression analysis to filter out reads with a very low mapping quality. We decided not to filter as stringently as MAPQ > 30 because very few reads would remain which might prevent us from identifying differentially expressed genes. We used two different statistical approaches to identify differentially expressed genes: DESeq2 (Love et al. 2014) and edgeR (Robinson M. D. et al. 2010) which are both available as R/Bioconductor packages.

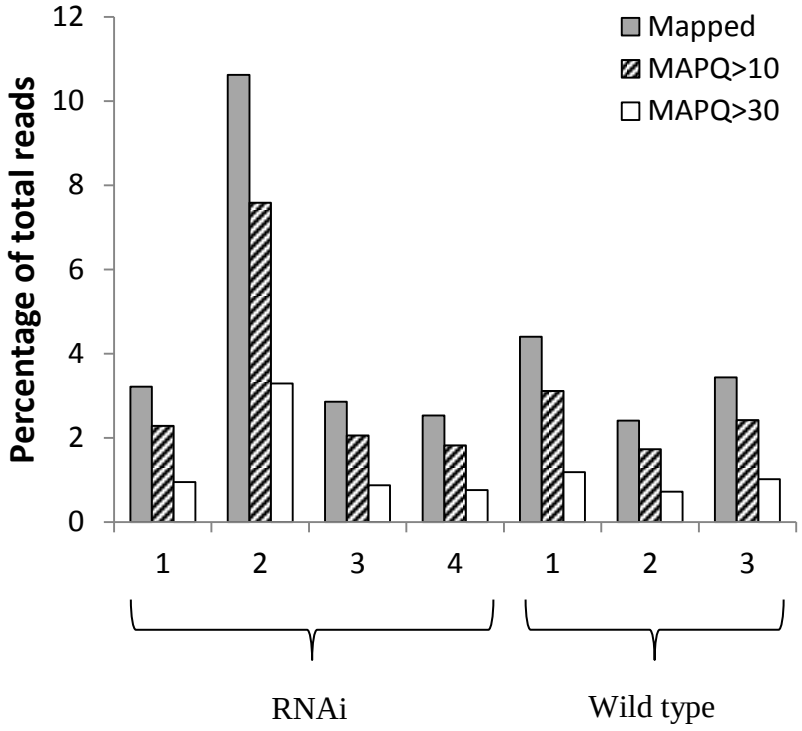


Figure 4.40. Read mapping quality allowing 1 mismatch during read alignment.

4.3.9.2 Differential expression analysis

Genes differentially expressed in wild type (WT) flag leaves compared to RNAi flag leaves were detected using DESeq2 and edgeR. The p-values generated by both DESeq2 and edgeR were adjusted for false discovery rates (FDR) across the multiple tests by using the Benjamini and Hochberg procedure as implemented in the R/Stats package. We decided to use a relatively relaxed cut-off value for significance (FDR < 0.1) because the aim of the differential expression analysis was to provide supporting information for the ChIP-seq experiment, therefore we wanted to identify as many

differentially expressed genes as possible. These differentially expressed genes would then be compared to the ChIP-seq data which would provide a stringent filtering step for true targets of NAM-B1.

Both DESeq2 and edgeR identified approximately equal numbers of differentially expressed genes (505 and 694 at FDR < 0.1) (**Figure 4.41**), of which 384 were common to both statistical methods. More genes were significantly down-regulated in WT plants than up-regulated (498 vs. 317) (**Figure 4.42**), when compared to RNAi plants. More than two times as many genes were identified by both DESeq2 and edgeR to be down-regulated than up-regulated (261 vs. 123). Many more unique genes were identified by edgeR than by DESeq2 for both down- and up-regulated genes, and this was also the case for genes with ≥ 2 fold differential expression.

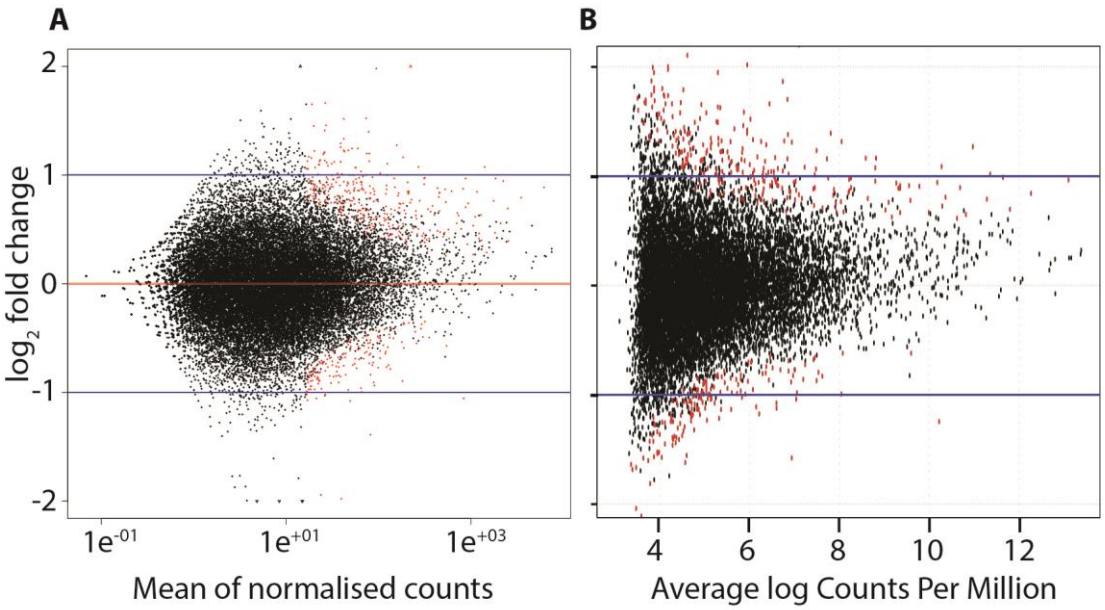


Figure 4.41. Expression differences between *NAM* RNAi and wild type flag leaves at 12 DAA. Detected by A) DESeq2, B) edgeR. Red dots are differentially expressed at FDR < 0.1. Blue lines represent a 2 fold change in expression level between genotypes.

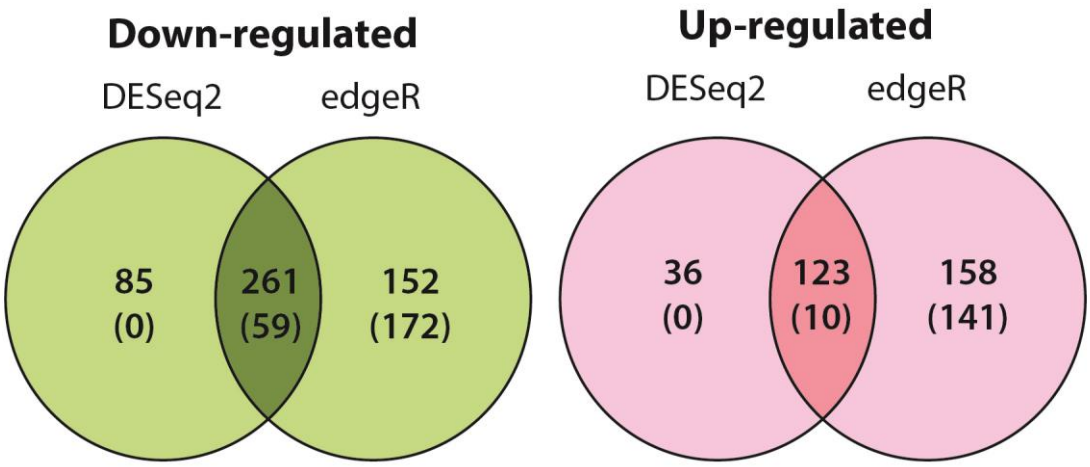


Figure 4.42. Comparison of differentially expressed genes identified by different statistical approaches. Venn diagrams show the unique and overlapping significantly down- and up-regulated genes identified by the R/Bioconductor packages edgeR and DESeq2. Values in brackets show genes with ≥ 2 fold differential expression.

4.3.9.3 GO enrichment in differentially expressed genes

The 384 genes which were found to be differentially expressed by both edgeR and DESeq2 were separated into down-regulated and up-regulated genes. Blast2GO was used to determine whether any specific GO terms were over or under-represented in these sets of genes. Out of the 123 genes up-regulated in WT plants only 87 were annotated in Blast2GO and a Fisher's exact test (FDR < 0.05) these showed no GO terms to be over or under-represented.

Out of the 261 down-regulated genes in WT plants, 229 were annotated in Blast2GO and 436 GO terms were identified to be over or under- represented. The 436 GO terms were divided into 3 categories: biological process (308), cellular compartment (64) and molecular function (64). The GO terms were condensed with similar GO terms by REVIGO (Supek et al. 2011) and plotted. The biological processes within down-regulated genes which were most over-represented were transport related genes including iron transport, water transport and hydrogen peroxide transport and responses to stimuli particularly abiotic stimuli such as iron, hydrogen peroxide, cadmium and temperature (**Figure 4.43**). Many biosynthetic processes were also over-represented (**Figure 4.44**) including primary metabolism, organic substance metabolism, nitrogen metabolism. Of the ten most significantly over-represented GO terms, nine were related to carbohydrate metabolism and photosynthesis (**Table 4.26**).

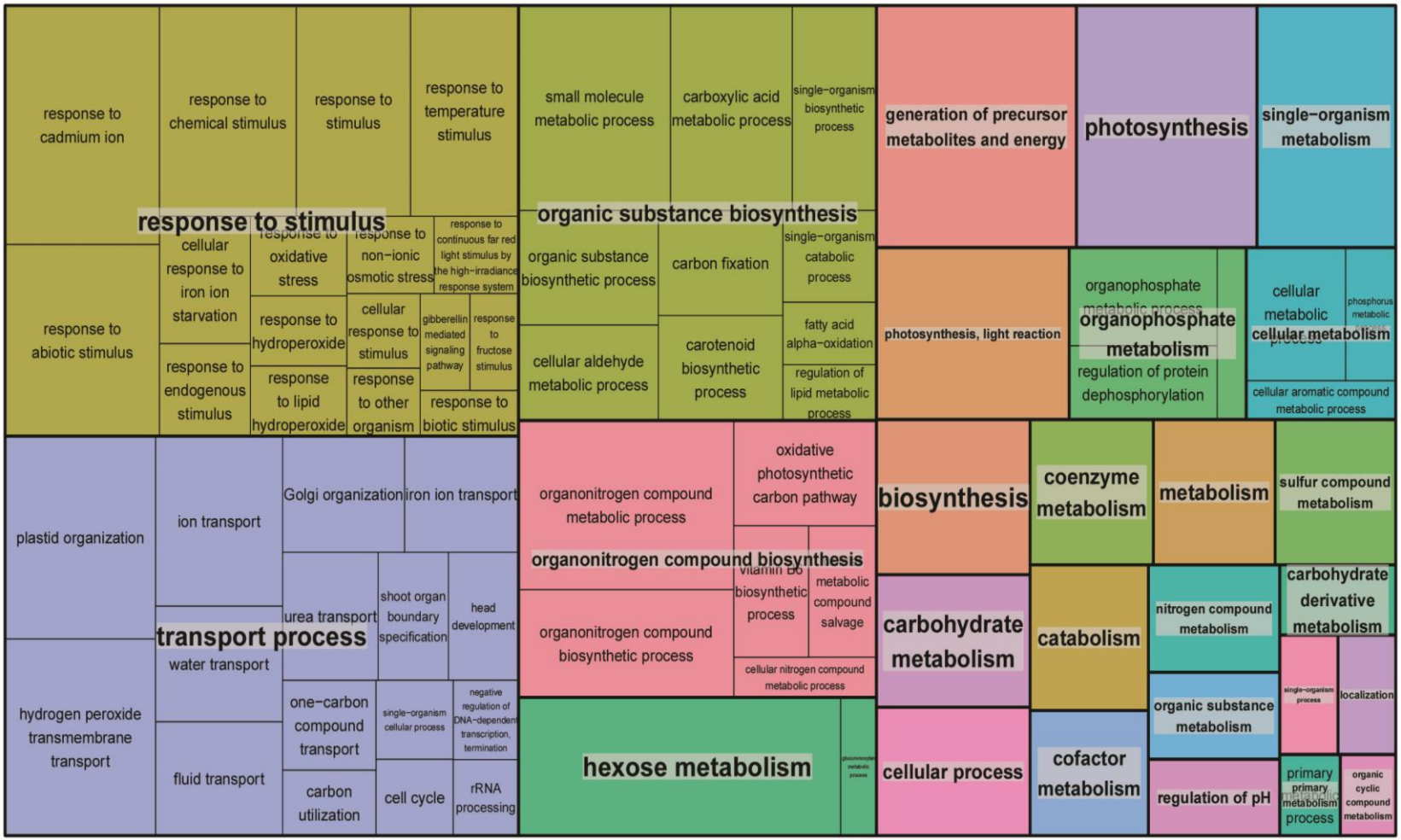


Figure 4.43. Representation of GO terms related to biological processes over-represented in down-regulated genes in WT compared to *GPC* RNAi plants. The size of the box represents the log(p-value adjusted for multiple testing).

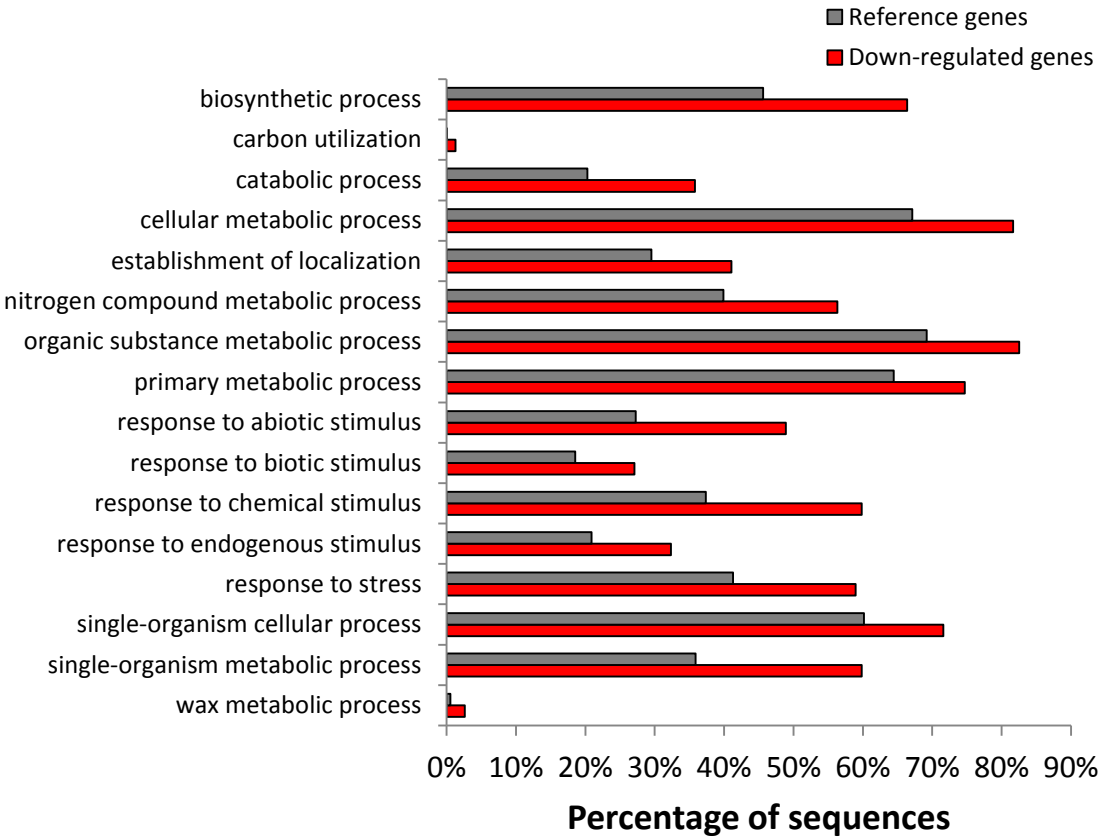


Figure 4.44. Level 3 biological process GO terms over-represented in down-regulated genes in WT compared to *GPC* RNAi plants.

GO number	GO name	Percentage of sequences in down-regulated genes	Percentage of sequences in reference	FDR
GO:0006091	generation of precursor metabolites and energy	22.71 %	5.54 %	6.67E-15
GO:0006006	glucose metabolic process	19.65 %	4.44 %	9.08E-14
GO:0019318	hexose metabolic process	20.09 %	4.66 %	9.11E-14
GO:0015979	photosynthesis	14.41 %	2.31 %	1.56E-13
GO:0006007	glucose catabolic process	17.03 %	3.57 %	1.02E-12
GO:0019320	hexose catabolic process	17.03 %	3.58 %	1.11E-12
GO:0046365	monosaccharide catabolic process	17.03 %	3.60 %	1.23E-12
GO:0005996	monosaccharide metabolic process	20.09 %	5.06 %	1.26E-12
GO:0046686	response to cadmium ion	20.52 %	5.68 %	1.58E-11
GO:0044724	single-organism carbohydrate catabolic process	17.47 %	4.17 %	1.88E-11

Table 4.26. The ten most significantly over-represented GO terms in down-regulated genes in WT flag leaves. FDR = false discovery rate.

Only 5 GO terms were under-represented in the down-regulated genes. Their molecular functions were related to (de)-phosphorylation during the cell cycle.

4.3.9.4 Functions of differentially expressed genes

A total of 384 genes were differentially expressed as detected by both edgeR and DESeq2. The most down-regulated genes in WT plants (i.e. during senescence) include a *NAM* transcription factor, two organelle translation related proteins (elongation factor Tu) a cytochrome b5-like protein and two genes related to secondary metabolism (zeaxanthin epoxidase and isoflavone reductase) (**Table 4.27**).

	DESeq2		edgeR		Rice BLAST		
mRNA ID	**FC**	**p-adj**	**FC**	**p-adj**	**Putative function**	**e-value**	**% ID**
mrna014085	0.1	6E-56	0.1	5E-42	#N/A	#N/A	#N/A
mrna121848	0.3	2E-11	0.2	1E-13	no apical meristem protein, putative, expressed	2E-174	85.0
mrna124230	0.3	4E-06	0.3	5E-07	elongation factor Tu, putative, expressed	7E-169	94.8
mrna093026	0.3	6E-04	0.2	1E-04	cytochrome b5-like Heme/Steroid binding domain containing protein, expressed	2E-99	89.1
mrna084138	0.3	4E-04	0.3	2E-05	zeaxanthin epoxidase, chloroplast precursor, putative, expressed	6E-136	88.0
mrna083482	0.3	2E-03	0.2	2E-05	expressed protein	8E-17	86.6
mrna081578	0.4	1E-05	0.3	8E-07	BTBZ2 - Bric-a-Brac, Tramtrack, and Broad Complex BTB domain with TAZ zinc finger and Calmodulin-binding domains, expressed	0	81.0
mrna071715	0.4	1E-04	0.3	3E-06	elongation factor Tu, putative, expressed	0	90.3
mrna083945	0.4	3E-04	0.3	6E-05	ki1 protein, putative, expressed	0	87.4
mrna075466	0.4	1E-03	0.3	2E-04	isoflavone reductase, putative, expressed	1E-104	77.8

Table 4.27. Top 10 down-regulated genes in WT compared to *NAM* RNAi flag leaves. FC = fold change, p-adj = p-value corrected for multiple testing, ID = identity.

Many of the most up-regulated genes in WT plants during senescence did not have a good BLAST hit (**Table 4.28**) and therefore their function remains to be found. Surprisingly one of the most up-regulated genes was the ribulose bisphosphate carboxylate (RuBisCO) small chain which is essential for the first step of carbon fixation, whereas we would expect that during senescence photosynthetic capability is down-regulated. Glutathione peroxidase catalyses the reaction of glutathione with hydrogen peroxide and may play a role in free radical scavenging during senescence (Strother 1988) to protect the leaf whilst its chloroplasts are disassembled.

	DESeq2		edgeR		Rice BLAST		
mRNA ID	**FC**	**p-adj**	**FC**	**p-adj**	**Putative function**	**e-value**	**% ID**
mrna046827	3.9	1E-10	4.8	4E-13	AIR9 protein, putative, expressed	0	84.0
mrna119498	3.8	3E-07	5.0	2E-10	mitosis protein DIM1, putative, expressed	9E-164	90.3
mrna124672	2.6	5E-05	3.0	8E-07	#N/A	#N/A	#N/A
mrna105961	2.1	7E-05	2.3	6E-05	#N/A	#N/A	#N/A
mrna094363	2.1	2E-02	2.5	6E-03	#N/A	#N/A	#N/A
mrna112976	2.1	6E-03	2.4	3E-03	#N/A	#N/A	#N/A
mrna106067	2.1	1E-02	2.4	2E-03	ribulose bisphosphate carboxylase small chain, chloroplast precursor, putative, expressed	2E-161	86.6
mrna086414	2.0	8E-03	2.3	4E-03	#N/A	#N/A	#N/A
mrna018403	2.0	2E-03	2.2	3E-04	#N/A	#N/A	#N/A
mrna050806	2.0	2E-02	2.3	5E-03	glutathione peroxidase, putative, expressed	6E-109	89.9

Table 4.28. Top 10 up-regulated genes in WT compared to *NAM* RNAi flag leaves. FC = fold change, p-adj = p-value corrected for multiple testing, ID = identity.

In conclusion, many genes which had functions related to nutrient remobilisation and senescence were differentially expressed in NAM RNAi plants compared to WT.

4.3.10 Hypothesis 8: Differentially expressed genes will contain shared DNA motifs in their promoters which facilitate co-expression

Co-expressed genes are often regulated by the same transcription factors. This regulation is mediated by specific DNA sequences in the co-expressed genes' promoters. Therefore we examined the promoter sequences of differentially expressed genes to see if any motifs were shared between them, which might indicate co-regulation.

4.3.10.1 Promoter motifs in differentially expressed genes

To assess whether specific promoter regions are enriched within differentially expressed genes, 1000 bp upstream of each of the 384 high-confidence differentially expressed genes was examined by MEME. Motif 3 (**Figure 4.45**) was present in 85 % of peaks and resembles motif 2 from the ChIP-seq sample L1B (**Figure 4.37**B) and motif 2 from the ChIP-seq 5 peduncle overlap sample (**Figure 4.37**E). Motifs 2 and 4 (**Figure 4.45**)

contain the repeated CxxC sequence which was previously identified by ChIP-seq (**Figure 4.37**B, motifs 1 and 5).

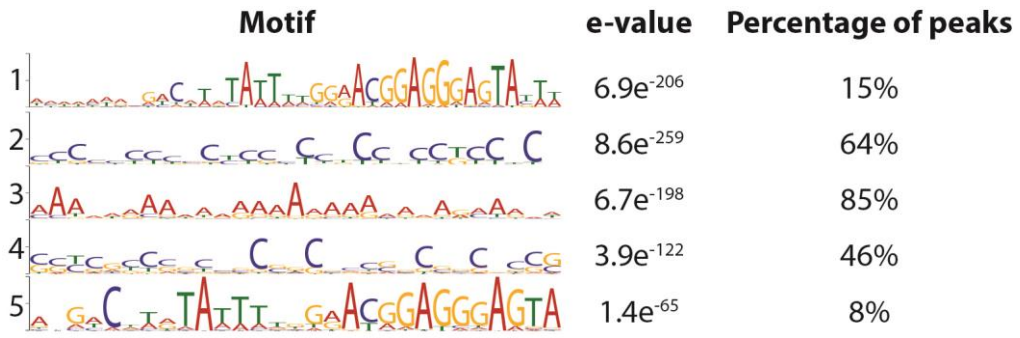


Figure 4.45. Motifs identified in promoter regions of differentially expressed genes in *NAM* RNAi lines.

4.3.10.2 Comparing ChIP-seq motifs to RNA-seq motifs

To do a more thorough comparison of motifs identified by ChIP-seq and RNA-seq we used MAST. The motifs identified by ChIP-seq were searched for in the contigs containing differentially expressed genes identified by RNA-seq. These motifs are also found in the contigs containing differentially expressed genes identified by RNA-seq (**Figure 4.46**). However these motifs are also found in contigs containing non-differentially expressed genes to a similar level. We might not have found a difference because the signal was lost by using the whole contigs, not just promoter regions. Alternatively no difference in motifs in differentially expressed and non-differentially expressed genes might be because the genes identified by RNA-seq are far downstream of the NAM-B1 binding event; therefore any traces of motif enrichment have been lost.

In conclusion, motifs identified by ChIP-seq were found in the contigs which contained differentially expressed genes, however the ChIP-seq motifs were also present in contigs containing non-differentially expressed genes.

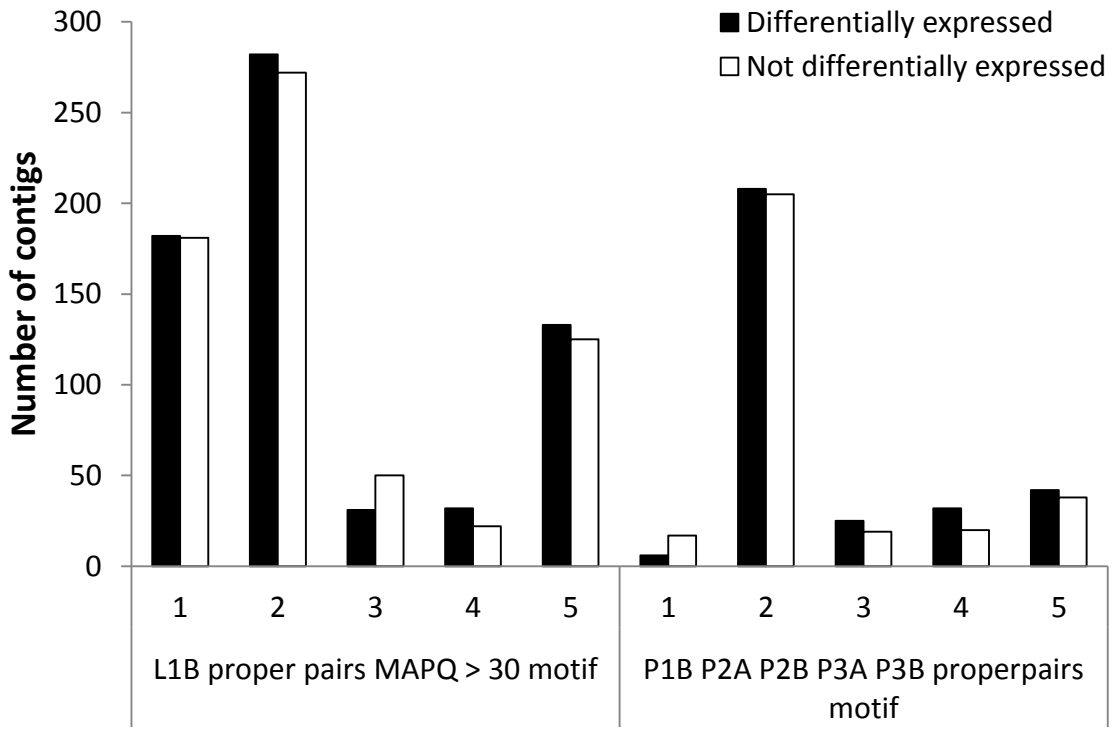


Figure 4.46. Number of ChIP-seq motifs in contigs identified to contain differentially or non-differentially expressed genes in *NAM* RNAi compared to WT plants.

4.3.11 Hypothesis 9: Direct targets of NAM-B1 will have their promoter sequences bound by NAM-B1 and will be differentially expressed in *NAM* RNAi plants

Combining the results of ChIP-seq and RNA-seq is a powerful technique to cross-validate target genes. ChIP-seq will identify direct targets genes of NAM-B1 by identifying DNA to which NAM-B1 directly binds. Assuming that the expression of the genes near to the binding sites is regulated by these binding interactions, these genes will also be differentially expressed if *NAM-B1* is removed (such as in *NAM* RNAi). Therefore comparing the ChIP-seq and RNA-seq experiments described above will identify high-confidence candidate target genes.

