Discovering genetic factors influencing cereal endosperm starch by exploiting natural variation in *Aegilops tauschii*

Rose McNelly

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

John Innes Centre

September 2024

Registration number 100316968

Abstract

Cereal endosperm starch is of great social and commercial importance, providing up to 50% of dietary calories globally, whilst also being used to produce paper and biodegradable polymers. Triticeae (e.g., wheat, barley and rye) grains have two types of starch granules: large lenticular A-type granules and small spherical B-type granules. Starch granule size distribution and composition influences nutritional and functional quality, but the mechanisms determining these remain poorly understood.

Here I aimed to discover genetic factors influencing starch granule size distributions in the Triticeae using natural variation in *Aegilops tauschii*, the wheat D-genome progenitor. I discovered significant variation in starch granule size distributions within a diversity panel of 117 *Ae. tauschii* accessions, which exceeded that observed in domesticated wheat. I exploited this variation in a genome wide association study, identifying ten novel genomic loci associated with B-type granules. By utilising publicly available datasets and bioinformatic predictions, I identified 13 gene candidates within these loci that might influence B-type granule formation. Interestingly, one of these encodes for Limit Dextrinase (LDA), a starch debranching enzyme. I discovered two LDA variants with increased enzymatic activity which are prevalent in *Ae. tauschii* accessions with large B-type granules. However, complete elimination of *LDA* in tetraploid wheat caused only minor effects on endosperm starch granules. Together, this suggests that gain-of-function mutations in *LDA* influence B-type granule formation, perhaps by affecting malto-oligosaccharide metabolism.

My work highlights the potential of *Ae. tauschii* to discover novel variation and new gene candidates involved in starch granule formation in the Triticeae, and reveals a role for LDA in influencing variation in starch granule traits. Further characterisation of the novel gene candidates will extend our mechanistic understanding of A-type and B-type granule formation. Furthermore, the loci could be integrated into breeding programs to improve grain quality by modulating starch granule size.

Table of Contents

List of abbreviations

3PGA – Glycerate-3-phosphate

ADP – Adenosine diphosphate

ADP-Glc – ADP-glucose

Ae – Amylose extender

AGPase – ADP-glucose pyrophosphorylase

ALD – Aldolase

ANOVA – Analysis of variance

ARC – Accumulation and replication of chloroplasts

BAC – Bacterial artificial chromosome

BE – Branching enzyme

BGC1 – B-granule content 1

bp – Base pairs

BSA – Bovine serum albumin

BT1 – Brittle 1

CBM – Carbohydrate binding module

cDNA – Complementary DNA

CIMMYT – International Maize and Wheat Improvement Centre

CMC4 – Curl mite colonisation 4

CNV – Copy number variation

cv. – Cultivar

DBE – Debranching enzyme

DNA – Deoxyribonucleic acid

DP – Degree of polymerisation

DPA – Days post anthesis

DPE2 – Disproportionating enzyme 2

DTT – Dithiothreitol

DUF – Domain of unknown function

EDTA – Ethylenediaminetetraacetic acid

ENY2 – Enhancer of yellow 2

ER – Endoplasmic reticulum

ESV1 – Early starvation 1

EWAS – Epigenome-wide association studies

F1,6BP – Fructose-1,6-bisphosphate

F6P – Fructose 6-phosphate

FBPase – Fructose-1,6-bisphosphatase

FK – Fructokinase

FLO6 – Floury endosperm 6

G1P – Glucose 1-phosphate

G6P – Glucose 6-phosphate

GBSS – Granule bound starch synthase

GCS1 – Generative Cell-Specific Protein 1

GH – Glycoside hydrolase

Glc – Glucose

GlcNAc – N-Acetylglucosamine

GRU – Germplasm Resource Unit

GT – Glycosyltransferase

GWAS – Genome wide association study

His – Histidine

Hz – Hertz

Indel – Insertion/deletion

IPTG – Isopropyl β-d-1-thiogalactopyranoside

ISA – Isoamylase

JIC – John Innes Centre

kDa – Kilo Dalton

KASP – Kompetitive allele specific PCR

L-L – log-normal - log-normal

L-N – log-normal - normal

LD – Linkage disequilibrium

LDA – Limit dextrinase

LDI – Limit dextrinase inhibitor

LDS – Lithium dodecyl sulfate

LESV – Like early starvation 1

MAGIC – Multi-parent advanced generation inter cross

Man – Mannose

Mb – Mega bases

MEX1 – Maltose exporter 1

MFP1 – Mar-binding filament protein 1

MOGS – Mannosyl-oligosaccharide glucosidase

MOPS – 3-[N-morpholino] propanesulfonic acid

MOS – Malto-oligosaccharides

MRC – Mysoin-resembling chloroplast protein

N-N – normal - normal

NEB – New England Biolabs

NIAB – National Institute of Agricultural Botany

NLR – Nucleotide binding/leucine-rich repeat

PAGE – Polyacrylamide gel electrophoresis

PARC6 – Paralog of ARC6

PCA – Principal component analysis

PCR – Polymerase chain reaction

PDB – Protein Data Bank

PGI – Phosphoglucose isomerase

PGM – Phosphoglucomutase

PHS1/PHO1 – Phosphorylase 1

Pi – Orthophosphate

PIN – Pin-formed

PPi – Inorganic pyrophosphate

PTST1 – Protein targeting to starch 1

PTST2 – Protein targeting to starch 2

PTST3 – Protein targeting to starch 3

PUL – Pullulanase

PullG6 – 4,6-O-Benzylidene-4-nitrophenyl-6^3-α-D-maltotriosyl-maltotriose

QTL – Quantitative trait loci

RAP – RNA-binding domain abundant in Apicomplexans

RNA – Ribonucleic acid

RNAi – RNA interference

RNAseq – RNA sequencing

RVA – Rapid viscoanalyser

S – Standard error of regression

SAGA – Spt-Ada-Gcn5 acetyltransferase

SBE – Starch branching enzyme

SBE – Starch branching enzyme

SDS – Sodium Dodecyl Sulfate

SDS–PAGE – Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SE – Standard error of the mean

SEM – Scanning electron microscopy

Ser – Serine

SHW – Synthetic hexaploid wheat

SNP – Single nucleotide polymorphism

SPPase – Sucrose phosphate phosphatase

SPSase – Sucrose-phosphate synthase

SS – Starch synthase

SSG4 – Substandard starch grain 4

SSG6 – Substandard starch grain 6

Sucrose–P – Sucrose-phosphate

SuSy – Sucrose synthase

TEMED – Tetramethylethylenediamine

TGW – Thousand grain weight

TILLING – Targeting-Induced Local Lesions In Genomes

TPM – Transcripts per million

TPT – Triose-phosphate/phosphate translocator

Triose–P – Triose-phosphate

TWAS – Transcriptome-wide association studies

UDP – Uridine diphosphate

UDP–Glc – UDP-glucose

UGPase – UDP-glucose pyrophosphorylase

UPGMA – Unweighted pair group method with arithmetic mean

UTR – Untranslated region

WT – Wild-type

List of figures

Main figures:

Supplementary figures:

List of tables

Main tables:

Supplementary tables:

Acknowledgements

My PhD has been an exciting, enjoyable and thoroughly rewarding experience. This is in part due to the people I have met, and the friends I have made along the way. Therefore, there are many people who I need to thank, as without them this thesis would not have been possible.

My biggest thanks must go to my supervisor David Seung who has been an enthusiastic, supportive and encouraging mentor. I remember first meeting David online via Zoom and even in this very first interaction he made me feel at ease and answered all my questions and concerns. This has continued throughout my PhD, there has never been anything which has too small or insignificant to ask him and he has always valued my concerns and ideas. I have always looked forward to our weekly meetings and each of these left me feeling excited and energised ready to start the next set of experiments or new analyses. In addition to the scientific knowledge I have learnt from David, he has also played a substantial part in my general career development, nominating me for opportunities and encouraging me to develop my skills beyond those needed in the lab. Overall, it has been a pleasure to be a PhD student in David's lab, and I would not have achieved as much as I have been able to without David's superb supervision.

I would also like to thank the remainder of my supervisory team who have constantly supported me and have been the cause of insightful and interesting discussions. Thanks to my secondary supervisor Cristobal Uauy whose knowledge on wheat, wheat genetics and GWAS have been immensely helpful. In fact, this thesis would not have been possible without Cristobal, as his lab kindly provided the *Aegilops tauschii* grain to start my project. Thanks to Claire Domoney who provided thoughtful suggestions and advice on my work and was still interested in my progress after she retired from my supervisory panel. I am grateful to both Alexander Watson-Lazowski and Burkhard Steuernagel who provided substantial support and guidance throughout my PhD, both in the form of helpful discussions during supervisory meetings and day-to-day help for aspects of my project.

A big thank you must go to the entire Seung lab, past and present, who warmly welcomed me into the group and quickly became my John Innes family. There have been so many laughs and smiles along the way, and I finish my PhD with loads of great memories including: celebrating birthdays with cake, visiting the seals, quizzes in the Rec and working together to complete the newspaper puzzles over lunch. A warm

thanks goes to Alexander Watson-Lazowski who was the first person I met in the Seung group and kindly supervised my rotation. Alex patiently taught me all the basic starch skills without which my project would not have been possible. Thanks to Brendan Fahy, who performed all the crossing to generate the *lda* mutants and has provided many an enjoyable conversation whilst sitting next to me in the office. On this note, thanks to Lara Esch who I had the joy of sitting next to for most of my PhD and always made me laugh and smile. Thanks to Qi Yang Ngai who taught me how to run native PAGE gels. Thanks to Nitin Kamble who is always helpful and supportive in the lab, and kindly provided extracted protein for some of my analysis. Thanks to Petros Zafeiriou who provided his method and top tips for the total starch assay, and then to Ruth Franklin who helped me to carry this out in the lab. Thanks to Sai Chennu for all his help with the *Ae. tauschii.* Thanks to Becca Testa for her efforts analysing the *lda* mutants. Thanks to Georgia Yiasoumi for her helpful feedback on the Python script for fitting distributions to Coulter counter traces. Thanks to Thomas Navarro, Ciara O'Brien and Farrukh Makhamadjonov for reading over their careful reading of this thesis. Additional thanks is given to Farrukh for teaching me how to express and purify proteins from *E. coli*.

Before joining the John Innes Centre, I had no experience with bioinformatics and was frankly quite scared by it. So, if someone had told me four years ago that I would end up conducting a genome wide association study (GWAS) as part of my PhD I would not have believed them. Luckily, I have been supported by many people who have patiently put up with my many questions and queries. I would like to thank Jesus Quiroz-Chavez, who introduced me to the GWAS pipeline; I spent many hours in his office as I grappled with running it for the first time and troubleshooting errors. Thanks to Burkhard Steuernagel, Sanu Arora and Max Jones who I have had many fruitful discussions about GWAS. And thanks to David Gilbert whose advice, support (and grains) have been invaluable.

I would also like to thank Amy Briffa who has been instrumental in developing the Python script for fitting distributions to Coulter counter traces. I enjoyed our countless conversations about this and would have had no idea where to start without her guidance and support.

I am extremely grateful for the John Innes Foundation for funding my PhD and would like to thank Ant Dodd for leading the John Innes Foundation Rotation Programme during my first year. It was strange, and scary, starting in 2020 but with the help of Ant

and my fellow rotation students – Emma, Max, Ruth and Guilia – I was able to settle into my PhD life. Thanks also to Janneke Balk and Myriam Charpentier for hosting me for rotations in their labs during my first year. In both rotations I was warmly welcomed into the group and learnt more than I would have thought possible during just ten weeks.

I would like to thank all the other friends, of which there are simply too many to mention, that I have met across the research park during my PhD. I am truly grateful to have met such kind, friendly and inspiring people and I will carry the memories we have made with me for life.

Finally, I want to thank the people who have supported and encouraged me during my PhD. Thanks to my boyfriend Ollie for his love and support, and an extra thank you for reading this thesis to check for typos. Most importantly thanks to my family – my parents, Julia and John, and my sisters, Kate and Alice. Through my PhD they have put up with me constantly talking about starch and wheat and have endured many a ‘Starch Trek’ presentation to support me. You four are my biggest support network and always encourage me to follow my dreams, so thank you for being ever so supportive and your unconditional love.

Chapter One – Introduction

1.1 Importance of starch

Starch is a polymer of glucose that is produced as semi-crystalline granules by plants, algae and some protozoa, including *Eimeria brunetti* (Ryley *et al.*, 1969) and the parasite *Toxoplasma gondii* (Guérardel *et al.*, 2005). The starch produced by plants is an important resource for humans, and the calories from starchy crops, most notably wheat (*Triticum aestivum*), rice (*Oryza sativa*) and potatoes (*Solanum tuberosum*), make up an estimated 50% of our daily calories (Huang *et al.*, 2021). Beyond food, starch has many uses in industries, including production of paper, textiles, adhesives, biofuels and biopolymers (Bertolini, 2009). Therefore, understanding starch synthesis in plants is a key priority to improve starch yield and quality for food and industrial uses.

During periods of light, green plants photosynthesize, fixing carbon in the atmosphere into sugars and starch. The sugars provide an immediate energy source, whilst starch is a storage carbohydrate for periods when photosynthesis is not possible. There are two major classes of starch based on their metabolic purpose – transitory starch and storage starch. Transitory starch is for short-term storage and is primarily found in leaves for use during dark periods. It is synthesized in chloroplasts during the day and degraded over the subsequent night. This process is highly regulated (Smith and Zeeman, 2020). Firstly, there are tight controls over how much photosynthate is used for transitory starch synthesis so that the plant can maximise carbon used for growth. In *Arabidopsis thaliana* approximately 80% of photosynthate is used for growth (Sulpice *et al.*, 2014). Secondly, the rate of nighttime degradation is carefully controlled such that the plant does not run out of energy before photosynthesis can recommence. This degradation rate is constant over the course of a night yet it is variable over different nights, such that if a plant experiences a shorter day and anticipates a longer night, the degradation rate will be slower to ensure starch lasts until dawn (Graf *et al.*, 2010). This tight regulation helps avoid starvation during the night and subsequent decreases in plant productivity and growth (Graf *et al.*, 2010).

Whilst transitory starch is important for short-term use, storage starch is a long-term energy storage molecule in tissues such as grains, pollen and tubers. It is synthesized in non-photosynthetic plastids called amyloplasts (Sadali *et al.*, 2019). Amyloplasts are distinct from chloroplasts as they lack the photosynthetic membranes and machinery. Storage starch is most abundant in grains and tubers, where it is utilised as an energy source during germination and sprouting. Defects in storage starch synthesis in maize

(*Zea mays*) (Tsai and Nelson, 1966), barley (*Hordeum vulgare*) (Morell *et al.*, 2003), wheat (Fahy *et al.*, 2022) and rice (Tang *et al.*, 2016) cause shrunken grains, demonstrating that starch is the major component of these grains. The amount of storage starch in these tissues of crops has been selected for during domestication. For example, the grain starch content of domesticated wheat, ~70% (Hannah, 2007), is higher than those of wild wheat relatives, ~42-55% (Brandolini *et al.*, 2015).

Other roles for starch include gravitropism in roots, whereby the sedimentation of starch-containing amyloplasts causes downstream repolarisation of PIN-FORMED (PIN) auxin transporters and differential growth (Chen *et al.*, 2023b), and in pollen, where it is essential for male fertility (Lee *et al.*, 2016).

1.2 Starch structure

At the molecular level, starch consists of two polymers – amylose and amylopectin. Amylopectin is the major component and is required for the formation of the granule matrix. Amylose is the minor constituent and makes up less than 10% of starch content in most leaf starches, and typically 20-30% in crop storage starches (Smith and Zeeman, 2020). Both amylose and amylopectin consist of chains of α-1,4-linked glucose subunits with α-1,6-linked branches, with the major difference being the frequency of branch points. In amylose, branch points occur infrequently, facilitating the formation of long, linear chains. Amylopectin has many more branch points resulting in a racemose-like structure. The distribution of branch points within amylopectin allows clusters of linear chains, with typically 12-15 glucose subunits each, to align. Neighbouring chains within the clusters form double helix structures, which interact with each other to create ordered crystalline arrays. Two different allomorphic crystalline arrays exist: the A-type, which is found in cereal starches and involves six helices forming a monoclinic unit with four water molecules, and the B-type which is found in leaf, root and tuber starches which also requires six helices, but instead forms a hexagonal unit with 36 water molecules (Bertoft, 2015). Interestingly pea seed starch contains the B-type allomorph at the centre of its granules and the A-type allomorph in surrounding regions. The presence of both A-type and B-type allomorphs is commonly referred to as C-type crystalline starch (Bogracheva *et al.*, 1998). The α-1,6-linked branch points of amylopectin, which do not partake in double helix formation, make up an amorphous lamella. The amorphous lamella alternates with the crystalline arrays, with a periodicity of 9-10 nm, to form the semi-crystalline amylopectin matrix (Smith and Zeeman, 2020) (Figure 1.1). Subsequent growth of the amylopectin matrix leads to the formation of a semicrystalline insoluble starch granule.

The specifics as to how the amylopectin matrix is arranged is unclear, with two opposing models: the cluster model and the building block backbone model (Figure 1.2). The first cluster models were proposed independently by Nikuni (1969) and French (1972). They are characterised by the α-1,6-linked branches being localised in clusters, rather than randomly throughout the starch granule, with non-reducing ends of the glucan chains pointing to the periphery of the granule. Advances in imaging amylopectin, in addition to biochemical assays and chain length profiling, has allowed the original cluster models to be refined. Nowadays the cluster model is characterised by clusters containing both short chains, with a degree of polymerisation (DP) of 6-35, and long chains, with a DP of ≥36. These are subsequently linked in a mainly tandem fashion, although rarely one cluster can branch into multiple clusters, and long chains connect the individual clusters. Crucially, both long and short chains are found in the amorphous lamellae (Nakamura and Kainuma, 2022). On the other hand in the building block backbone model, which was first formalised by Bertoft (2013), long chains form a backbone off which short chains branch. This backbone is located only in amorphous regions with the short chains protruding at right angles and forming crystalline arrays (Tetlow and Bertoft, 2020). A major difference between the models is the direction and timing of short and long chain synthesis. Both are synthesized at the same time and in the same direction in the cluster model, whereas short chains are synthesised after the synthesis of a long-chain backbone to facilitate their perpendicular placement under the backbone model (Nakamura and Kainuma, 2022). Critics of the cluster model argue that it has not been proven, as individual clusters have not been isolated and characterised (Bertoft, 2013). Meanwhile, critics of the building block backbone model argue that there must be complex mechanistic regulation for the temporal and spatial separation of long and short chain synthesis, for which there is currently no experimental evidence (Nakamura and Kainuma, 2022). Hence further experimental work, using the latest advances in high-resolution imaging, is required to prove or disprove these models.

In storage starches, the amylopectin matrix is further organised into periodic growth rings, which are thicker towards the centre of the starch granule (Bertoft, 2015). These rings are typically hundreds of nanometres in size with each ring containing tens of the aforementioned semi-crystalline repeats (Smith and Zeeman, 2020). Despite growth rings being described from the early eighteenth century (Bertoft, 2015), it is not known how they arise. One popular theory was that they reflected fluctuations in substrate supply, and subsequently granule growth, across light-dark cycles. However, growth rings persist under constant light in potato (Pilling and Smith, 2003) and barley

(Goldstein *et al.*, 2017), seemingly dismissing this hypothesis. Other theories focus on the circadian clock promoting differing rates of granule growth across a 24-hour period, or on the frequency of amylopectin branching changing as the granule grows (Smith and Zeeman, 2020). However, these theories require more thorough investigation. Interestingly, granules in leaves lack growth rings due to their rapid turnover rates and small size (Zeeman *et al.*, 2002).

Much like the uncertainties on how the amylopectin matrix is arranged, the precise location of amylose within the starch granule is also not known. X-ray scattering provides evidence that amylose is in the amorphous regions of the amylopectin matrix (Jenkins and Donald, 1995). This arrangement would minimise the disruption to amylopectin packaging in the crystalline regions. However, amylose is closer to amylopectin helices than other amylose chains in chemical crosslinking experiments with wheat starch (Jane *et al.*, 1992), suggesting the presence of at least some amylose in the crystalline amylopectin-rich regions. Furthermore, the distribution of amylose is not homogenous throughout the granule, with more amylose found at the periphery of starch granules from wheat (Morrison and Gadan, 1987) and maize (Pan and Jane, 2000). In potato, the situation is complex as some studies report increased amylose at the granule periphery (Jane and Shen, 1993), whilst others argue the opposite with more amylose at the granule core (Tatge *et al.*, 1999). The DP of amylose molecules varies across the granule, with molecules of lower DPs found at the granule core (Pérez *et al.*, 2009). The potential diversity in amylose location in the starch granule may not be surprising given the enormous diversity in starch granule morphology (Section 1.3). However, regardless of where it is located, amylose is not necessary for proper granule structure, as amylose-free mutants in *Arabidopsis thaliana (Seung et al., 2015)*, potato (Jacobsen *et al.*, 1989), wheat (Nakamura *et al.*, 1995) and rice (Zhang *et al.*, 2012) have no other structural defects in their starch, aside from the lack of amylose.

Beyond amylose and amylopectin themselves, another feature of starch granules is the hilum or core. This is unique to storage starch granules and absent from leaf starch granules for unknown reasons. The hilum is proposed to be the location of granule initiation (Section 1.5.1) and is potentially the least organised area of the granule, as it is highly susceptible to enzymatic degradation (Fuwa *et al.*, 1978). It has been proposed that both the hilum, and the previously mentioned growth rings, contain amorphous regions. This explains why the overall crystallinity of starch granules can be as low as 20% (Lopez-Rubio *et al.*, 2008).

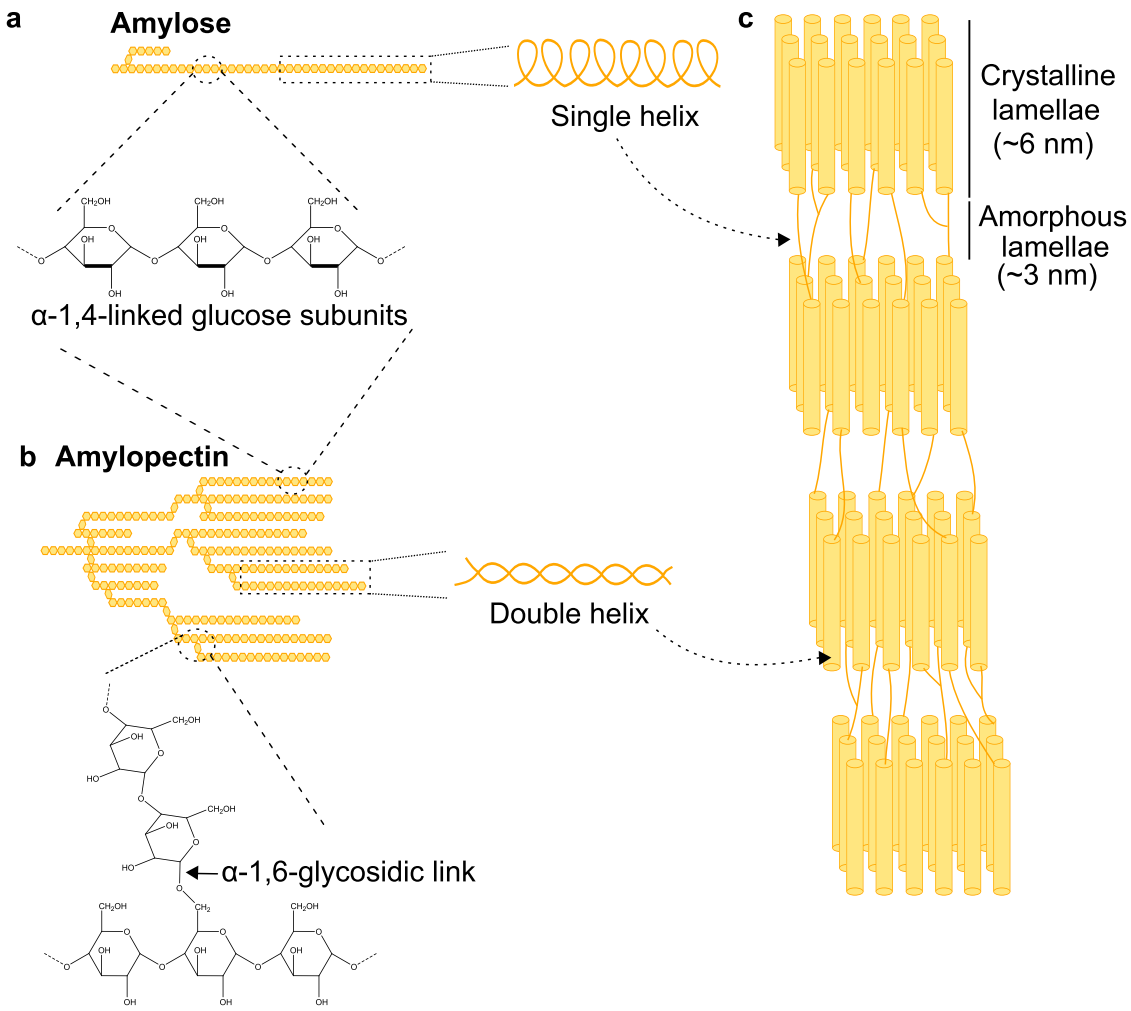


Figure 1.1 Amylose and amylopectin organisation. (a) Amylose consists of chains of α-1,4-linked glucose subunits with infrequent α-1,6-linked branches and adopts a single helix structure. (b) Amylopectin consists of chains of α-1,4-linked glucose subunits with frequent α-1,6-linked branches. Neighbouring chains within amylopectin clusters form double helix structures. (c) Amylopectin double helices (cylinders) create ordered crystalline arrays (~6 nm). The amorphous regions (~3 nm) contain the α-1,6-linked branch points of amylopectin and amylose single helices. The amorphous lamella alternates with the crystalline arrays, with a periodicity of 9-10 nm, forming the semi-crystalline amylopectin matrix.

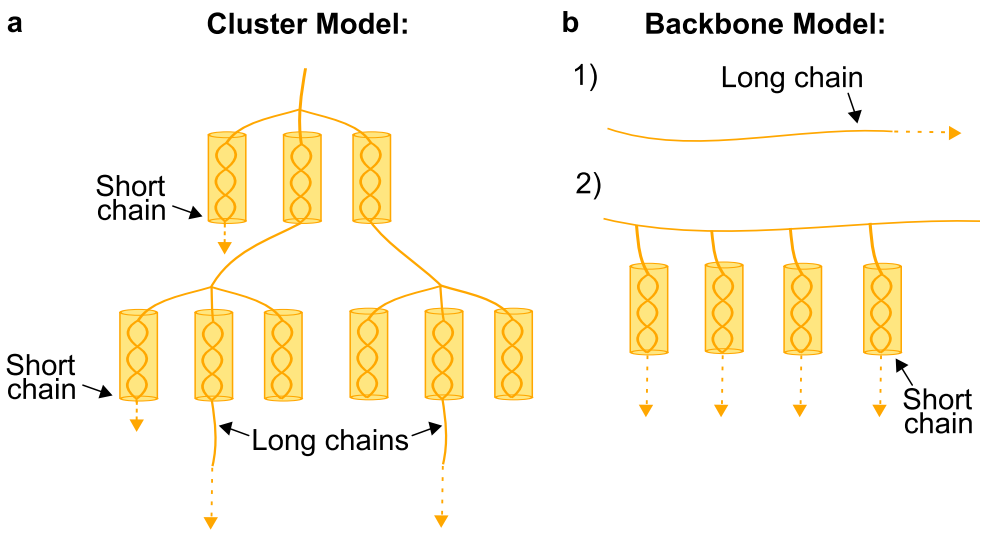


Figure 1.2 The cluster and backbone models for organisation of the amylopectin matrix. (a) Cluster model – long chains connect individual clusters of short chains; both are synthesized at the same time and grow in the same direction (arrows). (b) Backbone model – first, a long chain is synthesised (1), then short chains are synthesized and placed perpendicularly to the long chain (2).

1.3 Starch granule diversity

There is large variation in starch granule morphology and number both across tissues within a species, and between species.

There can be significant variation in transitory starch, both between chloroplasts of the same tissue and between tissues of the same plant. Much of this work has been conducted in *Arabidopsis thaliana*. When comparing chloroplasts of the same tissue, there is often variation in the number of granules. In *Arabidopsis thaliana* mesophyll chloroplasts, the mean number of granules per chloroplast is usually reported as five to seven (Crumpton-Taylor *et al.*, 2012). However, a recent study, using superior 3D imaging technologies, revealed a much greater range from eight to 32 granules detected in a single chloroplast (Bürgy *et al.*, 2021). Beyond variation in granule number, there are also differences in granule size between different tissues in *Arabidopsis thaliana,* with starch granules in guard cells being more than two-fold smaller, and also more numerous, than those in mesophyll cells (Liu *et al.*, 2021b). Whilst transitory starch has been greatly studied in *Arabidopsis thaliana*, how granules vary in photosynthetic organs of different species has not yet been investigated.

The diversity in starch composition and structure is exemplified in storage starch, where there is large interspecies diversity in starch granule morphology (Jane *et al.*, 1994). These starch granules can be broadly divided into three major morphological classes – simple, bimodal and compound (Matsushima *et al.*, 2013) (Figure 1.3) – each are characterised by different granule initiation patterns. Simple granules (Figure 1.3a) are produced when there is a single granule initiation per amyloplast, this granule morphology is found in major crops species including: maize (*Zea mays*) endosperm, yam (*Dioscorea* spp.) tubers, potato (*Solanum tuberosum*) tubers, and pea (*Pisum sativum*) seed embryos (Jane *et al.*, 1994; Chen *et al.*, 2021). Interestingly, even within species with simple starch granules, there are differences in granule size and shape. For instance, simple granules from potatoes are up to 75 μm in size and are ellipsoid in shape, whilst those from pea seed embryos are reported as both ellipsoid and disc-shaped and are up to 45 μm in diameter. Those from maize are different again, having a polyhedral morphology and being smaller, up to 20 μm (Chen *et al.*, 2021).

Whilst simple granules are found widely, bimodal granules (Figure 1.3b) are restricted to the Triticeae, which contains the major crop plants: barley, rye (*Secale cereale*) and wheat. Each individual granule is the result of one initiation, however, there are at least two waves of granule initiation. The first wave occurs in the body of the amyloplast at 4-6 DPA (days post anthesis) and leads to the production of A-type granules, which are large (20-30 μm) and lenticular. The second wave of granule initiation produces B-type granules which are smaller, 2-7 μm, and spherical (Bechtel *et al.*, 1990). This second wave of initiation is spatially and temporally separated from the first, occurring approximately 10 days later, at least partially in amyloplast extensions known as stromules (Bechtel *et al.*, 1990; Langeveld *et al.*, 2000; Matsushima and Hisano, 2019). Within species displaying a bimodal granule distribution, there is large diversity in the relative amounts and number of A-type and B-type granules. This diversity is exemplified in wild species which is discussed in Section 1.6. These different classes of granules have large influences on the end uses of the starch (Section 1.6.1), so understanding what controls the granule size distribution is a key objective.

Compound granules (Figure 1.3c), by contrast, are not formed from single initiations but are the result of multiple granules initiating in a single amyloplast and fusing together. The granules initiate at a similar time and as they grow the forces acting on the growing granules and the surrounding amyloplast membrane pushes the granules together into a centroidal Voronoi tessellation (Matsushima *et al.*, 2015). Interestingly,

the granules do not completely fuse as microscopy reveals the presence of junction sites (Buttrose, 1960) and the constituent granules easily fall apart during starch purification. Some studies have observed the presence of septa between the constituent granules, which is hypothesized to be the inner envelope membrane (Yun and Kawagoe, 2010; Kawagoe, 2013). There are many reports of grass species containing compound granules, with the most well-known being rice (*Oryza sativa*) (Chen *et al.*, 2021). Many early diverging grass species of the Poaceae have compound granules, perhaps pointing to them being the ancestral granule state in this family (Matsushima *et al.*, 2013). There is interspecies diversity in the number of granules contained within each compound granule, ranging from three (Tateoka, 1962) to more than 12 (Matsushima *et al.*, 2015).

Whilst these three classes of morphology explain most of the diversity seen in storage starch, some species do not fit neatly into these as they contain granules from different classes. Oat (*Avena sativa*) is a well described example of this, as its endosperm starch contains both simple and compound granules (Saccomanno *et al.*, 2017) (Figure 1.3d). This pattern of granule morphology is not exclusive to oat and is found in other genera including: *Miscanthus*, *Digitaria* and *Chrysopogon* (Tateoka, 1962). Amongst these genera, there is variation in the ratios of compound to simple granules, with more compound granules in *Digitaria*, more simple granules in *Chrysopogon*, and roughly equal proportions in *Miscanthus* (Tateoka, 1962).

Although the large diversity in starch granules has been widely documented, there is a poor understanding of why this diversity exists and the precise mechanisms of granule initiation which underpin the diversity.

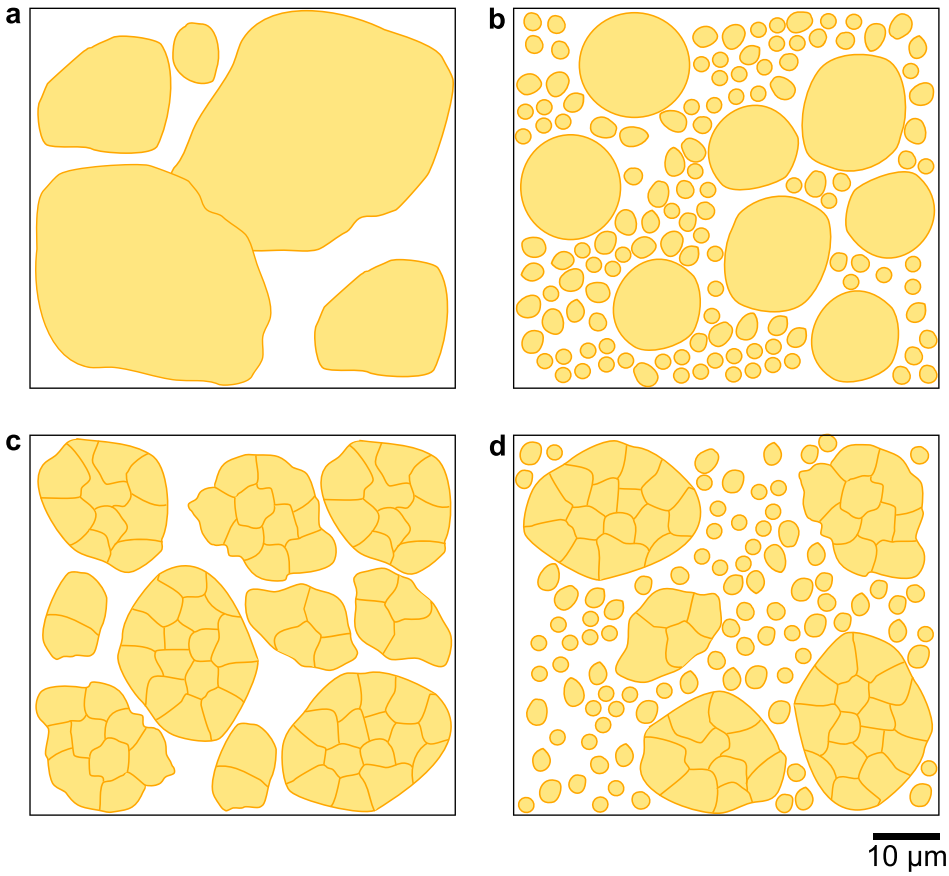


Figure 1.3 Variation in starch granule morphologies observed in storage starch. Diagrammatic representations of (a) simple starch granules, (b) bimodal starch granules, where the large granules are A-type granules and the smaller granules are B-type granules, (c) compound starch granules and (d) compound and simple granules. Scale bar = 10 µm for all parts of the figure.

1.4 Polymer synthesis

The biosynthesis of the starch polymers can be crudely divided into two parts. Firstly, the formation of the glycosyl donor adenosine diphosphate-glucose (ADP-Glc), and secondly the use of ADP-Glc to elongate growing amylose and amylopectin chains.

1.4.1 Synthesis of the glycosyl donor

In chloroplasts, starch synthesis is linked to carbon fixation in the Calvin cycle (Figure 1.4). The Calvin cycle intermediate, fructose 6-phosphate (F6P) is converted first into glucose 6-phosphate (G6P) and then into glucose 1-phosphate (G1P), by the action of phosphoglucose isomerase (PGI) and phosphoglucomutase (PGM), respectively. Both reactions are thermodynamically reversible. G1P is subsequently converted to the

nucleotide activated sugar ADP-Glc, catalysed by ADP-Glc pyrophosphorylase (AGPase) and releasing inorganic pyrophosphate (PPi). Unlike the previous reactions, this reaction is non-reversible as PPi is readily hydrolysed to produce orthophosphate (Pi) – a step that is thermodynamically unfavourable in the reverse direction.

ADP-Glc formation in non-photosynthetic tissues is not directly linked to the Calvin cycle, so carbon is instead provided from sucrose (Figure 1.4). Sucrose is transported into cells via plasma membrane sucrose transporters. In most tissues, sucrose is first metabolised into fructose and uridine diphosphate-glucose (UDP-Glc) by sucrose synthase. Fructose is converted to fructose 6-phosphate (F6P) by fructokinase and subsequently to G6P by PGI, whereas UDP-Glc is first converted to G1P, by UDP-Glc pyrophosphorylase. G1P can be imported into the plastid directly by a hexose-P transporter, or alternatively converted to G6P by PGM in the cytosol. G6P is transported into the plastid via the action of a plastid membrane localised Pi antiporter, where it is reconverted to G1P by PGM. G1P in the plastid is utilised by the plastid localised AGPase to produce ADP-Glc. This entire pathway is well described in a review by Tetlow (2011). Interestingly, this process is different in the endosperm of cereal grains where there is an additional form of AGPase localised in the cytosol (Denyer *et al.*, 1996; Thorbjørnsen *et al.*, 1996). The cytosolic AGPase provides 65-95% of AGPase activity (Tetlow *et al.*, 2004) and mutations specifically affecting the cytosolic form cause significant reductions in starch content (Johnson *et al.*, 2003; Wei *et al.*, 2017). Therefore, instead of hexose phosphates being transported into the plastid, ADP-Glc is transported directly via the ADP-Glc/ADP antiporter BRITTLE 1 (Shannon *et al.*, 1998; Patron *et al.*, 2004).

In both transitory and storage starch synthesis, the AGPase reaction is highly regulated as it is irreversible and thus, often referred to as the first committed step of starch synthesis. AGPase is a heterotetramer consisting of two catalytically active small and two regulatory large subunits. Expression of both subunits is controlled temporally and spatially, with the expression affected by different environmental factors, including sugar (Sokolov *et al.*, 1998) and phosphate availability (Nielsen *et al.*, 1998).

In addition to transcriptional control, there is also post-translational control of AGPase activity. There are differences between the post-translational control of AGPases in photosynthetic vs non-photosynthetic tissues. In photosynthetic tissues, there is strong allosteric control with AGPase activity enhanced by the metabolic intermediate

glycerate-3-phosphate (3PGA) and reduced by the binding of ADP or PPi (Ghosh and Preiss, 1966). This allows AGPase activity to be linked to photosynthesis, as photosynthesis results in a relative increase in the amount of 3PGA relative to PPi. In non-photosynthetic tissues the extent of allosteric controls varies between tissues and plants. On one hand, the AGPase in potato tubers experiences strong allosteric control by 3PGA (Ballicora *et al.*, 1995; Kavakli *et al.*, 2001), whereas AGPase from pea embryos is less responsive to 3PGA (Hylton and Smith, 1992). Similar to pea embryos, evidence from wheat and barley endosperm extracts suggests that AGPase activity, which is predominately due to the cytosolic form of the protein, is insensitive to 3PGA and PPi (Kleczkowski *et al.*, 1993; Gómez-Casati and Iglesias, 2002). Interestingly, when maize (Plaxton and Preiss, 1987) or rice (Sikka *et al.*, 2001) endosperm extracts are treated with 3PGA, an increase in AGPase activity is observed.

Another form of post-translational control of AGPase activity is the influence of redox state. Under oxidising conditions, a disulphide bridge forms between the two small subunits which causes dimer formation and prevents enzyme activity (Ballicora *et al.*, 2000; Tiessen *et al.*, 2002; Hädrich *et al.*, 2012). This phenomenon was first discovered in non-photosynthetic potato tubers but has since been demonstrated in photosynthetic tissues too (Hendriks *et al.*, 2003; Hädrich *et al.*, 2012). Interestingly, the key cysteine residues involved in disulphide bridge formation in potato tubers are absent in the cytosolic AGPase in cereals (Figueroa *et al.*, 2022). However, in rice alternative cysteine residues are instead involved in disulphide bridge formation and the enzyme is redox sensitive (Tuncel *et al.*, 2014).

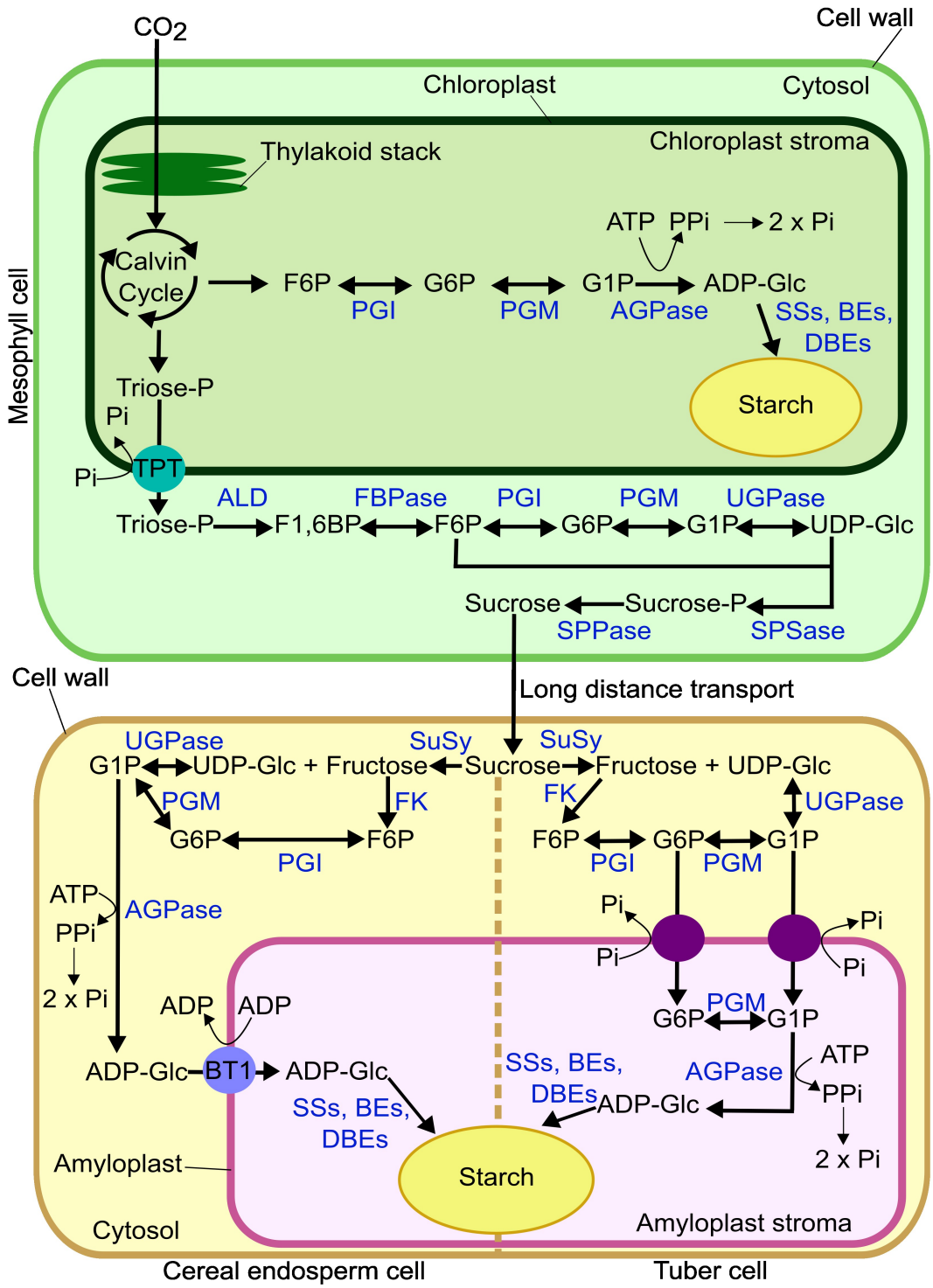


Figure 1.4 Starch synthesis in photosynthetic and non-photosynthetic cells. In photosynthetic cells (top), carbon is fixed from CO_2 in the Calvin cycle. Fructose 6-phosphate (F6P), a Calvin cycle intermediate, is converted to glucose 6-phosphate (G6P) by phosphoglucose isomerase (PGI) and then into glucose 1-phosphate (G1P), by phosphoglucomutase (PGM). G1P is converted to ADP-Glucose (ADP-Glc) catalysed by ADP-Glc pyrophosphorylase (AGPase) which can be utilised by starch synthases (SSs) to elongate growing starch polymers. The growing polymer is acted

on by branching enzymes (BEs) and debranching enzymes (DBEs) which ensure the production of the correct starch structure. Triose-phosphates (Triose-P) are also produced in the Calvin cycle and are exported into the cytosol via the triose-phosphate/phosphate translocator (TPT). This is subsequently converted to fructose-1,6-bisphosphate (F1,6BP) by aldolase (ALD), F6P by fructose-1,6-bisphosphatase (FBPase), G6P by PGI, G1P by PGM and uridine diphosphate-glucose (UDP-Glc) by UDP-glucose pyrophosphorylase (UGPase). UDP-Glc and F6P are funnelled into sucrose synthesis, first being converted into sucrose-phosphate (Sucrose-P) by sucrose-phosphate synthase (SPSase), which is metabolised to sucrose by sucrose-phosphate phosphatase (SPPase). Sucrose can be transported symplastically over long distances to non-photosynthetic tissues for the synthesis of storage starch (bottom). Once there, sucrose is metabolised to UDP-Glc and fructose, catalysed by sucrose synthase (SuSy). Alternatively, ADP-Glc can be produced instead of UDP-Glc (not shown). Fructose is converted to F6P by fructokinase (FK) and G6P and G1P by the actions of PGI and PGM respectively. UDP-Glc is converted directly to G1P by UGPase. In cereal endosperm cells (bottom left), a cytosolic AGPase catalyses AGP-Glc formation, which is imported into the amyloplast by BRITTLE1 (BT1), where it can be used by SSs, BEs and DBEs for starch synthesis. Whereas in other storage tissues, e.g. tubers (bottom right), G1P and G6P are imported into the amyloplast by hexose-P transporters (purple circles). In the amyloplast, G6P is converted to G1P by PGM which is converted to ADP-Glc by a plastid localised AGPase. As in cereal endosperms, ADP-Glc is then utilised by SSs, BEs and DBEs for starch synthesis.