4.3.11.1 Genes identified by both ChIP-seq and RNA-seq

Surprisingly very few genes overlapped between the RNA-seq and ChIP-seq data sets. From the RNA-seq data set we only used the genes which were detected to be differentially expressed by both DESeq2 and edgeR (FDR < 0.1) (384 genes). The only ChIP-seq sample which had any overlapping genes to the differentially expressed genes was L1B proper pairs MAPQ > 30. Out of the 1128 significant peaks only 447 were within 1 kb of the nearest gene on the same contig. Out of these 447 genes which are directly bound by NAM-B1 only 4 were differentially expressed (**Table 4.29**). A Chi-

squared test showed there was no significant association between the genes identified by ChIP-seq and by RNA-seq (p-value 0.87).

Peak ID	-LOG$_{10}$ (qvalue)	Distance from peak to mRNA (bp)	mRNA ID	DESeq2		edgeR		Rice BLAST hit	Rice BLAST % ID
				FC	p-adj	FC	p-adj		
387	3.3573	0	mrna078184	0.64	0.03	0.63	0.06	expressed protein	79.1
508	3.3573	243	mrna118131	0.54	0.09	0.48	0.06	ankyrin repeat domain containing protein	86.6
771	2.07805	0	mrna011521	0.60	0.03	0.58	0.06	vacuolar ATP synthase subunit C	81.0
807	6.43142	0	mrna086085	0.51	0.00	0.50	0.00	lectin protein kinase family protein	86.8

Table 4.29. Peaks identified in sample L1B proper pairs MAPQ > 30 which have a differentially expressed gene within 1 kb and on the same contig. FC = fold change, p-adj = p-value adjusted for multiple testing, ID = identity.

4.3.12 Hypothesis 10: NAM-B1 directly binds the DNA regions (peaks) identified by ChIP-seq

To validate the peak regions identified by ChIP-seq we have used a biochemical technique called electrophoretic mobility shift assay (EMSA). In this technique the purified NAM-B1 protein is incubated with a biotin-labelled double stranded DNA probe. When run on a native gel the protein-DNA interaction is maintained. This interaction causes the movement of the biotin labelled probe to be retarded, causing shift in the band size, compared to the free unbound probe. The biotin label on the probe can be detected using an antibody to biotin conjugated to horse-radish peroxidase which is detected by chemiluminescence.

The first step is to produce purified NAM-B1. We decided to use heterologous expression in *E. coli* to produce the protein because this would enable production of large quantities in an inducible system. Purifying NAM-B1 from our transgenic lines would be unfeasible because the expression levels are low and very large amounts of plant material would be required. We decided to clone the *NAM-B1* coding sequence into a vector containing an N-terminal His tag. The His tag would allow rapid and specific purification of the NAM-B1 protein from the *E. coli* lysate. The N-terminus

was favoured for the positioning of the tag because the C-terminus is disordered and the insertion of the tag at this position might cause problems with protein folding and protein-protein interactions.

4.3.12.1 Expression of NAM-B1 in *E. coli*

We cloned the full length coding sequence of *NAM-B1* into the PCR8 vector. Subsequently it was transferred into two expression plasmids containing an N-terminal His tag: pDEST17 and pETG-10A. We optimised expression levels for bacterial strain (tested BL21-CodonPlus, Rosetta-gami 2, SoluBL21), plasmid (pDEST17 and pETG-10A), length of induction, amount of IPTG (to induce NAM-B1 expression) and temperature of induction. An example is shown in **Figure 4.47** where a Western blot was carried out using an antibody to the His tag to detect NAM-B1. Expression was induced by addition of IPTG (final concentration 1 mM) and incubation overnight. The expected NAM-B1 protein is observed at ~48 kDa, however for many conditions many smaller degradation products which are also his-tagged are detected. This degradation is evident at both 28 °C and 18 °C for both constructs. Expression is seen in both the soluble (**Figure 4.47**A) and insoluble (**Figure 4.47**B) fractions. To use the NAM-B1 protein for EMSA it needed to be soluble and as pure as possible. Therefore we used the condition/construct combination which gave a single band in the soluble fraction with few degradation products: construct EB (pETG-10A colony B) at 18 °C (**Figure 4.47**B).

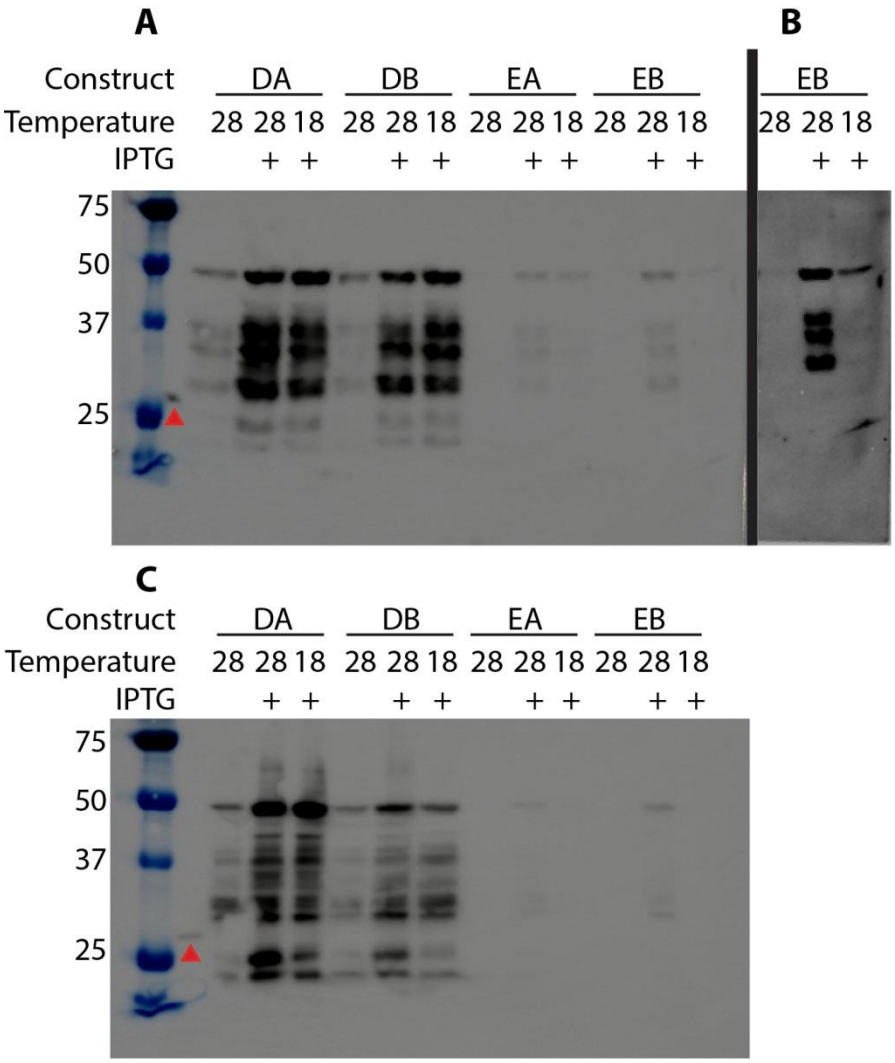


Figure 4.47. Western blot using anti-his antibody to detect NAM-B1 expression in soluBL21 *E. coli*. A) soluble fraction, B) final 3 lanes of soluble fraction with longer exposure, C) insoluble fraction. Temperature = induction temperature (°C). Constructs: DA = pDEST17 colony A, DB = pDEST17 colony B, EA = pETG-10A colony A, EB = pETG-10A colony B. Red arrowhead indicates his-tagged GFP positive control protein. Ladder size in kDa.

4.3.12.2 Purification of NAM-B1 from *E. coli* extracts

We expressed NAM-B1 in soluBL21 from the pETG-10A colony B construct at 18 °C overnight to produce a large amount of his-tagged NAM-B1. We used HPLC to purify the his-tagged NAM-B1 using a gradient elution from a His-trap column. The middle fractions of the elution using imidazole (e.g. * in **Figure 4.48**) had a strong protein band at the expected size of his-NAM-B1 (~48 kDa) (**Figure 4.48**A) and few other proteins. The faint protein bands which are smaller than NAM-B1 were mainly his-tagged as detected by Western blot (**Figure 4.48**B), indicating they were likely degraded pieces of NAM-B1.

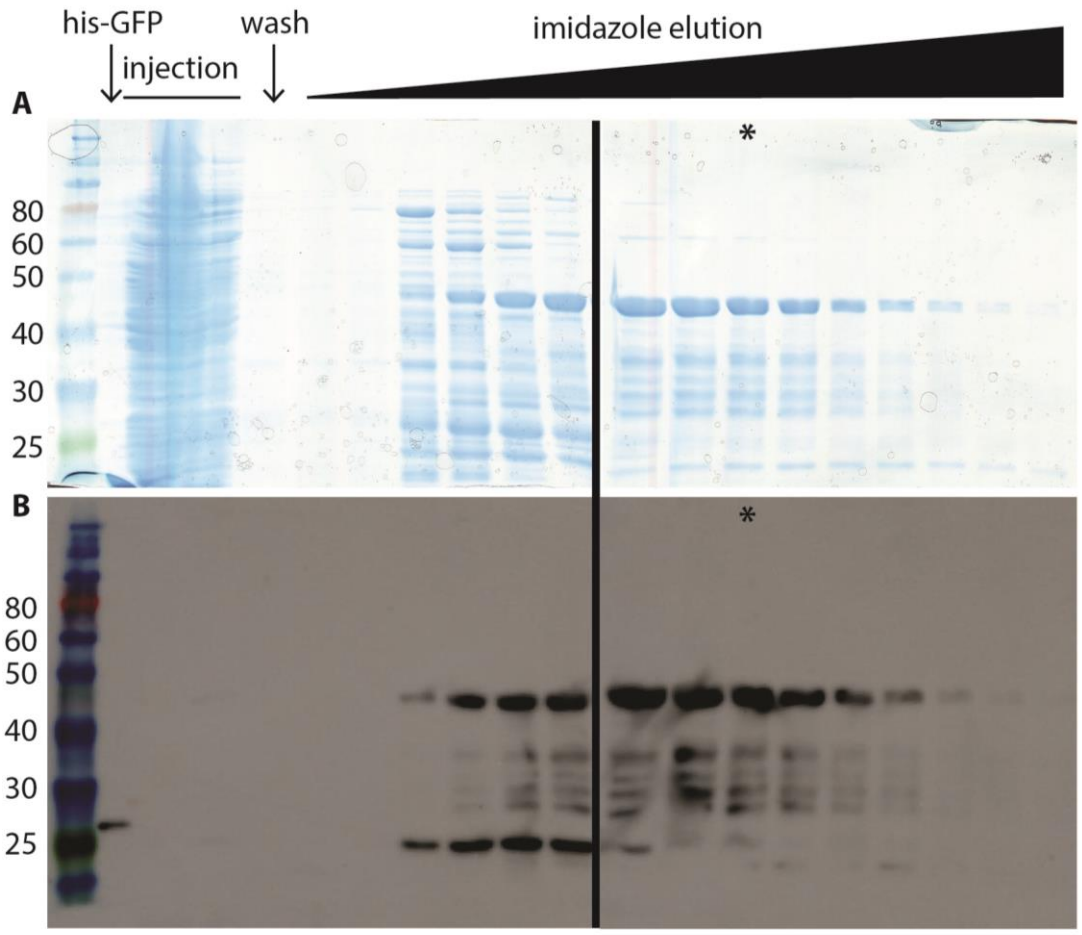


Figure 4.48. Purification of His-tagged NAM-B1 from the soluble fraction of *E. coli* extract using a His-trap column with gradient elution by imidazole. A) Protein gel stained with InstantBlue, B) Western blot using anti-his antibody to detect NAM-B1. Star (*) denotes fraction used in subsequent EMSA assays.

4.3.12.3 Validation of direct binding of ChIP-seq targets by electrophoretic mobility shift assay (EMSA)

To validate NAM-B1 protein binding to the DNA sequences identified by ChIP-seq in peak regions we carried out EMSA. We focussed on the motifs identified in L1B proper pairs MAPQ > 30 because these regions were non-repetitive so we hypothesised that they are more likely to have a biological function than repetitive regions. We used purified his-tagged NAM-B1 (* lane in **Figure 4.48**) and 25 bp oligonucleotides containing either motifs identified by MEME to be significantly enriched in peak regions or random DNA sequences. We found that NAM-B1 binds to **Figure 4.37**B motif 1, which was identified in L1B using proper pairs MAPQ > 30. This can been see in **Figure 4.49** where the free probe (lane 1) is shifted upon addition of NAM-B1 protein (lane 2). However addition of an excess of unlabelled probe (10,000 fold more than labelled probe) could not outcompete labelled probe binding by NAM-B1 (lane 3),

indicating that binding was non-specific. Furthermore addition of anti-his antibody to the reaction did not disrupt binding or cause a supershift (lane 4), which could mean that his-tagged NAM-B1 is not the protein responsible for the shifted band, or the his-epitope is not available after DNA binding by NAM-B1. Addition of increasing amounts of non-specific DNA (poly dIdC in lanes 2, 5 and 6) slightly decreased the size of the shifted band, indicating that the shifted band might represent non-specific binding rather than specific binding for the target motif. Addition of salmon sperm DNA (another non-specific DNA competitor) (lane 7) also reduced the size of the shifted band, indicating binding to be non-specific. Changes to the reaction buffer did not alter the binding pattern (lane 8).

To further characterise the non-specific binding characteristics of NAM-B1 we tested whether it would bind random DNA sequences, such as shown in lanes 9-15 of **Figure 4.49**. NAM-B1 bound to this random DNA sequence (lanes 9 and 10). Again binding was not disrupted by the addition of unlabelled probe (11). Addition of non-specific competitor DNA reduced the binding observed (lanes 12-14) and changing the buffer composition again made no difference to binding (lane 15).

We tested a number of other motif-containing probes, probes containing mutated motifs and random probes (**Table 4.30**). Every probe tested was bound by NAM-B1 which indicates the binding we observe is not sequence specific.

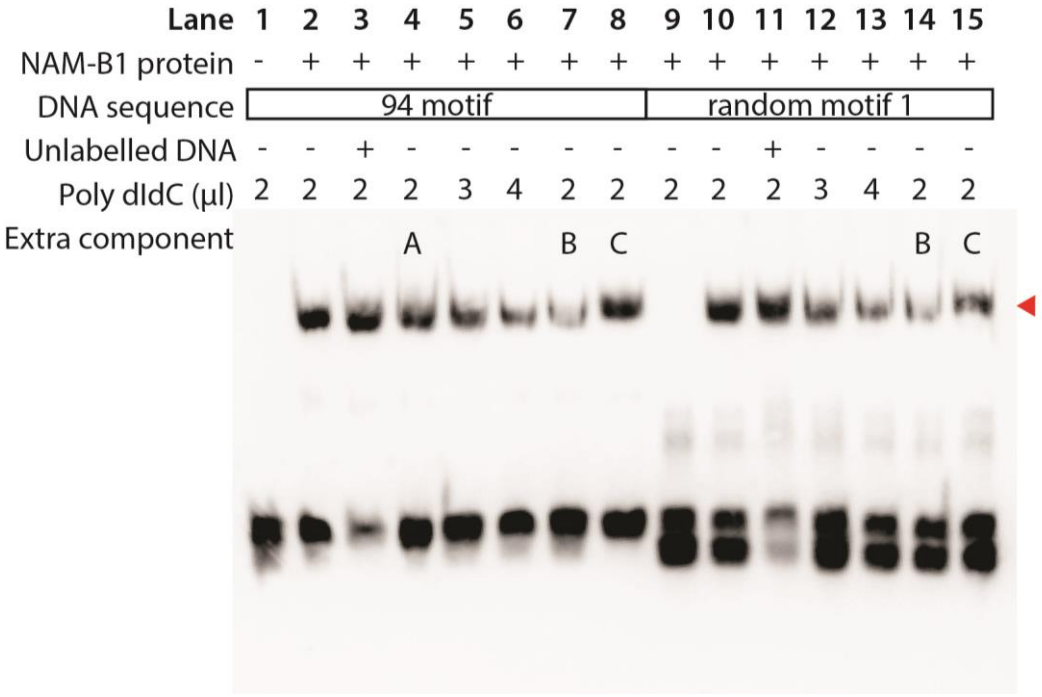


Figure 4.49. EMSA of his-tagged NAM-B1 binding to 25 bp oligonucleotides. Lanes 1-8: motif identified by MEME from L1B properpairs MAPQ > 30 (**Figure 4.37**B motif 1, the precise sequence used has the motif with 5 bp of random sequence either side: GCCGTCTCCTCCTCCTTCCCCTGTG), lanes 9-15: random 25 bp oligonucleotide (sequence: TGTTCTTGCTTAATGAGTTGCCGGA). Red arrowhead denotes shifted band. Extra components are as follows: A = anti-his antibody, B = salmon sperm DNA, C = glycerol, $MgCl_2$, NP-40 and KCl.

Motif ID	**Probe sequence (motif in red)**	**Motif number in Figure 4.37B**	**Gene also found by RNA-seq?**
Peak_498	TTGCGCTCGCCTCCCCCTCCAGACC	1	No
Peak_483	CACCCTCACCTGCGACGCCGAGCTG	5	No
Peak_904	GAGCCAGCATCGTCCTCACCTCTCC	1	No
Peak_94	GCCGTCTCCTCCTCCTTCCCCTGTG	1	No
Peak_11	GGCCGCCACCATGGCCTCCCTCACG	1	Yes
Peak_11_mutated	GGCCGTATATATATATATATTCACG	-	-
Peak_508	GCCCCTCCACGTCGCCCACGCTCGC	5	Yes
Peak_508_mutated	GCCCCATCAGGTAGCTCAGACTCGC	-	-
Peak_641	CCACTCCACATCTTCCTCCCGTGCA	1	Yes
Peak_641_mutated	CCACTGGAAGGCTCGTAGTAGTGCA	-	-
Random_1	TGTTCTTGCTTAATGAGTTGCCGGA	-	-
Random_2	ATCTGTGGACTGTCTGGTCCATGCA	-	-

Table 4.30. Probe sequences bound by NAM-B1. All peak motifs are based on motifs identified using L1B using proper pairs MAPQ > 30 (**Figure 4.37**B).

4.4 Discussion

4.4.1 Materials generated

Two main types of materials were developed to carry out ChIP-seq: the pseudomolecules which were used as the reference genome for read mapping and the transgenic lines which were the plant materials used.

4.4.1.1 Pseudomolecules

We have produced a gene-rich representation of the wheat genome, which contains 75,419 gene models. Building our own reference sequence was essential to carry out ChIP-seq because an ordered representation of the genome was required and was not available when we began this project. Within the last few years the sequence availability for wheat has greatly increased, however the sequences were generally in very small contigs and were all unordered. The availability of the IWGSC chromosome arm survey sequences (CSS) was essential for this project and they form the basis of the sequence for our pseudomolecules. To create an order within our pseudomolecules we used sequence data from a mapping population. This allowed us to order 36,923 gene-containing contigs within our pseudomolecules. To cover the maximum number of genes we also incorporated 40,872 contigs which were unordered but were assigned to the correct chromosome arm. This reference sequence enabled the detection of peaks within this pseudomolecule-reference for wheat.

A similar method to the one we used to order our contigs within the pseudomolecules has been applied to barley. This POPSEQ method (Mascher et al. 2013) differs from our method in the type of sequencing carried out on the mapping population: we used transcriptomic data, whereas POPSEQ uses whole genome sequencing. The POPSEQ method has recently been applied to wheat (IWGSC 2014) which produced a genetic order for 1,397,432 contigs of the IWGSC CSS containing 56,395 genes. The IWGSC has also used the GenomeZipper approach (Mayer et al. 2011) to order the wheat genes according to the syntenic order within related grass species. This created a virtual order in wheat of 66,085 genes (IWGSC 2014). Combining the POPSEQ and GenomeZipper approach allowed the positioning of a non-redundant set of 75,183 genes. This represents a more complete reference sequence than the pseudomolecules used for ChIP-seq, and reflects the current rapid progress in wheat genomics. The analysis of my ChIP-seq data was begun before these ordered representations were available and within the pseudomolecules 36,923 genes are ordered, which is only 2-fold lower than the

efforts of a large consortium. It would be of value to repeat the analyses presented within this chapter using the updated ordered reference sequences to increase the number of target genes of NAM-B1.

Despite the availability of these improved reference sequences, the pseudomolecule-reference sequence was still fit for purpose because it contained a large proportion of the genes within the wheat genome (>70 % of the estimated 106,000 functional protein coding genes (IWGSC 2014)). This reference was therefore suitable to use for ChIP-seq because we wanted to identify binding sites of NAM-B1 which are located near to genes, and therefore may have a biological function which can be further investigated using TILLING mutants or knowledge from other species.

We annotated the mRNAs in the pseudomolecules using Blast2GO to find GO terms. Over 45,000 mRNAs were fully annotated with GO terms which provide information about the biological processes, molecular function and cellular localisation associated with the mRNAs. However 35 % of the sequences could not be annotated due to poor BLAST hits. This may indicate that these mRNAs are unique to wheat and are not found in the species in the database against which BLAST was conducted. However the most comprehensive database available was used: UnitProtKB/Swiss-Prot (v2013-12) contained 541,954 high quality manually annotated protein sequences and the similarity at the protein level is generally well conserved between species. Therefore, it is perhaps more likely that many of these un-annotated sequences are not complete mRNAs, although high quality mRNA gene models were used as far as possible. To gain further information about the mRNA sequences the best hit from the rice cDNA (genome v7.0) was found by BLAST. This gave putative homologues of 36,659 mRNAs. However 83 % of the mRNAs which had strong rice BLAST hits (>70 % identity) were already annotated by Blast2GO, implying that 23,276 mRNAs (31 %) had no annotation information available during our analysis. It is possible that some of these non-annotated genes were long non-coding RNAs or the gene models are incorrect and therefore are missing regions which would be required to identify their homologues by BLAST.

4.4.1.2 Transgenic plants expressing HA or FLAG-tagged NAM-B1

Transgenic plants expressing HA or FLAG-tagged NAM-B1 were generated by NIAB. Several lines with single copy insertions were identified, however these lines had very low mRNA expression levels therefore were not used for the replicated ChIP-seq experiment. Three FLAG-tagged lines had mRNA expression levels very similar to the native promoter level, as would be expected because the construct contained the *NAM-B1* ~8.2 kb native promoter. Therefore we selected these lines to use even though they had multiple copy numbers. We considered it important to have mRNA expression levels similar to the native gene to increase the chance that NAM-B1 would interact with its normal target sequences.

The T_1 plants used for ChIP-seq had very variable mRNA expression of *NAM-B1* between individual plants. To examine the protein expression level of NAM-B1 in T_1 plants we used Western blots. However out of the 12 samples used for ChIP-seq the FLAG-tagged NAM-B1 protein was only detected in one peduncle sample (P2A), and this was only in the purified nuclei and not in the total input protein fraction or in the eluted fraction from immunoprecipitation (IP) where NAM-B1 would also be expected to be detected. This is concerning because NAM-B1 should be detected in the eluted fraction from the IP because this fraction was then de-crosslinked and sent for sequencing, therefore if NAM-B1 was not present in this fraction perhaps the IP did not work or the tagged-protein was not expressed. However, this lack of detection may be due to the very small sample amount used in the Western blot to check protein expression levels. Furthermore transcription factors are expected to be expressed at very low levels which may be hard to detect with Western blot. A much larger amount of sample was sent for sequencing (280 µl compared to 20 µl used for Western blot) which should have enhanced detection if the amount of NAM-B1 protein present was very low.

It would have been desirable to validate the function of the transgene by checking that the FLAG-tagged NAM-B1 could complement the mutant phenotype. This should be possible because the variety we transformed was a natural mutant (*nam-B1*), as are most elite varieties (Uauy et al. 2006b). However this was technically challenging because the effect of NAM-B1 is a quantitative trait, due to the presence of the 2 homoeologues and 3 paralogues in this genetic background, therefore effects would be subtle. From previous studies, introduction of *NAM-B1* accelerates senescence by approximately 1.5 days (Uauy et al. 2006a), which requires many replicates to detect. We grew our plants

in controlled environment rooms which unfortunately do not allow enough plants to be grown to detect these subtle differences. Another way to validate the function of the transgene would be to examine the protein content of the grain of the transgenic line, compared to control plants, and this experiment will be carried out in the future. The destructive nature of determining grain protein content prevented us from carrying it out in early generations of the transgenics because we needed to keep all the grain for future experiments. Again to detect the relatively small difference (1.5 % in grain protein content) (Uauy et al. 2006a) many plants would be required.

4.4.2 Hypothesis 1: The stringency of read mapping will affect the number of binding sites identified by ChIP-seq

4.4.2.1 The number of mismatches affects number of binding sites identified

The first step of ChIP-seq analysis is read mapping. We decided to use bwa (Burrow-Wheeler Alignment tool) which is a fast accurate aligner (Li Heng and Durbin 2009) which outputs mapping quality scores which can be used to select only high quality alignments. We used the algorithm bwa-aln which allows precise manipulation of the number of mismatches between the reads and the reference. We tested different read mapping stringencies, altering the number of mismatches allowed between the read and the reference sequence. As expected we found that as the number of mismatches increased, the number of mapped reads increased, and the number of peaks discovered increased. Increasing the number of mapped reads is known to increase the number of peaks identified (Bailey Timothy et al. 2013). Increasing the read depth allows more weak binding sites to be identified, sites which are only occupied in certain cell types within the non-homogeneous sample, or sites only occupied for short periods of time. Therefore it was important to maximise the number of mapped reads to identify the maximum number of binding sites. However, wheat's polyploid nature increases the difficulty of accurately mapping reads and we wanted to ensure that reads were mapping to the correct homoeologue. Therefore when using 100 bp paired end reads for our replicated experiment we decided to allow 2 mismatches which would ensure reads were 98 % similar to the reference: a trade-off between maintaining homoeologue specificity and losing valuable information by not mapping reads.

Using 2 mismatches (98 % similarity) in the T_1 samples on average 12.3 % of reads mapped in proper pairs which was very low compared to the >70 % of reads uniquely mapped reads in ChIP-seq in other organisms (mouse, human or Arabidopsis) (Bailey Timothy et al. 2013). This may be a cause for concern about the quality of the data (e.g.

inadequate read length or excessive amplification during PCR). However increasing the read length did not improve the percentage of reads mapped in the T_0 (only 6 % of reads were mapped using ~98 % similarity). Excessive amplification during PCR might have been a problem if the immunoprecipitation step was not very successful. To ensure the immunoprecipitation step was successful we used similar quantites of plant material (~0.5-1 g of leaf or peduncle material) as were recommended for Arabidopsis (Kaufmann et al. 2010) and used high quality commercial antibodies (Sigma). A likely reason for our low percentage of reads mapped is that our reference sequence only contained ~4 % of the total wheat genome (459 Mb of 17 Gb), therefore many regions where NAM-B1 may bind were not present in our reference, although the majority of these missing regions are likely to be repetitive elements which make up 80 % of the wheat genome. Furthermore proteins which bind to repetitive regions will lead to the sequencing of large portions of repetitive DNA which is not mappable. It appears in the case of NAM-B1 regions bound in the peduncle are indeed repetitive elements and this may further explain why the percentage of mapped reads is low because not only are these repetitive elements missing from our reference, reads originating from these regions are difficult to map.

4.4.2.2 The mapping quality of reads affects the number of binding sites identified

We also investigated whether using only reads with a higher mapping quality would affect the number of peaks identified. As a baseline we analysed the samples using properly paired reads, because reads which are not mapped with their pair are likely to be mapped in the wrong location. However the fragmented status of the contigs which were used to make the pseudomolecules (average size ~ 7 kb) may have prevented some read pairs from being able to map to the same contig.