1.4.2 Synthesis of amylose and amylopectin

Synthesis of starch polymers requires three classes of enzymes: starch synthases (SSs), starch branching enzymes (SBEs) and starch debranching enzymes (DBEs). Whilst the overall enzymatic activities are consistent between tissues and plants, there are differences in the specific isoforms present, the expression of the different isoforms, and how the proteins are targeted to the site of starch synthesis. In addition to these enzymes, some non-enzymatic proteins have also been reported to be involved in amylopectin synthesis - most notably the tryptophan rich proteins, EARLY STARVATION1 (ESV1) and LIKE EARLY STARVATION 1 (LESV) (Feike *et al.*, 2016; Yan *et al.*, 2024). ESV1 and LESV bind directly to glucans on the starch granule surface, which alters the crystallinity and effects the binding of DBEs and SSs (Singh *et al.*, 2022).

1.4.2.1 Starch Synthases

ADP-Glc acts as a substrate for SSs, which elongates growing starch polymers or free malto-oligosaccharides (MOS). There are at least five classes of SSs in plants, which all contain a glycosyltransferase (GT) 5 domain (Drula *et al.*, 2021) and most contain an additional GT1 domain (Pfister and Zeeman, 2016) (Figure 1.5).

One of the SSs, called GRANULE BOUND STARCH SYNTHASE (GBSS), is responsible for amylose synthesis. Plants with no or reduced GBSS activity have no or low amylose contents, respectively (Seung, 2020). This is true for both transitory starch in leaves (Seung *et al.*, 2015) and storage starch in tissues such as: potato tubers (Hovenkamp-Hermelink *et al.*, 1987), peas (Denver *et al.*, 1995), cassava roots (Ceballos *et al.*, 2007) and the endosperms of barley (Patron *et al.*, 2002), wheat (Nakamura *et al.*, 1995) and rice (Zhang *et al.*, 2012). Interestingly there are two paralogs of *GBSS* in monocots, coined *GBSSI* and *GBSSII*, which arose due to a whole genome duplication event around 251 million years ago (Cheng *et al.*, 2012). It is *GBSSII* that is analogous in function to *Arabidopsis thaliana GBSS* and is found in leaves, whilst *GBSSI* is an endosperm specific isoform (Cheng *et al.*, 2012). GBSS is unique as it is almost exclusively bound to the growing starch granule, targeted by the coiled-coil protein PROTEIN TARGETING TO STARCH 1 (PTST1) (Seung *et al.*, 2015). This targeting is necessary for proper GBSS activity. Knocking out *PTST1* or disrupting the GBSS-PTST1 interaction in *Arabidopsis thaliana* results in no or significantly reduced amylose, respectively (Seung *et al.*, 2015; Seung *et al.*, 2020). The role of PTST1 in cereal endosperms is less clear. In rice, PTST1 interacts with both GBSSI and GBSSII, but the amylose content of starch in rice *ptst1* mutants is only 2% lower than that of wild-type plants (Wang *et al.*, 2020). Therefore, PTST1 appears to only play a minor role in amylose synthesis in rice. In contrast, barley *ptst1* plants had no amylose and were completely starchless (Zhong *et al.*, 2018), suggesting that in barley PTST1 has an essential role in the starchy endosperm.

Three of the other SSs (SS1, SS2 and SS3) are involved in amylopectin synthesis whilst the others (SS4 and SS5) are linked to granule initiation (Section 1.5). SS1 acts on the shortest amylopectin chains (Delvallé *et al.*, 2005), while SS2 acts on intermediate length chains (Zhang *et al.*, 2008), and SS3 has roles in the production of both very short and long amylopectin chains (Wang *et al.*, 1993; Zhang *et al.*, 2005). However, the situation is rarely this simple: for instance in *Arabidopsis thaliana ss1* mutants, SS3 partially compensates for the lack of SS1 activity, suggesting that SS3 can use some of the SS1 substrates (Szydlowski *et al.*, 2011). Moreover, biochemical

studies reveal that some SSs can form complexes with each other and other enzymes which could influence their activity. In the maize endosperm, SS2a interacts with SS3, STARCH BRANCHING ENZYME (SBE) 2a and SBE2b (Hennen-Bierwagen *et al.*, 2008), while in cassava roots SS2 interacts with SBE2 and ISOAMYLASE (ISA) 2 (He *et al.*, 2022). To further add to the complexity, there are differences in the expression of the different SS isoforms both between plants, and between different tissues of the same plant. For example, SS1 is the major SS in potato leaves (Kossmann *et al.*, 1999), but the isoform is not expressed in maize leaves (Knight *et al.*, 1998); whereas in the maize endosperm SS1 and SS2 constitutes most of the SS activity (Cao *et al.*, 1999).

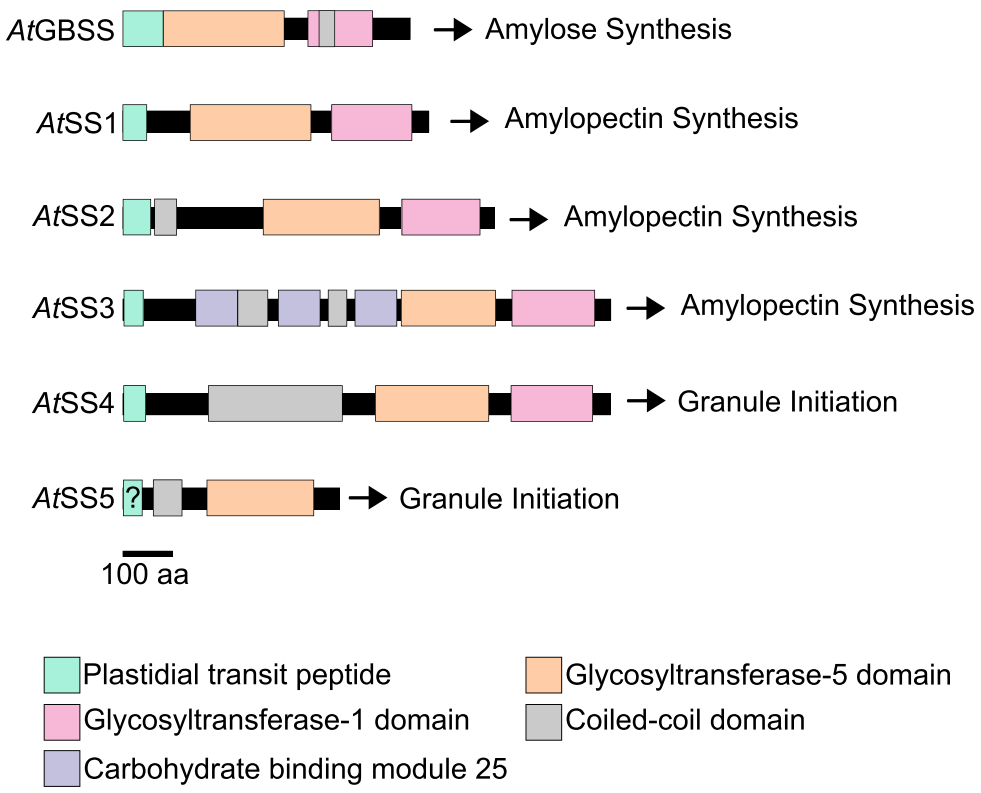


Figure 1.5 The six SSs in *Arabidopsis thaliana* have distinct domain structures and roles. Domains are represented as coloured boxes: turquoise = plastidial transit peptide, orange = glycosyltransferase-5 domain, pink = glycosyltransferase-1 domain, grey = coiled-coil domain and purple = carbohydrate binding module 25. Where the domain has not been characterised, it is marked with a '?'. Roles of the proteins in *Arabidopsis thaliana* are listed to the right. Scale bar = 100 amino acids. This figure is adapted from Pfister and Zeeman (2016) under a CC-BY license.

1.4.2.2 Starch branching enzymes

Whilst SSs can add ADP-Glc residues to the growing amylopectin chains, they cannot catalyse the formation of the α-1,6-linkages which form the branching points. This is carried out by starch branching enzymes (SBEs), which hydrolyse α-1,4-linkages and link the reduced end to a C6 hydroxyl group (Drummond *et al.*, 1972). SBEs have a distinct structure with a carbohydrate binding module (CBM) 48 towards their N-terminus and a glycoside hydrolase (GH) 13 domain at their C-terminus (Tetlow and Emes, 2014) (Figure 1.6). SBEs can be subdivided into two broad classes based on their amino acid sequences: SBE1s and SBE2s. Most plants have both classes although there is no SBE1 class enzyme in *Arabidopsis thaliana* (Goren *et al.*, 2018). The different SBE classes have varying preferences for their substrates. SBEIs have high activity for amylose, whilst SBE2s are more likely to use amylopectin as a substrate (Guan and Preiss, 1993; Morell *et al.*, 1997; Rydberg *et al.*, 2001). A further complexity in cereals is that the SBE2 class is subdivided into SBE2a and SBE2b. These two classes are differentially expressed, with different plants showing different patterns. In maize *SBE2b* is the major endosperm expressed SBE, with *SBE2a* being expressed in vegetative tissues and embryonic tissue (Gao *et al.*, 1997), in barley *SBE2a* is expressed highly in both the leaves and the endosperm (Sun *et al.*, 1998) and in wheat, *SBE2b* expression is low in the endosperm, whereas *SBE2a* is highly expressed in the endosperm (Regina *et al.*, 2005). However, in wheat the protein localisation also needs to be accounted for as SBE2b localises with endosperm starch granules, whilst SBE2a is mainly found in the amyloplast stroma (Regina *et al.*, 2005). Hence SBE2b may have a greater influence on amylopectin branching of starch granules than SBE2a. This highlights the importance of considering gene expression, protein abundance and localisation.

Mutations in *SBEs* cause defects in amylopectin. Perhaps the most famous *sbe* mutant is Mendel's wrinkled pea which has a transposon insertion in *SBE1*. This causes a reduction in the ratio of amylopectin to amylose (Bhattacharyya *et al.*, 1990). *SBE* mutations have also been described in other plants, including the *amylose extender* (a*e*) mutation in maize *SBE2b.* This causes high-amylose starch and has been selected for during maize domestication (Whitt *et al.*, 2002). Similar high-amylose, low-amylopectin phenotypes are seen in *sbe2b* mutants of rice (Nishi *et al.*, 2001), *sbe2a* mutants of tetraploid wheat (Hazard *et al.*, 2012), and barley lines where *SBE2a* and *SBE2b* have been silenced via RNAi (RNA interference) (Regina *et al.*, 2010).

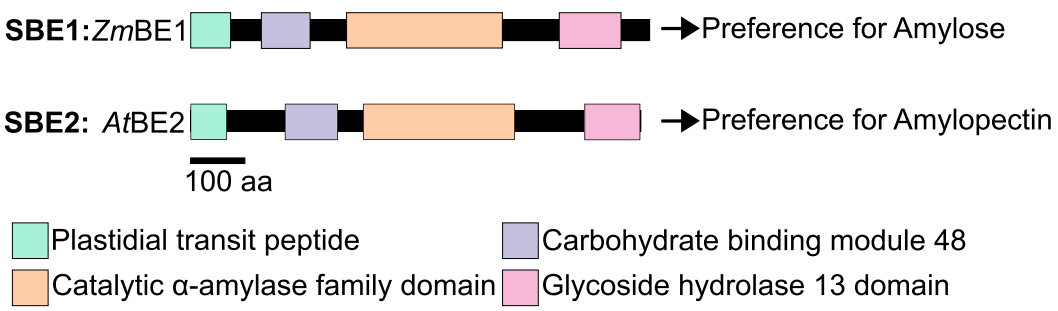


Figure 1.6 The two classes of SBEs have a similar domain structure. An example of a class 1 SBE (SBE1) and a class 2 SBE (SBE2) are shown, as *Arabidopsis thaliana* does not have a class 1 SBE, BE1 from maize is shown instead. Domains are represented as coloured boxes: turquoise = plastidial transit peptide, purple = carbohydrate binding module 48, orange = catalytic α-amylase family domains and pink = glycoside hydrolase (GH) 13 domain. The typical substrate preference of the proteins is listed to the right. Scale bar = 100 amino acids. This figure is adapted from Pfister and Zeeman (2016) under a CC-BY license.

1.4.2.3 Starch debranching enzymes

There is one additional class of enzymes involved in amylopectin synthesis - the starch debranching enzymes (DBEs). These cleave misplaced α-1,6-linkages to regulate overall amylopectin structure, in a process known as “trimming” (Smith and Zeeman, 2020). Like the SBEs, the DBEs also have a N-terminal CBM48 domain (Pfister and Zeeman, 2016) (Figure 1.7). DBEs can be separated into two types – the isoamylase (ISA) type and the limit dextrinase (LDA) type. The major distinguishing factor is that only the LDA type can degrade pullulan (Pfister and Zeeman, 2016); a compound produced by *Aureobasidium* fungi composed of maltotriose subunits joined with α-1,6-linkages (Wei *et al.*, 2021). The ISA class contains three distinct forms: ISA1, ISA2 and ISA3. LDA, sometimes called pullulanase (PUL), is the only example of a LDA type DBE in plants.

The key role of ISA1 in amylopectin synthesis is demonstrated by the accumulation of the highly branched soluble polymer phytoglycogen instead of starch in the *isa1* mutants of: *Arabidopsis thaliana* (Wattebled *et al.*, 2005), maize (James *et al.*, 1995) and barley (Burton *et al.*, 2002). Interestingly, ISA2 lacks the key residues required for catalytic activity (Hussain *et al.*, 2003). Nevertheless, ISA2 has an important role in some plants as *Arabidopsis thaliana isa2* plants are almost phenotypically identical to *isa1* plants (Delatte *et al.*, 2005). Thus, it has been proposed that ISA2 may be instead

important for ISA1 stability by forming a heteromultimer (Hussain *et al.*, 2003). However, the situation in the cereal endosperm is different. Maize *isa2* plants accumulate much less phytoglycogen than maize *isa1* plants (Kubo *et al.*, 2010). Likewise, in wild-type maize endosperm homomeric ISA1 complexes are detected (Kubo *et al.*, 2010). Thus, in cereal endosperms it has been suggested that ISA1 homomultimers can conduct most ISA functions, which contrasts with cereal leaves and potato tubers where ISA2 is also needed (Kubo *et al.*, 2010). Whilst ISA1 and ISA2 partake in amylopectin debranching, ISA3 and LDA are primarily involved in transitory starch degradation in *Arabidopsis thaliana* (Pfister and Zeeman, 2016). However, in cereals they may have a small influence on amylopectin branching. Rice *lda* mutants have longer amylopectin chains (approximately 3 glucose residues longer compared to wild-type) in their endosperm starch (Fujita *et al.*, 2009) and *isa3* mutants have more highly branched amylopectin (Yun *et al.*, 2011). Maize *lda* mutants have an accumulation of branched glucans (Dinges *et al.*, 2003). However, this is difficult to generalise more widely as reports of *lda* and *isa3* mutants are missing in many key species, including wheat.

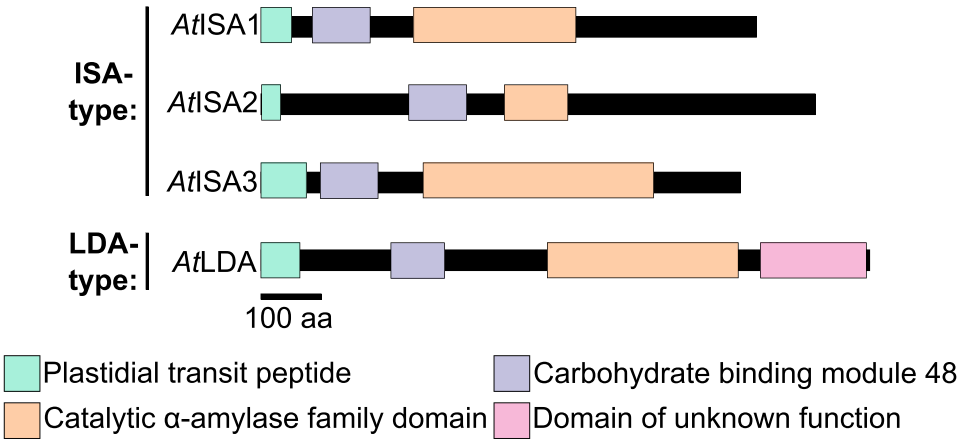


Figure 1.7 The two types of DBEs have a similar domain structure. An example of ISA-type and LDA-type DBEs from *Arabidopsis thaliana*. Domains are annotated according to *Arabidopsis thaliana* and are not identical to the annotations in cereals (Chapter Five). Domains are represented as coloured boxes: turquoise = plastidial transit peptide, purple = carbohydrate binding module 48, orange = catalytic α-amylase family domains and pink = domain of unknown function. Scale bar = 100 amino acids. This figure is adapted from Pfister and Zeeman (2016) under a CC-BY license.

1.5 Granule Initiation

1.5.1 The process of granule initiation

Whilst the synthesis of amylose and amylopectin is well understood, less is known about how starch granules initiate. This process is sometimes called ‘priming’ or ‘nucleating’ and influences granule number and morphology. At a mechanistic level, granule initiation does not rely on a glucosylated protein primer (Seung and Smith, 2019). Instead, it is a spontaneous event dependent on glucan substrates and various proteins (Section 1.5.2.4). Granule initiation is dependent on nucleation of glucan substrates, which can occur in a homogenous or heterogeneous manner (Geddes and Greenwood, 1969) (Figure 1.8). In homogenous nucleation, nucleation is carried out by the glucan substrates (or MOS) themselves independent of external factors. Glucan substrates are constantly moving in the plastid and structurally similar parts of the glucan chains align to generate a small, ordered structure. This has stability and can act as the starting point for granule growth. In contrast, during heterogenous nucleation an external factor (e.g., a protein), stabilises the glucan molecules directly allowing subsequent growth. Regardless of the nucleation mechanism, for growth to occur, there needs to be crystallisation of the starch polymers. This has been proposed to happen spontaneously from a coacervate, which is a liquid-liquid phase separated droplet composed of organic molecules (Badenhuizen, 1963). This process of nucleation and crystallisation has been studied *in vitro* and can be somewhat replicated by manipulating temperature (Ziegler *et al.*, 2005). However, the resulting structures, called spherulites, do not completely resemble starch granules as they lack growth rings and are difficult to form with branched glucan substrates. So, whilst *in vitro* studies provide insight to the chemical and physical changes that could occur within glucan substrates during granule initiation, it does not provide a model of these processes *in vivo* which is still a major gap in our understanding.

In plant tissues, granule initiation can be further divided into primary and secondary events (Figure 1.8). Primary starch granule initiation describes the initiation of a granule in plastids without pre-existing granules or granule-derived glucans. Secondary granule initiation occurs when starch granules initiate in plastids already containing starch granules or glucans, which can act as a source of glucan substrates. The frequency at which primary initiation occurs in chloroplasts is unknown, and secondary initiations may dominate. Within a single chloroplast in an *Arabidopsis thaliana* leaf, under normal day:night regimes, a few small starch granules are visible at the end of the night period. These most likely serve as a source of glucans for secondary granule initiations in the following light period (Bürgy *et al.*, 2021).

Therefore, during normal diurnal starch turnover, most granule initiations are secondary. Even in newly developed chloroplasts granule initiations are more likely to be secondary, as when chloroplasts divide by binary fission both daughter chloroplasts often contain a starch granule (Mai *et al.*, 2019). The frequency of daughter chloroplasts that do not inherit a starch granule has not been measured. Even so, these chloroplasts are likely to contain small glucans in some form which could act as substrates for granule initiation. In non-photosynthetic tissues, the amount of primary granule initiations is just as uncertain. Similar to leaf chloroplasts, amyloplasts of potato tubers divide by binary fission and each daughter amyloplast may receive a starch granule, such that any subsequent initiations in this amyloplast are considered secondary (Mingo-Castel *et al.*, 1991). Likewise in rice, whose amyloplasts divide at multiple places at once giving a 'beads-on-a-string' like appearance, newly formed amyloplasts may already possess starch granules (Yun and Kawagoe, 2009). In Triticeae, secondary initiations result in the formation of B-type granules (Section 1.3), although whether all A-type granules stem from primary initiations has not been demonstrated.

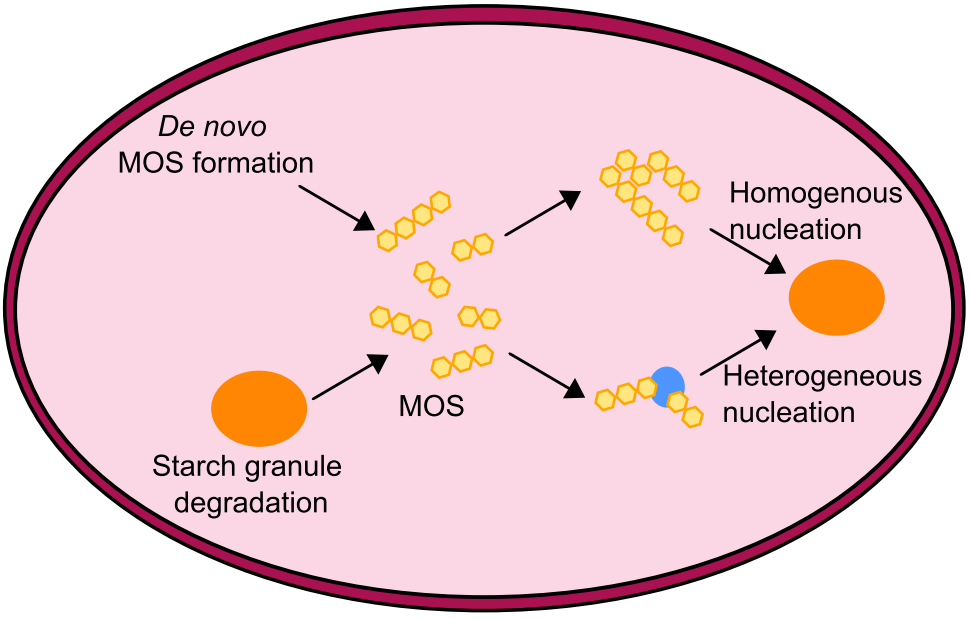


Figure 1.8 Granule initiation requires nucleation of glucan substrates.
A diagrammatic representation of a plastid undergoing granule initiation. Glucan substrates (MOS, yellow hexagons) can stem from *de novo* production (primary granule initiation) or from degradation of pre-existing starch granules (orange) (secondary granule initiation). Homogenous nucleation occurs when structurally similar parts of the glucan chains align to generate a small, ordered structure that acts as the starting point for granule growth. Heterogenous nucleation occurs when an external factor, for example a protein (blue), stabilises the glucan molecules directly facilitating growth of a starch granule.

1.5.2 Granule initiation in photosynthetic tissues

Much of the early work to understand how granules initiate was conducted on transitory starch in *Arabidopsis thaliana* leaves. However, recent work has started to investigate granule initiation in photosynthetic tissues of other species, with a focus on crop species such as wheat, barley and rice.

1.5.2.1 The role of starch synthases in granule initiation in Arabidopsis thaliana

The first protein to be implicated in granule initiation in *Arabidopsis thaliana* was SS4. *Arabidopsis thaliana ss4* mutants have mostly zero, or less frequently one or two, granules per chloroplast compared to the typical five to seven granules (Roldán *et al.*, 2007; Crumpton-Taylor *et al.*, 2013; Malinova *et al.*, 2017) (Figure 1.9a,b). Interestingly, the small number of granules in *ss4* mutants also have a different morphology, being more round in shape compared to normal lenticular morphology (Roldán *et al.*, 2007; Crumpton-Taylor *et al.*, 2013). Hence, SS4 not only influences granule number but also granule size and shape.

SS4 has a coiled-coil domain near its N-terminus in addition to its GT5 and GT1 domains (Figure 1.5), and these different domains have distinct functions. The C-terminal region (containing both GT domains) has a key role in promoting granule initiation. In *Arabidopsis thaliana*, expression of the *SS4* C-terminus in the *ss4* mutants partially rescues the granule number phenotype, increasing the number of chloroplasts containing multiple granules (Lu *et al.*, 2017). However, despite the increased number of granules, the granules still have a round rather than a lenticular morphology, hence suggesting that the N-terminal region containing the coiled-coil domain is important for determining granule morphology (Lu *et al.*, 2017). The N-terminus also has an additional role in ensuring the correct localisation of SS4 to distinct subchloroplastic puncta (Gámez-Arjona *et al.*, 2014). It has been hypothesized that SS4's punctate localisation is important for lenticular granule morphology by restricting SS4 catalytic activity to specific locations within the chloroplast (Lu *et al.*, 2017).

The small number of starch granules seen in *ss4* mutants may be initiated by SS3, as *ss3 ss4* double mutants are starchless (Szydlowski *et al.*, 2009). However, *ss3* single mutants have no defect in granule number (Zhang *et al.*, 2005). The mechanism by which SS3 carries out granule initiation in the absence of SS4 is unclear. SS3 has enzymatic activity (Busi *et al.*, 2008), but it is not reported to interact with any other proteins involved in starch granule initiation (Mérida and Fettke, 2021). Moreover, it is usually localised around the periphery of starch granules (Gámez-Arjona *et al.*, 2014;

Gámez-Arjona and Mérida, 2021) rather than the distinct punctate localisation of many of the starch granule initiation proteins, although its localisation in the *ss4* background has not been investigated.

In addition to SS4 and SS3, the non-catalytic SS5 also plays a role in granule initiation in *Arabidopsis thaliana*. *ss5* mutants have a reduced number of starch granules with an average of two to three per chloroplast (Abt *et al.*, 2020) (Figure 1.9c). SS5 is a noncanonical starch synthase, as it lacks the GT1 domain (Figure 1.5) and has no enzymatic activity (Abt *et al.*, 2020). Thus, it cannot be playing an enzymatic role in the granule initiation process. The *ss4 ss5* double mutant has a more extreme phenotype than either single mutant, with more chloroplasts containing no starch granules (Abt *et al.*, 2020). So, whilst the mechanistic role of SS5 is still unclear, it seems most likely that it does not function through SS4.

1.5.2.2 The role of coiled-coil proteins in granule initiation in Arabidopsis thaliana

Other coiled-coil proteins, aside from SS4, also have roles in granule initiation. PROTEIN TARGETING TO STARCH 2 (PTST2), a homolog of PTST1 (Section 1.4.2), has a coiled-coil domain and a carbohydrate binding module 48 (CBM48), and interacts with SS4. *Arabidopsis thaliana ptst2* mutants have, on average, zero or one large starch granule in their mesophyll chloroplasts, although this is less severe than the *ss4* phenotype (Seung *et al.*, 2017) (Figure 1.9d). Unlike *ss4* mutants, the morphology of the granules is discoid and flat much like the wild-type (Seung *et al.*, 2017; Liu *et al.*, 2021a). PTST2 interacts with MOS through its CBM48 domain. It has been proposed that PTST2 may provide MOS substrates to SS4 for elongation and formation of a granule initial. This explains how PTST2 has such strong effect on granule number despite having no enzymatic activity (Seung *et al.*, 2017).

PTST2 has a close homolog in some species, including *Arabidopsis thaliana*, named *PTST3*. *Arabidopsis thaliana ptst3* plants have an average of two granules per chloroplast (Seung *et al.*, 2017) (Figure 1.9e). This phenotype is less severe than the *ptst2* mutants. However, combining the *ptst3* mutation with the *ptst2* mutation increases the severity of the *ptst2* phenotype and *ptst2 ptst3 Arabidopsis thaliana* mutants have more chloroplasts with zero starch granules compared to *ptst2* plants (Seung *et al.*, 2017) (Figure 1.9f). This is also accompanied by changes in granule morphology with *ptst2 ptst3* mutants having rounded granules like the *ss4* mutant (Seung *et al.*, 2017). Unlike for PTST2, an interaction between SS4 and PTST3 has not been demonstrated so its precise role in granule initiation is unclear. Regardless,

PTST3 probably carries out its role with a lower efficiency than PTST2 given the strong reduction in granule number phenotype in *ptst2* plants. Instead, PTST3 may act with PTST2 directly to determine granule morphology, which is supported by the fact that PTST3 interacts with PTST2 (Seung *et al.*, 2017).

PTST2 binds to two other coiled-coil proteins: MAR-BINDING FILAMENT PROTEIN 1 (MFP1) and MYSOIN-RESEMBLING CHLROPLAST PROTEIN (MRC). Both are involved in granule initiation in chloroplasts as *Arabidopsis thaliana mfp1* and *mrc* mutants both have reduced granule number in mesophyll chloroplasts (Seung *et al.*, 2018; Vandromme *et al.*, 2019) (Figure 1.9g,h). Moreover, overexpressing *MFP1* in *Arabidopsis thaliana* causes significantly more granules per chloroplast (Sharma *et al.*, 2024). MFP1 has a distinct punctate localisation; it associates with the thylakoid membranes via its N-terminal transmembrane domain, whilst its C-terminal coiled-coil domain protrudes into the stroma (Jeong *et al.*, 2003). This localisation pattern is important for correct PTST2 and SS4 localisation. Ectopic placement of MFP1 to the inner envelope results in mislocalisation of both PTST2 and SS4 to the inner envelope, and the production of starch granules at the stromal periphery of this membrane (Sharma *et al.*, 2024). Whilst the role of MFP1 is starting to be understood, the specific role of MRC has not yet been deciphered. Nevertheless, MRC is likely to have an important role as the *mrc* phenotype is stark with an average of one granule per chloroplast (Figure 1.9h). Moreover, it localises to distinct puncta, like other granule initiation proteins, and it interacts with PTST2, SS4 and SS5 (Seung *et al.*, 2018; Vandromme *et al.*, 2019; Abt *et al.*, 2020).

It has been hypothesized that the five coiled-coil proteins aforementioned – SS4, PTST2, PTST3, MFP1 and MRC – form an ‘initiation complex’ (Seung and Smith, 2019). MFP1 would act to localise the other granule initial proteins, PTST2 would provide MOS which SS4 would elongate further ensuring that starch granule initiation occurs at specific controlled locations in the chloroplast. Much is still unknown about this complex, such as: the specific roles of MRC and PTST3, the relative stoichiometry of the proteins in the complex, whether any other proteins are involved in the complex and the stability of the complex.

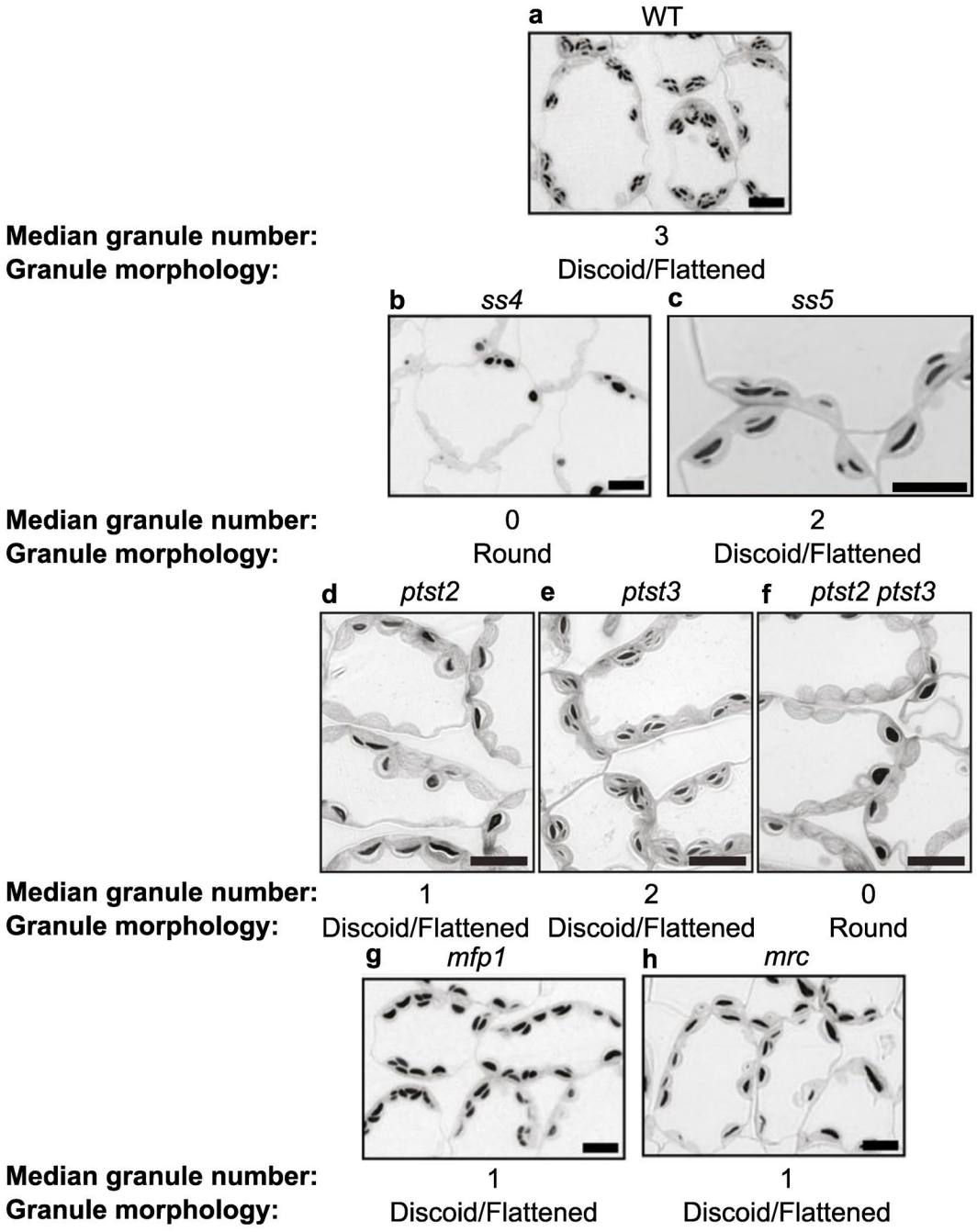


Figure 1.9 Mutating genes involved in starch granule initiation causes defects in granule number and morphology. Light microscopy sections of *Arabidopsis thaliana* leaves collected at the end of the day. Starch granules are stained black against grey chloroplasts. Median granule number per chloroplast section and granule morphology is reported below each image. (a) Wild-type (WT) Col-0, (b) *starch synthase 4*, (c) *starch synthase 5,* (d) *protein targeting to starch 2*, (e) *protein targeting to starch 3*, (f) *protein targeting to starch 2* and *protein targeting to starch 3*, (g) *mar-binding filament protein 1* and (h) *myosin-resembling chloroplast protein*. Scale bars = 10 μm. Images are adapted from Seung *et al.* (2017), Seung *et al.* (2018) and Abt *et al.* (2020) under CC-BY licenses. The wild-type looked comparable between all papers so for simplicity only one wild-type image is displayed here.

1.5.2.3 Other proteins implicated in granule initiation in Arabidopsis thaliana

In addition to SS4 there is another enzyme implicated in granule initiation in *Arabidopsis thaliana* - the plastidial α-glucan phosphorylase (PHS1). This enzyme catalyses a reversible reaction to extend glucan chains using G1P or degrade glucan chains to release G1P. PHS1 interacts with SS4 (Malinova *et al.*, 2018), however *Arabidopsis thaliana phs1* mutants have no differences in granule number per chloroplast (Malinova *et al.*, 2013). It is only when additional genes are knocked out in parallel with *PHS1*, such as maltose exporter 1 (*MEX1*) or the disproportionating enzyme 2 (*DPE2*) which is involved in cytosolic maltose metabolism, that a reduction in granule number per chloroplast is observed (Malinova *et al.*, 2013). Curiously, these phenotypes might be dependent on the environment as continuous light abolishes the phenotype in *dpe2 phs1* mutants and increases granule number per chloroplast (Malinova *et al.*, 2017; Malinova and Fettke, 2017). Therefore, it could be that PHS1 partakes in the granule initiation process but has either a minor role or an importance only under certain conditions.

1.5.2.4 Granule initiation in photosynthetic tissues of plants beyond Arabidopsis thaliana

Although much of the gene discovery has been conducted in *Arabidopsis thaliana*, the mechanisms appear to be largely conserved in the photosynthetic tissues of other plants. SS4 also promotes granule initiation in wheat leaves, with fewer granules per chloroplast in hexaploid wheat mutants of the D genome ortholog of *SS4* (Guo *et al.*, 2017). Similarly in knockout mutants of tetraploid durum wheat (*Triticum turgidum cv.* Kronos), 80% of chloroplasts had no starch granules (Hawkins *et al.*, 2021). Likewise, durum wheat *mrc* mutants had a significant reduction (~50%) in granule number per chloroplast, although it is less severe than that reported in *Arabidopsis thaliana* (Chen, 2022).

The granule initiation protein which has been characterised in the largest number of species is PTST2. Its roles in leaves have been investigated in: tetraploid durum wheat (where it is called B-GRANULE CONTENT 1, BGC1), barley (where it is called FLOURY ENDOSPERM 6, FLO6), *Brachypodium distachyon* (Watson-Lazowski *et al.*, 2022) and rice (where it is also called *FLO6*) (Zhang *et al.*, 2022). Like *Arabidopsis thaliana,* mutants in all these species have fewer granules per chloroplasts (Watson-Lazowski *et al.*, 2022). However, the phenotypes are not completely identical with differences in granules size and starch content, which has been attributed to differences in leaf sugar metabolism (Watson-Lazowski *et al.*, 2022).

There are some genes involved in granule initiation in *Arabidopsis thaliana* leaves that do not have orthologues in other plant families. Notably, there are no PTST3 or SS5 homologs in the grass subfamily of the Pooideae (Watson-Lazowski *et al.*, 2022). Hence, not everything learnt from *Arabidopsis thaliana* can be directly applied to other species, demonstrating the importance of studying granule initiation in a wide range of plants.

1.5.3 Granule initiation in non-photosynthetic tissues

Much of what is known about starch granule initiation in non-photosynthetic tissues centres around endosperms and tubers, with much less known about granule formation in pollen or roots. In both endosperm and tubers, granule initiation requires some of the same proteins involved in granule initiation in photosynthetic tissues. However, the specific roles or timing of expression of the genes often differs to give rise to the aforementioned diversity in granule morphology (Section 1.3).

1.5.3.1 Granule initiation in plants with compound granules

Much of the early work on granule initiation in non-photosynthetic tissues was conducted on rice. In rice there are two *SS4* paralogs: *SS4a* and *SS4b* (Dian *et al.*, 2005). Mutation of *ss4b* alone leads to no defects in endosperm starch (Toyosawa *et al.*, 2016). This could stem from the fact that *SS4b* is predominantly expressed in leaves whilst *SS4a* is the major endosperm SS4 (Dian *et al.*, 2005). The endosperm starch phenotypes of *ss4a* rice plants has not yet been characterised, perhaps due to the compromised growth of these plants (Jung *et al.*, 2018). Despite the lack of phenotype in *ss4b* mutants, combining the *ss4b* mutation with a *ss3a* mutation alters granule morphology from compound to spherical granules (Toyosawa *et al.*, 2016). This suggests the importance of SS3 and SS4 in determining correct granule morphology. However, from this mutant alone, it is difficult to dissect the roles of SS3 and SS4 so it would be important to examine endosperm starch in *ss4a ss4b* and *ss3a ss3b* mutants.

The role of *PTST2*, called *FLO6* in rice, has also been investigated. Rice *flo6* mutants have severe defects in compound granule formation, with smaller granules and altered morphology – they are more irregular in shape with rough surfaces (Peng *et al.*, 2014). Interestingly, this phenotype is similar to the rice *lesv* phenotype, which also has small granules that do not pack together tightly (Yan *et al.*, 2024). LESV interacts with FLO6 which is important for the targeting of ISA1 to starch granules (Yan *et al.*, 2024).

Whether this interaction is important for granule initiation is unclear and the storage starch phenotype of *lesv* mutants needs to be studied more broadly.

As in *Arabidopsis thaliana* leaves where PHS1 may have a conditional role, the effect of the *phs1* (also called *pho1*) mutation in rice seems to be dependent on temperature. When grown at 30ºC rice *phs1* grains were comparable to wild-type. Yet when grown at 20ºC, between 68-78% of grains had a shrunken endosperm (Satoh *et al.*, 2008). This was associated with a 20-fold reduction in endosperm starch content and the starch granules were smaller and more spherical (Satoh *et al.*, 2008). The temperature-dependency of the phenotype might be due to the existence of a factor that can only compensate for the lack of PHS1 at high temperatures, although such factor has not yet been discovered.

There are some proteins that influence granule initiation in compound granule formation in rice but have no reported roles in granule initiation in photosynthetic tissues. This includes *SUBSTANDARD STARCH GRAIN 4* (*SSG4*) and *SUBSTANDARD STARCH GRAIN 6* (*SSG6*), which when mutated cause larger starch granules in the compound granules of the endosperm (Matsushima and Hisano, 2019). This demonstrates how starch granule initiation varies across different tissues.

1.5.3.2 Granule initiation in plants with simple granules

Our understanding of initiation of simple granules is at a less advanced state compared to that described for compound granule formation in rice. This mainly stems from a lack of mutants in many of the granule initiation genes in plants like potato and maize. However, lines with altered expression can provide insight into the potential importance of these genes. Overexpressing the *Arabidopsis thaliana SS4* in potato increased potato tuber starch content but had no effect on granule size or shape. However, granule number was not measured, so it is difficult to conclude whether overexpressing *SS4* lead to a greater starch content via increasing the number of granule initiations (Gámez-Arjona *et al.*, 2011).