A set of guidelines issues by the ENCODE and modENCODE consortia recommend using uniquely mapped reads (Landt et al. 2012) to make sure peaks detected are reliable. However the definition of uniquely mapped reads is very hard to measure in practise due to the uncertainties that are inherent to probabilistic read mapping software and the uncertainties related to sequencing errors. Therefore to determine whether reads are mapped to the correct location we considered the mapping quality (MAPQ) which is calculated by bwa during read mapping:

$MAPQ = -10 \times \log_{10}(\text{probability the mapping position is wrong})$

This is a quantitative measure of the probability reads are mapped correctly, where:

Probability the mapping position is wrong	MAPQ
0.1 (1 in 10 reads)	10
0.01 (1 in 100 reads)	20
0.001 (1 in 1000 reads)	30

The guidelines in the bwa publication (Li Heng and Durbin 2009) suggest that confidently mapped reads will have a MAPQ > 10 therefore we have analysed our samples with this cut off. However this still allows 1 in 10 reads to be mapped incorrectly which over millions of reads will be many reads and we were concerned this might bias results. Therefore we also tested a stringent MAPQ > 30 cut off.

Increasing the MAPQ of reads dramatically reduced the percentage of mapped reads from 17.3 % for all mapped reads to 4.3 % for reads mapped in proper pairs with MAPQ > 30. We were concerned that such drastic reduction in read number would reduce the identification of peaks, which was seen, therefore we analysed all samples in parallel samples with 3 different MAPQ (proper pairs, proper pairs MAPQ > 10 and properpairs MAPQ > 30) to try to understand whether the peaks identified with less stringent mapping quality were noise or were biological targets of NAM-B1.

Comparing between samples it was noticeable that L1B had far more peaks identified than any other replicate, regardless of the mapping stringency. ChIP-seq studies of transcription factor binding sites using full reference genomes often report many thousands of peaks, and using stringent mapping quality L1B was the only sample which had over 1000 peaks. Therefore this sample was of particular interest during our subsequent analyses.

4.4.3 Hypothesis 2: True NAM-B1 binding sites will be detected in several replicates, the significance of binding sites will be correlated between samples and binding sites may be different between leaf and peduncle samples

We examined whether binding sites were found in more than one sample by looking for overlapping peak regions using bedops. In general there were very few overlapping peaks between leaf and peduncle samples. This may reflect biological differences in binding sites because NAM-B1 may regulate different genes to control tissue-specific processes.

In the peduncle using properly paired reads 5 out of 6 samples share over 50 % of their peaks. However when using more stringent read mapping conditions (MAPQ > 10 or MAPQ > 30) very few peaks are identified in peduncle samples, although 23 peaks are shared between P3A and P3B. This might suggest that the peaks identified are not true peaks and are artefacts of mapping. However since these peaks are identified in 5 independent samples, all of which have independent control samples, we hypothesise that there is a biological significance to these binding sites. The mapping quality of the reads in these regions is likely very low because they are repetitive elements (4.4.4) and therefore they are not detected with more stringent mapping parameters which discard reads with multiple mapping sites. If we can confirm biochemically that NAM-B1 binds to these repetitive regions it will be challenging to identify the biological function of this binding. To identify the genes which are being regulated will be difficult because there are so many copies of the binding sites within the genome. An additional worry about these samples is that the level of correlation of peak significance between the five peduncle samples is generally quite low (R^2<0.2) except for P3A and P3B which have a R^2 = 0.38 which is still not very high and does not indicate a strong relationship. We expected that the significance of peaks would be related between samples, however this does not appear to be the case, perhaps suggesting these peak regions identified are noise.

Amongst leaf samples there were fewer overlapping peaks than in the peduncle samples. Two leaf samples (L1A and L1B, which are biological replicates of the same transgenic line) however shared 330 peaks using properly paired reads, and using more stringent mapping (MAPQ > 10) these two samples shared 221 peaks and 8 peaks at MAPQ > 30. Regardless of stringency of mapping, the most significant peaks were most correlated between the two samples, with less significant peaks showing less correlation between samples. This may indicate that peaks further down the list are noise.

It is difficult to set a threshold for significance to choose peaks for further characterisation, because even only selecting peaks identified in several samples produces a list of over 100 regions to investigate. To set a threshold for true signals, information about reproducibility between replicates may be used. In further work we will use the statistical method called irreproducible discovery rate (IDR) (Li Qunhua et al. 2011). This method has been proposed to not only evaluate whether peaks are identified in more than one sample but also incorporates information about their relative

rankings of significance, even where the overall scale of the significance values differs between samples (e.g. due to technical differences between experiments). The IDR gives an indication of what proportion of peaks are true signal, and which are not reproducible (i.e. noise). Therefore calculating the IDR for different combinations of samples, for example for the five peduncle samples with shared peaks, will answer more definitively whether the peaks identified are true signal or are noise caused by read mapping uncertainties.

4.4.4 Hypothesis 3: NAM-B1 binding sites will be unique within the genome to ensure specific binding of target genes

Surprisingly many of the peak regions identified were highly repeated in the wheat genome. For the 465 regions which were common to five peduncle samples, the median number of BLAST hits with >99 % similarity was 250, which is higher than the median value for any one individual peduncle sample. It is possible that these highly repetitive regions are identified as an artefact in five out of six independent samples. However these regions represent over 50 % of peaks identified in these five samples, and all samples are completely independent because a separate control sample was used for each. Alternatively the reproducibility of these peak regions between many samples may reflect that they are true binding sites of NAM-B1. In mammalian systems it has been reported that transposable elements play an important role in the evolution of transcriptional control via the incorporation of transcription factor binding sites (TFBS) into transposable elements (Bourque et al. 2008). In particular the promoter regions are enriched for the Alu repeat, which is a dimer composed of a central A-rich region flanked by two similar sequence elements of about 130 bp (Polak and Domany 2006). Binding of transcription factors to binding sites within Alu repeats has been shown by electrophoretic mobility shift assays (Hambor et al. 1993) and specific classes of genes, e.g. biosynthetic genes, have Alu-associated TFBS, indicating a role in regulation of specific biological pathways. Therefore it is possible that NAM-B1 really does interact with repetitive regions of the wheat genome. Many of these repetitive regions had no hits to the TREP (Triticeae Repeat Sequence) database, which suggests they are not transposable elements and may be a novel type of repetitive element with an important biological function.

In contrast to the peduncle samples where repetitive regions were common binding sites, in the leaf sample L1B many unique sites were identified. In particular when using stringent read mapping quality (MAPQ > 30) 54 % of the 1128 peaks identified had no

other matches at over 99 % similarity, indicating they were unique sequences within the genome. It was surprising that this was the only sample which had a high proportion of unique sequences. This might reflect that this one sample is an outlier and something technically did not work, or instead it might be the only sample with true signal which is supported by the number of peaks (>1000) being more similar to other transcription factor studies reported, for example (Bolduc et al. 2012, Shamimuzzaman and Vodkin 2013), than the other samples which had relatively few peaks. However, it is possible that NAM-B1 would exert its effects through a few direct target genes, therefore identifying few peaks might not indicate failure of the other samples. Alternatively the identification of relatively few peaks might indicate the need for deeper sequencing to find all binding sites as has been previously reported (Landt et al. 2012).

4.4.5 Selecting peaks to further characterise

We examined all the samples using a variety of different metrics to try to determine which samples contained the most reliable peak regions. We used different read filtering stringencies, examined overlapping peaks between samples and examined the repetitiveness of peak regions. However, the binding sites identified were quite variable between leaf and peduncle tissues, not only in the precise regions identified but in the type (e.g. repetitive sequences) and number of peaks identified.

This is the first study to our knowledge of ChIP-seq in wheat and we were not sure what to expect in terms of peaks identified, especially considering the challenges associated with the large genome size and polyploid nature. Therefore we did not feel it was possible to identify one list of binding sites which contained true signal and instead carried out further analyses on several different lists detailed in the results section (0). The five lists were:

- Overlapping peaks between L1A and L1B proper pairs MAPQ > 10
- Overlapping peaks between L1A and L1B proper pairs
- Peaks from L1B proper pairs MAPQ > 30
- Overlapping peaks between P1B, P2A, P2B, P3A, P3B proper pairs
- Overlapping peaks between P3A and P3B proper pairs

4.4.6 Hypothesis 4: NAM-B1 will bind promoter regions to regulate gene expression

We hypothesised that NAM-B1 will bind promoter regions to regulate gene expression. We tested this hypothesis by identifying which genomic features (e.g. promoters) peaks were located in, and examining the distance from peaks to transcription start sites (TSS).

Sample L1B (proper pairs MAPQ > 30) had surprisingly few peaks within the promoter region (only 42 %) considering that 55 % of the total reference sequence was classified as promoter sequence. Many binding sites were located within intergenic regions, particularly 3' of the gene body which may indicate that enhancer or repressor elements are located downstream of these genes. Binding sites were also enriched within exons whereas binding was largely excluded from introns.

In peaks which were overlapping in samples L1A and L1B using either properly paired reads or properly paired reads with MAPQ >10 the majority of binding sites were in the promoter region as expected and very little binding occurred in the exons and introns. This localisation of binding might indicate that the binding sites are true regulatory regions, although unexpectedly binding patterns were almost symmetrical around gene bodies with binding sites 3' of the gene being nearly as frequent as sites 5' of the gene in the promoter. This may again indicate enhancer or repressor elements within the 3' sequence. Alternatively this enrichment in the 5' and 3' sites to equal degrees may be artefacts caused by our reference sequence. The reference sequence is composed of contigs which have median lengths 6,000 – 7,000 bp which we explain the enrichment up to these distances from genes.

Peaks which are detected in either five peduncle samples or in two peduncle samples have a similar profile relative to the TSS. These profiles have a large enrichment of binding sites immediately upstream of the TSS particularly in the 5'UTR and also a high proportion of binding sites are located in the 3' UTR. This pattern of binding is very similar to previously reported profiles for transcription factors (e.g. KN1 in maize (Bolduc et al. 2012)). Binding in the promoter regions or 5'UTR matches the expectation that transcription factors will bind 5' of genes to influence their expression.

The majority of the binding sites identified were located in promoter regions, as expected, but the sites in promoter regions were not more frequent than would be expected from the distribution of promoter sequences across the pseudomolecule.

However, it appears that the binding sites were non-random and reflected an enrichment of bound DNA because overall they were distributed to the genomic features in a different proportion than the overall pseudomolecules genomic features distribution. For example in all five lists of peaks, binding sites were nearly or completely absent in introns which contrasts greatly to the reported results for a soybean NAC transcription factor whose binding sites were frequently within introns (46 % of sites) (Shamimuzzaman and Vodkin 2013). In human cells the majority of interactions of transcription factors with DNA do not cause gene expression changes (Cusanovich et al. 2014). Therefore we hypothesise that many of the binding sites we have identified are not actively regulating genes, which might explain why many binding sites are located outside of promoter regions. Furthermore chromatin structure is very complex and therefore sites within positions other than the promoter may also regulate gene expression (Li Guoliang et al. 2012). These non-regulatory interactions or chromatin structure regulation may also explain why the binding site distributions were unique to each different sample, although this might also reflect differences between the function of NAM-B1 in leaves and peduncles.

4.4.7 Hypothesis 5: NAM-B1 binding will be mediated through a specific DNA sequence motif

The binding sites of a few other NAC transcription factors have been identified using a variety of methods including ChIP-seq (Shamimuzzaman and Vodkin 2013), 3D-structural modelling (Zhu Q. K. et al. 2014) and immuno-precipitation (Wang Xiao and Culver 2012). A consensus sequence which is bound by many NAC transcription factors including the senescence related *AtNAP* has been identified: (CGT[GA])(Welner et al. 2012, Zhang K. and Gan 2012). This motif was only identified in 11 % of the peaks in the sample L1A and L1B using proper pairs, and was not found in the other samples. The low detection of the NAC consensus binding site may also be caused by us only looking for five motifs per sample, we should repeat the analysis searching for many more motifs which might then identify the consensus binding site.

Motifs were found for all five samples examined. In the L1B proper pairs MAPQ > 30 sample, and the two peduncle samples at least one motif is identified in over 80 % of the peak regions. This is a much higher proportion of peak regions than for a random input of sequences. However in sample L1B proper pairs MAPQ > 30 a similar motif containing CxxC is identified to in the random sample, albeit at a much higher frequency. This indicates that the CxxC motif is present in most of the wheat genome,

however it is specifically enriched in regions to which NAM-B1 binds in the sample L1B proper pairs MAPQ > 30. In the peduncle samples the motifs identified had much more significant e-values (smaller than e^{-300}) and the sequences identified were different to in the leaf samples, with the CxxC motif not being significantly enriched. Instead a polyA motif was identified which was present in over 80 % of peaks in both the 5 peduncle overlap sample and the P3A-P3B overlap sample, and this polyA sequence was not found in the random input sequences, suggesting it is a specific NAM-B1 binding site.

The identification of different motifs in the leaf and in the peduncle give further support to NAM-B1 regulating different genes in these tissues and we hypothesise that binding specificity might depend on protein-protein interactions. It is known that NAC transcription factors form homodimers and heterodimers with other NAC transcription factors (Jeong et al. 2009, Olsen Addie Nina et al. 2005), and these interactions will depend on the tissue specific expression of interaction partners.

Although we have identified several promising motifs it will be necessary to try further motif searches using different algorithms for example allowing gapped alignments. For example CisFinder has the ability to find motifs which contain binding sites for two transcription factors (Sharov and Ko 2009) and GLAM2 detects motifs containing gaps (Frith et al. 2008). Identification of motifs which contain gaps or multiple binding sites will be an important next step because NAC transcription factors are reported to bind as dimers. The binding sites for the partners in the dimer are likely to be spaced somewhat apart within the DNA sequence as shown by computational modelling of AtNAC1 homodimers binding to DNA (Zhu Q. K. et al. 2014).

4.4.8 Hypothesis 6: NAM-B1 target genes will have common functions related to nutrient remobilisation and senescence

4.4.8.1 Leaf sample L1B proper pairs MAPQ > 30

In leaves, using sample L1B proper pairs MAPQ > 30, we found that the target genes of NAM-B1 were enriched for several GO terms, the majority of which are related to secondary metabolism specifically phenylpropanoid metabolism. Phenylpropanoids are a very diverse family of compounds which contribute to responses to both biotic and abiotic stresses (Vogt 2010) including pathogen attack, signalling and light stress responses. Phenylpropanoids are derived from intermediates of the shikimate pathway by combinations of reductases, oxygenases and transferases including cytochrome

P450-dependent monooxygenases which are heme-containing enzymes (Meyer K et al. 1996). Phenylpropanoids include lignins which are important for cell wall structure and flavonoids which are pigments. The red pigmented anthocyanins, a class of flavonoid, are known to be upregulated during leaf senescence, as seen in the red leaves of deciduous trees in autumn (Ougham et al. 2008) and protect leaves from photo-oxidative damage during this period of nutrient remobilisation (Feild et al. 2001). Although wheat leaves do not show such dramatic changes in colour, flavonoid biosynthetic genes are up-regulated in senescing flag leaves of wheat (Gregersen and Holm 2007) and barley (Jukanti Aravind K. et al. 2008). We hypothesise that *NAM-B1* up-regulates flavonoid biosynthesis to act as photo-protectants alongside accelerating senescence.

The enrichment of very-long chain fatty acid metabolism genes in NAM-B1 targets may reflect the macromolecular breakdown and nutrient translocation from leaves to the grain during senescence. Fatty acid breakdown begins at the onset of senescence in Arabidopsis (Yang Zhenle and Ohlrogge 2009) and in wheat, fatty acid catabolic enzymes are up-regulated during senescence (Gregersen and Holm 2007).

Relatively few GO terms were enriched within the ChIP-seq L1B sample targets, compared to the RNA-seq data. This may indicate that NAM-B1 mediates its effects on phenotype through several intermediate steps for example other transcription factors.

4.4.8.2 Genes identified by five peduncle samples

The genes which were identified as targets of NAM-B1 in five peduncle samples were significantly enriched for over 100 GO terms. These were mainly related to photosynthesis, generation of precursor metabolites and energy, and oxidation-reduction processes. Closer examination of the genes identified (using homology to rice) showed that nearly all these target genes encode chloroplast localised proteins including NADPH-dependent oxidoreductases, chloroplast ribosomal proteins and photosystem proteins.

This result shows that NAM-B1 binds to nuclear-encoded chloroplast genes which may alter the rate of senescence of the plant. We have noticed that in *NAM* RNAi plants the peduncle retains a green colour until the grain are fully mature, which supports the idea that *NAM-B1* regulates chloroplast lifespan. We hypothesise that *NAM-B1* represses the transcription of these chloroplast related genes. This repression would lead to chloroplast degradation which is one of the first steps in leaf senescence (Hörtensteiner

and Feller 2002). Control of chloroplast degradation also influences nitrogen export, for example nitrogen from the chloroplasts accounts for approximately 90 % of nitrogen exported from senescing rice leaves (Morita 1980). Therefore we hypothesise that *NAM-B1* causes the degradation of chloroplasts which releases nitrogen to be used in grain filling, explaining the increased level of nitrogen in grain wheat varieties containing a functional copy of *NAM-B1*. Consistent with this hypothesis previous work has shown that when *NAM-B1* homologues are down-regulated by RNAi, the percentage of nitrogen remaining in leaves and peduncle at maturity are increased (Waters et al. 2009), which indicates that *NAM-B1* plays a role in remobilisation and translocation of nitrogen, rather than uptake.

4.4.9 Hypothesis 7: Differential gene expression analysis will identify genes downstream of *NAM-B1* involved in nutrient remobilisation and senescence

We expected that genes regulated by *NAM-B1* would be differentially expressed in WT compared to *NAM* RNAi plants. Therefore we re-analysed an RNA-seq experiment carried out on flag leaves of WT and *NAM* RNAi plants at 12 DAA. We identified 384 differentially expressed genes (FDR < 0.1). As had previously been reported using this data set (Cantu et al. 2011) we found that many more genes were down-regulated in WT compared to RNAi than were up-regulated. This suggests that senescence requires the down-regulation of many processes, although some genes were up-regulated during senescence indicating that senescence is an active process.

Amongst the up-regulated genes surprisingly we did not find any GO term functional enrichment. This may be because only 87 up-regulated genes had GO term functions annotated which may have biased the analysis.

Amongst the down-regulated genes many GO terms were over-represented. In particular many biosynthetic processes were down-regulated including processes related to photosynthesis and monosaccharide metabolism, which correlates to the reduced ability to fix carbon during senescence. Processes related to nitrogen compound biosynthesis were also down-regulated which we hypothesise is related to the net export of nitrogen from leaves during senescence.

The functions of strongly up- or down-regulated genes have been inferred from a BLAST to rice. However our interpretation must be cautious because these functions might be incorrect and further analysis will be required. Additionally there are many

genes which are up- and down-regulated in WT compared to NAM RNAi plants and understanding the interactions and functional significance of each gene is a large task. One way to further investigate this would be to use gene regulatory network analysis. Computational modelling of networks can highlight the key genes within a network, and this would be an interesting future direction to which both the ChIP-seq and RNA-seq data could contribute. Nevertheless several hypotheses about the downstream targets of *NAM-B1* can be formulated from the gene function information obtained by differential gene expression analysis.

A very strongly down-regulated gene is a no apical meristem protein (NAM) which is in the same family of NAC transcription factors as NAM-B1. This gene (mrna121848) is located in chromosome 6AS which is the same chromosome arm as NAM-A1, a homoeologue of *NAM-B1*. However this down-regulated *NAM* gene is not *NAM-A1* because the opposite trend in expression for *NAM-A1* has been confirmed by qPCR (Cantu et al. 2011). Instead the strong down-regulation of this *NAM* gene during senescence, which was previously observed (Cantu et al. 2011), shows that different NAC transcription factors may play diverse roles during the senescence process.

Other genes which were strongly down-regulated included elongation factors which are involved in protein synthesis in the chloroplast. These have previously been reported to be down-regulated during drought-stress-induced senescence in Arabidopsis (Merewitz et al. 2011) and the lack of protein synthesis may contribute to chloroplast degradation during senescence.

No GO terms were significantly over- or under-represented amongst the genes up-regulated in WT compared to RNAi therefore it is hard to make generalisations about up-regulated genes. The two most highly up-regulated genes in WT plants were related to mitosis: AIR9 which is proposed to help ensure the correct localisation of the cortical division site (Buschmann et al. 2006) and DIM1 which was originally identified in yeast to be essential for cell cycle progression and is also important for RNA splicing (Simeoni and Divita 2007). Some of the up-regulated genes are known to be associated with senescence such as glutathione peroxidase, which may scavenge free radicals produced during chloroplast breakdown (Strother 1988). It was surprising that one of the most up-regulated genes was RuBisCO, because we expected carbon fixation to be reduced in WT plants during senescence. At the time-point sampled (12 DAA) visual signs of senescence have not appeared and photosynthesis is unaffected (**Figure 3.4**).

Moreover transcript level does not always correlate well with protein accumulation, and it has been shown that RuBisCO activity in maize seedlings is only transcriptionally regulated early in development and afterwards regulation acts via post-translational mechanisms (Loza-Tavera et al. 1990). Therefore the activity of RuBisCO may not differ between the two genotypes but this would need to be tested experimentally.

4.4.10 Hypothesis 8: Differentially expressed genes will contain shared DNA motifs in their promoters which facilitate co-expression

We hypothesised that differentially expressed genes may share common promoter elements which would permit their co-regulation. Several sequences were identified to be present in a high proportion of peaks including a polyA motif and a CxxC motif which were motifs identified by ChIP-seq. However these motifs were also identified in contigs containing non-differentially expressed genes at a similar frequency, which suggests that these motifs are common to many wheat genomic regions. It would be valuable to repeat this analysis using only promoter regions rather than whole contigs. Additionally not all of the differentially expressed genes contained the identified motifs. This may be because the differentially expressed genes may be several steps downstream of NAM-B1 in the gene regulatory network, and therefore not directly bound by NAM-B1. In the future it would be interesting to establish if these motifs are enriched close to the transcription start site as has been reported for other motifs (Ma et al. 2012).

4.4.11 Hypothesis 9: Direct targets of NAM-B1 will have their promoter sequences bound by NAM-B1 and will be differentially expressed in *NAM* RNAi plants

Overall there was no significant overlap between genes identified by ChIP-seq as direct targets of NAM-B1 and genes identified to be downstream of *NAM-B1* by differential expression analysis. The only genes identified as direct targets of NAM-B1 and differentially expressed were in the intersection between the RNA-seq data and the ChIP-seq sample L1B proper pairs MAPQ > 30. These contained one gene of unknown function, an ankyrin repeat domain protein with unknown biological function, a vacuolar ATP synthase subunit C gene which is involved in ion transport across membranes and a lectin protein kinase which may be involved in cell wall adhesion to the cell membrane (Gouget et al. 2006). We had initially expected that the strongest candidate genes would be identified by both differential gene expression analysis and ChIP-seq. However, the identities of these four genes, and the lack of significant

overlap made us reconsider this expectation, and consider genes identified by only one technique.

Despite very few genes being identified by both RNA-seq and ChIP-seq, there were many similarities between the classes of genes identified. This indicates that the same biological pathways controlled by NAM-B1 can be discovered by both differential expression analysis and ChIP-seq. For example many photosynthesis and primary metabolism genes were identified to be down-regulated in WT compared to RNAi plants, and the majority of genes identified as targets of NAM-B1 in the peduncle ChIP-seq samples were chloroplastic proteins such as photosystem components. However, the precise genes detected by differential gene expression are not those directly bound by NAM-B1 which could be due to time-point differences between the experiments (RNA-seq at 12 DAA and ChIP-seq at 20 DAA), differences in growing conditions altering the developmental processes or because differentially expressed genes are not necessarily direct NAM-B1 targets, but could instead be several steps downstream of NAM-B1.

Another reason why the same genes are not identified by ChIP-seq and RNA-seq could be that many direct binding interactions have little effect on transcription. In human cell lines it has been reported that the majority of the binding interactions between transcription factors and DNA do not cause measureable changes in gene expression (Cusanovich et al. 2014). Knockdown of gene expression by transfection with small interfering RNAs (siRNA) reduced the expression of target transcription factors by 50-90 % and differences in gene expression were detected by microarray. TFBS were detected by DNAse-seq (in which DNA bound by a transcription factor of interest is protected from DNAse treatment, and then sequenced to indicate binding sites) and ChIP-seq. Out of the transcription factors for which both binding and gene expression data was available only 12 out of 29 had a significant overlap between the differentially expressed genes and bound genes (identified by a binding site within 10 kb of the TSS). Only 32.3 % (median value) of differentially expressed genes were bound by the transcription factor. Surprisingly 88.9 % (median value) of binding sites were not associated with differentially expressed genes. This study indicates that in eukaryotes the binding interactions between transcription factors and DNA are complex, and that finding no significant overlap between differentially expressed genes in *NAM* RNAi lines and the ChIP-seq experiments is similar to findings in human cells. This emphasises that our understanding of gene regulation is still in its infancy and much work remains to be done.

4.4.12 Hypothesis 10: NAM-B1 directly binds the DNA regions (peaks) identified by ChIP-seq

To validate the NAM-B1 binding sites identified by ChIP-seq we decided to use electrophoretic mobility shift assays (EMSA). In this assay the direct binding of a purified protein to a specific DNA sequence is identified as a band shift on a gel. We first optimised the expression of the full length NAM-B1 protein in the heterologous host *E. coli* and purified the recombinant protein using an N-terminal his-tag.

We decided to focus on peak regions identified in the sample L1B proper pairs MAPQ > 30 because this sample contained non-repetitive peak regions and a well conserved motif in many peaks. We synthesised 25 bp DNA probes containing the CxxC motif identified in over 80 % of peak regions. Although binding of NAM-B1 to these probes was detected, the binding was non-specific and the recombinant protein bound to all probe sequences tested, including random DNA sequences. This non-specific binding could be caused by incorrect protein folding, which results in a non-functional protein incapable of specific DNA binding. Other NAC transcription factors have been successfully used in EMSA assays to show specific DNA binding, however in many cases only the N-terminal NAC domain of the protein which contains the DNA binding domain has been used, e.g. (Ernst et al. 2004, Zhong Ruiqin et al. 2007). This might be important because the C-terminal domain has a high intrinsic disorder which could cause aberrant binding to DNA or problems with protein folding. Additionally in other EMSA studies using NAC transcription factors a maltose-binding protein (MBP) tag was used instead of a his-tag. The MBP tag can aid protein solubility and may be important in stabilising the recombinant protein. Therefore re-cloning the NAC domain of NAM-B1 with an N-terminal MBP tag will be carried out. This new recombinant protein will be tested for use in EMSA to determine if the non-specific binding is eliminated when the C-terminal domain is removed. It will also be important to test binding of NAM-B1 to entire peak regions identified by ChIP-seq (which are 200-300 bp in length) and not only focus on very short motif sequences which may cause us to miss some binding interactions. Additionally we would like to test peak regions identified in the five peduncle samples to see whether both the sites identified in the leaf sample L1B and in the peduncle are real binding sites. This would indicate different functions for NAM-B1 in these two tissues.

4.4.13 Summary

The exploration of the targets of NAM-B1 using the next-generation sequencing based techniques of ChIP-seq and RNA-seq has been challenging but has revealed many biologically important putative targets of this master regulator of nutrient remobilisation and senescence.

We have developed a genetically ordered pseudomolecule-reference genome for wheat which has enabled our homoeologue specific analysis of NAM-B1 target genes. The bioinformatic analysis of the ChIP-seq data has enabled the testing of multiple hypotheses about the types of DNA regions and target genes which NAM-B1 regulates. We have found that target genes differ between the leaf and peduncle tissues. We have found that in the peduncle NAM-B1 binds to highly repetitive DNA sequences which are located near to chloroplast related genes. These binding sites are replicated in five out of six independent biological replicates showing that they are highly reproducible. Furthermore these binding sites are mainly located in the promoter regions and 5' UTRs of the target genes, indicating a regulatory role and in similar locations to other transcription factor binding sites. The target genes of NAM-B1 in the peduncle are associated with roles in the chloroplast, which could explain how NAM-B1 regulates nutrient remobilisation and senescence. Chloroplast degradation is one of the earliest steps of senescence and degradation of chloroplasts releases nitrogen for grain filling, which NAM-B1 is also known to affect. The genes identified to be direct targets of NAM-B1 in the peduncle were not identified to be differentially expressed in flag leaves, this may be due to differences in the tissue type, the time-point of the experiment or because differentially expressed genes detected were several steps downstream of NAM-B1. However genes with shared functions were identified in both the peduncle ChIP-seq samples and by differential gene expression including many photosynthesis and primary carbon metabolism related genes.