Altered expression has also been used to assess the role of PHS1 in maize and potato. The *shrunken4* maize mutant, which has a 70% reduction in *PHS1* activity, has one third of the starch of a wild-type grain (Tsai and Nelson, 1969). Hence PHS1 may play an important role in starch formation, and possibly granule initiation in maize. However, a more in-depth characterisation of starch granule morphology and number in the *shrunken4* plants is needed to confirm or refute this. *PHS1* has also been

knocked down in potato tubers to 8-12% of wild-type activity levels. Under low temperature conditions (15ºC), these plants have changes in starch granule structure, but no differences in starch granule number (Orawetz *et al.*, 2016). A similar knock down study also in potato tubers, revealed a role for PHS1 in MOS metabolism, suggesting that this could influence granule initiation through the generation of the nucleation structure (Flores-Castellanos and Fettke, 2022). Recently, a *phs1* mutant in potato was isolated, which had a greater number of starch granules than the wild-type and the granules had reduced size (Sharma *et al.*, 2023). Together this suggests that *PHS1* is involved in granule initiation in potato tubers, although a mechanism has not yet been proposed.

An example of a complete knockout of a starch granule initiation gene in a plant that produces simple granules is *Brachypodium distachyon bgc1* (*ptst2*) mutants. *Brachypodium distachyon* is a grass outside the Triticeae, but within the Pooideae. Loss of *bgc1* causes small compound granules to form in the endosperm instead of simple granules (Watson-Lazowski *et al.*, 2022). It will be interesting to see if this phenotype is consistent in other plants such as maize and potato, or whether BGC1 adopts different roles in different species.

1.5.3.3 Granule initiation in plants with bimodal granules

Like in leaves, SS4 also plays a key role in granule initiation in the wheat endosperm. Durum wheat *ss4* mutants have abnormal granule initiation patterns, resulting in multiple granules initiating per amyloplast and these fuse to form compound granules (Hawkins *et al.*, 2021) (Figure 1.10a,b). The compound granules appear during early grain development and are detected from at least 8 DPA. Hence in wild-type endosperm tissue, SS4 must be important in restricting A-type granule initiation to one initiation per amyloplast. The occurrence of multiple initiations in the absence of *SS4* in the endosperm contrasts the phenotype in leaves where absences of *SS4* results in reduced initiations. How SS4 restricts granule initiation in the endosperm is unknown.

In addition to SS4, the role of SS3 in endosperm starch formation in the Triticeae has also been investigated. Hexaploid wheat has two *SS3* paralogs – *SS3a* and *SS3b*. It is *SS3a* that is more strongly expressed in the endosperm, peaking at approximately 12 DPA (Li *et al.*, 2000). When *ss3a* is knocked out in hexaploid wheat, A-type and B-type granules are still observed, but the A-type granules are significantly smaller (Fahy *et al.*, 2022). Interestingly, the morphology of A-type granules is also altered, and many

have irregular shapes or protrusions. Notably, no compound granules were observed in the *ss3* plants.

The durum wheat *ss4* endosperm starch phenotype is strikingly similar to the full knockout phenotype of *bgc1* in durum and hexaploid wheat. This is characterised by abnormal starch granules with diverse shapes and sizes (polymorphous starch) due to the initiation and then fusion of multiple granules (Chia *et al.*, 2017; Chia *et al.*, 2020; Hawkins *et al.*, 2021) (Figure 1.10c). In hexaploid wheat, *BGC1* has a dosage effect, as plants lacking the A and D genome copies (--BB--) do not have wild-type or polymorphous starch phenotype, but instead exhibit fewer B-type granules. Combining the phenotypes of the double and triple mutants suggests that BGC1 promotes B-type granule formation. It also has a role in A-type granule initiation, where it acts in conjunction with SS4 to ensure there is only one A-type granule per amyloplast thus preventing the formation of polymorphous starch (Chia *et al.*, 2020; Hawkins *et al.*, 2021). This is also consistent in barley where full knockouts of *BGC1* also have polymorphous starch (Saito *et al.*, 2018).

Whilst *BGC1* has roles in both A- and B-type granule initiation, other proteins involved in granule initiation in leaves appear to have specific roles in the initiation of either granule type. Expression analysis revealed that *MRC* is specifically expressed during early grain development, up to 10 DPA, which coincides with A-type granule formation (Chen *et al.*, 2023a; Chen *et al.*, 2024). Durum *mrc* mutants have smaller A-type granules and a higher percentage volume of B-type granules (Figure 1.10d). The smaller A-type granules corresponded with the earlier emergence of B-type granules in the mutant, which were detected from 10 DPA. Thus *MRC* has been co-opted to control B-type granule initiation during endosperm starch formation by inhibiting B-type granule formation during early grain development (Chen *et al.*, 2024). This is in direct contrast to its role in photosynthetic tissue where it promotes granule initiation. As coiled-coil proteins are involved in protein-protein interactions, it has been proposed that the differing effects of MRC on granule initiation in wheat leaf chloroplasts versus endosperm amyloplasts might be a result of interactions with different proteins (Chen, 2022). However, the interaction partners of MRC in the endosperm need to be identified to confirm this.

PHS1 exemplifies a protein with either no, or a very minor, role in granule initiation in *Arabidopsis thaliana* and durum wheat leaf chloroplasts (Malinova *et al.*, 2013; Kamble *et al.*, 2023). However, it is important in the endosperm. Durum wheat *phs1* mutants

have fewer and larger B-type granules in the endosperm compared to the wild-type (Kamble *et al.*, 2023) (Figure 1.10e). Yet there is no effect on A-type granules, demonstrating a unique role for PHS1 in B-type granule formation. It has been hypothesized to elongate MOS substrates released from A-type granules during B-type granule initiation (Kamble *et al.*, 2023). Whilst no *phs1* mutant has been characterised in barley, the PHS1 protein abundance in barley endosperm is strikingly similar to that of wheat (Cuesta-Seijo *et al.*, 2017). This could suggest that the role of PHS1 is conserved across the Triticeae.

Reverse genetics has also been used to identify genes involved in granule formation in barley using missense TILLING (Targeting-Induced Local Lesions In Genomes) mutants. Mutations in *LIMIT DEXTRINASE (LDA)* or *SS1* resulted in significantly altered percentage of B-type granules with an increase in the *lda* mutant and both increases and decreases in the *ss1* mutants, depending on the specific mutation (Sparla *et al.*, 2014). These findings might seem contradictory; however, the TILLING strategy can generate both loss of function and gain of function alleles. Therefore, analysis of a single mutant phenotypes makes it impossible to conclude the specific effect of either LDA or SS1 on granule initiation, apart from the fact that they likely participate in the process. No full knockouts have been described for either *LDA* or *SS1* in barley or any other Triticeae species, so how these fit into our model of starch granule initiation is not clear.

Granule initiation in the Triticeae occurs in a spatiotemporal pattern, which involves changes in amyloplast morphology. There is therefore interest in how plastid morphology affects granule initiation. One protein of particular interest is PARALOG OF ARC6 (PARC 6; where ARC stands for Accumulation and Replication of chloroplasts) which is a component of the plastid division machinery (Glynn *et al.*, 2009). Consistent with its role in plastid division, the *parc6* knockout mutants of durum wheat have larger amyloplasts in the endosperm (Esch *et al.*, 2023). This is accompanied by larger A-type and B-type granules compared to the wild-type plants (Esch *et al.*, 2023). Additionally, *parc6* plants have an increase in B-type granule content, which is due to both the increase in the size of the B-type granules and an increase in the relative number of B-type granules (Esch *et al.*, 2023). Hence, altering amyloplast size alters both the number of granules initiated and how big these granules can grow.

Through these genetic studies, a working model of granule initiation in the Triticeae has been developed (Figure 1.11). In this model A-type granules are initiated soon after anthesis with involvement from SS4 and BGC1. Whether the A-type granule initiation is a primary or secondary initiation (Section 1.5.1) is unknown, as *de novo* MOS production has not been demonstrated. B-type granule initiation is repressed during early grain development by MRC. As MRC expression decreases, B-type granules start to initiate in the main body of the amyloplast and stromules. This process relies on BGC1 and PHS1. The initiation of B-type granules are secondary initiation events (Section 1.5.1) as they rely on pre-existing MOS primers, which are most likely released directly from A-type granules through the activity of DBEs. However, there are still large gaps in our understanding of granule initiation. One of the most obvious gaps is which enzymes generate the MOS primers for B-type granule formation, and what controls the overall ratio of A-type to B-type granules?

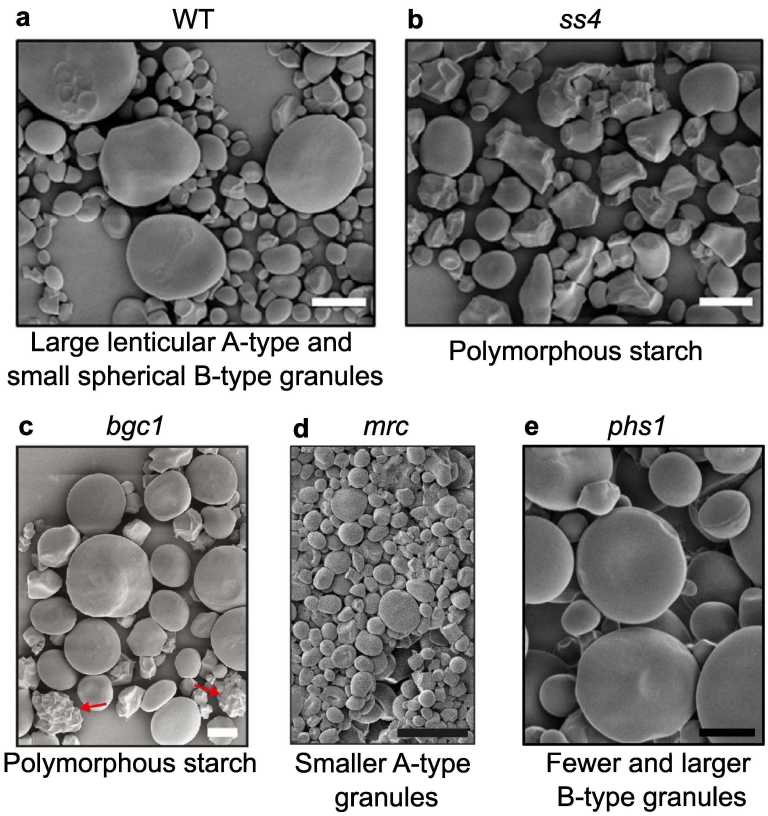


Figure 1.10 Mutations in key granule initiation genes disrupts normal A-type and B-type formation in the endosperm of durum wheat. Scanning electron micrographs of purified endosperm starch from (a) wild-type durum (*T. turgidum*) wheat, (b) *starch synthase 4*, (c) *b-granule content 1*, (d) *myosin-resembling chloroplast protein* and (e) *phosphorylase 1*, where all mutants are in the durum wheat background. Some polyhedral granules in the *bgc1* starch are highlighted with red arrows. Descriptions of the starch granule phenotype are provided underneath each micrograph. Scale bars = 10 μm in (a), (b), (c) and (e) and 40 μm in (d). Images are adapted from Hawkins *et al.* (2021), Chen *et al.* (2024) and Kamble *et al.* (2023) under CC-BY licenses. The wild-type looked comparable between all papers so for simplicity only one wild-type image is displayed here.

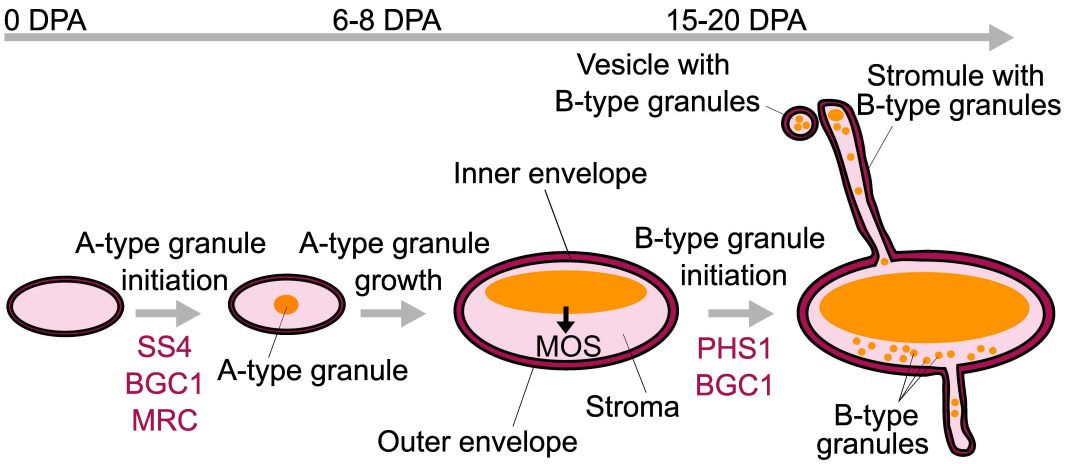


Figure 1.11 Current model of granule initiation in the Triticeae. Model based on Figure 9 from Kamble *et al.* (2023). A-type granules are initiated early, soon after 0 DPA, with involvement from SS4 and BGC1, whilst MRC represses B-type granule development during this stage. At approximately 15-20 DPA, B-type granules are initiated in the main body of the amyloplast, stromules and surrounding vesicles. This relies on PHS1 and BGC1 and may be primed by MOS released from A-type granules.

1.6 Studying granule initiation in the Triticeae

Previous studies have been effective in identifying some genes involved in the initiation of A- and B-type granules in wheat, but there are still many unanswered questions. For example, what determines granule size and morphology? And what affects the levels of MOS that fuel granule initiations? Addressing these unknowns is where my attention will turn for the remainder of the thesis.

1.6.1 Why study granule initiation in the Triticeae?

Identifying genes involved in the granule initiation process is of high priority especially in the endosperm of the Triticeae, given the important implications of A- and B-type granules on grain quality. The majority, approximately 67%, of Triticeae starch is used directly as food (BeMiller and Whistler, 2009). The rest is funnelled into feed for livestock or industrial uses, including the production of paper, adhesives and biodegradable polymers (BeMiller and Whistler, 2009). Whilst A-type and B-type granules do not appear to have any physiochemical differences (Saccomanno *et al.*, 2022), the overall granule size distribution of Triticeae starch influences the end use as high or low B-type granule contents are desirable for specific applications.

In the food industry, the granule size distribution affects breadmaking and noodle production, with high B-type granule contents being negatively correlated with overall

quality (Soulaka and Morrison, 1985; Park *et al.*, 2005; Guo *et al.*, 2014). Conversely, pasta firmness, a key quality trait, increases when made with reconstituted flours containing a high proportion of B-type granules (Soh *et al.*, 2006). Beyond food, B-type granule content is important for the brewing industry as B-type granules are resistant to digestion during malting and can gelatinise. This affects wort viscosity and results in a hazy beer (Bathgate and Palmer, 1972; Tillett *et al.*, 1993).

Outside the food industry, B-type granule content can also affect the mechanical characteristics of biodegradable films, which are a potential alternative to plastics. Starch films made from only A-type granules had 14% greater elongation than those made from only B-type granules (Montaño-Leyva *et al.*, 2008). In papermaking, small B-type granules are advantageous for coating paper as they can fill pores on the paper's surface (Shevkani *et al.*, 2017). In contrast, for the production of carbonless copy paper large, ≥ 22 μm, A-type granules can act as a protective material; they are desirable for this as they are approximately 2.5 times the size of the micro-encapsulated ink particles (Bond *et al.*, 1975; BeMiller and Whistler, 2009).
As granule size distribution has such a large impact on starch end uses, dissecting the genetic determinants of this trait remains a high priority. This knowledge could allow creation of novel size distributions in commercially relevant cultivars through genetic modification or engineering techniques. Alternatively, naturally occurring variation in these traits could be introduced from wild relatives by introgression or synthetic lines. Either way, these could be powerful approaches to generate industrially useful starch granule size distributions.

1.6.2 Natural diversity as a resource for studying starch formation

While the studies above have used induced mutations to elucidate components involved in granule initiation; natural diversity is a relatively unexplored resource to find novel genes and variation that affect starch granule formation. The Triticeae is a tribe of the Poaceae grass family, diverging around 35 million years ago (Huang *et al.*, 2002) and contains both undomesticated and domesticated species. It can be subdivided into two subtribes: the Hordeineae and the Triticineae, which both contain nine genera (Feldman and Levy, 2015). The Hordeineae is referred to as the barley lineage as its major crop species is barley, whereas Triticineae contains both wheat and rye. Whilst barley and rye are diploid, wheat has diploid, tetraploid and hexaploid species. The only domesticated diploid species of wheat is *Triticum monococcum* (genome composition = AA), often referred to as einkorn wheat (Ahmed *et al.*, 2023); however it is worth noting that undomesticated *T. monococcum* varieties also exist.

The evolution of tetraploid and hexaploid wheat was driven by allopolyploidisation between different diploid species (Figure 1.12). The initial hybridisation event occurred between *Triticum urartu* (genome composition = AA), a sister species of *T. monococcum*, and an extinct relative of *Aegilops speltoides* (genome composition = BB) approximately 0.5-0.8 million years ago (Marcussen *et al.*, 2014). This produced the tetraploid wheat *Triticum turgidum* (AABB), which is sometimes referred to as emmer or durum wheat (Marcussen *et al.*, 2014). Around 10,000 years ago, *Triticum turgidum* underwent hybridisation with *Aegilops tauschii* (genome composition = DD) to produce hexaploid *Triticum aestivum* (AABBDD), which is also known as bread wheat (Marcussen *et al.*, 2014). These hybridisation events are thought to have occurred in the Fertile Crescent, and this is now referred to as the domestication centre for modern day wheat (Salamini *et al.*, 2002).

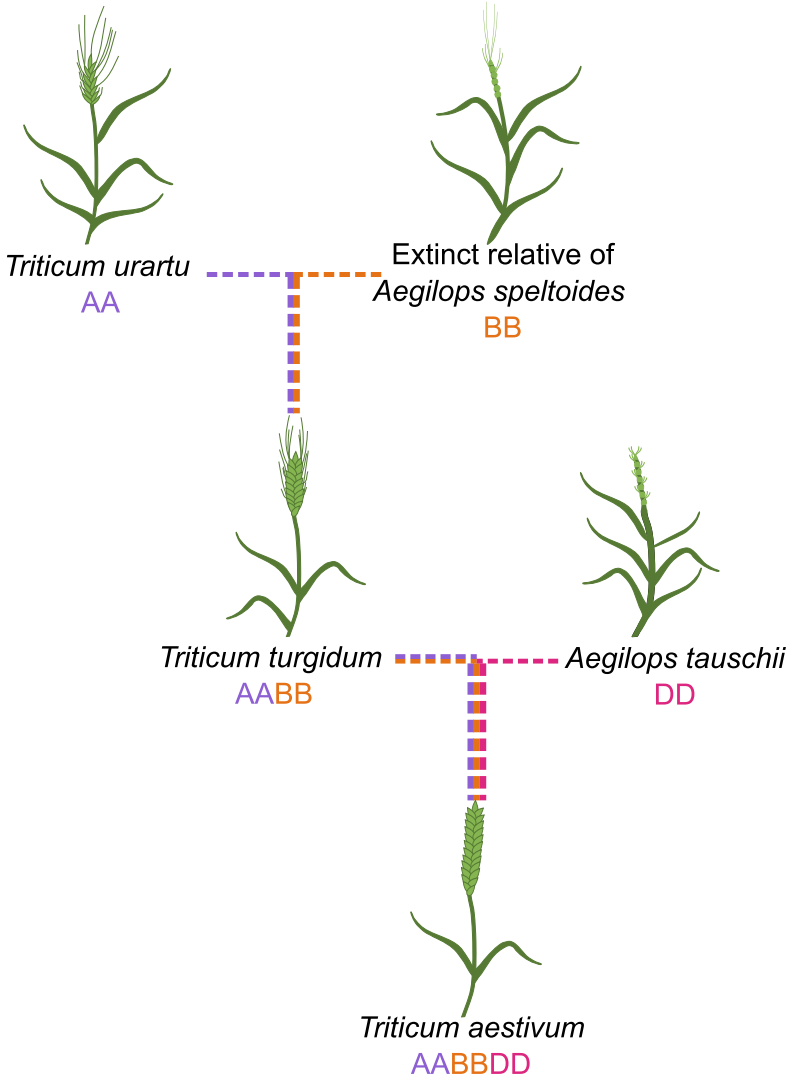


Figure 1.12 The evolution of *Triticum aestivum* involved two hybridisation events. Diagrammatic representations of the *T. aestivum* ancestors. Hybridisation between *T. urartu* and an extinct relative of *Ae. speltoides* lead to the formation of *T. turgidum*, this subsequently hybridised with *Ae. tauschii* forming *T. aestivum*.

There is natural diversity in starch granule size distribution within the Triticeae, particularly in the B-type granule content by volume, which is the percentage of the total volume of starch granules made up by B-type granules (Table 1.1). Between species, there is variation in B-type granule content, with barley, on average, having lower B-type granule contents than rye and wheat. Within modern wheat cultivars there is also variation in B-type granule content. However, few lines have been reported to have low B-type granule contents; even in Zhang *et al.* (2016) there were fewer than 10 lines where the B-type granule content was below 10%. However, in the wild grass species *Ae. peregrina* there are lines with an extremely low, or even zero, B-type granule content (Stoddard and Sarker, 2000). At the other end of the spectrum, some undomesticated species of *T. monococcum* have greater B-type granule contents than reported in wheat cultivars or landraces (Stoddard, 1999a). This reveals the importance of wild grass species for the discovery of novel starch granule size distributions.

In addition to variation in granule size and content, there are studies which report variation in amylose content (Table 1.1). What is striking about the variation, is that there are many lines with very little amylose, these are referred to as *waxy* and have mutations in *GBSS* (Nakamura *et al.*, 1995), but only a few lines with extremely high amylose contents. The highest recorded amylose among wild species is 35% in *Ae. peregrina* (Stoddard and Sarker, 2000), and 34% in *Ae. tauschii* (Mohammadkhani *et al.*, 1998), which is no higher than amylose content in domesticated wheat, 36% (Stoddard and Sarker, 2000).