The leaf sample L1B was very different from the other ChIP-seq samples. Many more binding sites were identified in this sample than in any other leaf samples or than in the peduncle samples. The regions bound by NAM-B1 in this sample were unique within the genome, which is in contrast to the repetitive regions bound by NAM-B1 in the peduncle. This uniqueness may be because these binding sites were often located within the exons of genes, which is surprising because transcription factors usually bind to promoter regions. The majority of binding within exons was towards the 5' end of the gene so perhaps these binding interactions are still relevant to gene regulation. A

strongly enriched motif was identified in NAM-B1 binding sites: CxxC. The consensus NAC transcription factor binding site was not identified in these binding sites which may indicate that like the Arabidopsis NAC transcription factor *VOZ2*, *NAM-B1* has an altered DNA binding specificity (Lindemose et al. 2014). Different motifs were identified to be bound by NAM-B1 in peduncle and leaf samples, this may be because the binding specificity of NAM-B1 depends on its dimerization partner or other protein-interactions which are different between tissues. The target genes of NAM-B1 in the leaf have a diverse range of putative functions including secondary metabolism which could be related to nutrient remobilisation and senescence, although this is speculative and further experiments are required.

In conclusion, we have identified putative target genes of NAM-B1 using a combination of ChIP-seq and RNA-seq. However, further validation of the binding of NAM-B1 to the promoter regions will need to be carried out to determine whether the differences in target genes observed between leaf and peduncle samples are biologically relevant. This will help us to identify candidate genes to functionally characterise using the newly available sequenced TILLING population. The identification of many putative downstream targets genes of NAM-B1 deepens our understanding of how NAM-B1 can regulate the remobilisation of nutrients from senescing leaves into the developing grain and opens the door into understanding the control of these agronomically important traits at the molecular level.

5 Chapter 5 – General discussion

The aim of this thesis was to understand the mechanisms by which the transcription factor *NAM-B1* influences grain nutrient content, nutrient remobilisation and monocarpic senescence in hexaploid wheat. To achieve this aim we have used a range of methods to characterise *NAM-B1*. We have examined expression patterns of *NAM-B1*, the effects of *NAM-B1* on nutrient transport, photosynthetic capacity and grain filling, and finally identified putative downstream targets of *NAM-B1*.

In this chapter I will summarise our findings, discuss how these shape our understanding of the function of *NAM-B1* and suggest future directions for this work.

5.1 NAM genes are expressed in vascular tissues but their precise role in transport is not known

5.1.1 *NAM-B1* is expressed in stems, leaves and ears after anthesis

Previously *NAM-B1* expression had been reported in flag leaves after anthesis where it plays a role in nutrient remobilisation (Uauy et al. 2006b). We found that *NAM-B1* was expressed more broadly after anthesis in not only flag leaves but also internode, peduncle, ear tissues and in the developing grain. In general expression of *NAM-B1* increased to a maximum at 20 days after anthesis (DAA), however in the grain the maximum level of expression is likely delayed compared to the other tissues. These results suggest that *NAM-B1* may also regulate the transport of nutrients from the flag leaf to the ear and control nutrient loading into the grain itself. We found that the homologues of *NAM-B1* showed similar expression patterns across tissues and time-points to *NAM-B1*, which helps explain how these homologues can play similar roles in senescence and nutrient remobilisation (Avni et al. 2014).

5.1.2 *NAM* genes are expressed in vascular bundles and lignified tissues

Using a probe which detected expression of all *NAM-B1* homologues (*NAM* genes) we found that in flag leaves *NAM* genes were expressed strongly at 20 DAA in vascular bundles (mestome sheath and xylem), indicating a role in nutrient remobilisation and transport of nutrients from the leaves. In stem tissues, expression of *NAM* genes was also strong at 20 DAA in the vascular bundles, but expression was also present in epidermal and parenchyma cells, suggesting that *NAM* genes may have a broader role in

these tissues. In both stem and leaf tissues *NAM* genes were expressed in cells with secondary cell walls such as xylem cells and sclerenchyma cells.

5.1.3 Xylem and phloem transport are unaffected in *NAM* RNAi plants

Grain nutrient concentrations are known to be altered in *NAM* RNAi plants (Uauy et al. 2006b, Waters et al. 2009) which we hypothesized to be due to differences in xylem and phloem transport. Our RNA *in situ* hybridisations supported this hypothesis because *NAM* genes were expressed strongly around the vascular bundles in the mestome sheath indicating a role in nutrient transport. Strong *NAM* expression in the xylem suggested that down-regulation of *NAM* genes in RNAi lines might reduce xylem transport to the grain. The lack of *NAM* gene expression in the phloem suggested that phloem transported nutrients should be unaffected. These differences in phloem and xylem specific expression of *NAM* genes were seen not only in the flag leaf, at the site of nutrient remobilisation, but also in peduncle tissues which transport the nutrients from the leaf to the grain, and are a site of nutrient remobilisation in their own right.

In support of our hypothesis that *NAM* genes are necessary for xylem transport, we found that some xylem transported nutrients (zinc, iron and copper) were reduced in RNAi plants' ears. However other xylem transported elements (magnesium, manganese and phosphorous) were not altered, but this may be due to differences only accumulating at later stages of grain filling. The late effect of *NAM* genes on grain nutrient composition is supported by our expression data which shows that *NAM* genes are likely to be expressed later in the grain than in other tissues.

To directly test xylem and phloem transport in RNAi plants we fed phloem-mobile rubidium (Rb) and xylem-mobile strontium (Sr) to cut tillers. We did not see any difference in the percentage of Rb or Sr that was transported to the ears in RNAi plants compared to controls. This may be because *NAM* genes do not affect xylem and phloem transport or due to limitations to our experiment. The limitations included only incubating the tillers for 72 h with the Rb and Sr solution, and treating the ears as a homogeneous tissue rather than dividing them into their components (e.g. rachis, glumes and grain). Repeating this experiment using a hydroponic system and dividing the ears into component tissues would allow a longer and more accurate feeding experiment to be carried out.

5.2 NAM genes affect photosynthesis, grain filling and water soluble carbohydrate accumulation

5.2.1 Extended photosynthesis did not increase grain mass in *NAM* RNAi plants

A paradigm in wheat breeding has been that extending green canopy duration (i.e. delaying senescence), will extend the grain filling period and increase grain yield. Therefore we tested whether the delay in senescence in RNAi plants would increase the grain mass due to an extended period of photosynthesis. We confirmed that photosynthesis is maintained in RNAi plants for approximately 10 days longer than in control plants, however this increase in carbon fixation (approximate 2 g glucose equivalent per plant) does not translate to a grain mass increase in RNAi plants. The finding that *NAM* gene regulation of delayed senescence does not increase yield has been previously reported in both *NAM* RNAi plants (Waters et al. 2009) and in plants with mutations in *NAM-A1*, *NAM-B1* and *NAM-D1* (Avni et al. 2014, Uauy et al. 2006a). The effect of stay-green to increase yield frequently depends on the environmental conditions, with a strong association with drought stress (Borrell et al. 2000, Lopes and Reynolds 2012). However in the case of *NAM* genes neither drought nor heat stress gave stay-green *NAM* RNAi plants a yield advantage (Guttieri et al. 2013).

The ability of delayed senescence to enhance yield depends not only on the source of photosynthate but also on the capacity of the sink to utilise it. We hypothesised that the grain of the *NAM* RNAi plants may not have been able to use the extra photosynthate and tested this by examining the rate of development and the activity of key starch synthesis enzymes.

5.2.2 Grain maturation in *NAM* RNAi plants is decoupled from leaf senescence

To understand more about the source-sink relationship we studied whether grain development was coupled to flag leaf senescence. We examined grain moisture content and the activity of key enzymes important for starch synthesis in the grain of RNAi and control plants. We found that although flag leaf senescence was delayed by approximately 10 days in RNAi plants compared to controls, the grain reached physiological maturity only 3 days later than control plants. This suggests that the effects of *NAM-B1* on development are not equal in all tissues. The expression of *NAM-*

B1 is likely to peak later in the grain than in the flag leaves; this might be why the effects of *NAM-B1* are weaker in the grain. We found that during grain filling the activity of the enzymes AGPase and starch synthase were similar between control and RNAi plants, except at 20 DAA when the RNAi grain had lower enzyme activity. 20 DAA is usually around the time of maximum grain filling and this implies that starch synthesis rate might be lower in RNAi grain. This reduction in enzyme activity might have prevented the RNAi grain from being able to accelerate starch synthesis to use the extra photosynthate produced during the delayed senescence which occurred from 20 to 30 DAA.

5.2.3 Carbon fixed during the stay-green period in *NAM* RNAi plants was stored as water soluble carbohydrates in stem tissues

The extra photosynthate produced during the stay-green period of the RNAi plants does not translate into a yield gain. To understand the fate of this photosynthate we examined the water soluble carbohydrate (WSC) content of stem tissues which are known to act as a temporary store of photosynthate during grain filling. We found that WSC, particularly in the form of fructan, were accumulated by RNAi plants from anthesis until maturity, whereas control plants remobilised WSC from anthesis to maturity. In RNAi plants the fructans were mainly accumulated in internode 1. *NAM-B1* expression levels continued to increase in internode 1 from 20 to 30 DAA, whereas in all other tissues they stabilised or began to fall. These high expression levels later after anthesis may be linked to the accumulation of WSC to a higher level in this tissue, compared to other stem tissues. Furthermore in stem tissues, including internode 1, expression of *NAM* genes was present in the parenchyma cells which are the sites of WSC storage (Leegood 2008, Scofield et al. 2009), whereas in leaf tissues expression was restricted to vascular bundles and sclerenchyma. Overall the extra WSC in RNAi plants accounted for 28 % of the extra photosynthate fixed per flag leaf during the stay-green period. The accumulation of WSC indicates that the grain were not able to use all the photosynthate available from the leaves, supporting our hypothesis that the capacity of the grain limited the yield in the RNAi plants. From our experiments it was not possible to establish the fate of the rest of the extra photosynthate, and ^{14}C tracer studies might prove useful to determine whether extra photosynthate is stored elsewhere in the plant or whether increased rates of respiration in RNAi plants might be accountable.

5.3 Identification of direct target genes of NAM-B1

5.3.1 Putative NAM-B1 binding sites differ between leaf and peduncle tissues

To understand the pathways through which *NAM-B1* regulates grain nutrient content, nutrient remobilisation and senescence we used ChIP-seq to identify direct target genes of NAM-B1. We identified putative target genes in both peduncle and flag leaf tissues. We found that the putative NAM-B1 binding sites were different between peduncle and leaf tissues; this suggests that NAM-B1 may have different target genes in these tissues and a different biological function. Not only were the putative binding sites in different locations they had different properties.

The binding sites identified in peduncle tissues were highly repetitive within the wheat genome but were reproducible with over 300 sites common to five out of six samples. These binding sites were mainly located in the 5'UTR and promoter regions of target genes; this indicates a regulatory role and is similar to binding locations reported for other transcription factors. We identified an A-rich motif which was present in all binding sites and other less frequent putative binding motifs. The putative target genes in the peduncle are related to chloroplast function. We have shown that knockdown of *NAM* genes prolongs photosynthetic ability during senescence, so we hypothesise that *NAM-B1* may target chloroplast genes to cause degradation. During chloroplast degradation, nitrogen and transition metals such as iron are released and can be remobilised to the grain; the lack of chloroplast degradation may explain why the grain of RNAi plants have reduced protein and iron content (Uauy et al. 2006b).

Amongst the flag leaf samples one sample in particular stood out (L1B). In this leaf sample there were over 1000 putative binding sites identified even using stringently mapped reads, whereas in most other samples at this stringency very few binding sites were identified (<100). The regions identified in this sample were unique within the genome, in contrast to the highly repetitive regions identified in the peduncle. The unique nature of these binding sites may be due to their localisation within exons, which was surprising because transcription factors usually bind to promoter regions. However the binding within exons was generally towards the 5' end of the gene and might have a regulatory role. A strongly enriched motif (CxxC) was identified in putative NAM-B1 binding sites in the leaf sample L1B. Electrophoretic mobility shift assays were used to test whether NAM-B1 directly interacts with DNA sequences containing the CxxC

motif, however the binding was not specific to this sequence and further work is required to validate whether NAM-B1 binds the regions identified in either the leaf sample or in the peduncle samples discussed above.

The putative target genes in the L1B leaf sample have a wide range of functions and are enriched for phenylpropanoid metabolism compared to the genes present in the pseudomolecule reference sequence. Phenylpropanoids include lignin which is located in secondary cell walls, and *NAM-B1* expression was frequently detected by RNA *in situ* hybridisation in cells containing secondary cell walls; perhaps *NAM-B1* plays a role in these specialised cells. Another type of phenylpropanoid are the flavonoids which may act as photo-protectants; we hypothesise that they may be important during the down-regulation of photosynthesis which is regulated by *NAM-B1*. Amongst the top 15 most significantly bound target genes there are many genes involved in metabolism, including nitrogen and carbon metabolism, and also transport proteins such as an amino acid transporter; these functions fit with the known role of *NAM-B1* in affecting grain nutrient content, nutrient transport and senescence.

The genes identified by ChIP-seq as putative direct targets of NAM-B1 constitute a step towards understanding the mechanisms by which *NAM-B1* regulates grain nutrient content, nutrient remobilisation and senescence. However, validation of the putative binding sites is still required and further work will be needed to understand why different targets are identified in peduncle and leaf tissues.

5.3.2 Putative NAM-B1 direct target genes are not differentially expressed in *NAM* RNAi plants

We compared the genes identified by ChIP-seq with genes which were differentially expressed in *NAM* RNAi plants. We hypothesised that direct target genes of NAM-B1 would also be differentially expressed. Surprisingly, very few common genes were identified and the overlap was not statistically significant for either flag leaf or peduncle tissues. This may be because the differentially expressed genes are downstream of NAM-B1 but are not direct targets or because the time-points used in the two experiments were different. In the light of work carried out in human cells the lack of overlap is not surprising: in a study of 29 human transcription factors only 41 % had overlap between directly bound target genes and differentially expressed genes (Cusanovich et al. 2014). However in our experiment the differential expression analysis did identify many genes with functions similar to those identified by ChIP-seq

and perhaps represent downstream steps of the processes initiated by direct binding of NAM-B1. For example in peduncle samples NAM-B1 has putative direct targets which are chloroplast-related and many photosynthesis related genes were differentially expressed in *NAM* RNAi plants.

5.4 Novel mechanisms for NAM-B1 action

As summarised above we have obtained new information about several roles of *NAM-B1* in relation to expression patterns, nutrient transport, source-sink relations and putative targets. The combination of these observations has led to several insights into the mechanisms by which *NAM-B1* acts to control grain nutrient content, nutrient remobilisation and senescence; these are described below.

5.4.1 Different roles in flag leaf and peduncle tissues

Based on several results including expression, WSC accumulation and putative target genes we hypothesise that *NAM-B1* has different roles in flag leaf and peduncle tissues. We have found that NAM-*B1* expression is higher in peduncle tissues than in flag leaves. We have shown in flag leaves *NAM* expression is mainly observed in vascular bundles and in sclerenchyma, however in the peduncles expression is more widespread and is observed in epidermal and parenchymal cells in addition to the vascular and sclerenchyma cells. This expanded domain of expression might reflect the importance of *NAM* genes for WSC accumulation and remobilisation which occurs in the parenchyma cells of stem tissues including the peduncle. In both the flag leaves and the peduncle *NAM-B1* is expressed in vascular tissues and we hypothesise that *NAM-B1* regulates nutrient remobilisation and export in these tissues. The hypothesis that *NAM-B1* is important in the peduncle tissues, which has not been emphasised previously, is supported by micronutrient accumulation to high levels in the peduncles of *NAM* RNAi plants (Waters et al. 2009). This accumulation in the peduncle suggests that an important transport point between the peduncle and grain may be regulated by *NAM* genes, or feedback from the grain prevents the import of micronutrients from the peduncle; this potential transport blockage requires further investigation.

The difference in roles of *NAM-B1* in peduncle and flag leaf tissues is also supported by the putative target genes identified in these two tissues. In the peduncle the target genes are mainly chloroplast related which may be related to the extension of photosynthesis. We showed that photosynthesis was extended in *NAM* RNAi plants' flag leaves,

however the peduncles also stay-green and we hypothesise that these green peduncles would also remain photosynthetically active. In the flag leaf the putative target genes of *NAM-B1* had a diverse range of functions including primary carbon and nitrogen metabolism, phenylpropanoid metabolism and nutrient transporter. Further experiments to validate putative target genes in both peduncle and flag leaf tissues are required before strong conclusions can be drawn from the ChIP-seq results. We hypothesise that the different expression patterns observed in the flag leaf and peduncle tissues may cause NAM-B1 to interact with different protein interaction partners and this may explain why NAM-B1 targets different genes in these two tissues and ultimately may have different biological functions.

5.4.2 Function related to photosynthesis

NAM-B1 was previously reported to affect senescence (Distelfeld et al. 2007, Uauy et al. 2006a) and in our work using ChIP-seq we have found a more direct connection of *NAM-B1* to photosynthesis whose down-regulation is a hallmark of senescence. We have measured photosynthesis and found a 10 day extension of photosynthesis in *NAM* RNAi plants. We hypothesise that the regulation of photosynthesis may be directly controlled by NAM-B1 because putative target genes of NAM-B1 in the peduncle are mainly chloroplast-related, for example components of photosystems. In the flag leaf putative direct target genes include flavonoid biosynthesis genes which may be important for photo-protection during the down-regulation of photosynthesis. Differential expression analysis also supports the idea that *NAM-B1* regulates photosynthesis with many photosynthesis and monosaccharide biosynthesis related genes differentially expressed in *NAM* RNAi plants compared to control plants.

The extension of photosynthesis by preventing chloroplast degradation may also explain how *NAM-B1* can influence grain nutrient content. Chloroplasts contain a large proportion of total leaf nitrogen and iron, therefore delaying senescence causes these nutrients to be retained within the photosynthetic tissues, as has been seen in the flag leaves of *NAM* RNAi plants (Waters et al. 2009). The reduced remobilisation and transport to the grain may explain why grain nutrient contents are decreased in plants with delayed senescence, as we have observed in *NAM* RNAi plants. This trend is seen across the *NAM* gene mutants with grain protein content being negatively correlated with delayed senescence (**Figure 5.1**).

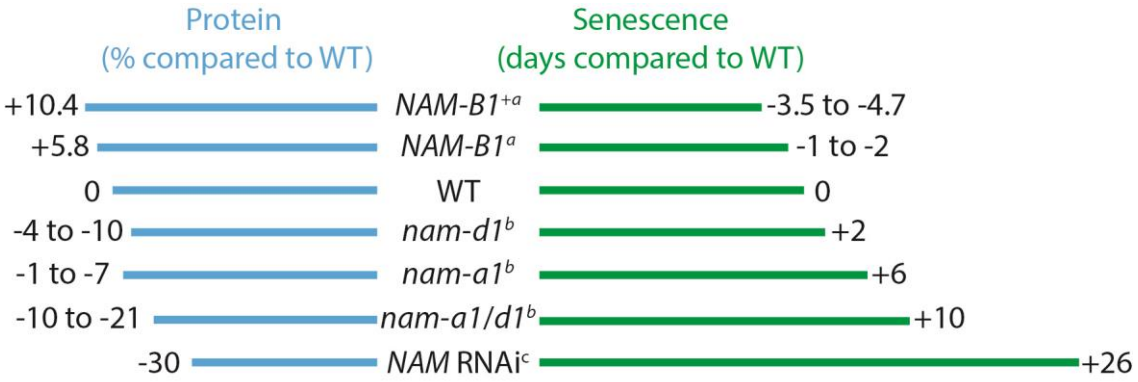


Figure 5.1. Comparison of the effects of *NAM-B1* homologues on protein and senescence; in all cases no significant change to yield was detected compared to wild type (WT). WT genotype was *NAM-A1*/*nam-b1*/*NAM-D1*/*NAM-A2*/*NAM-B2*/*NAM-D2*. All data are from field-grown plants except *NAM* RNAi plants which were grown in the glasshouse. All mutations are in a hexaploid background except $^{+}$. Data are from ^{a} (Brevis and Dubcovsky 2010), ^{b}(Avni et al. 2014) and ^{c}(Uauy et al. 2006b, Waters et al. 2009).

5.4.3 Function related to secondary cell walls

A function which has been highlighted by this work, but is less obvious in how it could contribute to nutrient remobilisation and senescence, is the role of *NAM-B1* in secondary cell walls. In both flag leaf and stem tissues the expression of *NAM* genes was high in cell types containing secondary cell walls such as sclerenchyma and xylem. The putative target genes of *NAM-B1* in the leaf, identified by ChIP-seq, were enriched for phenylpropanoid metabolism which is the pathway which produces lignin which gives strength to secondary cell walls. Several NAC transcription factors which regulate secondary cell wall development have been identified in model species such as Arabidopsis (Yamaguchi et al. 2008, Zhong R. Q. et al. 2010), Brachypodium (Valdivia et al. 2013) and rice (Yoshida et al. 2013) and although these are not the closest homologues of *NAM-B1* perhaps the function is conserved.

5.4.4 Water soluble carbohydrates (WSC) in parenchyma cells

We discovered that down-regulation of *NAM* genes by RNAi prolonged flag leaf photosynthesis, however the extra photosynthate was not used for grain filling, perhaps because the enzyme activity within the RNAi grain was too low and the grain development was not delayed enough to take advantage of this extra photosynthate coming from the late senescing leaves. We found that a significant proportion of this extra photosynthate (28 %) was accumulated in the stem tissues of RNAi plants as WSC. This may indicate a role of *NAM-B1* in regulating WSC accumulation. The

expression of *NAM* genes in the parenchyma tissues in stems supports this hypothesis, as does the continued increase in *NAM* gene expression from 20 to 30 DAA in the internode 1 tissue where the levels of WSC accumulation are highest. Down-regulation of *NAM* genes alters carbon metabolism in leaves and we hypothesise that *NAM-B1* also influences carbon metabolism, particularly fructan biosynthesis in the stems.

5.4.5 Roles in the grain

Previous studies have focussed on understanding the effects of *NAM-B1* on nutrient remobilisation from vegetative tissues (Guttieri et al. 2013, Waters et al. 2009) and on the final yield effects (Brevis and Dubcovsky 2010, Brevis et al. 2010, Tabbita et al. 2013), however studies on the effects of *NAM-B1* in the grain themselves have been lacking. To address this we examined *NAM* gene expression in the grain and we found that *NAM* genes (*NAM-A1*, *NAM-B1* and *NAM-B2*) were expressed at similar levels in the grain to the flag leaves, although the expression in the grain reached a maximum later than in the flag leaves. This result suggested that the effects of *NAM* genes on the grain (e.g. micronutrient content) might be directly regulated within this tissue. We also examined whether two key enzymes of starch synthesis were altered in *NAM* RNAi plants. We found that both AGPase and starch synthase were similar in *NAM* RNAi and control plants at early and late stages of grain filling, however at 20 DAA, when maximum grain filling usually occurs, the activities were significantly lower in *NAM* RNAi plants. We hypothesise that lower enzyme activity in RNAi grain may have prevented the acceleration of starch synthesis to take advantage of the extra photosynthate from the stay-green period which occurs at this time-point.

Viewing this result from the opposite angle we hypothesise that *NAM-B1* might permit starch synthesis rate to increase, which might explain how, despite accelerated senescence, grain yield is maintained. This may also explain how *NAM-B1* is able to escape the negative correlation between grain protein content (GPC) and yield (Brevis and Dubcovsky 2010, Kumar et al. 2011, Tabbita et al. 2013). To make this tentative conclusion more concrete it would be necessary to examine enzyme activity, grain filling rate and gene expression in plants grown at the same time in the same conditions: this would allow true integration of datasets and the relative timings could be examined.

5.4.6 Roles in transport

No difference was found in xylem and phloem transport in *NAM* RNAi compared to control plants, which suggests that differences in nutrient remobilisation (Waters et al. 2009) may be due to individual transporters. For example one of the top putative targets of NAM-B1 in leaf tissue (sample L1B) is an amino acid transporter. The roles of individual transporters will need to be evaluated on a case by case basis to further understand nutrient transport. The finding that *NAM* genes are expressed not only in flag leaves, but also in stem tissues and the grain suggests that transport could be regulated by *NAM* genes at several different steps. Differential gene expression analysis of *NAM* RNAi and control plants found enrichment for genes related to transport, and again these require further evaluation on an individual basis.

5.4.7 Implications for use of *NAM-B1* in agriculture

Several aspects of this work have implications for the application of *NAM-B1* in wheat breeding and agriculture. Firstly our work on effects of *NAM-B1* on carbohydrate metabolism has suggested several possible explanations for how *NAM-B1* can increase GPC but also maintain yield. We have found that reduction of *NAM* gene expression is associated with accumulation of WSC in the stems at maturity, and reduced activity starch synthesis enzymes. In the opposite case where *NAM-B1* is introduced into germplasm lacking a functional copy, such as many elite varieties, *NAM-B1* accelerates senescence leading to an increase in GPC. We hypothesise that alongside this previously identified function, *NAM-B1* promotes the remobilisation of WSC from the stems and permits up-regulation of the rate of starch synthesis to maintain yield in the face of accelerated senescence. This hypothesis suggests that it may be possible to introduce *NAM-B1* into a wide range of germplasm to enhance GPC, without negatively affecting yield, as has been shown in several cases (Brevis and Dubcovsky 2010, Kumar et al. 2011, Tabbita et al. 2013).

Secondly, extension of this work may allow the separation of the traits affected by *NAM-B1* which include grain protein content, micronutrient content and monocarpic senescence. From the perspective of human nutrition it would be valuable to increase particular micronutrients, for example iron and zinc. Our work has indicated that *NAM-B1* alters micronutrient content not by large scale alterations to xylem and phloem transport, but instead we hypothesise its effects are caused by changes to remobilisation and transport of individual micronutrients. Independent regulation of specific

micronutrients is supported by the non-uniform change in micronutrient content in the grain of *NAM* RNAi plants. Alongside our physiological studies, our work has identified many putative downstream targets of *NAM-B1*, several of which have transport related functions. These genes may be a useful target for marker assisted selection to enhance the content of a particular micronutrient. An alternative strategy would be to use genetic modification or genome editing (Voytas and Gao 2014) which may be more acceptable to the public in the coming years. Many of the putative *NAM-B1* target genes could be of interest in agriculture, for example genes involved in nitrogen remobilisation and photosynthesis. Our results also suggest that *NAM-B1* plays different roles in different tissues; this knowledge could be useful for genetic modification to enable specific targeting of *NAM-B1* functions within tissues. For example, if our hypothesis that *NAM-B1* permits an enhanced rate of starch synthesis in the grain is correct, the expression of *NAM-B1* in the grain may enhance yields, without affecting nutrient remobilisation from vegetative tissues.

5.5 Summary and future directions

5.5.1 Summary

In summary we have deepened our understanding of the mechanisms by which *NAM-B1* acts at both the molecular and physiological level. We have found that *NAM-B1* has a broader range of roles than previously suspected and may act in a tissue specific manner. For example *NAM-B1* expression in the grain may regulate enzyme activity, expression in the internode tissues may influence water soluble carbohydrate accumulation and putative direct target genes of NAM-B1 are different between flag leaves and peduncles. We have shown that *NAM* genes play an important role in regulating photosynthesis during senescence, although the extended photosynthesis does not translate to a grain mass increase. In all of these cases it is difficult to distinguish direct effects *NAM-B1* within the tissue itself and effects caused by the influence of *NAM-B1* in other tissues. To understand the whole system of nutrient remobilisation and senescence is challenging but the identification of putative direct target genes of *NAM-B1* opens the door to understanding the regulatory pathway in more detail.

5.5.2 Future directions

An essential next step is the validation of the binding of NAM-B1 to the putative target genes. This will reveal whether target genes are different in peduncle and flag leaf tissues and whether all regions identified are binding sites. This validation will be carried out by electrophoretic mobility shift assay and ChIP-qPCR.