Table 1.1 Variation in endosperm starch granules in the Triticeae. Naturally occurring variation in B-type granule content and amylose content in the Triticeae was identified through a literature search. Table is arranged according to species and is subdivided into cultivars and wild populations. Different populations, even if they are from the same study, are on different rows. Within each species the data are presented in chronological order based on publication date. Where information was provided on the population the information is included in the table. The number of individuals in each population is given in brackets after the population description or, where the numbers were different for different starch properties in brackets, after the starch granule property. * = when examined under SEM no B-type granules are observed, so could be as low as 0 %, n/a = not measured.

Species	Description of population (number of individuals)	B-type granule content range (%)	Amylose content range (%)	Reference
Triticum aestivum – cultivars	Varieties available in New Zealand in 1979 (59)	21 – 40	n/a	Dengate and Meredith (1984)
	Japanese cultivars (4)	n/a	17 – 24	Hayakawa *et al.* (1997)
	Australian varieties (130)	23 – 50	n/a	Stoddard (1999a)
	Australian varieties (2)	n/a	28 – 30	Black *et al.* (2000)
	Hard red winter wheat (98)	13 – 49	n/a	Park *et al.* (2009)
	Soft red spring wheat (99)	34 – 58	n/a	Park *et al.* (2009)
	Winter wheat (7)	35 – 47	13 – 15	Dai *et al.* (2009)
	Chinese varieties - *waxy* (3)	n/a	1 – 2	Wang *et al.* (2015)
	Chinese varieties	7 – 28 (345)	30 – 33 (4)	Zhang *et al.* (2016)
	Varieties and advanced lines from Yellow and Huai Valley Facultative Wheat Region (166)	16 – 30	n/a	Li *et al.* (2017)

Table 1.1 Variation in endosperm starch granules in the Triticeae. (continued)

Species	Description of population (number of individuals)	B-type granule content range (%)	Amylose content range (%)	Reference
Triticum aestivum – cultivars (continued)	Chinese cultivars, founder parents and breeding lines (205)	n/a	16 – 23	Chen *et al.* (2019)
	No additional information provided – *waxy* (2)	n/a	1.7 - 2	Li *et al.* (2020)
Triticum aestivum – landraces	Landraces from Turkey, Iran, India, China and Israel-Jordan-Syria (100)	17 – 48	n/a	Stoddard (1999a)
	Landraces from Afghanistan, China, Egypt, Ethiopia, India, Iran, Syria, and Turkey (133)	26 – 40	24 – 30	Black *et al.* (2000)
Triticum aestivum – cultivars or landraces unknown	Cultivars or landraces (no distinguishing between the two in data analysis) from Australia and the fertile crescent, India and China (200)	n/a	18 – 36	Stoddard and Sarker (2000)
	Recombinant inbred line	12 – 28	n/a	Feng *et al.* (2013)
Triticum turgidum – cultivars or landraces unknown	Subspecies *dicoccoides* and *turgidum* (351)	n/a	19 – 31	Mohammadkhani *et al.* (1998)
	Subspecies *dicoccoides* (50)	18 – 47	n/a	Stoddard (1999a)
	Subspecies *carthlicum, dicoccum, durum, polonicum* and *turanicum* (81)	17 – 40	n/a	Stoddard (1999a)
	Durum wheats (12)	29 – 39	24 – 34	Konik-Rose *et al.* (2009)

Table 1.1 Variation in endosperm starch granules in the Triticeae. (continued)

Species	Description of population (number of individuals)	B-type granule content range (%)	Amylose content range (%)	Reference
Triticum monococcum	Subspecies *monococcum, boeolicum, aegilopoides, thaoudar* and *urartu* (247)	n/a	15 – 28	Mohammadkhani *et al.* (1998)
	Subspecies *aegilopoides, boeoticum, monococcum, sinskajae,* and *thaoudar* (104)	23 – 62	n/a	Stoddard (1999a)
Aegilops tauschii	Subspecies includes *strangulata, anathera, meyeri,* and *typica* (252)	n/a	21 – 34	Mohammadkhani *et al.* (1998)
	Subspecies includes *strangulata, eusquarrosa, anathera, meyeri,* and *typica* (68)	15 – 38	n/a	Stoddard (1999a)
	No additional information provided (17)	21 – 36	24 – 30	Konik-Rose *et al.* (2009)
Aegilops peregrina	No additional information provided (6)	5* –13	23 – 35	Stoddard and Sarker (2000)
Hordeum vulgare – cultivars or landraces unknown	Japanese varieties – *waxy* (8)	n/a	1 – 13	Banks *et al.* (1970)
	No additional information provided (29)	6 – 31	n/a	Goering *et al.* (1973)
	No additional information provided (14)	4 – 8	n/a	Oliveira *et al.* (1994)
	No additional information provided – *waxy* (5)	n/a	0 – 5	Asare *et al.* (2012)
	No additional information provided (10)	25 – 45	24 – 27	Jaiswal *et al.* (2014)
	European spring varieties (254)	n/a	17 – 31	Shu and Rasmussen (2014)

Table 1.1 Variation in endosperm starch granules in the Triticeae. (continued)

Species	Description of population (number of individuals)	B-type granule content range (%)	Amylose content range (%)	Reference
Hordeum vulgare – cultivars or landraces unknown (continued)	No additional information provided (100)	n/a	16 – 31	Li *et al.* (2021)
Secale cereale – cultivars or landraces unknown	No additional information provided (233)	n/a	12 – 28	Mohammadkhani *et al.* (1998)
	No additional information provided (50)	21 – 39	n/a	Stoddard (1999a)
	No additional information provided (unknown)	10 – 40	22 – 26	Németh and Tömösközi (2021)

1.6.3 Association genetics to discover genomic loci and regions influencing endosperm starch

Association genetics involves correlating phenotypic differences with genetic polymorphisms to identify genomic regions associated with the trait of interest. Association genetics has been used in plants for years to study complex traits, such as yield (Stuber *et al.*, 1987), and recent developments in genome sequencing have made it easier and more powerful.

A common form of association genetics is quantitative trait loci (QTL) mapping. This relies on a segregating population. Historically, biparental mapping populations have been generated by crossing individuals with contrasting phenotypes and then analysing the F_2 generation. However, these populations often result in poor mapping resolution due to low recombination rates. To increase the amount of recombination, more complex crossing and selfing schemes can be used to produce near isogenic lines (Muehlbauer *et al.*, 1988), recombinant inbred lines (Pollard, 2012) and advanced intercross lines (Darvasi and Soller, 1995) (Figure 1.13a-c). These populations provide better resolution for QTL mapping, but this comes at the expense of the additional generations needed to produce these populations (Jamann *et al.*, 2015). A limitation of all these populations is that they stem from just two individuals so the allele diversity is narrow. To overcome this, it is possible to extend the ideas from biparental mapping populations to include more parents. This approach has been employed in the production of multi-parent advanced generation inter-cross (MAGIC) populations (Scott *et al.*, 2020) (Figure 1.13d). Once the segregating population has been generated, genotyping is conducted to characterise the allele patterns across the genome of each individual. In early QTL mapping studies, restriction fragment length polymorphisms were used as genetic markers and successfully identified QTLs influencing fruit mass and pH in tomatoes (Paterson *et al.*, 1988). More recently, advances in genome sequencing have enabled easier identification of SNPs (single nucleotide polymorphisms), which are exploited as genetic markers in SNP arrays (Ganal *et al.*, 2014). Once genotyped, this information is correlated with phenotypic data for the trait of interest to identify markers, and hence QTLs, associated with the trait. There are numerous examples of QTL mapping being used to identify QTLs associated with endosperm starch features in the Triticeae (Table 1.2). However, identification of the causative gene within these QTLs, which are often large in size due to low recombination rates, can be challenging and most studies do not progress this far. One exception is Howard *et al.* (2011), included in Table 1.2, which performed QTL mapping on a population produced from a cross between *Aegilops peregrina*,

which lacks B-type granules (Table 1.1), and a synthetic tetraploid *Aegilops*, with normal A-type and B-type granules. They found a QTL on chromosome 4S and identified the causal gene to be *BGC1* (Chia *et al.*, 2020; Hawkins *et al.*, 2021). This demonstrates the potential of QTL mapping, especially when additional variation can be introduced from wild, undomesticated species.

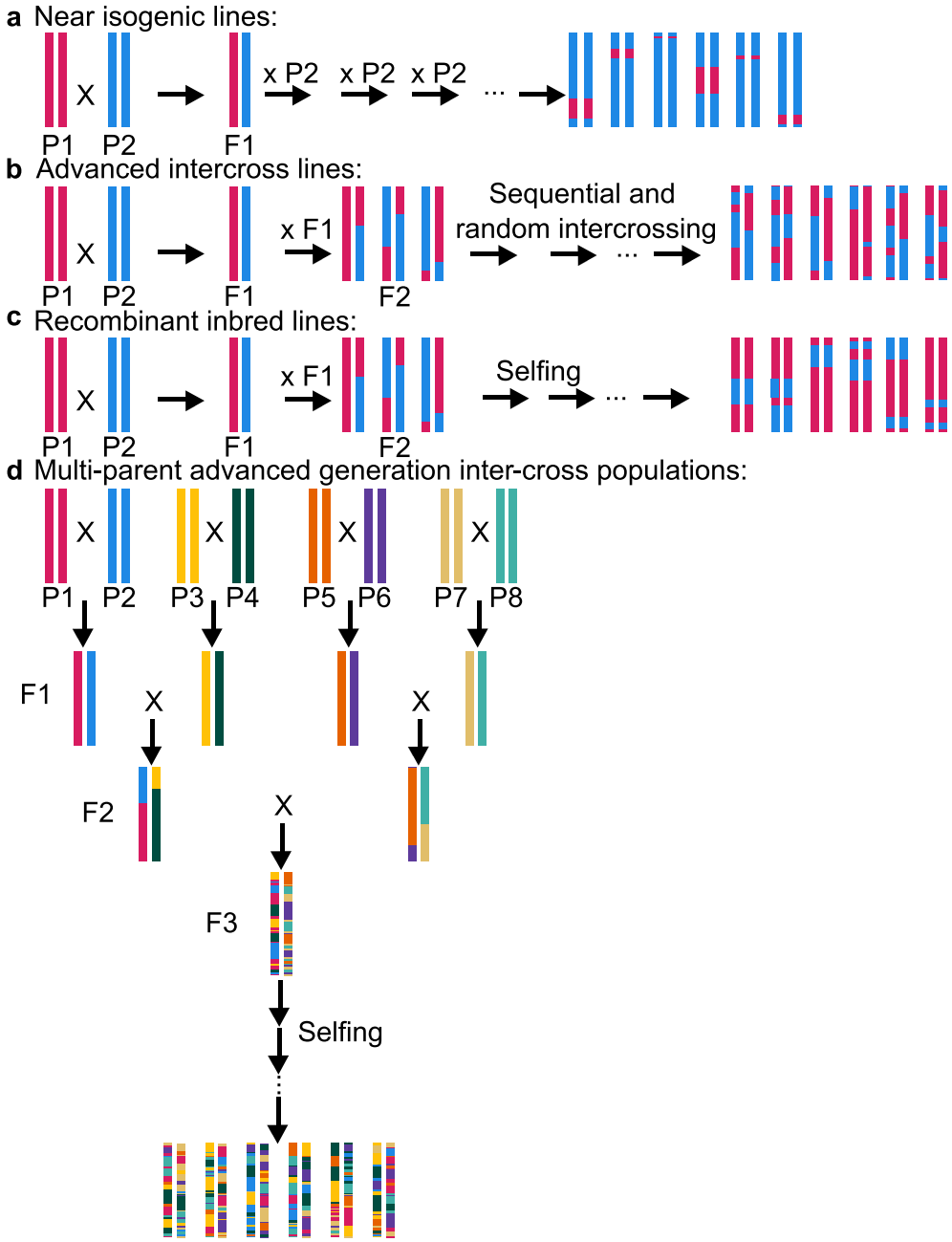


Figure 1.13 Populations for quantitative trait loci mapping. Diagrams showing how: (a) near isogenic lines, (b) advanced intercross lines, (c) recombinant inbred lines, (d) multi-parent advanced generation inter-cross populations, are generated through crossing. Crosses are represented by arrows, with multiple crosses represented by ellipses and chromosomes from different parents are represented with different colours.

Table 1.2 QTLs associated with different endosperm starch properties, in the Triticeae, as determined by QTL mapping. A literature search identified QTLs with links to endosperm starch in the Triticeae. The starch properties arranged in alphabetical order, are named as reported in the original references, with individual studies for each starch property listed in order of publication.

Endosperm starch property	Species and description of population (if available)	Number of loci and chromosome located on	Reference
A:B ratio	*Triticum aestivum* – doubled haploid progeny of two crosses	1 QTL on chromosome 4B	Batey *et al.* (2001)
Amylose content	*Triticum aestivum* – recombinant inbred line	2 QTLS across chromosomes 2A and 2D	Sun *et al.* (2008)
	Triticum aestivum – doubled haploid population of a cross between two elite Chinese cultivars	11 QTLs on chromosomes 1B, 3A, 3B, 4A, 5D and 7B	Tian *et al.* (2015)
	Triticum aestivum – doubled haploid population	3 QTLs across chromosomes 3A, 3B and 4A	Deng *et al.* (2018)
	Triticum aestivum – recombinant inbred line	4 QTLs across chromosomes 3B, 4A, 7A and 7D	Deng *et al.* (2018)
	Triticum aestivum – $F_{2:3}$ population from a cross of a high amylose what variety and a high yielding wheat variety	36 QTLs across chromosomes 1A, 2A, 2B, 3A, 3B, 4A, 4B, 4D, 5B, 6B, 7A, 7B, and 7D	Mishra *et al.* (2021)
	Triticum aestivum – recombinant inbred line	7 QTLs across chromosomes 1B, 1D, 3B, 4A, 7A and 7D	Guo *et al.* (2023)

Table 1.2 QTLs associated with different endosperm starch properties, in the Triticeae, as determined by QTL mapping. (continued)

Endosperm starch property	Species and description of population (if available)	Number of loci and chromosome located on	Reference
A-type granule content	*Hordeum vulgare* – doubled haploid lines from a F_1 generation	1 QTLs on chromosome 2	Borém *et al.* (1999)
	Triticum aestivum – recombinant inbred line	3 QTLs across chromosomes 1D, 4A and 7B	Feng *et al.* (2013)
A-type granule size	*Hordeum vulgare* – doubled haploid lines from a F_1 generation	2 QTLs across chromosomes 2, 7	Borém *et al.* (1999)
	Triticum aestivum - recombinant inbred lines	4 QTLs across chromosomes 1B, 4D and 7A	Igrejas *et al.* (2002)
B-type granule content	*Aegilops* – F_2 generation of crosses between *Aegilops peregrina* and synthetic tetraploids	1 QTL on chromosome 4	Howard *et al.* (2011)
B-type granule shape	*Hordeum vulgare* – doubled haploid lines from a F_1 generation	2 QTLs across chromosomes 2, 4	Borém *et al.* (1999)
B-type granule size	*Triticum aestivum* - recombinant inbred lines	3 QTLs across chromosome 4D	Igrejas *et al.* (2002)
Total starch content	*Triticum aestivum* – soft winter wheat, advanced breeding lines	3 QTLs across chromosomes 1D, 5A and 5B	Reif *et al.* (2011)
	Triticum aestivum – doubled haploid population from a cross between two elite Chinese cultivars	7 QTLs across chromosomes 2A, 3A, 3B, 4A and 5D	Tian Bin *et al.* (2011)

Table 1.2 QTLs associated with different endosperm starch properties, in the Triticeae, as determined by QTL mapping. (continued)

Endosperm starch property	Species and description of population (if available)	Number of loci and chromosome located on	Reference
Total starch content (continued)	*Triticum aestivum* – doubled haploid population	1 QTL on chromosome 3A	Deng *et al.* (2015)
	Triticum aestivum – recombinant inbred line from a cross of two winter wheat cultivars	6 QTLs on chromosomes 1B, 1D, 4A, 5b, 7B and 7D	Deng *et al.* (2015)
	Triticum aestivum – doubled haploid population of a cross between two elite Chinese cultivars	14 QTLs across chromosomes 1B, 2A, 3A, 3B, 4A and 5D	Tian *et al.* (2015)
	Triticum aestivum – doubled haploid population	1 QTL on chromosome 4A	Deng *et al.* (2018)
	Triticum aestivum – recombinant inbred line	3 QTLs across chromosomes 4A, 7A and 7D	Deng *et al.* (2018)

The aforementioned drawbacks of QTL mapping and the ‘boom’ in genomic technologies, including the reduction in the cost of whole genome sequencing, have led to the development of other association mapping approaches, most notably genome wide association studies (GWAS). These have become an increasingly popular association genetics methodology since their use in plants was first demonstrated in *Arabidopsis thaliana* (Aranzana *et al.*, 2005). The strategy has subsequently been applied to crop plants, with the first example being the identification of loci associated with oleic acid content in maize (Beló *et al.*, 2008). GWAS are, in principle, like QTL mapping studies, with the major difference being the population used. A GWAS uses a diverse panel of individuals, or pre-existing breeding populations, rather than a structured population which provides multiple advantages (Zhu *et al.*, 2008). Firstly, these populations are usually naturally occurring, so it does not require multiple generations of crossing, which can save considerable experimental time. Secondly, as the panel does not originate from a limited number of parents and are a result of years of natural crossing, they contain large amounts of historical recombination and are allelically diverse. This can be exploited to map regions to a greater resolution compared to QTL mapping, where there are often relatively few recombination events during the generation of the population.

Much like QTL mapping, during a GWAS, the population is phenotyped and genotyped. The genotyping step often relies on SNPs, as they can be detected easily in high throughput SNP-arrays and are found across the genome; the 2007 *Arabidopsis thaliana* HapMap detected a SNP approximately once every 166 base pairs (bp) (Clark *et al.*, 2007). Following genotyping, statistical analysis is conducted to identify regions associated with the trait in question. This is more complex than the corresponding stage in QTL mapping, as population structure must be accounted for to reduce the chance of false positives due to inter-relatedness. Common ways of achieving this include the STRUCTURE algorithm which relies on the genotyping data (Pritchard *et al.*, 2000), or a principal component analysis (PCA) (Tibbs Cortes *et al.*, 2021).

GWAS methodologies have developed immensely since the original studies in plants. One clear difference is that there is no longer a reliance on SNP markers for genotyping, this is beneficial as SNPs have disadvantages, including an inability to detect copy number variation (CNV), insertions and deletions (indels) and translocations (Karikari *et al.*, 2023). An alternative genotypic marker is the presence or absence of k-mers (Gupta, 2021); these are nucleotide sequences of length k,

which are derived from raw sequencing reads. K-mers offer a distinct advantage because, as well as identifying SNPs, they enable structural variants to be discovered (Voichek and Weigel, 2020). They also reduce the reliance on the availability of a reference genome as no prior knowledge is needed to generate or use them. While there has been an advancement in the genomic markers that can be used, nowadays genomic markers are not the only option and other markers can be used instead. This includes using epigenomic marks, such as methylation or histone modifications, in an epigenome-wide association study (EWAS) (Gahlaut *et al.*, 2020). Alternatively, gene expression data can be used in transcriptome-wide association studies (TWAS) (Tibbs Cortes *et al.*, 2021). Combining results from multiple association studies can increase power compared to conducting a GWAS alone, while also making candidate gene identification easier (Kremling *et al.*, 2019).

In recent years, many of the association studies that have investigated endosperm starch have used the traditional SNP-based GWAS approaches in crop plants (Table 1.3). The apparent inconsistencies between the studies could stem from the use of different populations with different genetic compositions. Whilst the studies in Table 1.3 highlight the success of GWAS in the Triticeae, this is not exclusive, and GWAS have also been insightful beyond the Triticeae. GWAS have been used to identify loci associated with total starch content in: maize (Liu *et al.*, 2016; Hu *et al.*, 2021; Duan *et al.*, 2023), potato (Schönhals *et al.*, 2017), sweet potato (Haque *et al.*, 2023) and *Sorghum bicolor* (Sapkota *et al.*, 2020). Likewise, GWAS in rice (Praphasanobol *et al.*, 2023), maize (Li *et al.*, 2018b) and sweet potato (Nie *et al.*, 2023) have identified loci associated with amylose content. The vast majority of these published GWAS focus on either amylose content or total starch content, with limited studies investigating starch granule traits (size, morphology). Furthermore, most of these studies have been conducted on domesticated crops or panels with limited diversity. It is widely known that starch has been intensely selected for during domestication of many key species, including wheat and barley. This has led to reductions in allelic diversity in genes related to starch (Chapter Three; Section 3.1.3). Thus, a promising strategy would be to perform these studies on an undomesticated species which is likely to have greater variation in starch-related genes.

Table 1.3 Genomic loci associated with endosperm starch properties identified via GWAS. A literature search was conducted to identify studies that have investigated endosperm starch properties in the Triticeae via a GWAS approach. The number of loci that significantly associated with the starch properties are listed; alongside the chromosome they are located on. The starch properties are arranged in alphabetical order and are named as reported in the original references, with individual studies for each starch property listed in order of publication.