Following validation of putative target genes of NAM-B1 we will characterise these genes and their biological functions. This objective will be greatly aided by the exome-sequencing of a tetraploid EMS mutagenized TILLING population. Once the data are available (early 2015) it will be possible to identify a mutation in a gene of interest *in silico* and order the seed online. This will bring an unprecedented ease of access to mutants in wheat and facilitate a step change in the understanding of gene function. Once a knock-out mutation is identified in both the A and B genome copies, a single cross will create double mutants and one generation of self-pollination will produce homozygous double mutants suitable for phenotypic characterisation. Depending on the type of gene, we will examine the effect of the mutation on expression patterns, senescence, plant morphology, yield and grain nutrient content. If the gene is predicted to have a specific biochemical function (e.g. an iron transporter) we will use

biochemical assays or heterologous expression to test the function (e.g. yeast complementation).

Another area which would benefit from further study is the role of *NAM-B1* in the grain. We found that *NAM-B1* is detected in the grain by qPCR, and it would be of great interest to extend this further to carry out RNA *in situ* hybridisations to establish which cell types *NAM* genes are expressed in. This might indicate how *NAM* genes affect the grain, for example expression in the endosperm would suggest direct regulation of starch synthesis genes, or expression in the modified aleurone would suggest a role to regulate transport into the grain. In parallel it would be useful to examine the activity of starch synthesis enzymes and the moisture content of grain from the same batch of plants. This would enable the temporal relationships between expression of *NAM* genes in the grain and putative downstream effects to be established. It would also be interesting to examine the distribution of micronutrients and proteins within the grain which could be achieved using staining techniques such as diphenyl thiocarbazone for zinc (Ozturk et al. 2006), Perls' Prussian blue solution for iron (Choi et al. 2007) and modified Bradford reagent for protein (Ozturk et al. 2009) or energy-dispersive X-ray fluorescence spectrometry for iron and zinc (Paltridge et al. 2012).

Whilst *NAM-B1* is an important regulator of nutrient remobilisation and senescence it does not act in a vacuum. One way to further understand this protein would be to look for protein interaction partners. It is known that NAC transcription factors act as transcriptional regulators in the form of homodimers and heterodimers (Jeong et al. 2009), and transcription factors often interact with many proteins within cells to enable specific binding to regions of DNA. An unbiased approach to identify proteins which interact with NAM-B1 would be to carry out a yeast 2 hybrid screen. In this technique wheat cDNA fragments are expressed in individual yeast strains to produce a library of proteins; the interaction of NAM-B1 with these proteins can then be tested. This would reveal novel components of the pathway regulating nutrient remobilisation and senescence and their relationship to NAM-B1. The function of these proteins could then be studied using the TILLING population, as described above.

NAM-B1 is not alone in affecting nutrient remobilisation and senescence, its homoeologues *NAM-A1* and *NAM-D1* have a similar function (Avni et al. 2014) and the paralogous *NAM-A2*, *NAM-B2* and *NAM-D2* remain to be fully characterised. We found that many of these homologues share a similar pattern of expression which suggests

they may have similar functions. Several lines of investigation could elucidate their shared or distinct functions. Single, double and triple mutant lines could be grown at the same time in controlled or field conditions, and characterised for their effect on senescence, nutrient content and yield. This would give an indication of their biological functions and whether the effects seen were stronger in some homologues. These genes also present an opportunity to study how transcription factor function can diverge in polyploid species. For example carrying out ChIP-seq on all *NAM-B1* homoeologues would enable comparison of target genes between homoeologues.

The aim of studying *NAM-B1* is to understand the regulation of grain nutrient content, nutrient remobilisation and senescence. To fully understand these developmental processes it will be necessary to look for components which are unrelated to *NAM-B1*. One approach to achieve this is screening for mutants with altered phenotypes, although this is challenging in polyploid wheat due to functional redundancy between homoeologues. However the advent of the sequenced TILLING population will be an excellent resource to use for forward screens, because if a phenotype can be identified, the mutations within the identified line will be known, and creation of an F_2 mapping population will enable validation of which of the mutations present is the causal one. A second approach is to use next-generation sequencing to identify the underlying genes controlling grain nutrient content, nutrient remobilisation and senescence. A time-course could be carried out examining how gene expression alters across different tissues (e.g. grain, peduncle and flag leaves) during the grain filling and monocarpic senescence period. Differences between tissues would indicate tissue-specific regulation. The data collected could be used for gene-regulatory network modelling which can help to identify the most important regulatory genes within a large set of differentially expressed genes. It is possible to take advantage of time-courses in different tissues to produce a more reliable gene network and allow deeper understanding of causal relationships between gene expression (Penfold et al. 2012). The application of these methods to wheat is an exciting prospect which has only become a possibility due to the recent expansion of sequence resources in wheat.

The explosion of availability of genomic and genetic resources is enabling the in-depth study of developmental processes in wheat at the molecular level for the first time. These advances will help us to understand important biological pathways, such as nutrient remobilisation and senescence, and will contribute to breeding improved wheat varieties fit to feed our world's growing population.

6 Appendix

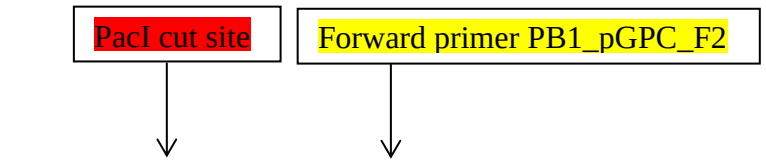
6.1 Sequences

6.1.1 Alignment of homologues of *NAM-B1* to the *NAM* probe (*NAM-B1*) used for RNA *in situ* hybridisation

```
NAM-A2          CGGGGATCAGCAGAGGAGCATGGAGTGCGAGGACTCCGTGGAGGACGCCGTCACCGCATA 60
NAM-D2          CGGGGATCAGCAGAGGAGCATGGAGTGCGAGGACTCCGTGGAGGACGCCGTCACCGCATA 60
NAM-B2          CGGGGATCAGCAGAGGAGCATGGAGTGCGAGGACTCCGTGGAGGACGCCGTCACCGCATA 60
NAM-A1          CGGAGATCAGCAGAGGAGCACGGAGTGCGAGGACTCCGTGGAGGACGCGGTCACCGCGTA 60
NAM-B1          CGGCGATCAGCAGAGGAACACGGAGTGCGAGGACTCCGTGGAGGACGCGGTCACCGCGTA 60
NAM-D1          CGGAGATCAGCAGAGGAGCATGGAGTGCGAGGACTCCGTGGAGGACGCGGTCACCGCGTA 60
                *** ************* ** ************************** ******** **

NAM-A2          CCCGCCCTATGCCACGGCGGGCATGACCGGTGCAGGGGCGCATGGCAGCAACTACGATTC 120
NAM-D2          CCCGCCCTATGCCACGGCGGGCATGACCGGTGCAGGGGCGCATGGCAGCAACTACGATTC 120
NAM-B2          CCCGCTTTATGCCACGGCGGGCATGACCGGTGCAGGGGCGCATGGCAGCAACTACGATTC 120
NAM-A1          CCCGCTCTATGCCACGGCGGGCATGGCCGGTGCAGGTGCGCATGGCAGCAACTACGCTTC 120
NAM-B1          CCCGCTCTATGCCACGGCGGGCATGACCGGTGCAGGTGCGCATGGCAGCAACTACGCTTC 120
NAM-D1          CCCGCTCTATGCCACGGCGGGCATGGCCGGTGCAGGTGCGCATGGCAGCAACTACGCTTC 120
                *****  ****************** ********** ****************** ***

NAM-A2          A------CTGCTCCATCACCTGGACAGCCACGAGGACAACTTCCTGGACGGCCTGCTCAC 174
NAM-D2          A------CTGCTCCATCACCAGGACAGCCACGAGGACAACTTCCTGGACGGCCTGCTCAC 174
NAM-B2          A------CTGCTCCATCACCAGGACAGCCACGAGGACAACTTCCTGGACGGCCTGCTCAC 174
NAM-A1          ACCTTCACTGCTCCATCATCAGGACAGCCAT---------TTCCTGGAGGGCCTGTTCAC 171
NAM-B1          ACCTTCACTGCTCCATCATCAGGACAGCCAT---------TTCCTGGACGGCCTGTTCAC 171
NAM-D1          ATCTTCACTGCTCCATCATCAGGACAGCCAT---------TTCCTGGACGGCCTGTTCAC 171
                *      *********** * *********          ******** ****** ****

NAM-A2          AGCAGAGGACGCCGGCCTCTCGGCGGGN----CCTCGCTGAGCCACCTGGCCGCGGCGGC 230
NAM-D2          AGCAGAGGACGCCGGCCTCTCGGCGGGCGCCACCTCGCTGAGCCACCTAGCCGCGGCGGC 234
NAM-B2          AGCAGAGGACGCCGGCCTCTCGGCGGGCCCCACCTCGCTGAGCCACCTAGCCGCGGCGGC 234
NAM-A1          AGCAGACGACGCCGGCCTCTCGGCGGGCGCCACCTCGCTGAGCCACCTGGCCGCGGCGGC 231
NAM-B1          AGCAGACGACGCCGGCCTCTCGGCGGGCGCCACCTCGCTGAGCCACCTAGCAGCGGCGGC 231
NAM-D1          AGCAGACGACGCCGGCCTCTCGGCGGGCGCCACCTCGCTGAGCCACCTGGCCGCGGCGGC 231
                ****** ******************     **************** ** ********

NAM-A2          GAGGGCGAGCCCGGCTCCGACCAAACAGTTTCTCGCCCCGTCGTCTTCAACCCCGTTCAA 290
NAM-D2          GAGGGCGAGCCCGGCTCCGACCAAACAGTTCCTGGCCCCGTCGTCTTCAACCCCGTTCAA 294
NAM-B2          GAGGGCGAGCCCGGCTCCGACCAAACAGTTTCTGGCCCCGTCGTCTTCAACCCCGTTCAA 294
NAM-A1          GAGGGCGAGCCCGGCTCCGACCAAACAGTTTCTCGCCCCGTCGTCTTCAACCCCGTTCAA 291
NAM-B1          GAGGGCAAGCCCGGCTCCGACCAAACAGTTTCTCGCCCCGTCGTCTTCGACCCCGTTCAA 291
NAM-D1          GAGGGCGAGCCCGGCTCCGACCAAACAGTTTCTCGCCCCGTCGTCTTCAACCCCGTTCAA 291
                ****** ********************** ** ************** ************

NAM-A2          CTGGCTCGATGCGTCAACCGTTGGCATCCTCCCGCAGGCAAGGAATTTTCCTGGGTTTAA 350
NAM-D2          CTGGCTCGATGCGTCAACCGTTGGCATCCTCCCACAGGCAAGGAATTTTCCTGGGTTTAA 354
NAM-B2          CTGGCTCGATGCGTCAACCGTTGGCATCCTCCCACAGGCAAGGAATTTCCCTGGGTTTAA 354
NAM-A1          CTGGCTCGATGCGTCACCCGCCGGCATCCTGCCACAGGCAAGGAATTTCCCTGGGTTTAA 351
NAM-B1          CTGGCTCGATGCGTCACCAGTCGGCATCCTCCCACAGGCAAGGAATTTTCCTGGGTTTAA 351
NAM-D1          CTGGCTCGAGGCGTCAGCCGCTGGCATCCTGCCACAGGCAAGGAATTTCCCTGGGTTTAA 351
                ********* ****** * *  ******** ** ************** ************

NAM-A2          CAGGAGCAGAAATGTCGGCAACATGTCGCTGTCGTCGACGGCCGACATGGC------GGT 404
NAM-D2          CAGGAGCAGAAATGTCGGCAACATGTCGCTGTCGTCGACGGCCGACATGGC------GGT 408
NAM-B2          CAGGAGCAGAAATGTCGGCAACATGTCGCTGTCGTCGACGGCCGACATGGC------GGT 408
NAM-A1          CAGGAGCAGAAACGTCGGCAATATGTCGCTGTCATCGACGGCCGACATGGCTGGCGCGGC 411
NAM-B1          CAGGAGCAGAAACGTCGGCAATATGTCGCTGTCATCGACGGCCGACATGGCTGGCGCAGT 411
NAM-D1          CAGGAGCAGAAACGTCGGCAATATGTCGCTGTCATCGACGGCCGACATGGCTGGCGCGGC 411
                ************ ******** *********** *****************       *

NAM-A2          GGACAACGGCGGGGGCAATGCGATAAACGCCATGCCTCCATTTATGAATCATCTACCCGT 464
NAM-D2          GGACAACGGCGGCGGCAATGCGATAAACGCCATGCCTCCATTTATGAATCATCTTCCCGT 468
NAM-B2          GGACAACGGCGGGGGCAATGCGATAAACACCATGCCTCCATTCATGAATCATCTTCCCAT 468
NAM-A1          CG------------GCAATGCGGTGAACGCCATGTC------------------------ 435
NAM-B1          GGACAACGGTGGAGGCAATGCGGTGAACGCCATGTCTAC--CTAT-------CTTCCCGT 462
NAM-D1          CG------------GCAATGCGGTGAACGCCATGTCTGCATTTATGAATCCTCTTCCCGT 459
                 *            ******** * *** ***** *

NAM-A2          GCAAGA 470
NAM-D2          GCAAGA 474
NAM-B2          GCAAGA 474
NAM-A1          ------
```

```
NAM-B1          GCAAGA 468
NAM-D1          GCAAGA 465
```

6.1.2 *NAM-B1* promoter used for transgenics in ChIP-seq



TTTTTTTTAATTAATACTTTTACACCCGGTCACGGTCAGTATCCGATCTCTCCCGACAAAAATGTCT
CCGTTTCTGGCCCAATTTTGGCCCGAGAAGAAATGTTTAGGGCCTTCCCTGTACCGCCCAAGCTTGT
CACCATAATCCAATCCGAGAGATTTTCTAACCAATCCAAAAGACCGCACGCTGTGTATGAAGAAACA
TTTTTGAAATTTGTTCTTGGTGTTGGCTAACTAACAGCTAGTTGGTCCCACAGAAAAAAAAAACTTG
GCAGCCAGTTGACTCAGGAACTTACTAGCAGCCAGACAAAGAGAATCATGATACTCCATTTAACTTG
CTCATCGCGGCTCGTACGCAAATACACATGCGCATGCACGTTACCTGCGCATGGTTTCTCAGGCAAC
GATCGGATCGGATCACGAGGTATTCCACCACCTCCTAAATTATCCACTGTATACATCTCTGATGTAT
GTTGTCCACTGTCCAGCTATAAATAGCACTGGCTAGATTATCCGTCGATCCGCACCACTGGAACAGT
AGCACCTTGCTATTAGCTGGCTTTGTGAATTGTGAATTAGCCTGCCCCATCTCCGGCCAACGACAAT
GGCTCCCTGCGGGGCCGGTGCCACCGCTGCCACCAAGGTGACCGTCGCTGTCGCCGCGGTCTGCATC
CTGCTAGCCATGCACGTCGCGACGACGGTGGCGGCTGCCGGGGACTGCGGGCGCTGTATGCAGCAGC
TGCGCTGTCCGACCTCCTGCAGCTGCGTGGCCTGGTGCCGAGACCGTGGGCCTGAGTGCGAGTTGTG
CATGGAACTAGAAGGGCTGTGCCTCATCCATTGCTCCCAGCGCTGTCCCTGCATGGACATACGCCTG
CAGTCGTCTCGAGAACTGCATCTGCATGCATGAGCTCAGTGCAAGAAGCATGCATGGGATGAATGCT
AGACGTAGTTGTGTTGTTTCTTCGTGTGGATCTATGGATGTATCTATATTCTATAGGTATAGCCATG
TTTAGATGTGGAAGACAATTTTGAGGAGAGGCTTTTAGTCAGTGCAAGCTTGTGTTCCTCAGTGTAG
CCTCTGACCTTTCTTAGTACTACAACCGAATGAACTATTCTTTTTTAAACACAATTGTGCTTTTCAC
TTTTTCTGGAAGCATGGTCGTGTGTTTTCCCTTTTTTGGAGTTGTGCTTCTTAGGGCATCTCCACCC
GTGCCCCCAGACGCCCTCCAGGAGCGTTTTTTTAGCACCGGCGAGTAAAAAACGTCCCAGCCGCGAT
CCCAGAGTCTCGTTTCGCGCCGGATCTGATGGCCGGCGCACGCAATGCAAACCAGGAAAACGGGGGG
GGGGGGGGGGGGAGCTGGGGGCGCCGGCGCTTTGTTTTAGCCGTGAAAGGCCCACTGCCAGCCTCTC
TGTCTCTCTCTCTCTCTCTCCATCTTTTCCTCTCTAGGCCCACCTAATGCATGCAAAAAAAGCCAAT
CTCTGTTCTCTCTCTCCCGCATGCACGCCGCCCTGCTCCTCGCTCCTCTGCGCTCTCTCCTCTCCCC
CCTGCTCCTCGCTCCGGCGCCCGTCGCTGGCCCCGTGCCGCCGCTGCGCCGGCCCCGAGCATGCAAG
CACGGCCGGCCCCCAGCGCGAAGGCCACGAGCGGCCCTGGCCCGGAGCCATGCACCTGGTGCCATGC
ATGCGCCCCGGCCGGGGAGCAGCAGCTGGCCTTCCACGAGGCTGCCGCCGAGGGCTCCGCAACCTCG
AGCTCCTCGCCGTGTCGCTCTCCACCGCCGTAGCACTGCATGTGGAGGAAGAAGCCGGCGGGGGGCT
GGCCCTTGCTGGCCAGCTCCTCGTCCATCCACACGCGGCCGCCGTCCGCCGGCCCGGGCCCGCGCGC
CTTCACCATGAACTGCGCCGCGCCGAGCCCCCCGGCGAACCACCAGTCGTGCACGTCCCACATCACC
TTCACCACGGCGCCGTCGCCGAGGACGGCCGTGCGGTTGCCGCGGAAGTTCCACCCCACGCGCCGCG
CCTCCACTATGCTCTTCCCGTCCATGCTCACCGCCACCTCCCCGTCGCCGTGCTCGCCCCTCGACGC
CGCGCCCCGCCGGTCGCCGCGGCAGCGCACCGTCACCTCGTGCTCCGCCCCGCCGTCCCGGAACCTG
CACCGCGTGCTGTGCGCCACCGAGGCCGCCTCCACGGACGACGGCGCGCGGAGCTGCTCGCGCCGGC
TGACCAGGGCCGAGAAGGGGCGTGGCGACGCGGCGGGACAGCTCCTCCCGCCCACGATTCGCGCCGG
CCCCATGCGTCGCCGTCCGGCCTCCTCCGCTTCGCGGCGGCCCCCTCCGCTCCACGCCGGCCCCGCT
CGCCGCCTCGCGCCAGCCGGTTTCCCGCTCGCCGGCTGGTCGGCTGCCGCCTCGGGCACCCAGACCG
CGCTCGCTGGCTGGTCAGCGGCCCGGACTGCGTCGGGTCGGCTCCGCCTCGCCTCGAGCTCGCCATC

GGCCTTGCTCCGCCCCCCCGCCTCGCGCCTGCACCCCGCGCACGCCGGCTCCACAGCTCGGCCCCGC
TCCTAGCAGCGTCGCGGCTTCCGCCTTCCCTTGCGCTGTCCCAGTCGGCGGTGGGGACGGGCACGAC
GGAGACGGAGGGGAGCTCGAGGAAGAGGAGGAAGCAGAGGTACCCCGCCGCCGCGGAGCGCCAGCCG
GTTCGGCGGTTCGCGTGGTGGTGGTGGCGGGGATACGTGCGTGGGTGGCGAAACGAATGGATATTTT
CTCGATCGCAGGCGAAAAATTTCGCCGGCAGCCCCCCACGGGGCGATAAAAATGCCTCCTGGGGGCC
TCTACGGCTGGAGATGCCCTTAGGGTGTGTCTAGTAGAGTGCATGGAGGGTGCATGAGCTCAAGACT
TCTCTCCCCGACCCATTTTTGTGTGTTTGGTAGTTTGCACGGGGCATCTTCAGCATGCACGAACTCA
ACCCAAACACTCTCTTAGCATGCAGGAAGAGAGAACACCCCAACAGCGGTTCGCTGGTCAGCATGCA
TAAGCGAAGCGAGCGACAATGTGTTTTAAAAAATGTTCGAATTTCTAAAAAATAATTTAATTTTCAG
AAAATGTTGTGGATACAAAAAATTATTCAAATTTTCAAAAAAGACCATATTTTCAAAATATTCAAAA
ATTACGAAAAATGGCCAAATTTTCAAAATTTGTTAAGATTTTCAAAAAATGTTCAGAATTTTAAATT
TTTGTCGCAATTAAAAAATCAAATCTTGAAACTATTTGATTTTCCAAAATTATTTCAGATTTCAAAT
ATTGTTTATTTTTTAAAAAAATTCAAACTTTCAAATTTGTTTACTTTTTTTTAAAAAAAACTTTAGAA
TATCAAGAAATGTTTAATTTCCAAAAATGTTTTTTCATGTTTCAAAAAATTGTTCACATTTTTGAAA
TGTTTTGTCACTTATTTGAAAATCTCGGTTGCTACAGTGCGGATCGACATGTCTAAACAGATTCAAG
CTGCCAAACAGAGTCTCAGCTTAGCATGTCTCAGCACATAGACCGCTACCAAACATCAGCAAATGCA
TGTACTCAGCATGTCTACACTGAGCATATATCAGAGAAGAACAACCATGCTAGGTTGGGACTGCTAC
CAAACACACACGGAAAGCACAATTTTTGTTTCTTTTTCGAGAAGCACCGGCTATGCCTTTTCTTTTC
TGGGAAGCACAACTATGTTTTTCTTTTTCTTTTTGGTGCTTCTCATGGGAGCACAACTATGCTTTTA
TATTTTCGAGAAGCACAACTCAGGAAGCACGGACTATTTCTTTCTCTTTTTGAAAAGCATAAGTATG
TTTTTCGCTCGCGGAAAAGCACCGCTTCCAAATGTTTTTATACCAAAACCAAGCAAAAACTGAAAAG
CTAAAAGAACCAACAAACAAAATTAAAAAAACGAAAACATGCGTCGAGAAGGAAAAAAAAAAGAAAT
CTTGAGGGTGTGCCCAGCACACGTGGTGGCTGTTGGGCGCATCATTGAGGTGCTTCTAGGTAAAAAA
ACATGACCTCTGCGGGAACCCTCCAAGTGGTACCGTTAATAAGTTGCTCCCTAGCAGCGCCCGATTT
TGGCCATGTCACTCTGGACTCCTAAGTTATAGTAACAAGTGATTTTGGCCCAACAAAAATTGTCGTT
GTTTCGTCAGTATCCAGGGAGTCACTAACGAATGGTGACTCCATTCGAGGAACCTCGAGGGGTCACT
TTTGGTGGTCTGGAGCGCGCTAAGTGGTGCCTCCAGCCGATACCATATGATGCCGCCGTCACATGTC
CCATGCTTGAGGTTCCCTCTGAATTTCAGTTTTTTATTTTTTTGTACGCAATTTCGGATTTTTGTTT
TTTTCGATTTTGCTTGGTTTTCCTAAGTTTTAAAGAGAAAAAAAAATATTTTTTGCGAAGCACATCT
GTGTGTGTGTGTTTCTAATGGAAATGGAAACACATATACGCTTCCATGAGAAGCACAACCCATGTTT
CCCCTAAGACCATCGCATGTGCTTTCACGAGAAACACAATCGGTGCTTCCCGAGAAAATAAAAGGAA
AACACAGCACGTGCTCCCCCCCACACCCAAAGTGAAAAACCTTAACCTTGCTTTCGGACTCCGGTTT
TTTCTTCCGTTTTCTCGGTTGCCAGTTTGTCTGATTTTTTTGGATTCATTTTTTGTTTCCTTTTTTC
CTGGTTTCTAATTTTTGGGGGTTTTTTCATGGGAAAAAATGTTTGTTGAAACCTATTAATAAGGGAT
CTTGTTTCGAAGATCTCGATGCGAGAAATCCAACGATTAAAACAGTTTGATAGTTGGAGACATGGTT
CAAGAGATAAAACATTTTAAATAAATGAATCTATGAAAAAAGAAAAAAAAATCCAGGTTGCGATAAG
TGTGCGCACAACAGCGTTCACTTCTCTGAGAAGGTGAGGAGTGACCTTTGTAAGCGGTAACCCTTAA
AATAGTGATTTAACGGTATTCGAATTCTGAGCGAAGAATTATTTTTGCGTTCCTGGACAATTTTGGC
CCGATAAGAAATGCTTAGGGCTTTCGTGACGCCCAAGCTTGTTACGATAATCCAATTCGAGAGATTT
TCCAACCAATCCAAAAGACCGCTCGCCGGATGAAGAAACATTTTTGAAAATTCCTTCTTTGTGTTGG
CTAACTAGATGCAAGTTGTTTCGGCCGGAAAAAAAACTAGCATCTAGTTGTTCCCGCAGAAAAAAAA
AACTAGCAGCCAATTGACTCAGGAACTTTCTAGCAGCCAGACACGAAAGCCCTATGAAAAAAGAAAA

AAAAATCCAGGTTGTGATAAGTGTGCGCACAACAGCGTTCACTTCTCTGAGAAGGTGAGGAGTGACC

TTTGTAAGCGGTAACCCTTAAAATAGTGATTTAACGGTATTCGAATTCTGAGCGAAGAATTATTTTT

GTGTTCCTGGACAATTTTGGCCCGATAAGAAATGCTTAGGGCTTTCGTGACGCCCAAGCTTGTTACG

ATAATCCAATTCGAGAGATTTTCCAACCAATCCAAAAGACCGCTCGCCGGATGAAGAAACATTTTTG

AAAATTCCTTCTTCGTGTTGGCTAACTAGATGCAAGTTGTTTCGGCCGGAAAAAAAACTAGCATCTA

GTTGTTCCCGCAGAAAAAAAAAACTAGCAGCCAATTGACTCAGGAACTTTCTAGCAGCCAGACAAAC

ATAATCATGATAATCCAGGAATAATATAAGCATCAATATTAAGTTTAGGACTTGAACTCTGATGGGC

TAGGGATACATGTCCGGCATATCCTATTTGGCGCTTATGCGCCGGTTCGCCAACCGGCGCCTGCAGG

CATTTCTTATTGGGCCGGCCCGTTACGTCCTTTTTTCCTGGTTTTGAACCAGCAAAATTTGTGTGCC

GATGGGGATTCGAACACACGACCAAGTTGTTGAAAGCGCACTGCTGTAGACACCAGGGACACCCGCT

CAAGCGTGTAAAGTTAGCTCGTTTTTTTCTTTTTTACTTTGTTCTTCAGTTCGGTTTCTTTTCTATT

CTTTTTCTTTTTTCTTTGTTTACTTTAGTGCACGGTTTTTTTTTATTCCATGAACTTTTTTCAGAAC

CTATGAACTTTTTCAGAATCAATAACTTATTTTCAAATTCGATGAATTTCTTTTAATTGAATGAACT

TTTTTGAAATTTTATGTACTTTTTTTTGCAAAATCGATGAACTTTTTCAAAAACCCATGAACTTTTTT

CCATATCCGATGAACTTTTTTCAAATCCGATGAACTTTTTTAAAATTGTTGTGAACTTTTTGCAAAA

AAAAAATCAAATCCGATGAACTTTTTTGAAAATTTGGTGAACTTTTTCGAAATTTGATGAACTATTT

TTCAAATCTAATGAACTTTTTTGAAATTCAATGAACTTTTTCCAAATTTCAGTGAACTTTTTTTCAA

ATCCATGAACTTATTTCAAACTCATGGTCATTTTTCATTGTTGTGAACTGTTTTTTCAAAATTATTG

CACTTTTTTCATTTCATGAATTTTTTTTTGAAAGTCAATGAACGTTTTCTTCCAAAATGAAGAACTT

TTTTCAAAATCACGATCTTTTTTTCCAAATAATTCAAAAAAATTAAATGGTACCATAGATGAATGAGT

ATGTTCAATAGAATAGCACTGTGGCCTTGTTATTAAGCTGCAGCCTCTGTTAGTAGTCATAGATAGC

AAGTAGCCCGTGTGGCTTGTGGCACCTGTTTGTTAGCATGCGGTGGCGAGTTCGAGTCCTGAGGAGA

GCATTATTTTTACCCGTTTTTGGTTTTGCGAGCGTTTTATGTTCTCTGGTTTGCACTCTTCATGGGC

CGGCCCAGCGCGGGGCGCTGCAGGCGTCAGGAGCGCCAACGGGCGCCTGCAGCGCCGTATAGGAGCT

CCCTACATGTCCTTCTAACCATCCAATCACAGATTGGTTTGCTTTTTTTGACGAACTGATTTGTTTT

TAATTGTTTGAATTGCTTACCACATGTTTCACGTGGTCTGACGAACTGGTTGGGTTCATAAGGTGTG

TGAAATGAGGTCACGATGTTAAAAGAGGTACCCTTTGATTAGTTGCTTCCCGAGAAACATCTGTGCT

AAAAAAGAGAGCAAGAAGAGGTCACGATGTTGGTATCATGCGGCCCACCAGAATATGCTGAAATGGA

GCTCAGCACGGGCCAACAAAATACCTAGCCCAGGAAGCTACGAAGCCGGTGCAAACTACCAGGCCTT

AAGCCGAAGGATGGGGCAAGGTGAACCGTGTCCGGTCCGGCTCCCCCGCGTCCGCCCGCATAGAAGG

AAGTGGCGGAATACTCTTCCCATCCCAGAAGAAAAAATAAGGTAGGAAGCGGAATGGGTGGCCGTGG

TCGCGCGAGCTTGCGCCGTCCGGTGGCGATCTGACACGCGGTACGAGCGGCGGCCGGCATACGTGTC

CAGCGGCGACGGGCCCGCGGCCCCGGGCTAGGTACAACGGTGCCGTATTTCTCCCCTGCTCTCCTCG

CCACGGTTTCACAGCCGCCCCGAAACCACCAGAGCTCCCACAGCAGATCACTCGCCCGCCTCTCCTC

TCCTTCCTCCCAACCGTCGCTGTAACAAATCCCACTCCGTTCCTTCCTTCACACTACCTAGAAGCTT

TGGCAGTTGAGTTAGGTGCCCACCACAGGGGTGCTCTGGTGGGATCATCTGGTGTGTTTGTTGGTAG

CTAGCTAGCTAGGGGAAGAAGATCTGATGAGGTCC

Promoter region highly conserved (85%) with rice

Rsr2 cut site

Reverse primer PB1_pGPC_R2

6.1.3 Gene synthesis 3xHA-NAM-B1

HA tag highlighted in pink, exons highlighted in green, start and stop codon highlighted in blue, untranslated regions and splice junctions in red text.

CTGCAACTTAAGCCGAAGGATGGGGCAAGGTGAACCGTGTCCGGTCCGGCTCCCCCGCGTCCGCCCGCAT
AGAAGGAAGTGGCGGAATACTCTTCCCATCCCAGAAGAAAAAATAAGGTAGGAAGCGGAATGGGTGGCCG
TGGTCGCGCGAGCTTGCGCCGTCCGGTGGCGATCTGACACGCGGTACGAGCGGCGGCCGGCATACGTGTC
CAGCGGCGACGGGCCCGCGGCCCCGGGCTAGGTACAACGGTGCCGTATTTCTCCCCTGCTCTCCTCGCCA
CGGTTTCAC**AGCCGCCCCGAAACCACCAGAGCTCCCACAGCAGATCACTCGCCCGCCTCTCCTCTCCTTC**
CTCCCAACCGTCGCTGTAACAAATCCCACTCCGTTCCTTCCTTCACACTACCTAGAAGCTTTGGCAGTTG
AGTTAGGTGCCCACCACAGGGGTGCTCTGGTGGGATCATCTGGTGTGTTTGTTGGTAGCTAGCTAGCTAG
GGGAAGAAGATCTGATGAGGTCCATGGGGTTGATTAACATCTTTTACCCATACGATGTTCCTGACTATGC
GGGCTATCCCTATGACGTCCCGGACTATGCAGGATCCTATCCATATGACGTTCCAGATTACGCTGCTCAG
TGCAGCGGCAGCTCCGACTCAT**CT**TCCGGCTCGGCGCAAAAAGCAACGCGGTATCACCATCAGCATCAGC
CGCCGCCTCCGCAGCGGGGCTCGGCGCCGGAGCTCCCGCCGGGCTTCCGGTTCCACCCGACGGACGAGGA
GCTGGTGGTGCACTACCTCAAGAAGAAGGCCGACAAGGCGCCGCTCCCCGTCAACATCATCGCCGAGGTG
GATCTCTACAAGTTCGACCCATGGGAGCTCCCCG**GT**ATGTTATGTCTATCTCGTCGGCCGGCCGTGCTTA
CTTTATCAAGCGCCGCAAATTTT**C**GGTGCAATTAAATAATCGAATAATCCATCCATCTCATGCTTATACT
CCTGTGCACAAGTAGTATTTTTATATTCTTCCAGTACACATGTGTGTAGATGGTTTATGTATGTGATCCT
GTCGTGCTTGTTCATGCGCTCGGGATCCGGATCCATC**AG**AGAAGGCGACCATCGGGGAGCAGGAGTGGTA
CTTCTTCAGCCCGCGCGACCGCAAGTACCCCAACGGCGCGCGGCCGAACCGGGCGGCGACGTCGGGGTAC
TGGAAGGCCACCGGCACGGACAAGCCTATCCTGGCCTCGGGGACGGGGTGCGGCCTGGTCCGGGAGAAGC
TCGGCGTCAAGAAGGCGCTCGTGTTCTACCGCGGGAAGCCGCCCAAGGGCCTCAAAACCAACTGGATCAT
GCACGAGTACCGCCTCACCGACGCATCTGGCTCCACCACCGCCACCAACCGACCGCCGCCGGTGACCGGC
GGGAGCAGGGCTGCTGCCTCTCTCAGG**GT**ACGTACACGTGTCGATCGCACGGTCTAGCAGTATTTAATTG
CTCTCCAGCTTAATTAGGGTATTGTTGATGGTTGATGAAGTTAATTATGTACCGTCGTCTCATC**AG**TTGG
ACGACTGGGTGCTGTGCCGCATCTACAAGAAGATCAACAAGGCCGCGGCCGGCGATCAGCAGAGGAACAC
GGAGTGCGAGGACTCCGTGGAGGACGCGGTCACCGCGTACCCGCTCTATGCCACGGCGGGCATGACCGGT
GCAGGTGCGCATGGCAGCAACTACGCTTCACCTTCACTGCTCCATCATCAGGACAGCCATTTCCTGGACG
GCCTGTTCACAGCAGACGACGCCGGCCTCTCGGCGGGCGCCACCTCGCTGAGCCACCTAGCAGCGGCGGC
GAGGGCAAGCCCGGCTCCGACCAAACAGTTTCTCGCCCCGTCGTCTTCGACCCCGTTCAACTGGCTCGAT
GCGTCACCAGTCGGCATCCTCCCACAGGCAAGGAATTTTCCTGGGTTTAACAGGAGCAGAAACGTCGGCA
ATATGTCGCTGTCATCGACGGCCGACATGGCTGGCGCAGTGGACAACGGTGGAGGCAATGCGGTGAACGC
CATGTCTACCTATCTTCCCGTGCAAGACGGGACGTACCATCAGCAGCATGTCATCCTCGGCGCTCCGCTG
GTGCCAGAAGCCGCCGCCGCCACCTCTGGATTCCAGCATCCCGTTCAAATATCCGGCGTGAACTGGAATC
CC**TGA**TCAAATGATATGAACACCACATACGGGCATGCACGCACGCATGCATAACTTTTGCAAGTCGTACA
ACATTGCTAGCCAGTAGTTGTTGCAGTTTGTGGTAGTCCCTTTCAGTGAGCACTGAGTAGTTGCATGCAC
ATCACCACTGCATGGATATATATGGCTGCATTGCACCATGGGCACGTACTTGTGCGAACTTGCTAGCCAT
ATATATAGTAGTACAATAGCTAGGAGTATTTTCGAAGTACAAAAAAATCATAACATAGTACTCCATATAT
ATTGTAATAAATATATTTTTTAGATATATATGAGTATGATAATAGCTTCATATGCATATACTATATTGTA
GTGCATATAGTAAATATATATGCAGTTGATTACAAACCCCAGTCAGATACAATTTTTGGCCGGGTCTAGC
TACTTATCATGTATAAATAGGGGTGTATGTAAATGAGCAACGCATGGTAACTGCATGCACGCGTATCATT

GATCTATATGGACGTACACTGTATAACTTTTTTTTTCATGAATAAGATTTGCGAACTAGTATGTCGCAGT
GGTTTAGTTATTATGTGAATGGCCCTTATATATATATATATATATATATATATATATAGATATCTATA
TATATATATATATATATATATATATATATATATATGTTCGAAAAATGTGTGTGTCGACCGTCAA

6.1.4 Gene synthesis 3xFLAG-NAM-B1

FLAG tag highlighted in yellow, exons highlighted in green, start and stop codon highlighted in blue, untranslated regions and splice junctions in red text.

CTGCAACTTAAGCCGAAGGATGGGGCAAGGTGAACCGTGTCCGGTCCGGCTCCCCCGCGTCCGCCCGCAT
AGAAGGAAGTGGCGGAATACTCTTCCCATCCCAGAAGAAAAAATAAGGTAGGAAGCGGAATGGGTGGCCG
TGGTCGCGCGAGCTTGCGCCGTCCGGTGGCGATCTGACACGCGGTACGAGCGGCGGCCGGCATACGTGTC
CAGCGGCGACGGGCCCGCGGCCCCGGGCTAGGTACAACGGTGCCGTATTTCTCCCCTGCTCTCCTCGCCA
CGGTTTCAC**AGCCGCCCCGAAACCACCAGAGCTCCCACAGCAGATCACTCGCCCGCCTCTCCTCTCCTTC**
CTCCCAACCGTCGCTGTAACAAATCCCACTCCGTTCCTTCCTTCACACTACCTAGAAGCTTTGGCAGTTG
AGTTAGGTGCCCACCACAGGGGTGCTCTGGTGGGATCATCTGGTGTGTTTGTTGGTAGCTAGCTAGCTAG
GGGAAGAAGATCTGATGAGGTCCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATG
ACATCGACTACAAGGATGACGATGACAAGGGCAGCTCCGACTCAT**CT**TCCGGCTCGGCGCAAAAA
GCAACGCGGTATCACCATCAGCATCAGCCGCCGCCTCCGCAGCGGGGCTCGGCGCCGGAGCTCCCGCCGG
GCTTCCGGTTCCACCCGACGGACGAGGAGCTGGTGGTGCACTACCTCAAGAAGAAGGCCGACAAGGCGCC
GCTCCCCGTCAACATCATCGCCGAGGTGGATCTCTACAAGTTCGACCCATGGGAGCTCCCCG**GT**ATGTTA
TGTCTATCTCGTCGGCCGGCCGTGCTTACTTTATCAAGCGCCGCAAATTTT**C**GGTGCAATTAAATAATCG
AATAATCCATCCATCTCATGCTTATACTCCTGTGCACAAGTAGTATTTTTATATTCTTCCAGTACACATG
TGTGTAGATGGTTTATGTATGTGATCCTGTCGTGCTTGTTCATGCGCTCGGGATCCGGATCCATC**AG**AGA
AGGCGACCATCGGGGAGCAGGAGTGGTACTTCTTCAGCCCGCGCGACCGCAAGTACCCCAACGGCGCGCG
GCCGAACCGGGCGGCGACGTCGGGGTACTGGAAGGCCACCGGCACGGACAAGCCTATCCTGGCCTCGGGG
ACGGGGTGCGGCCTGGTCCGGGAGAAGCTCGGCGTCAAGAAGGCGCTCGTGTTCTACCGCGGGAAGCCGC
CCAAGGGCCTCAAAACCAACTGGATCATGCACGAGTACCGCCTCACCGACGCATCTGGCTCCACCACCGC
CACCAACCGACCGCCGCCGGTGACCGGCGGGAGCAGGGCTGCTGCCTCTCTCAGG**GT**ACGTACACGTGTC
GATCGCACGGTCTAGCAGTATTTAATTGCTCTCCAGCTTAATTAGGGTATTGTTGATGGTTGATGAAGTT
AATTATGTACCGTCGTCTCATC**AG**TTGGACGACTGGGTGCTGTGCCGCATCTACAAGAAGATCAACAAGG
CCGCGGCCGGCGATCAGCAGAGGAACACGGAGTGCGAGGACTCCGTGGAGGACGCGGTCACCGCGTACCC
GCTCTATGCCACGGCGGGCATGACCGGTGCAGGTGCGCATGGCAGCAACTACGCTTCACCTTCACTGCTC
CATCATCAGGACAGCCATTTCCTGGACGGCCTGTTCACAGCAGACGACGCCGGCCTCTCGGCGGGCGCCA
CCTCGCTGAGCCACCTAGCAGCGGCGGCGAGGGCAAGCCCGGCTCCGACCAAACAGTTTCTCGCCCCGTC
GTCTTCGACCCCGTTCAACTGGCTCGATGCGTCACCAGTCGGCATCCTCCCACAGGCAAGGAATTTTCCT
GGGTTTAACAGGAGCAGAAACGTCGGCAATATGTCGCTGTCATCGACGGCCGACATGGCTGGCGCAGTGG
ACAACGGTGGAGGCAATGCGGTGAACGCCATGTCTACCTATCTTCCCGTGCAAGACGGGACGTACCATCA
GCAGCATGTCATCCTCGGCGCTCCGCTGGTGCCAGAAGCCGCCGCCGCCACCTCTGGATTCCAGCATCCC
GTTCAAATATCCGGCGTGAACTGGAATCCC**TGA**TCAAATGATATGAACACCACATACGGGCATGCACGCA
CGCATGCATAACTTTTGCAAGTCGTACAACATTGCTAGCCAGTAGTTGTTGCAGTTTGTGGTAGTCCCTT
TCAGTGAGCACTGAGTAGTTGCATGCACATCACCACTGCATGGATATATATGGCTGCATTGCACCATGGG
CACGTACTTGTGCGAACTTGCTAGCCATATATATAGTAGTACAATAGCTAGGAGTATTTTCGAAGTACAA

AAAAATCATAACATAGTACTCCATATATATTGTAATAAATATATTTTTTAGATATATATGAGTATGATAA
TAGCTTCATATGCATATACTATATTGTAGTGCATATAGTAAATATATATGCAGTTGATTACAAACCCCAG
TCAGATACAATTTTTGGCCGGGTCTAGCTACTTATCATGTATAAATAGGGGTGTATGTAAATGAGCAACG
CATGGTAACTGCATGCACGCGTATCATTGATCTATATGGACGTACACTGTATAACTTTTTTTTTCATGAA
TAAGATTTGCGAACTAGTATGTCGCAGTGGTTTAGTTATTATGTGAATGGCCCTTATATATATATATATA
TATATATATATATATATAGATATCTATATATATATATATATATATATATATATATATATATGTTCGAA
AAATGTGTGTGTCGACCGTCAA

6.2 Scripts

6.2.1 Repetitiveness of peaks

6.2.1.1 bed_to_fasta.pl - Perl script to convert bed files to fasta files using the reference genome

```
#!/nbi/software/production/perl/5.16.2/x86_64/bin/perl/ -w

use strict;
use diagnostics;
use lib "/nbi/software/testing/bioperl/1.6.1/src/BioPerl-1.6.1/";
use Bio::Perl;
use Bio::AlignIO;

##NB#########################################
# Author: Philippa Borrill philippa.borrill@jic.ac.uk
# Version edited 15.8.14
# Input files (peak containing files) in bed format (5 column)
# make sure $dir is the directory containing bed files you want
extract fasta files for

my $path = "/nbi/group-data/ifs/NBI/Research-Groups/Cristobal-
Uauy/NAM-B1/";
my $dir = "/2nd_chip_seq/chip-seq/pseudomol_v3.3/properpairs";

# put all the peak containing files into an array
push my @files, `ls -1 $path/$dir/[1-9]*peaks.bed`;

# extract the names of all the files in the array (remove the full
path name) and
#loop through all the files carry out the steps below to launch BLAST
for my $file (@files){
        print "file =  $file\n";
        chop $file;
        my($inputfile) = $file =~ /([\.\w\+\-\:]+)peaks.bed/;
        print "inputfile = $inputfile\n";
        my ($substr) = substr($inputfile, 0, 26);

# make 1 output directory for each file in the original directory
my $output_dir = "$substr"."_blast_iwgsc";

#move directory and make the first output directory
chdir ("$path/$dir")or die;
mkdir ("$output_dir") or die;

# using samtools faidx extract regions in the peak file (.bed) from
wheat the pseudomolecule v3.3 (.fasta) and put these sequences into a
new fasta file
# creates output of fasta file with all regions and their sequence in
one file

# open the first peak containing file
open(POSITIONS,"$inputfile"."peaks.bed") or die "Could not open
$inputfile"."peaks.bed: $!";

# move into the output directory
chdir ("$path/$dir/$output_dir")or die;
```

```
# read through the peak containing file, extract the information about
the peak location for each line
while (my $line = <POSITIONS>){
        my    ($chromName, $begin, $end, $peakname, $qvalue) =
split('\s+', $line);

# use samtools to extract the peak locations from the pseudomolecules
for each peak in the peak containing file
  open(SAMTOOLS,"samtools faidx $path/wheat_pseudo_v3.3.fasta
$chromName:$begin-$end |");
    while(my $line = <SAMTOOLS>){
        my $outputfile = "extracted_promoters.txt";
        open(OUTPUT, ">>$outputfile") or die "Cannot open file
$outputfile to write to\n\n";
        print $line;
        print OUTPUT $line;
        close(OUTPUT);
        }
    close(SAMTOOLS);
}

close(POSITIONS);

print "\nregions extracted successfully from bed file\n";

# split fasta into separate fasta for each region (new files named as
the regions e.g. 1A:11144-11256)

open (PERL, "perl /usr/users/metbio/borrillp/split_fasta.pl
extracted_promoters.txt|");

print "\nextracted regions split into separate fasta files\n";

}

# the programme then goes back to the next peak containing file in the
array of input files and extracts the sequences for these regions
# until the sequences for all regions in all peak containing files
have been extracted
```

6.2.1.2 split_fasta.pl - Perl script to split a fasta file containing multiple sequences into one fasta file per sequence (written by John Nash)

```
#! /nbi/software/production/perl/5.16.2/x86_64/bin/perl/
use warnings;
use strict;
use Text::Wrap;

## Program Info:
#
# Name:  split_fasta
#
# Function:  Takes a multiple fasta file and removes a set of
#    sequences to makes a second fasta file.  Useful for pulling
#    subsets of sequences from entire genomes.
#
# Author: John Nash
#  Copyright (c) National Research Council of Canada, 2000-2001,
#  all rights reserved.
```

```
#
# Licence: This script may be used freely as long as no fee is charged
#    for use, and as long as the author/copyright attributions
#    are not removed.
#
# History:
#   Version 1.0 (August 1, 2002): first non-beta release.
#   Version 1.1 (3 Nov, 2003): cleaned up code.
#   Version 1.2 (4 Nov 2003): fixed an oops
#           1.3 (10 Dec 2003): option to spit to STDOUT
#           1.4 (2 Feb, 2012): remove | from filenames
##

my $title = "split_fasta.pl";
my $version = "1.4";
my $date = "2012-02-02";
$Text::Wrap::columns = 65;
my $error_msg = "Type \"$title -h\" for help";

# Get and process the command line params:
my @cmdline = process_command_line();

my $pipeout = $cmdline[0];

## Handle input parameters:
## Handle input parameters:
## Does the input sequence exist:  handle errors
# If $ARGV[0] is not blank, test for file's existence:
if (defined $ARGV[0]) {
    unless (-e $ARGV[0]) {
        die("\nError: Sequence file \'$ARGV[0]\' does *not* exist. \n",
            $error_msg, "\n");
    }
    open (FILE, $ARGV[0]);
}

# If it has come in from a redirection or pipe,
#  check it is bigger than 0:
else {
    my $fh = *STDIN;
    unless ((-p $fh) or (-s $fh)) {
        die("\nError: Piped sequence file does *not* exist. \n",
            $error_msg, "\n");
    }
    *FILE = *STDIN;
}


# Massage each sequence from the FASTA file to a string:
my (%seq_name, %seq_str);

# read in the sequence from a FASTA file, from stdin:
my $count = 0;

while (<FILE>) {

# Substitutes DOS textfile carriage returns with Unix ones:
    s/\r\n/\n/g;
    chomp;

# If the line begins with a ">", it's a comment field.
# Therefore we need the name:
    if (/^>/)  {
```

```
# Split on the SPACE character or else we could have HUGE file names:
        my (@temp) = split / /;

# Replace the "|" character with "_".
        $temp[0] =~ s/\|/\_/g;
# Remove trailing underscores.
        $temp[0] =~ s/\_\Z//g;

# Add it to the array/
        $seq_name{$count} = $temp[0];
        $count++;
    }
    else {
        $seq_str{$count - 1} .= $_;
    }
} # end of while

my @dupes = ();

# sort each value alphabetically, and look for duplicate names:
my @namelist = sort { $a cmp $b } values %seq_name;
for ($count = 1; $count <= $#namelist; $count++)  {
    if ($namelist[$count] eq $namelist[$count - 1])  {
        push @dupes, $namelist[$count];
    }
}

# remove the additional duplicate names:
my %dupes = ();
foreach (@dupes) {
    $dupes{$_}++;
}
@dupes = keys %dupes;

# find the key/value pair of each copy of the duplicate, and change
it:
foreach my $match (@dupes)  {
    $count = 0;
    foreach my $item (keys %seq_name)  {
        if ($seq_name{$item} eq $match)  {
            $count++;
            $seq_name{$item} = "$seq_name{$item}" . "_" . "$count";
        }
    }
}

# It's a quick fix, we can be elegant later:
# Print it all out:

foreach (sort {$a <=> $b} keys %seq_name)  {
    if ($pipeout eq "no")      {
        open (OUTFILE, ">$seq_name{$_}.fa") or
            die("\nError: Cannot open requested file ($!) \n");
    }
    else {
        *OUTFILE = *STDOUT;
    }
    print OUTFILE $seq_name{$_}, "\n";
    print OUTFILE wrap('', '', "$seq_str{$_}\n");
}

exit;
```

```
### end of main:

sub help {
print <<EOHelp;
$title $version ($date)

Syntax:  $title \"fasta_file\"
   or    $title -h for help
   or    $title -P \"fasta_file\" to send output to STDOUT

   \"$title fasta_file\" will split a multiple fasta file into
   many, many, many files, each containing one sequence, in fasta
   format.  It will use the first word after the \">\" sign in the
   comment field to name each file. It will rename duplicate entries
by
   appending \"\_1\", \"\_2\", etc, to the filename.

   Genbank delimits fields in fasta headers with the "|" character.
This
   breaks computers as it is an illegal file name.  These get changed
   to a "_" charchter.

Use the \"-P\" switch to use the program to pipe data out

EOHelp
die ("\n");
} # end of sub help

sub process_command_line {
# Variables:
    my %opts = ();     # command line params, as entered by user
    my @cmd_line;      # returned value
    my @list;          # %opts as an array for handling
    my $cmd_args;           # return value for getopts()
    my $item;
    my $pipeout="no";

# Get the command=line parameters:
    use vars qw($opt_h);
    use Getopt::Std;
    $cmd_args = getopts('hP', \%opts);

# Die on illegal argument list:
    if ($cmd_args == 0) {
        die ("Error: Missing or incorrect command line
parameter(s)!\n",
             $error_msg, "\n");
    }

# Make the hashes into an array:
    @list = keys %opts;

# Do a quick check for "help" and the compulsory parameters:
    foreach $item (@list)  {
# Help:
        if ($item eq "h")  { help(); }
# Pipe out:
        if ($item eq "P")  { $pipeout="yes"; }
    }
    @cmdline=($pipeout);
    return @cmdline;
} #end of sub process_command_line()
```

6.2.1.3 blast_promoter_regions_to_chroms_lsf.pl - Perl script to launch computation jobs which BLAST each peak against all the IWGSC chromosome arm survey sequences, one chromosome at a time

```
#!/nbi/software/production/perl/5.16.2/x86_64/bin/perl/ -w

# Author: Philippa Borrill philippa.borrill@jic.ac.uk
# Version edited 15.8.14

# This scripts blasts all fasta files (.fa) in all directories called
$output_dir inside directories in @blast_dir
# against the iwgsc chromosome arm contigs, it does this separately
for the chromosomes named in @chrom_db and makes a
# .txt file for each .fa file against each chromosome with hits over a
certain percentage e.g. 99%

use strict;
use diagnostics;

# input parameters
my $path = "/nbi/group-data/ifs/NBI/Research-Groups/Cristobal-
Uauy/NAM-B1/";
my $dir = "/2nd_chip_seq/chip-seq/pseudomol_v3.3/properpairs";

# chromosomes to BLAST against
my @chrom_db = qw(1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B
6D 7A 7B 7D);
my $output_dir = "blast_batch99";
my $final_output = "99_average_hits_per_chromosome.txt";

# sample folders containing sequence files to use as queries in BLAST
my @blast_dir = qw(1AL_BAMPE__blast_iwgsc 1BL_BAMPE__blast_iwgsc
2AL_BAMPE__blast_iwgsc 2BL_BAMPE__blast_iwgsc 3AL_BAMPE__blast_iwgsc
3BL_BAMPE__blast_iwgsc 1AP_BAMPE__blast_iwgsc 1BP_BAMPE__blast_iwgsc
2AP_BAMPE__blast_iwgsc 2BP_BAMPE__blast_iwgsc 3AP_BAMPE__blast_iwgsc
3BP_BAMPE__blast_iwgsc);

# for each of the folders containing query sequences make a new
temporary file which will contain the BLAST job
# each BLAST job will take one query sequence and BLAST it against 1
chromosome

# go through all the sample folders containing query sequences in turn
and for each:
for my $blast_dir(@blast_dir){
print "blast_dir = $blast_dir\n";
print "path/dir/ = $path/$dir\n";

# move into the folder containing the query sequence
chdir ("$path/$dir/$blast_dir")or die;

# make an output folder
mkdir ("$output_dir") or die;

# make the template header for the temporary file which will be
submitted to bsub (the header is needed to correctly submit the job to
the cluster)
my $bsub_header = <<"LSF";
#!/bin/bash
#
```

```
# LSF batch script to launch parallel bwa tasks
#
#BSUB -q testing
#BSUB -o email_JOBNAME.txt
#BSUB -J BLAST_JOBNAME
#BSUB -R "rusage[mem=100000]"
LSF

# for each peak sequence in the folder extract the peak sequence
identifier

push my @files, `ls -1 $path/$dir/$blast_dir/*.fa`;
for my $file (@files){
        chop $file;
        my($inputfile) = $file =~ /([\w\+\-\:]+).fa/;
                for my $chrom_db(@chrom_db){
                my $output_file = "$inputfile"."_against_"."$chrom_db";

# make a temporary file for the individual peak sequence and the
chromosome it will be BLASTed against
                my $tmp_file =
"/tmp/blast_lsf.$inputfile._against_$chrom_db";
                  open (BSUB, ">$tmp_file") or die "Couldn't open temp
file\n";
                  my $bsub_header = $bsub_header;

# alter the template header to contain this specific BLAST name
                  $bsub_header =~
s/JOBNAME/"$inputfile._against_$chrom_db"/;

# print the header into the temporary file
                  print BSUB "$bsub_header\n\n";

# print to the temporary file:
# 1) the command to change directory in the relevant directory
                  print BSUB "\ncd $path/$dir/$blast_dir\n";
# 2) the command to source blast+ (v2.2.28)
                  print BSUB "source blast+-2.2.28\n";
# 3) the BLAST command searching for hits to the peak sequence in the
relevant chromsome, keeping hits with over 99% identity, giving a
tabular format output
                  print BSUB "blastn -db $path/2nd_chip_seq/chip-
seq//iwgcs/$chrom_db -query $inputfile.fa -num_threads 1  -
perc_identity 99 -outfmt 6 -out
$path/$dir/$blast_dir/$output_dir/$output_file.txt\n";
# stop printing to the temporary file
                  close BSUB;

# submit the temporary file as a job to the cluster
                  system("bsub < $tmp_file");

#delete the temporary file
                  unlink $tmp_file;
                        }

# repeat for all chromsomes
                }

# repeat for all peak regions from that sample
}

# repeat for all samples
```