Endosperm starch property	Species and description of population (if available)	Number of loci and chromosome located on	Reference
A-type to B-type ratio	*Triticum aestivum* – advanced lines from Yellow and Huai Valley Facultative Wheat Region	25 loci across chromosomes 2A, 2B, 2D, 3A, 3D, 4A, 4B, 4D, 5A, 6A, 6B, 7A and 7B	Li *et al.* (2017)
Amylopectin content	*Triticum aestivum* – Chinese cultivars, founder parents and breeding lines	23 loci across chromosomes 2A, 2B, 3A, 3B, 4A, 6A, 6B and 7D	Chen *et al.* (2019)
Amylose content	*Hordeum vulgare* – spring barley	20 loci across chromosomes 1, 5, 6 and 7	Shu and Rasmussen (2014)
	Triticum aestivum – Chinese cultivars, founder parents and breeding lines	15 loci across chromosomes 2A, 2B, 3A and 4A	Chen *et al.* (2019)
	Hordeum vulgare	2 loci on chromosome 6	Li *et al.* (2021)
	Triticum aestivum – cultivars, landraces and breeding lines	2 loci across chromosomes 1B and 5A	Guo *et al.* (2023)
Amylose to Amylopectin ratio	*Triticum aestivum* – Chinese cultivars, founder parents and breeding lines	18 loci across chromosomes 1B, 2A, 3B, 3D, 4A, 5B, 6A, 6B and 7B	Chen *et al.* (2019)

Table 1.3 Genomic loci associated with endosperm starch properties identified via GWAS. (continued)

Endosperm starch property	Species and description of population (if available)	Number of loci and chromosome located on	Reference
Percentage volume of A-type or B-type granules	*Triticum aestivum* – advanced lines from Yellow and Huai Valley Facultative Wheat Region	23 loci across chromosomes 2A, 2B, 2D, 3A, 3D, 4A, 4B, 4D, 5A, 5D, 6A, 6B, 7A, 7B and 7D	Li *et al.* (2017)
Total starch content	*Hordeum vulgare* – spring barley	25 loci across chromosomes 1, 2, 3, 4, 5, 6 and 7	Pasam *et al.* (2012)
	Triticum aestivum – Chinese cultivars, founder parents and breeding lines	22 loci across chromosomes 2A, 2B, 3A, 3B, 4A, 6A and 6B	Chen *et al.* (2019)
	Triticum aestivum – winter wheat advanced breeding lines	1 locus on chromosome 5B	Tsai *et al.* (2020)
	Hordeum vulgare – spring barley advanced breeding lines	1 locus on chromosome 4	Tsai *et al.* (2020)
	Triticum aestivum – European winter wheat	29 loci across chromosomes 2B, 3A, 3B, 6A	Muqaddasi *et al.* (2020)
	Hordeum vulgare	13 loci across chromosomes 1, 2, 3, 4, 5, 6 and 7	Li *et al.* (2021)
	Secale cereale – advanced inbred lines	87 loci across chromosomes 1, 2, 3, 4, 5, 6 and 7	Siekmann *et al.* (2021)
	Triticum aestivum – cultivars, breeding lines and landraces	3 loci on chromosomes 1B, 3A and 6A	Lou *et al.* (2021)

1.7 Aims of my thesis

The overall aim of my thesis is to use the potential of natural diversity and association mapping to identify novel genes and variation that influence endosperm starch synthesis in the Triticeae. Forward and reverse genetics have been useful in identifying many factors involved in this process, however there are many aspects of granule initiation and granule morphology that are poorly understood, including what is generating the MOS for B-type granule initiation, and what controls the overall number, size and ratio of A-type to B-type granules.

I have used natural variation in *Aegilops tauschii* to answer some of these outstanding questions. Chapter Three introduces the *Ae. tauschii* diversity panel and performs a thorough analysis of the endosperm starch, discovering variation in starch granule properties and expands our understanding of starch granule variation in this species. In Chapter Four, I utilise this variation in a GWAS and identify novel loci associated with B-type granule number and diameter. Using publicly available gene expression data and bioinformatic predictions, I identify 13 strong candidate genes which may be influencing B-type granule formation. Finally in Chapter Five, I investigate a candidate identified in the GWAS, *LDA*. I transfer this knowledge over to tetraploid wheat (*Triticum turgidum*) and characterise endosperm starch in *lda* TILLING mutants. I also characterise allelic variation in *LDA* in the *Ae. tauschii* diversity panel and investigate how variation affects enzyme activity using recombinantly expressed proteins. Overall, my results shed new light on the effect of domestication on starch granule traits, the potential of natural populations as a source of useful variation for crop breeding and the role of LDA in MOS metabolism during granule initiation.

Chapter Two – Materials and Methods

The composition of standard components are listed in Appendix One; Table S1.

2.1 Plant material and growth

2.1.1 *Aegilops tauschii*

Grains of 117 accessions (Table 2.1) of *Aegilops tauschii spp. strangulata* (lineage two) were obtained from Cristobal Uauy's group (John Innes Centre; JIC), from the same batch of grains analysed in Gaurav *et al.* (2021). These were grown as follows: grains were stratified at 4°C on damp filter paper in the dark for two days, and then germinated for four days in daylight and at room temperature (~20°C). Germinated grains were sown into 96 cell trays filled with JIC cereal mix (Table S1) and grown in glasshouse conditions with no supplemental lighting and natural vernalisation conditions with a randomised block design. After nine weeks of growth, plants were transplanted into 2 L pots containing JIC cereal mix. Grains were harvested from the first five emerged spikes at maturity with a minimum of three plants per accession. These grains were used for starch analyses described in Chapter Three.

The TOWWC193 accession was re-grown so that grains could be harvested across grain development. For this, grains were stratified and sown into 40 cell trays as described above. They were placed in a controlled environment room (CER) set to provide 16 hours of light at 400 µmol photons/m^2/s (fluorescent lamps supplemented with LED panels) at a temperature of 20°C and 8 hours of dark at 16°C for 2 weeks. They were transferred to a CER set to vernalisation conditions (5°C, with 8 hours of light (70 µmol photons/m^2/s) and 0% humidity) for 8 weeks, before being transferred back to the original CER and transplanted into 2 L pots containing JIC cereal mix. Plants were assessed for anthesis. Anthesis was determined based on emergence of the first anther from the first spike – this was recorded as day 0. Entire spikes were harvested by cutting the spike at its base and freezing in liquid nitrogen at 10, 15, 19 and 22 DPA. Between five and eight spikes were collected from each plant. A minimum of four biological replicates per time point were collected, each from independent plants. Anthesis and harvesting was conducted at ~10:00 every morning.

All 117 accessions were also grown for five weeks for further analysis in Chapter Five. For this, grains were stratified, germinated and sown as described above for the TOWWC193 accession.

Table 2.1 *Aegilops tauschii* accessions used in this study. Table arranged in alphabetical order based on name, going down each column. To conserve space, columns are shown adjacent to each other rather than on separate pages.

Accessions used			
TOWWC002	TOWWC058	TOWWC117	TOWWC162
TOWWC003	TOWWC059	TOWWC118	TOWWC163
TOWWC004	TOWWC060	TOWWC119	TOWWC164
TOWWC005	TOWWC061	TOWWC120	TOWWC166
TOWWC007	TOWWC063	TOWWC123	TOWWC167
TOWWC008	TOWWC064	TOWWC124	TOWWC168
TOWWC009	TOWWC069	TOWWC125	TOWWC169
TOWWC010	TOWWC070	TOWWC126	TOWWC171
TOWWC011	TOWWC083	TOWWC127	TOWWC172
TOWWC012	TOWWC086	TOWWC129	TOWWC173
TOWWC013	TOWWC087	TOWWC130	TOWWC176
TOWWC016	TOWWC088	TOWWC131	TOWWC177
TOWWC017	TOWWC089	TOWWC133	TOWWC178
TOWWC020	TOWWC090	TOWWC134	TOWWC179
TOWWC021	TOWWC092	TOWWC135	TOWWC180
TOWWC022	TOWWC095	TOWWC137	TOWWC182
TOWWC023	TOWWC096	TOWWC138	TOWWC183
TOWWC025	TOWWC097	TOWWC139	TOWWC186
TOWWC026	TOWWC098	TOWWC140	TOWWC187
TOWWC027	TOWWC099	TOWWC141	TOWWC191
TOWWC028	TOWWC100	TOWWC143	TOWWC193
TOWWC034	TOWWC103	TOWWC144	
TOWWC040	TOWWC104	TOWWC145	
TOWWC042	TOWWC105	TOWWC147	
TOWWC043	TOWWC106	TOWWC148	
TOWWC044	TOWWC108	TOWWC149	
TOWWC045	TOWWC109	TOWWC153	
TOWWC046	TOWWC110	TOWWC154	
TOWWC050	TOWWC112	TOWWC155	
TOWWC051	TOWWC113	TOWWC156	
TOWWC056	TOWWC114	TOWWC157	
TOWWC057	TOWWC115	TOWWC160	

2.1.2 *Triticum aestivum*

Grains of 16 *Triticum aestivum* cultivars (Table 2.2) were obtained from Cristobal Uauy's group (JIC). These had been sown into 96 cell trays containing JIC cereal mix and grown for 2 weeks in glasshouse set to 16 hours of light at 20°C and 8 hours of dark at 16°C. They were then transferred to the vernalisation CER for ~8 weeks. The plants were then transferred to 9 cm pots containing JIC cereal mix and moved back to the glasshouse. Grains were harvested at plant maturity.

Table 2.2 *Triticum aestivum* used in this study. Table arranged in alphabetical order based on name. The origin, growth habit and whether it is a cultivar or landrace is provided in the description column with the reference for this information in the righthand column. CIMMYT = International Maize and Wheat Improvement Centre.

Name	Description	Reference
ArinaLrFor	Swiss winter wheat, cultivar	Walkowiak *et al.* (2020)
Baj	Spring wheat from CIMMYT, cultivar	Yazdani *et al.* (2023)
Cadenza	British spring wheat, cultivar	Walkowiak *et al.* (2020)
CDC Landmark	Canadian spring wheat, cultivar	Walkowiak *et al.* (2020)
CDC Stanley	Canadian spring wheat, cultivar	Walkowiak *et al.* (2020)
Chinese Spring	Chinese spring wheat, landrace	Sears and Miller (1985)
Claire	British winter wheat, cultivar	Walkowiak *et al.* (2020)
Jagger	American winter wheat, cultivar	Walkowiak *et al.* (2020)
Julius	German winter wheat, cultivar	Walkowiak *et al.* (2020)
Lancer	Australian spring wheat, cultivar	Walkowiak *et al.* (2020)
Mace	Australian spring wheat, cultivar	Walkowiak *et al.* (2020)
Norin61	Japanese facultative spring wheat, cultivar	Walkowiak *et al.* (2020)
Paragon	British spring wheat, cultivar	Walkowiak *et al.* (2020)
Robigus	British winter wheat, cultivar	Walkowiak *et al.* (2020)
SY Mattis	French winter wheat, cultivar	Walkowiak *et al.* (2020)
Weebil	Spring wheat from CIMMYT, cultivar	Walkowiak *et al.* (2020)

2.1.3 Triticum turgidum – lda mutants

Brendan Fahy used the wheat *in silico* TILLING database (https://dubcovskylab.ucdavis.edu/wheat-tilling (Krasileva *et al.*, 2017)) to select *Triticum turgidum cv.* Kronos plants with mutations in the A (K2607) or B (K3912 and K983) genome copy of *LDA* (*TRITD7Av1G039520*, *TRITD7Bv1G013060*). These were obtained from the JIC Germplasm Resource Unit (https://www.seedstor.ac.uk/). The K2607 line was crossed with K3912 and K983 to produce the *lda-1* and *lda-2* lines respectively. Deoxyribonucleic acid (DNA) extraction and Kompetitive Allele Specific PCR (KASP) genotyping was conducted by Richard Goram at the JIC genotyping platform. The KASP markers used are listed in Table 2.3. In the F_2 generation, for both crosses, plants were obtained which were wild-type segregants (AABB), single homeolog mutants (*aa*BB or AA*bb*) and double homeolog mutants (*aabb*).

Grains were stratified at 4ºC on damp filter paper in the dark for 3 days. They were transferred to room temperature (~21ºC) and daylight for a further 3 days to germinate. Germinated grains were sown into 96 cell trays containing JIC cereal mix and placed into a CER. They were grown for 4 weeks and then transplanted into 9 cm pots filled with JIC cereal mix. The CERs were set up to provide 16 hours of light at 400 µmol photons/m^2/s (fluorescent lamps supplemented with LED panels) at a temperature of 20ºC and 8 hours of dark at 16ºC. The relative humidity was constant throughout the light-dark cycle at 65%. Plants were harvested ~4 months after sowing, with plant height and tiller number measured concurrently.

Table 2.3 KASP primers used to genotype *lda* TILLING mutants. Primer sequences are given in the 5'-3' direction. A VIC/HEX tail (5'-GAAGGTCGGAGTCAACGGATT) was added to the 5' end of the primers designed to identify the wild-type allele and a FAM tail (5'- GAAGGTGACCAAGTTCATGCT) was added to the 5' end of primers designed to identify the mutant allele. Nucleotides represented in lower case letters are those which distinguish the mutated base (wild-type and mutant alleles), or homoeologous SNPs (common primer).

Gene	Line	Primer sequences
LDA-A1	K2607	Wild-type allele: CGTCGAATGGCGGCATCc Mutant allele: CGTCGAATGGCGGCATCt Common: TCAGGTTCAGGCTGAGAGCg
LDA-B1	K983	Wild-type allele: TGGGATCCATTTATTCCACGTTg Mutant allele: TGGGATCCATTTATTCCACGTTa Common: CATTGCCACAGGTATGGTGAa
	K3912	Wild-type allele: ACTTACGCTGACCAAATACAGg Mutant allele: ACTTACGCTGACCAAATACAGa Common: GCAAATTAATATTCAACCAGTGGGc

2.2 Endosperm starch characterisation

Grain from three individual plants were used for all analyses in the following section.

2.2.1 Purification of endosperm starch

Mature grains (two per plant) were cracked and soaked overnight in 0.5 M NaCl at 4°C. The soaked grains were ground using a ball mill (MM300, Retsch) at 25 hertz (Hz) for 15-20 minutes with five 3 mm glass beads until the homogenate was cloudy and the grain broken up. The homogenate was passed through a 70 µm filter (pluriStrainer Mini, Pluriselect), transferred to 96-deep well plates (Fisher Scientific), and the filtered homogenates were frozen at -20°C until ready for further processing. The homogenates were centrifuged at 3,000*g* for 5 minutes, and the resulting pellets were resuspended in 90% (*v/v*) Percoll, 50 mM Tris-HCl, pH 8. The samples were centrifuged at 2500*g* for 5 minutes and the supernatants discarded. Starch pellets were washed once with 50 mM Tris-HCl, pH 6.8, 10 mM ethylenediaminetetraacetic acid (EDTA), 4% (*v/v*) sodium dodecyl sulfate (SDS), 10 mM dithiothreitol (DTT), once with ddH_2O and once with 100% (*v/v*) ethanol. The resulting starch pellets were air-dried overnight.

2.2.2 Coulter counter for measurement of granule size distributions

Purified starch was resuspended in Isoton II (Beckman Coulter) electrolyte. A Multisizer 4e Coulter counter (Beckman Coulter) fitted with a 70 μm aperture was used to generate relative volume (%) versus diameter plots. The Coulter counter was programmed to measure a minimum of 50,000 particles (starch granules) per sample and all measurements were performed using logarithmic bin spacing. All Coulter counter traces shown here have been transformed to account for differences in bin size and are represented on a linear scale.

2.2.2.1 Fitting mathematical distributions to Coulter counter traces

To calculate mean granule diameters and the relative proportion of B-type granules per volume (henceforth called B-type granule content), curves were fitted to the Coulter counter volume (%) versus diameter plots. A script was written in Python 3.11.5 on a Jupyter notebook (version 6.5.4) to achieve this. Prior to fitting, bin adjustment was conducted by dividing each differential volume data point by its bin diameter to give an adjusted differential volume. This was readjusted to a percentage by dividing by the sum of all adjusted differential volumes.

The normal - normal (N-N) function was defined as:

$$\frac{A_1}{\sigma_1\sqrt{2\pi}}exp(-\frac{(x-\mu_1)^2}{2{\sigma_1}^2}) + \frac{A_2}{\sigma_2\sqrt{2\pi}}exp(-\frac{(x-\mu_2)^2}{2{\sigma_2}^2})$$

with initial values of $A_1 = 2.0, \mu_1 = 2.5, \sigma_1 = 2.2, A_2 = 10.0, \mu_2 = 20.0, \sigma_2 = 6.0.$

The log-normal - log-normal (L-L) function was defined as:

$$\frac{A_1}{x\,\sigma_1\sqrt{2\pi}}exp(-\frac{(\ln x-\mu_1)^2}{2{\sigma_1}^2}) + \frac{A_2}{x\,\sigma_2\sqrt{2\pi}}exp(-\frac{(\ln x-\mu_2)^2}{2{\sigma_2}^2})$$

with initial values of $A_1 = 25.0, \mu_1 = 1.8, \sigma_1 = 0.4, A_2 = 55.0, \mu_2 = 3.1, \sigma_2 = 0.3.$

The log-normal - normal (L-N) function was defined as:

$$\frac{A_1}{x\,\sigma_1\sqrt{2\pi}}\exp(-\frac{(\ln x-\mu_1)^2}{2{\sigma_1}^2}) + \frac{A_2}{\sigma_2\sqrt{2\pi}}\exp(-\frac{(x-\mu_2)^2}{2{\sigma_2}^2})$$

with initial values of $A_1 = 20.0, \mu_1 = 1.8, \sigma_1 = 0.4, A_2 = 50.0, \mu_2 = 20.0\,, \sigma_2 = 6.0.$

For all fittings, the optimise.curve_fit function from the scipy 1.11.1 package was used to fit the distributions, given the previously stated initial values as the starting points. The mean values of the mathematical curves gave the mean diameter of the starch

granules. The integrate.quad function, from scipy 1.11.1, was used to calculate the areas under the curves, and B-type granule content (%) was calculated as:

$$\frac{100 * Area\ under\ B\text{-}type\ granule\ peak}{Area\ under\ A\text{-}type\ granule\ peak + \ Area\ under\ B\text{-}type\ granule\ peak}$$

For all fittings, total uncertainty was calculated as sum of the square root of diagonal of the covariance matrix of errors produced by optimise.curve_fit and standard error of regression (S) was calculated as:

$$\sqrt{\frac{1}{n-2} * \frac{\sum(y_i - \hat{y}_i)^2}{\sum(x_i - \bar{x})^2}}$$

where n is the total number of data points, y_i the measured y value, $\hat{y}_i$ the y value calculated from the fitting, x_i the measured x value and $\bar{x}$ the mean x value. The scripts and instructions for use are freely available online at https://github.com/DavidSeungLab/Coulter-Counter-Data-Analysis.

2.2.2.2 Calculation of B-type granule number

The number of starch granules with a diameter <10.07 µm was calculated directly from the Coulter counter output. This was divided by the total number of starch granules counted and expressed as a percentage.

2.2.3 Amylose content

The amylose content of extracted starch was determined by adapting the method of Washington *et al.* (2000). Briefly, 1 mg of extracted starch was dissolved in 200 µL water and 200 µL of 2 M NaOH and left to incubate at room temperature overnight. The solution was neutralised to pH 7 with 1 M HCl. The solution (5 µL) was diluted in 220 µL of water and 25 µL Lugol's iodine solution (Sigma Life Science) and incubated at room temperature for 10 minutes. The absorbance at 535 nm and 620 nm was measured on a BMG Omega Plate Reader and the apparent amylose content was estimated using the same equation as outlined by Washington *et al.* (2000):

$$Amylose\ content = 1.4935 * \ e^{-2.7029 * \frac{Absorbance\ at\ 620\ nm}{Absorbance\ at\ 535\ nm}}$$

2.2.4 Total starch content

Mature grains (two per plant) were ground to wholemeal flour using a ball mill (MM300, Retsch) at 30 Hz for 3 minutes. Flour (5-10 mg) was dispersed in 20 µL of 80% (*v/v*) ethanol and incubated with 500 µL of thermostable α-amylase in 100 mM sodium acetate buffer (pH 5.0) at 80ºC for 20 minutes with regular shaking of the samples to digest the starch into maltodextrins. Amyloglucosidase was added and samples were

incubated at 5ºC for 35 minutes to digest the maltodextrins into glucose. The samples were centrifuged at 3220*g* for 10 minutes. The supernatant (5 µL) was used in a spectrophotometric hexokinase/glucose-6-phosphate dehydrogenase assay to measure glucose. Calibration curves using glucose standards (ranging from 0-100 nmol) were produced to convert the absorbance into the amount of glucose equivalents, which was subsequently converted into total starch content using the known volumes and flour weights. All the enzymes and reagents for these steps were from the Total Starch Assay Kit (K-TSHK, Megazyme).

2.3 Imaging

2.3.1 Scanning electron microscopy

Purified starch was resuspended in ddH_2O to ~5 mg/mL, and 2.5 µL of the suspension was placed onto a glass cover slip attached to a scanning electron microscopy (SEM) stub and left to air dry. Stubs were sputter coated with gold (8.18 - 8.21 nm) and imaged at 3 kV with a spot size of 3 on a Nova NanoSEM 450 (FEI) SEM instrument.

2.3.2 Photography

Images of the *lda-1* and *lda-2* plants was conducted by Phil Robinson when plants were ~8 weeks old. Photographs of *lda-1* and *lda-2* grains were taken using an Epson Perfection V750 pro scanner and the background has been darkened to make it solid black.

2.4 Bioinformatics

2.4.1 Geographical origins of collection

The locations of origin for 93 of the *Ae. tauschii* accessions was obtained from the JIC Germplasm Resource Unit (GRU) https://www.seedstor.ac.uk/search-browseaccessions.php?idCollection=38.

2.4.2 Phylogenetic analysis

A phylogenetic tree of the *Ae. tauschii* accessions was generated using a Python script from Gaurav *et al.* (2021) (https://github.com/wheatgenetics/owwc/tree/master/phylogenetics). This used 100,000 k-mers to produce an unweighted pair group method with arithmetic mean (UPGMA) phylogenetic tree with 100 bootstraps. The tree was edited to remove the L1 and L3 lineages using the Interactive Tree Of Life interactive web platform (https://itol.embl.de/).

2.4.3 *Genome wide association study*

A k-mer based association mapping method was run as outlined in Gaurav *et al.* (2021) using the published k-mer presence/absence matrix and associated pipeline (https://github.com/wheatgenetics/owwc). The presence/absence matrix contains all k-mers (51 nucleotides) which were observed in a minimum of two accessions with presence represented by 1 and absence represented by 0. Phenotype values from the 117 non-redundant accessions (Table 2.1) were entered into the pipeline as mean values from the three biological replicates. The pipeline was run with between 3,000,000,000 – 5,000,000,000 k-mers and each k-mer had to be present in a minimum of four accessions to be included in the analysis. All parameters were set to the defaults used in Gaurav *et al.* (2021), this included a PCA dimensions value of three which was chosen to account for population structure based on prior analysis with the STRUCTURE algorithm (Gaurav *et al.* 2021). Pre-filtering was included such that the Pearson's correlation coefficient between the k-mers presence/absence and the phenotype had to be ≥ 0.2 to be included in the analysis. For each starch parameter, the pipeline was run against the AL8/78 (TOWWC193) (Wang *et al.*, 2021a), TOWWC106 and TOWWC112 (Gaurav *et al.*, 2021) reference genomes. The significance threshold was set to 9.3 (black dashed line on all Manhattan plots) which was the stringent threshold for this population determined by Gaurav *et al.* (2021) and is based on the Bonferroni-adjusted -log *P* value, where $P = 0.05$. Manhattan plots were produced with Python using the scripts available at https://github.com/wheatgenetics/owwc/tree/master/kGWAS/src. This script splits the *x* axis into 10 kb blocks and k-mers within these blocks are plotted at the same *x* coordinate.