6.2.1.4 calc_promoter_region_hits_to_chroms.pl – Perl script to summarise the number of hits to the IWGSC chromosome arm sequences per peak region (as determined by BLAST)

```
#!/nbi/software/production/perl/5.16.2/x86_64/bin/perl/ -w

# Author: Philippa Borrill philippa.borrill@jic.ac.uk
# Version edited 15.8.14

# This script counts the number of hits per chromosome for each
promoter region and puts this into a separate file for each promoter
region.
# It then makes a file containing each promoter region and the total
number of blast hits for that region in 2 columns.
use strict;
use diagnostics;
use lib "/nbi/software/testing/bioperl/1.6.1/src/BioPerl-1.6.1/";
use Bio::Perl;
use Bio::AlignIO;

my $path = "/nbi/group-data/ifs/NBI/Research-Groups/Cristobal-
Uauy/NAM-B1/";
my $dir = "/2nd_chip_seq/chip-
seq/pseudomol_v3.3/results/randomised_peaks_1BP_2AP_2BP_3AP_3BP";
my @chrom_db = qw(1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B
6D 7A 7B 7D);
my $output_dir = "blast_batch99";
my $final_output = "99_average_hits_per_chromosome.txt";
my @blast_dir = qw(randomised_1BP_2AP_2BP_3AP_blast_iwgsc);

for my $blast_dir(@blast_dir){

chdir ("$path/$dir/$blast_dir/$output_dir")or die;

#for each file count number of hits per chromosome and count total
number of hits


#go through each peak region in turn
push my @files, `ls -1 $path/$dir/$blast_dir/*.fa`;

for my $file (@files){
        chop $file;
        my($inputfile) = $file =~ /([\w\+\-\:]+).fa/;

#get all files relevant to that input file into an array (i.e. all the
different chromsomes it was blasted against)
        push my @blast_output, `ls -1 $inputfile*.txt`;
                for my $blast_output (@blast_output){
                chop $blast_output;
                my ($blasted_chrom) = $blast_output =~ /([\w\+\-
\:]+).txt/;
                my $specific_chrom = substr ($blasted_chrom, -2);

#open the blast results and count number of lines (i.e. number of
hits), push into array and put into txt file specific to that original
region
                open(FILE, "<", "$blasted_chrom.txt") or die "Could not
open file: $!";
```

```
                my($lines)="0";
                while (<FILE>){
                        $lines ++;
                        }
# print the number of hits for that specific chromosome in a file
which colllates all the different chromosomes for that peak region
                my @blast_array = ( "$specific_chrom", "$lines\n" );
                my $count_output = "count_$inputfile.txt";
                open(OUTPUTFILE, '>>', $count_output) or die "Couldn't
open: $!";
                print OUTPUTFILE join("\t",@blast_array);
                close(OUTPUTFILE);
                }

}
# repeat for all different peak regions so you have 1 file per peak
region which has a list of the chromosome name and the number of hits



# add up the hits across all chromsomes for each peak region
push my @count_files, `ls -1 count_*.txt`;
                for my $count_files (@count_files){
                chop $count_files;
                my ($file) = $count_files =~ /count_([\w\+\-\:]+).txt/;
                    my $col2 = 0;
                    my $rows = 0;
                    open (COUNT_FILE, "count_$file.txt") || die
"couldn't open the file!";
                    while (my $record = <COUNT_FILE>) {
                    my ($var1, $var2) = split(/\t/, $record);
                    $col2 += $var2;
                    $rows=$rows+1;
                    }
                    close(COUNT_FILE);

# print the total for that peak region to the final output file
my @average_array = ( "$file", "$col2\n" );

                open(OUTPUTFILE, '>>', $final_output) or die "Couldn't
open: $!";
                print OUTPUTFILE join("\t",@average_array);
                close(OUTPUTFILE);
                }

}
# repeat for all peak regions to have a final file which contains one
column of peak region names, and one column of the number of hits each
peak region had over 99% similarity
```

6.2.1.5 R_graphs_blast_promoters.pl – Perl script to launch an R script to make graphs of the number of hits to the IWGSC chromosome arm survey sequences for each sample

```
#!/nbi/software/production/perl/5.16.2/x86_64/bin/perl/ -w

# This script takes the files of blast hits in the genome and makes
graphs using R script loop_plot_no_blast_hits.r
# Author: Philippa Borrill philippa.borrill@jic.ac.uk
# Version edited 15.8.14

use strict;
use diagnostics;
use lib "/nbi/software/testing/bioperl/1.6.1/src/BioPerl-1.6.1/";
use Bio::Perl;
use Bio::AlignIO;
use File::Copy;

my $path = "/nbi/group-data/ifs/NBI/Research-Groups/Cristobal-
Uauy/NAM-B1/";
my $dir = "/2nd_chip_seq/chip-
seq/pseudomol_v3.3/results/randomised_peaks_1BP_2AP_2BP_3AP_3BP";
my $output_dir = "blast_batch99";
my $final_output = "99_average_hits_per_chromosome.txt";
my @blast_dir = qw(randomised_1BP_2AP_2BP_3AP_blast_iwgsc);

# go through each directory containing the number of BLAST hits for
different peak regions and launch an R script which will make a graph
of the number of hits per sample
for my $blast_dir(@blast_dir){
chdir ("$path/$dir/$blast_dir/$output_dir")or die;
system ("Rscript /nbi/group-data/ifs/NBI/Research-Groups/Cristobal-
Uauy/NAM-B1/2nd_chip_seq/R_scripts/loop_plot_no_blast_hits.r");

# specify to rename the pdf file of graphs produced by R to the same
name as the directory it came from (ie. the sample name)
my $rename = $blast_dir;

# rename the pdf file of graphs to the sample name and move it to a
final directory of graphs for all samples
move("$path/$dir/$blast_dir/$output_dir/whole_genome_blast_hits_99.pdf
", "$path/$dir/$rename"."_whole_genome_blast_hits_99.pdf");
```

6.2.1.6 loop_plot_no_blast_hits.r – R script to plot the number of blast hits for a particular sample as a box plot and two histograms with specific axes

```
#plot_histogram_promoter_%ID.r

# 07-11-13
# Author: Philippa Borrill philippa.borrill@jic.ac.uk
# Version edited 15.8.14

# read in the table of number of hits per peak region
blast.data<-
read.table("99_average_hits_per_chromosome.txt",sep="\t",header=F)

# check the data looks ok (print to output email)
head(blast.data)
summary(blast.data)
dim(blast.data)
mean(blast.data[,2])
median(blast.data[,2])

# make a pdf file to contain the graph
pdf(file='whole_genome_blast_hits_99.pdf', height=5, width=3)

# make a graph containing 3 panels and specify margin sizes
par(mfrow=c(3,1))
par(cex=0.8)
par(mar = c(3, 3, 0, 0), oma = c(0, 0, 3, 1))
par(mgp = c(1.75, 0.5, 0))

# further data checks (print to output email)
max(blast.data[,2])
print (blast.data[,2])

# make a box plot of the data (writes to the pdf)
boxplot(blast.data[,2], horizontal = TRUE)

# make histogram up to 250 blast hits
hist(blast.data[,2], main=NULL,
     col="grey", prob=F, xlim =c(0,250),  xlab=NULL, ylab="No. of
regions",
breaks=c(0,20,40,60,80,100,120,140,160,180,200,220,240,300,400,500,200
00))

# make histogram up to 20 blast hits (to focus on single copy regions)
hist(blast.data[,2], main=NULL,
     col="grey", prob=F, xlim =c(0,20),  xlab="No. of hits >99%
identity", ylab="No. of regions",
breaks=c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,30,40,50
,60,70,80,90,100,200,300,400,500,20000))

# print the title to the pdf
title("Hits to IWGSC contigs
      >99% identity", outer=TRUE, cex=1.5)

# close the pdf and save
dev.off()
```

References

Aoki N, Whitfeld P, Hoeren F, Scofield G, Newell K, Patrick J, Offler C, Clarke B, Rahman S, Furbank R. 2002. Three sucrose transporter genes are expressed in the developing grain of hexaploid wheat. *Plant Molecular Biology* 50: 453-462.

Araus JL, Brown HR, Febrero A, Bort J, Serret MD. 1993. Ear photosynthesis, carbon isotope discrimination and the contribution of respiratory CO_2 to differences in grain mass in durum wheat. *Plant, Cell & Environment* 16: 383-392.

Asplund L, Hagenblad J, Leino MW. 2010. Re-evaluating the history of the wheat domestication gene *NAM-B1* using historical plant material. *Journal of Archaeological Science* 37: 2303-2307.

Asplund L, Bergkvist G, Leino MW, Westerbergh A, Weih M. 2013. Swedish spring wheat varieties with the rare high grain protein allele of *NAM-B1* differ in leaf senescence and grain mineral content. *Plos One* 8.

Avivi L. 1978. High protein content in wild tetraploid Triticum dicoccoides Korn. Paper presented at 5th Internation Wheat Genetics Symposium.

Avni R, Zhao RR, Pearce S, Jun Y, Uauy C, Tabbita F, Fahima T, Slade A, Dubcovsky J, Distelfeld A. 2014. Functional characterization of *GPC-1* genes in hexaploid wheat. *Planta* 239: 313-324.

Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, Madrigal P, Taslim C, Zhang J. 2013. Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS Comput Biol* 9: e1003326.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* 37: W202-W208.

Balazadeh S, Riaño-Pachón DM, Mueller-Roeber B. 2008. Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. *Plant Biology* 10: 63-75.

Balazadeh S, Kwasniewski M, Caldana C, Mehrnia M, Zanor MI, Xue GP, Mueller-Roeber B. 2011. *ORS1*, an H_2O_2-responsive NAC transcription factor, controls senescence in *Arabidopsis thaliana*. *Molecular Plant* 4: 346-360.

Bancal M-O, Roche R, Bancal P. 2008. Late foliar diseases in wheat crops decrease nitrogen yield through N uptake rather than through variations in N remobilization. *Annals of Botany* 102: 579-590.

Barron C, Surget A, Rouau X. 2007. Relative amounts of tissues in mature wheat (*Triticum aestivum* L.) grain and their carbohydrate and phenolic acid composition. *Journal of Cereal Science* 45: 88-96.

Berdugo CA, Steiner U, Dehne HW, Oerke EC. 2012. Effect of bixafen on senescence and yield formation of wheat. *Pesticide Biochemistry and Physiology* 104: 171-177.

Blacklow WM, Darbyshire B, Pheloung P. 1984. Fructans polymerised and depolymerised in the internodes of winter wheat as grain-filling progressed. *Plant Science Letters* 36: 213-218.

Blum A. 1998. Improving wheat grain filling under stress by stem reserve mobilisation. *Euphytica* 100: 77-83.

Bogard M, Allard V, Martre P, Heumez E, Snape JW, Orford S, Griffiths S, Gaju O, Foulkes J, Le Gouis J. 2013. Identifying wheat genomic regions for improving grain protein concentration independently of grain yield using multiple inter-related populations. *Molecular Breeding* 31: 587-599.

Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E, Hake S. 2012. Unraveling the *KNOTTED1* regulatory network in maize meristems. *Genes Dev* 26: 1685-1690.

Borrás L, Slafer GA, Otegui MaE. 2004. Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal. *Field Crops Research* 86: 131-146.
Borrell AK, Hammer GL. 2000. Nitrogen dynamics and the physiological basis of stay-green in sorghum. *Crop Sci.* 40: 1295-1307.
Borrell AK, Hammer GL, Henzell RG. 2000. Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. *Crop Science* 40: 1037-1048.
Borrell AK, Van Oosterom EJ, Hammer GL, Jordan DR, Douglas A. 2003. The physiology of "stay-green" in sorghum. Paper presented at Australian Agronomy Conference, Geelong, Victoria, Australia.
Borrell AK, van Oosterom EJ, Mullet JE, George-Jaeggli B, Jordan DR, Klein PE, Hammer GL. 2014a. Stay-green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns. *New Phytol* 203: 817-830.
Borrell AK, Mullet JE, George-Jaeggli B, van Oosterom EJ, Hammer GL, Klein PE, Jordan DR. 2014b. Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *Journal of Experimental Botany*.
Borrill P, Connorton J, Balk J, Miller T, Sanders D, Uauy C. 2014. Biofortification of wheat grain with iron and zinc: integrating novel genomic resources and knowledge from model crops. *Frontiers in Plant Science* 5.
Bourque G, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research* 18: 1752-1762.
Breeze E, et al. 2011. High-resolution temporal profiling of transcripts during *Arabidopsis* leaf senescence reveals a distinct chronology of processes and regulation. *Plant Cell* 23: 873-894.
Brenchley R, et al. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491: 705-710.
Brevis JC, Dubcovsky J. 2010. Effects of the chromosome region including the *Gpc-B1* locus on wheat grain and protein. *Crop Science* 50: 93-104.
Brevis JC, Morris CF, Manthey F, Dubcovsky J. 2010. Effect of the grain protein content locus *Gpc-B1* on bread and pasta quality. *Journal of Cereal Science* 51: 357-365.
Brusslan JA, Alvarez-Canterbury AMR, Nair NU, Rice JC, Hitchler MJ, Pellegrini M. 2012. Genome-wide evaluation of histone methylation changes associated with leaf senescence in *Arabidopsis*. *Plos One* 7: 13.
Buschmann H, Chan J, Sanchez-Pulido L, Andrade-Navarro MA, Doonan JH, Lloyd CW. 2006. Microtubule-associated *AIR9* recognizes the cortical division site at preprophase and cell-plate insertion. *Curr Biol* 16: 1938-1943.
Cakmak I, Pfeiffer WH, McClafferty B. 2010. Biofortification of durum wheat with zinc and iron. *Cereal Chemistry* 87: 10-20.
Calderini DF, Abeledo LG, Slafer GA. 2000. Physiological maturity in wheat based on kernel water and dry matter. *Agronomy Journal* 92: 895-901.
Calderini DF, Reynolds MP, Slafer GA. 2006. Source–sink effects on grain weight of bread wheat, durum wheat, and triticale at different locations. *Australian Journal of Agricultural Research* 57: 227-233.
Canny MJ. 1986. Water pathways in wheat leaves. III. The passage of the mestome sheath and the function of the suberised lamellae. *Physiologia Plantarum* 66: 637-647.
Cantu D, Pearce SP, Distelfeld A, Christiansen MW, Uauy C, Akhunov E, Fahima T, Dubcovsky J. 2011. Effect of the down-regulation of the high *grain protein content*

(*GPC*) genes on the wheat transcriptome during monocarpic senescence. *BMC Genomics* 12: 492.
Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J. 2003. Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). *Theoretical and Applied Genetics* 107: 931-939.
Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Haselkorn R, Gornicki P. 2008. *Acc* homoeoloci and the evolution of wheat genomes. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9691-9696.
Chee PW, Elias EM, Anderson JA, Kianian SF. 2001. Evaluation of a high grain protein QTL from *Triticum turgidum L.* var. *dicoccoides* in an adapted durum wheat background. *Crop Science* 41: 295-301.
Choi EY, Graham R, Stangoulis J. 2007. Semi-quantitative analysis for selecting Fe- and Zn-dense genotypes of staple food crops. *Journal of Food Composition and Analysis* 20: 496-505.
Christopher JT, Manschadi AM, Hammer GL, Borrell AK. 2008. Developmental and physiological traits associated with high yield and stay-green phenotype in wheat. *Australian Journal of Agricultural Research* 59: 354-364.
Conklin PL, Barth C. 2004. Ascorbic acid, a familiar small molecule intertwined in the response of plants to ozone, pathogens, and the onset of senescence. *Plant, Cell & Environment* 27: 959-970.
Cusanovich DA, Pavlovic B, Pritchard JK, Gilad Y. 2014. The functional consequences of variation in transcription factor binding. *PLoS Genet* 10: e1004226.
Davies PJ, Gan S. 2012. Towards an integrated view of monocarpic plant senescence. *Russian Journal of Plant Physiology* 59: 467-478.
Denyer K, Dunlap F, Thorbjornsen T, Keeling P, Smith AM. 1996. The major form of ADP-glucose pyrophosphorylase in maize endosperm is extra-plastidial. *Plant Physiology* 112: 779-785.
DePauw RM, Townley-Smith TF, Humphreys G, Knox RE, Clarke FR, Clarke JM. 2005. Lillian hard red spring wheat. *Canadian Journal of Plant Science* 85: 397-401.
DePauw RM, et al. 2011. New breeding tools impact Canadian commercial farmer fields. *Czech Journal Genetic Plant Breeding* 47: S28-S34.
Derkx AP, Orford S, Griffiths S, Foulkes MJ, Hawkesford MJ. 2012. Identification of differentially senescing mutants of wheat and impacts on yield, biomass and nitrogen partitioning. *J Integr Plant Biol* 54: 555-566.
Dimmock PRE, Gooding MJ. 2002. The effects of fungicides on rate and duration of grain filling in winter wheat in relation to maintenance of flag leaf green area. *The Journal of Agricultural Science* 138: 1-16.
Distelfeld A, Avni R, Fischer AM. 2014. Senescence, nutrient remobilization, and yield in wheat and barley. *Journal of Experimental Botany*.
Distelfeld A, Uauy C, Fahima T, Dubcovsky J. 2006. Physical map of the wheat high-grain protein content gene *Gpc-B1* and development of a high-throughput molecular marker. *New Phytologist* 169: 753-763.
Distelfeld A, Uauy C, Olmos S, Schlatter AR, Dubcovsky J, Fahima T. 2004. Microcolinearity between a 2-cM region encompassing the grain protein content locus *Gpc-6B1* on wheat chromosome 6B and a 350-kb region on rice chromosome 2. *Functional & Integrative Genomics* 4: 59-66.
Distelfeld A, Korol A, Dubcovsky J, Uauy C, Blake T, Fahima T. 2008. Colinearity between the barley grain protein content (GPC) QTL on chromosome arm 6HS and the wheat *Gpc-B1* region. *Molecular Breeding* 22: 25-38.

Distelfeld A, Cakmak I, Peleg Z, Ozturk L, Yazici AM, Budak H, Saranga Y, Fahima T. 2007. Multiple QTL-effects of wheat *Gpc-B1* locus on grain protein and micronutrient concentrations. *Physiologia Plantarum* 129: 635-643.
Distelfeld A, Pearce S, Avni R, Scherer B, Uauy C, Piston F, Slade A, Zhao R, Dubcovsky J. 2012. Divergent functions of orthologous NAC transcription factors in wheat and rice. *Plant Molecular Biology* 78: 515-524.
Donovan GR, Lee JW, Hill RD. 1977. Compositional changes in the developing grain of high and low protein wheats. II. Starch and protein synthetic capacity. *Cereal Chemistry* 54: 646-656.
Dubcovsky J, Dvorak J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316: 1862-1866.
Dupont FM, Altenbach SB. 2003. Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *Journal of Cereal Science* 38: 133-146.
Duval M, Hsieh TF, Kim SY, Thomas TL. 2002. Molecular characterization of *AtNAM*: a member of the *Arabidopsis* NAC domain superfamily. *Plant Molecular Biology* 50: 237-248.
Dvorak J, Luo MC, Yang ZL, Zhang HB. 1998. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. *Theoretical and Applied Genetics* 97: 657-670.
Emes MJ, Bowsher CG, Hedley C, Burrell MM, Scrase-Field ESF, Tetlow IJ. 2003. Starch synthesis and carbon partitioning in developing endosperm. *Journal of Experimental Botany* 54: 569-575.
Erenoglu EB, Kutman UB, Ceylan Y, Yildiz B, Cakmak I. 2011. Improved nitrogen nutrition enhances root uptake, root-to-shoot translocation and remobilization of zinc (65Zn) in wheat. *New Phytologist* 189: 438-448.
Ernst HA, Olsen AN, Larsen S, Lo Leggio L. 2004. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO Rep* 5: 297-303.
Espinoza C, Medina C, Somerville S, Arce-Johnson P. 2007. Senescence-associated genes induced during compatible viral interactions with grapevine and *Arabidopsis*. *Journal of Experimental Botany* 58: 3197-3212.
FAO. 2009. Global agriculture towards 2050.
Farquhar GD, von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta* 149: 78-90.
Feild TS, Lee DW, Holbrook NM. 2001. Why leaves turn red in autumn. The role of anthocyanins in senescing leaves of red-ossier dogwood. *Plant Physiology* 127: 566-574.
Feldman M. 2001. The origin of cultivated wheat. 3-56. The wheat book: A history of wheat breeding. Paris: Lavoisier Publishing.
Feller U, Herren T, Holzer R. 1992. Nutrient accumulation in wheat grains: selective transfer of solutes from the xylem to the phloem in the stem. Wellesbourne, UK: European Society for Agronomy.
Fischer AM. 2012. The complex regulation of senescence. *Critical Reviews in Plant Sciences* 31: 124-147.
Foulkes MJ, Hawkesford MJ, Barraclough PB, Holdsworth MJ, Kerr S, Kightley S, Shewry PR. 2009. Identifying traits to improve the nitrogen economy of wheat: Recent advances and future prospects. *Field Crops Research* 114: 329-342.
Frith MC, Saunders NFW, Kobe B, Bailey TL. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* 4: e1000071.
Gaju O, Allard V, Martre P, Le Gouis J, Moreau D, Bogard M, Hubbart S, Foulkes MJ. 2014. Nitrogen partitioning and remobilization in relation to leaf

senescence, grain yield and grain nitrogen concentration in wheat cultivars. *Field Crops Research* 155: 213-223.
Garnett TP, Graham RD. 2005. Distribution and remobilization of iron and copper in wheat. *Annals of Botany* 95: 817-826.
Gong YH, Zhang J, Gao JF, Lu JY, Wang JR. 2005. Slow export of photoassimilate from stay-green leaves during late grain-filling stage in hybrid winter wheat (*Triticum aestivum* L.). *Journal of Agronomy and Crop Science* 191: 292-299.
Gouget A, Senchou V, Govers F, Sanson A, Barre A, Rougé P, Pont-Lezica R, Canut H. 2006. Lectin receptor kinases participate in protein-protein interactions to mediate plasma membrane-cell wall adhesions in *Arabidopsis*. *Plant Physiology* 140: 81-90.
Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech* 29: 644-652.
Grant EH, Fujino T, Beers EP, Brunner AM. 2010. Characterization of NAC domain transcription factors implicated in control of vascular cell differentiation in *Arabidopsis* and *Populus*. *Planta* 232: 337-352.
Gregersen PL, Holm PB. 2007. Transcriptome analysis of senescence in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal* 5: 192-206.
Guo YF, Gan SS. 2006. *AtNAP*, a NAC family transcription factor, has an important role in leaf senescence. *Plant Journal* 46: 601-612.
Guo YF, Gan SS. 2014. Translational researches on leaf senescence for enhancing plant productivity and quality. *Journal of Experimental Botany* 65: 3901-3913.
Guttieri MJ, Stein RJ, Waters BM. 2013. Nutrient partitioning and grain yield of *TaNAM*-RNAi wheat under abiotic stress. *Plant and Soil* 371: 573-591.
Haider N. 2013. The origin of the B-genome of bread wheat (*Triticum aestivum* L.). *Genetika* 49: 303-314.
Hambor JE, Mennone J, Coon ME, Hanke JH, Kavathas P. 1993. Identification and characterization of an Alu-containing, T-cell specific enhancer located in the last intron of the human *CD8 alpha* gene. *Molecular and Cellular Biology* 13: 7056-7070.
Hands P, Kourmpetli S, Sharples D, Harris RG, Drea S. 2012. Analysis of grain characters in temperate grasses reveals distinctive patterns of endosperm organization associated with grain shape. *J Exp Bot* 63: 6253-6266.
Hargreaves JA, Aprees T. 1988. Turnover of starch and sucrose in roots of *Pisum sativa*. *Phytochemistry* 27: 1627-1629.
Harris K, Subudhi P, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, Mullet J. 2007. Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. *Journal of Experimental Botany* 58: 327-338.
Hawker J, Jenner C. 1993. High temperature affects the activity of enzymes in the committed pathway of starch synthesis in developing wheat endosperm. *Functional Plant Biology* 20: 197-209.
Hay RKM. 1995. Harvest index: a review of its use in plant breeding and crop physiology. *Annals of Applied Biology* 126: 197-216.
He XJ, Mu RL, Cao WH, Zhang ZG, Zhang JS, Chen SY. 2005. *AtNAC2*, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant Journal* 44: 903-916.
Hedden P. 2003. The genes of the Green Revolution. *Trends Genet* 19: 5-9.
Herren T, Feller URS. 1997. Transport of cadmium via xylem and phloem in maturing wheat shoots: comparison with the translocation of zinc, strontium and rubidium. *Annals of Botany* 80: 623-628.
Heun M, Schäfer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F. 1997. Site of Einkorn wheat domestication identified by DNA fingerprinting. *Science* 278: 1312-1314.

Hocking PJ. 1994. Dry-matter production, mineral nutrient concentrations, and nutrient distribution and redistribution in irrigated spring wheat. *Journal of Plant Nutrition* 17: 1289-1308.
Hörtensteiner S. 2009. Stay-green regulates chlorophyll and chlorophyll-binding protein degradation during senescence. *Trends in Plant Science* 14: 155-162.
Hörtensteiner S, Feller U. 2002. Nitrogen metabolism and remobilization during senescence. *Journal of Experimental Botany* 53: 927-937.
Housley TL, Kirleis AW, Ohm HW, Patterson FL. 1981. An evaluation of seed growth in soft red winter wheat. *Canadian Journal of Plant Science* 61: 525-534.
IWGSC. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345.
Jamar C, Loffet F, Frettinger P, Ramsay L, Fauconnier M-L, du Jardin P. 2010. *NAM-1* gene polymorphism and grain protein content in *Hordeum*. *Journal of Plant Physiology* 167: 497-501.
James MG, Denyer K, Myers AM. 2003. Starch synthesis in the cereal endosperm. *Current Opinion in Plant Biology* 6: 215-222.
Jantasuriyarat C, Vales MI, Watson CJW, Riera-Lizarazu O. 2004. Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 108: 261-273.
Jenner C. 1979. Grain-filling in wheat plants shaded for brief periods after anthesis. *Functional Plant Biology* 6: 629-641.
Jenner C, Denyer K, Hawker J. 1994. Caution on the use of the generally accepted methanol precipitation technique for the assay of soluble starch synthase in crude extracts of plant tissues. *Functional Plant Biology* 21: 17-22.
Jenner CF, Rathjen AJ. 1972. Factors limiting the supply of sucrose to the developing wheat grain. *Annals of Botany* 36: 729-741.
Jensen MK, Kjaersgaard T, Nielsen MM, Galberg P, Petersen K, O'Shea C, Skriver K. 2010. The *Arabidopsis thaliana* NAC transcription factor family: structure–function relationships and determinants of *ANAC019* stress signalling. *Biochemical Journal* 426: 183-196.
Jeong J, Park Y, Jung H, Park S-H, Kim J-K. 2009. Rice NAC proteins act as homodimers and heterodimers. *Plant Biotechnology Reports* 3: 127-134.
Jia JZ, et al. 2013. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496: 91-95.
Jiang H, Li M, Liang N, Yan H, Wei Y, Xu X, Liu J, Xu Z, Chen F, Wu G. 2007. Molecular cloning and function analysis of the stay green gene in rice. *The Plant Journal* 52: 197-209.
Jolliffe PA, Tregunna EB. 1968. Effect of temperature, CO_2 concentration, and light intensity on oxygen inhibition of photosynthesis in wheat leaves. *Plant Physiology* 43: 902-906.
Joppa LR, Du C, Hart GE, Hareland GA. 1997. Mapping gene(s) for grain protein in tetraploid Wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. *Crop Sci.* 37: 1586-1589.
Jordan DR, Hunt CH, Cruickshank AW, Borrell AK, Henzell RG. 2012. The relationship between the stay-green trait and grain yield in elite sorghum hybrids grown in a range of environments. *Crop Sci.* 52: 1153-1161.
Jukanti AK, Fischer AM. 2008. A high-grain protein content locus on barley (*Hordeum vulgare*) chromosome 6 is associated with increased flag leaf proteolysis and nitrogen remobilization. *Physiologia Plantarum* 132: 426-439.
Jukanti AK, Heidlebaugh NM, Parrott DL, Fischer IA, McInnerney K, Fischer AM. 2008. Comparative transcriptome profiling of near-isogenic barley (*Hordeum vulgare*) lines differing in the allelic state of a major grain protein content locus

identifies genes with possible roles in leaf senescence and nitrogen reallocation. *New Phytologist* 177: 333-349.
Kade M, Barneix AJ, Olmos S, Dubcovsky J. 2005. Nitrogen uptake and remobilization in tetraploid 'Langdon' durum wheat and a recombinant substitution line with the high grain protein gene *Gpc-B1*. *Plant Breeding* 124: 343-349.
Kaufmann K, Muino JM, Osteras M, Farinelli L, Krajewski P, Angenent GC. 2010. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nature Protocols* 5: 457-472.
Keeling PL, Bacon PJ, Holt DC. 1993. Elevated temperature reduces starch deposition in wheat endosperm by reducing the activity of soluble starch synthase. *Planta* 191: 342-348.
Khush GS. 2001. Green revolution: the way forward. *Nat Rev Genet* 2: 815-822.
Kichey T, Hirel B, Heumez E, Dubois F, Le Gouis J. 2007. In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. *Field Crops Research* 102: 22-32.
Kihara H. 1944. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. *Agric Hortic* 19: 13-14.
Kim B-H, Cai X, Vaughn J, von Arnim A. 2007. On the functions of the h subunit of eukaryotic initiation factor 3 in late stages of translation initiation. *Genome Biology* 8: R60.
Kim Y-S, Kim S-G, Park J-E, Park H-Y, Lim M-H, Chua N-H, Park C-M. 2006. A membrane-bound NAC transcription factor regulates cell civision in *Arabidopsis*. *The Plant Cell Online* 18: 3132-3144.
Kitajima M, Butler WL. 1975. Quenching of chlorophyll fluorescence and primary photochemistry in chloroplasts by dibromothymoquinone. *Biochim Biophys Acta* 376: 105-115.
Krasileva K, et al. 2013. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology* 14: R66.
Kristiansson E, Thorsen M, Tamás MJ, Nerman O. 2009. Evolutionary forces act on promoter length: identification of enriched cis-regulatory elements. *Molecular Biology and Evolution* 26: 1299-1307.
Kühbauch W, Thome U. 1989. Nonstructural carbohydrates of wheat stems as influenced by sink-source manipulations. *Journal of Plant Physiology* 134: 243-250.
Kumar J, et al. 2011. Introgression of a major gene for high grain protein content in some Indian bread wheat cultivars. *Field Crops Research* 123: 226-233.
Kunieda T, Mitsuda N, Ohme-Takagi M, Takeda S, Aida M, Tasaka M, Kondo M, Nishimura M, Hara-Nishimura I. 2008. NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in *Arabidopsis*. *Plant Cell* 20: 2631-2642.
Kuo J, O'Brien TP, Canny MJ. 1974. Pit-field distribution, plasmodesmatal frequency, and assimilate flux in the mestome sheath cells of wheat leaves. *Planta* 121: 97-118.
Landt SG, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* 22: 1813-1831.
Lawlor DW. 1987. Photosynthesis: metabolism, control and physiology. Harlow, UK: Longman Scientific and Technical.
Lee IC, Hong SW, Whang SS, Lim PO, Nam HG, Koo JC. 2011. Age-dependent action of an ABA-inducible receptor kinase, *RPK1*, as a positive regulator of senescence in *Arabidopsis* leaves. *Plant and Cell Physiology* 52: 651-662.

Leegood RC. 2008. Roles of the bundle sheath cells in leaves of C3 plants. *Journal of Experimental Botany* 59: 1663-1673.
Li G, et al. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148: 84-98.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754-1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079.
Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. 1752-1779.
Lindemose S, Jensen MK, de Velde JV, O'Shea C, Heyndrickx KS, Workman CT, Vandepoele K, Skriver K, Masi FD. 2014. A DNA-binding-site landscape and regulatory network analysis for NAC transcription factors in *Arabidopsis thaliana*. *Nucleic Acids Research* 42: 7681-7693.
Ling HQ, et al. 2013. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496: 87-90.
Loladze I. 2014. Hidden shift of the ionome of plants exposed to elevated CO_2 depletes minerals at the base of human nutrition.
Loneragan J. 1981. Distribution and movement of copper in plants. 165–188 in JF Loneragan AR, RD Graham, ed. Copper in soils and plants. Sydney: Academic Press.
Lopes MS, Reynolds MP. 2012. Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. *Journal of Experimental Botany*.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2.
Loza-Tavera H, Martínez-Barajas E, Sánchez-de-Jiménez E. 1990. Regulation of ribulose-1,5-bisphosphate carboxylase expression in second leaves of maize seedlings from low and high yield populations. *Plant Physiology* 93: 541-548.
Luo MC, Yang ZL, You FM, Kawahara T, Waines JG, Dvorak J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theoretical and Applied Genetics* 114: 947-959.
Luo PG, Zhang HY, Shu K, Wu XH, Zhang HQ, Ren ZL. 2009. The physiological genetic effects of 1BL/1RS translocated chromosome in "stay green" wheat cultivar CN17. *Canadian Journal of Plant Science* 89: 1-10.
Ma S, Bachan S, Porto M, Bohnert HJ, Snyder M, Dinesh-Kumar SP. 2012. Discovery of stress responsive DNA regulatory motifs in *Arabidopsis*. *PLoS ONE* 7: e43198.
Marcussen T, et al. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345.
Martin P. 1982. Stem xylem as a possible pathway for mineral retranslocation from senescing leaves to the ear in wheat. *Functional Plant Biology* 9: 197-207.
Martre P, Porter JR, Jamieson PD, Triboï E. 2003. Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. *Plant Physiology* 133: 1959-1967.
Mascher M, et al. 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *The Plant Journal* 76: 718-727.
Masclaux C, QuillerÉ I, Gallais A, Hirel B. 2001. The challenge of remobilisation in plant nitrogen economy. A survey of physio-agronomic and molecular approaches. *Annals of Applied Biology* 138: 69-81.
Maxwell K, Johnson GN. 2000. Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany* 51: 659-668.

Maydup ML, Antonietta M, Guiamet JJ, Graciano C, López JR, Tambussi EA. 2010. The contribution of ear photosynthesis to grain filling in bread wheat (*Triticum aestivum* L.). *Field Crops Research* 119: 48-58.
Mayer KFX, et al. 2011. Unlocking the barley genome by chromosomal and comparative genomics. *The Plant Cell Online* 23: 1249-1263.
McClean P. 1998. Nucleic acid hybridizations. (19/09/2014; http://www.ndsu.edu/pubweb/~mcclean/plsc731/dna/dna6.htm)
McCormick AJ, Cramer MD, Watt DA. 2006. Sink strength regulates photosynthesis in sugarcane. *New Phytologist* 171: 759-770.
McMaugh SJ, et al. 2014. Suppression of starch synthase I expression affects the granule morphology and granule size and fine structure of starch in wheat endosperm. *Journal of Experimental Botany*.
Merewitz EB, Gianfagna T, Huang B. 2011. Protein accumulation in leaves and roots associated with improved drought tolerance in creeping bentgrass expressing an *ipt* gene for cytokinin synthesis. *Journal of Experimental Botany*.
Mesfin A, Frohberg RC, Anderson JA. 1999. RFLP markers associated with high grain protein from *Triticum turgidum* L. var. *dicoccoides* introgressed into hard red spring wheat. *Crop Science* 39: 508-513.
Meyer FD, Smidansky ED, Beecher B, Greene TW, Giroux MJ. 2004. The maize *Sh2r6hs* ADP-glucose pyrophosphorylase (AGP) large subunit confers enhanced AGP properties in transgenic wheat (*Triticum aestivum*). *Plant Science* 167: 899-911.
Meyer K, Cusumano JC, Somerville C, Chapple CC. 1996. Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases. *Proceedings of the National Academy of Sciences* 93: 6869-6874.
Mochida K, Uehara-Yamaguchi Y, Takahashi F, Yoshida T, Sakurai T, Shinozaki K. 2013. Large-scale collection and analysis of full-length cDNAs from *Brachypodium distachyon* and integration with *Pooideae s*equence resources. *PLoS ONE* 8: e75265.
Morita K. 1980. Release of nitrogen From chloroplasts during leaf senescence in rice (*Oryza sativa* L.). *Annals of Botany* 46: 297-302.
Nalam VJ, Vales MI, Watson CJ, Kianian SF, Riera-Lizarazu O. 2006. Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum* L.). *Theoretical and applied genetics* 112: 373-381.
Nelson OE, Rines HW. 1962. The enzymatic deficiency in the *waxy* mutant of maize. *Biochemical and Biophysical Research Communications* 9: 297-300.
Neph S, et al. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28: 1919-1920.
Nuruzzaman M, Sharoni AM, Kikuchi S. 2013. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Frontiers in Microbiology* 4.
Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S. 2010. Genome-wide analysis of NAC transcription factor family in rice. *Gene* 465: 30-44.
O'Brien T, Kuo J. 1975. Development of the suberized lamella in the mestome sheath of wheat leaves. *Australian Journal of Botany* 23: 783-794.
Olmos S, Distelfeld A, Chicaiza O, Schlatter AR, Fahima T, Echenique V, Dubcovsky J. 2003. Precise mapping of a locus affecting grain protein content in durum wheat. *Theoretical and Applied Genetics* 107: 1243-1251.
Olsen AN, Ernst HA, Leggio LL, Skriver K. 2005. NAC transcription factors: structurally distinct, functionally diverse. *Trends in Plant Science* 10: 79-87.
Olsen AN, Ernst HA, Lo Leggio L, Johansson E, Larsen S, Skriver K. 2004. Preliminary crystallographic analysis of the NAC domain of ANAC, a member of the

plant-specific NAC transcription factor family. *Acta Crystallogr D Biol Crystallogr* 60: 112-115.
Olsen OA. 2001. Endosperm development: cellularization and cell fate specification. *Annu Rev Plant Physiol Plant Mol Biol* 52: 233-267.
Ougham H, Thomas H, Archetti M. 2008. The adaptive value of leaf colour. *New Phytologist* 179: 9-13.
Ozturk L, Altintas G, Erdem H, Gokmen OO, Yazici A, Cakmak I. 2009. Localization of iron, zinc, and protein in seeds of spelt (Triticum aestivum ssp. spelta) genotypes with low and high protein concentration. Paper presented at The Proceedings of the International Plant Nutrition Colloquium XVI, California.
Ozturk L, Yazici MA, Yucel C, Torun A, Cekic C, Bagci A, Ozkan H, Braun H-J, Sayers Z, Cakmak I. 2006. Concentration and localization of zinc during seed development and germination in wheat. *Physiologia Plantarum* 128: 144-152.
Pajoro A, et al. 2014. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biology* 15: R41.
Palta J, Fillery I. 1995. N application enhances remobilization and reduces losses of pre-anthesis N in wheat grown on a duplex soil. *Australian Journal of Agricultural Research* 46: 519-531.
Paltridge NG, Milham PJ, Ortiz-Monasterio JI, Velu G, Yasmin Z, Palmer LJ, Guild GE, Stangoulis JCR. 2012. Energy-dispersive X-ray fluorescence spectrometry as a tool for zinc, iron and selenium analysis in whole grain wheat. *Plant and Soil* 361: 261-269.
Parrott DL, McInnerney K, Feller U, Fischer AM. 2007. Steam-girdling of barley (*Hordeum vulgare*) leaves leads to carbohydrate accumulation and accelerated leaf senescence, facilitating transcriptomic analysis of senescence-associated genes. *New Phytologist* 176: 56-69.
Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America* 101: 9903-9908.
Patrick JW, Offler CE. 2001. Compartmentation of transport and transfer events in developing seeds. *Journal of Experimental Botany* 52: 551-564.
Paux E, et al. 2008. A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322: 101-104.
Pearson JN, Rengel Z, Jenner CF, Graham RD. 1995. Transport of zinc and manganese to developing wheat grains. *Physiologia Plantarum* 95: 449-455.
Penfold CA, Buchanan-Wollaston V, Denby KJ, Wild DL. 2012. Nonparametric Bayesian inference for perturbed and orthologous gene regulatory networks. *Bioinformatics* 28: i233-i241.
Perchorowicz JT, Raynes DA, Jensen RG. 1981. Light limitation of photosynthesis and activation of ribulose bisphosphate carboxylase in wheat seedlings. *Proceedings of the National Academy of Sciences* 78: 2985-2989.
Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* 276: 73-81.
Polak P, Domany E. 2006. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7: 133.
Pottier M, Masclaux Daubresse C, Yoshimoto K, Thomine S. 2014. Autophagy as a possible mechanism for micronutrient remobilization from leaves to seeds. *Frontiers in Plant Science* 5.
Puranik S, Sahu PP, Srivastava PS, Prasad M. 2012. NAC proteins: regulation and role in stress tolerance. *Trends in Plant Science* 17: 369-381.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. *Nature* 457: 843-848.
Rathjen JR, Strounina EV, Mares DJ. 2009. Water movement into dormant and non-dormant wheat (*Triticum aestivum* L.) grains. *Journal of Experimental Botany* 60: 1619-1631.
Ray DK, Mueller ND, West PC, Foley JA. 2013. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* 8: e66428.
Rellán-Álvarez R, Giner-Martínez-Sierra J, Orduna J, Orera I, Rodríguez-Castrillón JÁ, García-Alonso JI, Abadía J, Álvarez-Fernández A. 2010. Identification of a tri-iron(III), tri-citrate complex in the xylem sap of iron-deficient tomato resupplied with iron: new insights into plant iron long-distance transport. *Plant and Cell Physiology* 51: 91-102.
Ricachenevsky FK, Menguer PK, Sperotto RA. 2013. kNACking on heaven's door: how important are NAC transcription factors for leaf senescence and Fe/Zn remobilization to seeds? *Frontiers in Plant Science* 4.
Ricardi M, et al. 2014. Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato *ASR1*, a drought stress-responsive transcription factor. *BMC Plant Biology* 14: 29.
Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotech* 29: 24-26.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.
Roccaro M, Somssich IE. 2011. Chromatin immunoprecipitation to identify global targets of WRKY transcription factor family members involved in plant immunity. 45-58 in McDowell JM, ed. Plant Immunity: Methods and Protocols, vol. 712. Totowa: Humana Press Inc.
Rösti S, Rudi H, Rudi K, Opsahl-Sorteberg H-G, Fahy B, Denyer K. 2006. The gene encoding the cytosolic small subunit of ADP-glucose pyrophosphorylase in barley endosperm also encodes the major plastidial small subunit in the leaves. *Journal of Experimental Botany* 57: 3619-3626.
Sablowski RWM, Meyerowitz EM. 1998. A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*. *Cell* 92: 93-103.
Safar J, et al. 2004. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant Journal* 39: 960-968.
Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet* 3: 429-441.
Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C. 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell Online* 20: 11-24.
Sanchez-Bragado R, Molero G, Reynolds MP, Araus JL. 2014. Relative contribution of shoot and ear photosynthesis to grain filling in wheat under good agronomical conditions assessed by differential organ $\delta^{13}C$. *Journal of Experimental Botany*.
Schnyder H. 1993. The role of carbohydrate storage and redistribution in the source-sink relations of wheat and barley during grain filling — a review. *New Phytologist* 123: 233-245.
Scofield GN, Ruuska SA, Aoki N, Lewis DC, Tabe LM, Jenkins CLD. 2009. Starch storage in the stems of wheat plants: localization and temporal changes. *Annals of Botany* 103: 859-868.

Serrago RA, Alzueta I, Savin R, Slafer GA. 2013. Understanding grain yield responses to source–sink ratios during grain filling in wheat and barley under contrasting environments. *Field Crops Research* 150: 42-51.
Shamimuzzaman M, Vodkin L. 2013. Genome-wide identification of binding sites for NAC and YABBY transcription factors and co-regulated genes during soybean seedling development by ChIP-Seq and RNA-Seq. *BMC Genomics* 14: 477.
Sharov AA, Ko MSH. 2009. Exhaustive search for over-represented DNA sequence motifs with CisFinder. *DNA Research* 16: 261-273.
Shewry PR. 2009. Wheat. *Journal of Experimental Botany* 60: 1537-1553.
Shewry PR, Mitchell RAC, Tosi P, Wan Y, Underwood C, Lovegrove A, Freeman J, Toole GA, Mills ENC, Ward JL. 2012. An integrated study of grain development of wheat (cv. Hereward). *Journal of Cereal Science* 56: 21-30.
Shikanai T, Muller-Moule P, Munekage Y, Niyogi KK, Pilon M. 2003. *PAA1*, a P-type ATPase of Arabidopsis, functions in copper transport in chloroplasts. *Plant Cell* 15: 1333-1346.
Shure M, Wessler S, Fedoroff N. 1983. Molecular identification and isolation of the *Waxy* locus in maize. *Cell* 35: 225-233.
Simeoni F, Divita G. 2007. The Dim protein family: from structure to splicing. *Cellular and Molecular Life Sciences* 64: 2079-2089.
Simmonds NW. 1995. The relation between yield and protein in cereal grain. *Journal of the Science of Food and Agriculture* 67: 309-315.
Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai Y-S, Gill BS, Faris JD. 2006. Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172: 547-555.
Simpson RJ, Lambers H, Dalling MJ. 1983. Nitrogen redistribution during grain growth in wheat (*Triticum aestivum* L.) : IV. Development of a quantitative model of the translocation of nitrogen to the grain. *Plant Physiology* 71: 7-14.
Sinclair TR, de Wit CT. 1975. Photosynthate and nitrogen requirements for seed production by various crops. *Science* 189: 565-567.
Slafer GA, Savin R. 1994. Source—sink relationships and grain mass at different positions within the spike in wheat. *Field Crops Research* 37: 39-49.
Slafer GA, Andrade FH, Feingold SE. 1990. Genetic improvement of bread wheat (*Triticum aestivum* L.) in Argentina: relationships between nitrogen and dry matter. *Euphytica* 50: 63-71.
Smidansky ED, Martin JM, Hannah LC, Fischer AM, Giroux MJ. 2003. Seed yield and plant biomass increases in rice are conferred by deregulation of endosperm ADP-glucose pyrophosphorylase. *Planta* 216: 656-664.
Smith EL. 1938. Limiting factors in photosynthesis: light and carbon dioxide. *The Journal of general physiology* 22: 21-35.
Spano G, Di Fonzo N, Perrotta C, Platani C, Ronga G, Lawlor DW, Napier JA, Shewry PR. 2003. Physiological characterization of 'stay green' mutants in durum wheat. *Journal of Experimental Botany* 54: 1415-1420.
Sperotto RA, Ricachenevsky FK, Duarte GL, Boff T, Lopes KL, Sperb ER, Grusak MA, Fett JP. 2009. Identification of up-regulated genes in flag leaves during rice grain filling and characterization of *OsNAC5*, a new ABA-dependent transcription factor. *Planta* 230: 985-1002.
Stitt M. 1991. Rising CO_2 levels and their potential significance for carbon flow in photosynthetic cells. *Plant, Cell & Environment* 14: 741-762.
Strother S. 1988. The role of free radicals in leaf senescence. *Gerontology* 34: 151-156.
Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6: e21800.

Sweetlove LJ, Hill SA. 2000. Source metabolism dominates the control of source to sink carbon flux in tuberizing potato plants throughout the diurnal cycle and under a range of environmental conditions. *Plant, Cell & Environment* 23: 523-529.
Tabbita F, Lewis S, Vouilloz JP, Ortega MA, Kade M, Abbate PE, Barneix AJ. 2013. Effects of the *Gpc-B1* locus on high grain protein content introgressed into Argentinean wheat germplasm. *Plant Breeding* 132: 48-52.
Thomas H, Stoddart JL. 1975. Separation of chlorophyll degradation from other senescence processes in leaves of a mutant genotype of meadow fescue (*Festuca pratensis* L.). *Plant Physiology* 56: 438-441.
Thomas H, Howarth CJ. 2000. Five ways to stay green. *Journal of Experimental Botany* 51: 329-337.
Thomas H, Ougham H. 2014. The stay-green trait. *Journal of Experimental Botany* 65: 3889-3900.
Thorbjornsen T, Villand P, Denyer K, Olsen OA, Smith AM. 1996. Distinct isoforms of ADPglucose pyrophosphorylase occur inside and outside the amyloplasts in barley endosperm. *Plant Journal* 10: 243-250.
Thorne JH. 1985. Phloem unloading of C-assimilates and N-assimilates in developing seeds. *Annual Review of Plant Physiology and Plant Molecular Biology* 36: 317-343.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14: 178-192.
Tilman D, Balzer C, Hill J, Befort BL. 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* 108: 20260-20264.
Triboi E, Triboi-Blondel A-M. 2002. Productivity and grain or seed composition: a new approach to an old problem—invited paper. *European Journal of Agronomy* 16: 163-186.
Tuncel A, Okita TW. 2013. Improving starch yield in cereals by over-expression of ADPglucose pyrophosphorylase: Expectations and unanticipated outcomes. *Plant science : an international journal of experimental plant biology* 211.
Uauy C, Brevis JC, Dubcovsky J. 2006a. The high grain protein content gene *Gpc-B1* accelerates senescence and has pleiotropic effects on protein content in wheat. *Journal of Experimental Botany* 57: 2785-2794.
Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. 2006b. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314: 1298-1301.
Valdivia ER, Herrera MT, Gianzo C, Fidalgo J, Revilla G, Zarra I, Sampedro J. 2013. Regulation of secondary wall synthesis and cell death by NAC transcription factors in the monocot *Brachypodium distachyon*. *J Exp Bot* 64: 1333-1343.
Valluru R, Van den Ende W. 2008. Plant fructans in stress environments: emerging concepts and future prospects. *Journal of Experimental Botany* 59: 2905-2916.
Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3: research0034.0031 - research0034.0011.
Vijn I, Smeekens S. 1999. Fructan: more than a reserve carbohydrate? *Plant Physiology* 120: 351-360.
Vogt T. 2010. Phenylpropanoid biosynthesis. *Molecular Plant* 3: 2-20.
Voytas DF, Gao C. 2014. Precision genome engineering and agriculture: opportunities and regulatory challenges. *PLoS Biol* 12: e1001877.

Vrinten PL, Nakamura T. 2000. Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues. *Plant Physiol* 122: 255-264.
Wang X, Culver J. 2012. DNA binding specificity of ATAF2, a NAC domain transcription factor targeted for degradation by Tobacco mosaic virus. *Bmc Plant Biology* 12: 157.
Wang Z, Fu J, He M, Tian Q, Cao H. 1997. Effects of source/sink manipulation on net photosynthetic rate and photosynthate partitioning during grain filling in winter wheat. *Biologia Plantarum* 39: 379-385.
Wang Z, Chen X, Wang J, Liu T, Liu Y, Zhao L, Wang G. 2007. Increasing maize seed weight by enhancing the cytoplasmic ADP-glucose pyrophosphorylase activity in transgenic maize plants. *Plant Cell, Tissue and Organ Culture* 88: 83-92.
Waters BM, Uauy C, Dubcovsky J, Grusak MA. 2009. Wheat (*Triticum aestivum*) NAM proteins regulate the translocation of iron, zinc, and nitrogen compounds from vegetative tissues to grain. *Journal of Experimental Botany* 60: 4263-4274.
Welner DH, Lindemose S, Grossmann JG, Møllegaard NE, Olsen AN, Helgstrand C, Skriver K, Lo Leggio L. 2012. DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. *Biochemical Journal* 444: 395-404.
Wheat Initiative. 2012. National food security. (11/09/2014; http://www.fao.org/docs/eims/upload/306175/Briefing%20Paper%20%283%29-Wheat%20Initative%20-%20H%C3%A9l%C3%A8ne%20Lucas.pdf)
Wu X-Y, Kuai B-K, Jia J-Z, Jing H-C. 2012. Regulation of leaf senescence and crop genetic improvement. *Journal of Integrative Plant Biology* 54: 936-952.
Yamaguchi M, Kubo M, Fukuda H, Demura T. 2008. *VASCULAR-RELATED NAC-DOMAIN7* is involved in the differentiation of all types of xylem vessels in *Arabidopsis* roots and shoots. *The Plant Journal* 55: 652-664.
Yang JC, Zhang JH, Huang ZL, Zhu QS, Wang L. 2000. Remobilization of carbon reserves is improved by controlled soil-drying during grain filling of wheat. *Crop Science* 40: 1645-1655.
Yang JC, Zhang JH, Wang ZQ, Zhu QS, Liu LJ. 2003. Involvement of abscisic acid and cytokinins in the senescence and remobilization of carbon reserves in wheat subjected to water stress during grain filling. *Plant, Cell & Environment* 26: 1621-1631.
Yang JC, Zhang J, Wang Z, Xu G, Zhu Q. 2004. Activities of key enzymes in sucrose-to-starch conversion in wheat grains subjected to water deficit during grain filling. *Plant Physiology* 135: 1621-1629.
Yang Z, Ohlrogge JB. 2009. Turnover of fatty acids during natural senescence of *Arabidopsis*, *Brachypodium*, and switchgrass and in *Arabidopsis* β-oxidation mutants. *Plant Physiology* 150: 1981-1989.
Yant L, Mathieu J, Dinh TT, Ott F, Lanz C, Wollmann H, Chen X, Schmid M. 2010. Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor *APETALA2*. *The Plant Cell Online* 22: 2156-2170.
Yoshida K, et al. 2013. Engineering the *Oryza sativa* cell wall with rice NAC transcription factors regulating secondary wall formation. *Front Plant Sci* 4: 383.
Young TE, Meeley RB, Gallie DR. 2004. *ACC synthase* expression regulates leaf performance and drought tolerance in maize. *Plant Journal* 40: 813-825.
Zee SY, O'Brien TP. 1970. A special type of tracheary element associated with xylem-discontinuity in floral axis of wheat *Australian Journal of Biological Sciences* 23: 783-791.
Zhang K, Gan SS. 2012. An abscisic acid-*AtNAP* transcription factor-SAG113 protein phosphatase 2C regulatory chain for controlling dehydration in senescing *Arabidopsis* leaves. *Plant Physiol* 158: 961-969.

Zhang Y, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* 9: R137.
Zheng Y, Wang Z, Sun X, Jia A, Jiang G, Li Z. 2008. Higher salinity tolerance cultivars of winter wheat relieved senescence at reproductive stage. *Environmental and Experimental Botany* 62: 129-138.
Zhong R, Richardson EA, Ye Z-H. 2007. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *The Plant Cell Online* 19: 2776-2792.
Zhong RQ, Lee CH, Ye ZH. 2010. Global analysis of direct targets of secondary wall NAC master switches in *Arabidopsis*. *Molecular Plant* 3: 1087-1103.
Zhou Y, Huang W, Liu L, Chen T, Zhou F, Lin Y. 2013. Identification and functional characterization of a rice NAC gene involved in the regulation of leaf senescence. *BMC Plant Biology* 13: 132.
Zhu J-Y, Sun Y, Wang Z-Y. 2012. Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq). *Methods in molecular biology (Clifton, N.J.)* 876: 173-188.
Zhu QK, Zou JX, Zhu ML, Liu ZB, Feng PC, Fan GT, Wang WJ, Liao H. 2014. In silico analysis on structure and DNA binding mode of AtNAC1, a NAC transcription factor from *Arabidopsis thaliana*. *Journal of Molecular Modeling* 20.