2.4.4 *Identification of genes under peaks on Manhattan plots*

Genes under the peaks were identified using the gene annotations for the AL8/78 v5 reference genome on JBrowse (http://aegilops.wheat.ucdavis.edu/jbrowse/index.html?data=Aetv5) and NCBI (https://www.ncbi.nlm.nih.gov/genome/?term=Aegilops+tauschii). As there were discrepancies between these annotations, the genes were classed as genes if there were annotated in both, or one of these and the AL8/78 v4 genome (Luo *et al.*, 2017). Low confidence genes, pseudo-genes and genes encoding ncRNAs were excluded.

2.4.5 *Identification of orthologs*

The protein sequences of the D genome homeologs of *T. aestivum* starch related genes (from Chen *et al.* (2023a)) were obtained from Ensembl plants

https://plants.ensembl.org/index.html). These were used in BLASTp searches against the *Ae. tauschii* v4 genome (Luo *et al.*, 2017). For each BLASTp search, the hit with the highest percentage identify was classed as the *Ae. tauschii* starch gene ortholog. To identify the positions of these genes in the *Ae. tauschii* v5 genome (Wang *et al.*, 2021a), the nucleotide sequence from v4 was used in BLASTn searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the *Ae. tauschii* v5 genome.

To identify wheat orthologs of the genes under the GWAS peaks, the protein sequences of the *Ae. tauschii* genes were used as the query sequences in BLASTp searches against the *T. aestivum* Chinese Spring and the *T. turgidum* genomes using Ensembl Plants. The hit with the highest percentage identity was classed as the putative ortholog. In some cases, there were multiple hits which had equally high percentage identities. For these, the gene trees available on Ensembl plants were used to verify true orthologous relationships.

2.4.6 Prediction of protein domains

Protein sequences encoded by the canonical transcript, based on the Ensembl plants annotations, were used as inputs for the InterPro tool (https://www.ebi.ac.uk/interpro/).

2.4.7 Prediction of protein localisation

The canonical transcript of the *Ae. tauschii* genes was identified using the Ensembl Plants canonical transcript labelling. The amino acid sequences encoded by the canonical transcripts were used as inputs for the WoLF PSORT protein subcellular localisation prediction tool (Horton *et al.*, 2007) (https://wolfpsort.hgc.jp/). The sub cellular location which had the most predictions was classed as the predicted location.

2.4.8 Retrieval of gene expression data

Expression of *T. aestivum* orthologs was examined using the wheat expression browser (http://www.wheat-expression.com/) (Borrill *et al.*, 2016; Ramírez-González *et al.*, 2018). The data were filtered based on tissue, and the expression in the endosperm of the grain was extracted. Genes were classed as endosperm expressed if the mean transcripts per million (TPM) value was ≥ 1.

Endosperm expression data for canonical transcripts of *T. turgidum* genes was retrieved from Chen *et al.* (2023a). Genes were classed as being endosperm expressed if the mean TPM value was ≥ 1 for at least one timepoint. For cases where

there were multiple A or B genome *T. turgidum* orthologs for the same *Ae. tauschii* gene, the expression was averaged.

2.4.9 Clustering gene expression

The *Clust* package (Abu-Jamous and Kelly, 2018) was used for the endosperm expressed *T. turgidum* orthologs of genes under the GWAS peaks, and the *T. turgidum* orthologs of: *BGC1* (*TRITD4Av1G198830*, *TRITD0Uv1G034540*), *PHS1* (*TRITD5Av1G205670*, *TRITD5Bv1G201740*), *MRC* (*TRITD6Av1G081580*), *MFP1.1* (*TRITD1Av1G054690*, *TRITD1Bv1G062760*), *MFP1.2* (*TRITD3Av1G038460*, *TRITD3Bv1G047250*) and *SS4* (*TRITD1Av1G202430* and *TRITD1Bv1G193680*). The TPM values were used as inputs for the package with each biological replicate inputted separately. The standard, inbuilt normalisation methods (log_2, Z-score and quartile normalisation) were implemented alongside a minimum cluster size of two. Different tightness values were tested, with a tightness of 0.4 being chosen for the final analyses, as it resulted in the greatest number of clusters.

2.4.10 ClustalW alignment of amino acid sequences

Amino acid sequences of LDA were either generated by sequencing (Section 2.5.3), or were obtained from UniProt (https://www.uniprot.org/) using the ID identifiers in Supplementary Table 1 of Andersen *et al.* (2020). ClustalW alignments were conducted with the MEGA-X software (Kumar *et al.*, 2018) with default parameters. The alignment figures displayed in Chapter Five and Appendix Three were generated using ESPript 3.0 (Robert and Gouet, 2014).

2.5 Cloning of LDA

2.5.1 RNA extraction and cDNA synthesis

Leaf samples (50-100 mg of tissue, 3-week-old plants), were collected, frozen in liquid nitrogen and ground using a ball mill (MM300, Retsch) at 25 Hz for 90 seconds with three 5 mm glass beads. RNA was extracted from the homogenates with the RNeasy Plant Mini kit (Qiagen). The method was adapted to include an on-column DNase digestion step (room temperature, 15 mins) prior to washing with RW1 buffer. To assess RNA quality, RNA concentrations, $A_{260/230}$ and $A_{260/280}$ values were measured with a spectrophotometer (DeNovix DS-11 FX), and all were of verified to be of sufficient quality for cDNA (complementary DNA) synthesis. cDNA synthesis was conducted with the Promega GoScript Reverse Transcription kit with the provided Oligo(dT) primers and 4 µL of RNA.

2.5.2 Amplification of LDA

Primer design for amplification of *LDA* cDNA was based on that from (Vester-Christensen *et al.*, 2010b) but adapted to match the *LDA* sequence from *Ae. tauschii* (Table 2.4, RM122 and RM123). These primers produce a 2695 bp polymerase chain reaction (PCR) product which encodes amino acids 79-963 of LDA. The PCR was set up using the Q5® high fidelity polymerase (NEB, New England Biolabs) with the GC enhancer. The cycling conditions were: 2 minutes at 94ºC, then 35 cycles: 94ºC (15 seconds), 55ºC (30 seconds), 72ºC (3 minutes), followed by a final elongation at 72ºC (7 minutes). The 25 µL reactions were run on a 1% (*w/v*) agarose gel. Bands representing *LDA* were extracted from the gel using the Monarch® DNA Gel Extraction Kit (NEB) according to the manufacturer's instructions, with the final elution carried out in 12 µL of ddH_2O.

Alongside this, *LDA* cDNA from accession TOWWC177 was also amplified using a second set of primers (Table 2.4, RM92 and RM93). These primers produce a 3262 bp PCR product which encodes amino acids 1-963 of LDA. The PCR was set up using the Q5® high fidelity polymerase (NEB) with the GC enhancer included. The cycling conditions were: 30 seconds at 98ºC, then 35 cycles: 98ºC (10 seconds), 65ºC (30 seconds), 72ºC (128 seconds), followed by a final elongation at 72ºC (2 minutes). A small sample (5 µL) of the PCR reactions were run on a 1% (*w/v*) agarose gel. The Zymo DNA Clean and Concentrate kit (Zymo Research) was used, as per the manufacturer's instructions, to purify PCR products from the remaining 20 µL of PCR reaction.

Table 2.4 Primers used for amplification of *LDA* from cDNA. Primers are all shown in the 5'-3' direction.

Name	Primer sequences	Direction
RM122	CCAGGCGTTCATGCCGGA	Forward
RM123	TCAACACCGAGGTTCAACAAAGACT	Reverse
RM92	ATGCCAATGCCGATGCGA	Forward
RM93	GAGGCATCAACACCGAGGTT	Reverse

2.5.3 Blunt end cloning into pJET1.2 and Sanger sequencing

All PCR products were cloned into the pJET1.2 blunt vector using the CloneJET PCR Cloning Kit (Thermo Fisher Scientific). The cloning step method was adapted from the protocol (Table 2.5) and left at 22ºC for 45 minutes. The cloning reaction was subsequently transformed into *Escherichia coli* DH5α cells (Library Efficient

Competent Cells, NEB) using a heat shock method. Briefly, 5 µL of cloning reaction was added to 50 µL of DH5α cells and left on ice for 30 minutes. The reaction was heat shocked at 42°C using a water bath for 45 seconds and left on ice for 2.5 minutes. SOC media (200 µL) was added, and the reaction was incubated at 37°C for 60 minutes at 300 rpm. The entire reaction was plated onto LB agar with 100 µg/mL carbenicillin plates and incubated at 37°C overnight. For each cloning reaction, three individual colonies were picked, each was used to inoculate 5 mL LB with 100 µg/mL carbenicillin and left to grow at 37°C overnight. Plasmids were isolated from the overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Sequences of the *LDA* were determined by Sanger sequencing (Azenta Genewiz) with the primers listed in Table 2.6, and all three primers were needed to cover the *LDA* cDNA sequence. The *LDA* cDNA sequence was determined as the consensus sequence from the three fully sequenced clones. The protein sequence was produced from the *LDA* cDNA sequence using the translate function on Geneious.

Table 2.5 CloneJET ligation reaction.

Component	Volume (µL)
2X buffer	10
Gel extracted PCR product	6 (regardless of the concentration)
T4 ligase	1
ddH_2O	2
pJET blunt	1

Table 2.6 Primers for sequencing *LDA* cDNA in pJET1.2. Primers are all shown in the 5'-3' direction.

Primer Name	Primer sequences	Direction
pJET forward	CGACTCACTATAGGGAGAGCGGC	Forward
RM96	CCTTCTCTGACATAACCATCTACGA	Forward
pJET reverse	AAGAACATCGATTTTCCATGGCAG	Reverse

2.5.4 Generation of expression plasmids

The pET-28a vector was synthesized to contain the *Ae. tauschii LDA* cDNA sequence (henceforth referred to as pET-28a-LDA-Ref), which encodes amino acids 58-973, by Genscript Biotech (UK) (Appendix Two: Synthesized DNA Two). Site directed mutagenesis (Q5® Site-Directed Mutagenesis Kit, NEB) was used to produce variants of the construct which encode for the TD, A140T, E248D and V271I variants. The kit was used to the manufacturer's instructions, with the exception that the ligation time was increased to 2 hours. The primers used and the corresponding annealing temperatures for the PCR are listed in Table 2.7 and were designed with NEBaseChanger® (https://nebasechanger.neb.com/). For the TD construct, two rounds of site directed mutagenesis were conducted, the first using the primers which generate the A140T mutation, then with the primers which generate the E248D mutation. The pET-28a plasmids were transformed into *E. coli* DH5α cells (Library Efficient Competent Cells, NEB) by heat shock as described in Section 2.5.3, with the exception that 28ºC was used instead of 37ºC and 50 µg/mL kanamycin was used for selection. Sanger sequencing (conducted by Azenta Genewiz) with the T7 primer (5'-TAATACGACTCACTATAGGG-3') was used to confirm the correct mutations had been introduced.

To generate the pET-28a vector which expresses the 177-like variant of *LDA,* traditional restriction enzyme cloning was used. The pET-28a-LDA-Ref plasmid contains a NdeI site (5'-CATATG-3') upstream of the inserted *LDA* cDNA, and a NotI (5'-GCGGCCGC-3') site downstream of the stop codon of the *LDA* cDNA. To clone into this, NdeI and NotI sites had to be generated in the 177-like *LDA* pJET1.2 vector, flanking the coding sequence. These sites were added using site directed mutagenesis (Q5® Site-Directed Mutagenesis Kit, NEB), using primers that generated a NdeI site immediately before the residues which encode for amino acid 58 and a NotI site immediately after the stop codon (Table 2.8). Sanger sequencing (Azenta Genewiz) with the pJET forward and pJET reverse primers (Table 2.6), confirmed the restriction sites had been added correctly – this plasmid is henceforth referred to as pJET-177-NdeI-NotI. Restriction enzyme digestions (Table 2.9) were set up with the pET-28a-*LDA*-Ref and pJET-177-NdeI-NotI plasmids for 4 hours at 37ºC, followed by 20 minutes at 65ºC. The restriction enzyme digests were run on a 1% (*w/v*) agarose gel. Bands representing the pET-28a backbone and the 177-like *LDA* cDNA were extracted from the gel using the Monarch® DNA Gel Extraction Kit (NEB), according to the manufacturer's instructions with the final elution carried out in 12 µL of ddH_2O. The pET-28a backbone (5301 bp, 24 ng) and the 177-like *LDA* cDNA (2731 bp, 37 ng)

were used in a ligation reaction (Table 2.10) at 16ºC for 16 hours followed by 65ºC for 10 minutes. Products from the ligation reaction were transformed into *E. coli* DH5α cells (Library Efficient Competent Cells, NEB) by heat shock as described in Section 2.5.3, with the exception that 28ºC was used instead of 37ºC and 50 µg/mL kanamycin was used for selection. Whole plasmid sequencing (conducted by Plasmidsaurus) was used to confirm the 177-like *LDA* cDNA had been incorporated into the pET-28a backbone correctly.

Plasmids were transformed independently into *E. coli* BL21 cells ΔglgAP cells – these lack the endogenous glycogen synthase gene and hence minimises the risk of contaminating glycogen and glycogen-active enzymes. Heat shock was used for transformation as described in Section 2.5.3, with the exception that 28ºC was used instead of 37ºC and 50 µg/mL kanamycin and 34 µg/mL chloramphenicol was used for selection. Glycerol stocks (25 % (*v/v*) glycerol) were made for each transformed bacterial strain.

Table 2.7 Primers used for site directed mutagenesis of the pET-28a expression vector. Primers are all shown in the 5'-3' direction. Lower case letters represent the sites which are altered during site directed mutagenesis.

Nucleotide substitution	Amino acid change	Primer sequences	Annealing temperature (ºC)
G -> A	A140T	Forward: ACCGGAGAGCaCCGGGCTCCCGG Reverse: TGCAGCTCAACCTTGGAGTCGTAGCC	72
G -> T	E248D	Forward: GGTTCAGCTCGAtGAGTCAAATG Reverse: GTCTCCAGCACAGGG	61
G -> A	V271I	Forward: TTTGTATGAAaTCGACGTGTATC Reverse: AATACCGGTTTTCCCAC	61

Table 2.8 Primers used for site directed mutagenesis for generation of restriction enzyme sites in the pJET1.2. Primers are all shown in the 5'-3' direction. Lower case letters represent the sites which are altered during site directed mutagenesis. Underlining represents the restriction site which will be introduced.

Nucleotide changes	Restriction site introduced	Primer sequences	Annealing temperature (°C)
CC -> AT	NdeI	Forward: CGGAGGACGCatATGGCGGCCGGCG Reverse: GGCCCCGGGCGCGGT	72
GATGCCTCATC TTT -> GAGCGGCCGC CTTT	NotI	Forward: ccgcCTTTCTAGAAGATCTCCTAC Reverse: ccgcTCAACACCGAGGTTC	56

Table 2.9 Components for restriction enzyme digests.

Component	Volume (µL)
Plasmid (100 – 130 ng/µL)	5
10 X rCut Smart Buffer (B6004S, NEB)	5
NotI-HF (R3189S, NEB)	1
NdeI (R0111S, NEB)	1
ddH_2O	38

Table 2.10 Components for ligation reactions.

Component	Volume (µL)
10 X T4 DNA Ligase Reaction Buffer (B0202A, NEB)	2.0
pET-28a backbone (5301 bp, 24 ng)	8.0
177-like *LDA* cDNA (2731 bp, 37 ng)	6.4
T4 ligase (M0202M, NEB)	1.0
ddH_2O	2.6

2.6 Biochemistry methods

2.6.1 Crude protein extraction from leaves and grain

Leaf tissue (~2 cm in length, 5-week-old plants) was harvested, immediately frozen in liquid nitrogen and ground using a ball mill (MM300, Retsch) at 30 Hz for 90 seconds with three 3 mm glass beads. To the powder, 100 µL of ice-cold extraction buffer (100 mM 3-[N-morpholino] propanesulfonic acid (MOPS) pH 7.2, 1 mM DTT, 1 mM EDTA, 10% (*v/v*) glycerol and EDTA-free Protease Inhibitor cocktail (Roche)) was added and samples were left on ice for one hour. Samples were centrifuged at 20817*g* for 15 minutes at 4°C, the supernatant was retrieved and centrifuged at the same conditions. Protein concentration was determined using the Bradford's assay.

For protein extraction from developing grains, grains were selected from the central spikelet of 10, 15, 19 and 22 DPA spikes (Section 2.1.1). Endosperms were manually dissected and homogenised using a micropestle in 50 µL of ice-cold extraction buffer (as above). The remainder of the steps were the same as outlined for leaf protein. For the *T. turgidum* developing grain protein extracts, this method was conducted by Nitin Uttam Kamble.

2.6.2 Bradford protein assay

The Bradford assay was conducted using a BMG Omega Plate Reader with 96-well clear flat bottom plates (Cytiva). For each sample, three 200 µL reactions were set up with 160 µL of Bradford's reagent (Bio-Rad Protein Assay) and 40 µL of protein and ddH_2O. The absorbance was measured on the plate reader at 595 nm with a 5 second pre-shake. Standard curves were included in every plate and were constructed with known amounts of bovine serum albumin (BSA) (Melford) (Table 2.11). Protein concentrations of samples were calculated using the standard curve. For cases where the measured absorbance fell outside the range of the standard curve, the assay was repeated with diluted protein.

Table 2.11 Standard curve for the Bradford's assay.

Amount of BSA in 200 µL reaction (µg)	Volume to add of 100 µg/mL BSA (µL)	Volume to add of H_2O (µL)
0	0	40
0.6	6	34
1.2	12	28
1.8	18	22
2.4	24	16
3.0	30	10

2.6.3 LDA native gels

Red pullulan containing polyacrylamide gels were cast according to the recipe in Table 2.12. Protein (35-100 µg) was loaded onto the gels with 10 X loading buffer (50% (*v/v*) glycerol and 0.005% (*w/v*) bromophenol blue). All lanes were loaded on an equal protein basis; the amount of protein loaded was different between gels and is specified in the figure legends. The gels were run at 100V at 4ºC using a Bio-Rad Mini-PROTEAN II electrophoresis tank until the loading buffer reached the bottom of the gel tank (~2.5 hours). Gels were washed once in 50 mL of incubation buffer (100 mM Tris-HCl pH 7.0, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 5 mM DTT) and left overnight in a fresh 50 mL of incubation buffer at room temperature. Gels were imaged the following morning with an Epson Perfection V550 scanner. Gel images have been separated so that only the green channel, which displays the best contrast, is displayed. For easy visual identification of the bands the colour has then been inverted and the brightness and contrast adjusted.

Table 2.12 Components for red pullulan containing polyacrylamide gels.

Component	Volume for 1 x 1.5 mm gel
Resolving gel:	
30% (*w/v*) Acrylamide (Severn Biotech Ltd)	2.5 mL
1.5M Tris-HCl (hydroxymethyl)aminomethane (Tris), pH 8.8	2.5 mL
Red Pullulan (Megazyme)	100 mg
ddH_2O	5 mL
10% (*w/v*) Ammonium persulphate (APS)	80 µL
Tetramethylethylenediamine (TEMED)	5 µL
Stacking gel:	
30% (*w/v*) Acrylamide (Severn Biotech Ltd)	0.5 mL
0.5M Tris-HCl (hydroxymethyl)aminomethane (Tris), pH 6.8	0.5 mL
ddH_2O	3 mL
10% (*w/v*) Ammonium persulphate (APS)	32 µL
TEMED	3 µL

2.6.4 Recombinant protein expression in E. coli

2.6.4.1 Solubility test

The glycerol stock (5 µL) of the *E. coli* strain containing pET-28a-LDA-Ref (Section 2.5.4) was used to inoculate 5 mL LB with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin. This was grown overnight at 28°C at 200 rpm. LB (2 x 100 mL) containing 34 µg/mL chloramphenicol and 50 µg/mL kanamycin were each inoculated with 1 mL of overnight culture and grown at 28°C at 200 rpm. The OD_{600} of these cultures were monitored every hour until an OD_{600} of 0.6-0.8 was reached. Protein expression was induced (I) in one culture with 1 mM IPTG (Isopropyl β- d-1-thiogalactopyranoside), and the other culture was left uninduced (U). Both were grown at 18°C at 200 rpm overnight. Cultures were centrifuged at 3,000*g* at 4°C for 10 minutes, the supernatant was discarded, and the pellet frozen at -80°C. Lysis buffer (10 mL of 500 mM Tris-HCl pH 7.5, 300 mM NaCl, 40 mM Imidazole, EDTA-free Protease Inhibitor cocktail (Roche), 2 mM DTT, 1 mg/mL lysozyme) was used to resuspend the frozen pellets, before sonication with a Soniprep 150 Plus (MSE) sonicator fitted with an exponential probe for 3 minutes with an amplitude of 5.8. The samples were incubated at 4°C on an orbital shaker (Denley) for 60 minutes and then centrifuged at 20,000*g* at 4°C for 10 minutes. A 100 µL fraction of the supernatant was reserved (soluble protein fraction,

S), the pellet was resuspended in 10 mL of lysis buffer and 100 µL of this was reserved (insoluble protein fraction, I).

The US, II and IS fractions (3 µL of each), were heated for 70ºC for 10 mins with 2.5 µL of LDS (lithium dodecyl sulfate) sample buffer (Invitrogen), 1 µL of NuPAGE® (polyacrylamide gel electrophoresis) reducing agent (Invitrogen) and 3.5 µL of ddH_2O. The UI fraction was too viscous to load and was omitted from the gel. Samples were run on a premade 1mm NuPAGE® Bis-Tris Mini Protein Gel (4-12% gradient, Invitrogen) in an Invitrogen Mini Cell electrophoresis tank for 75 minutes at 150 V. NuPAGE® MOPS SDS Running buffer (Invitrogen) was used to run the gels, with 0.25% (*v/v*) NuPAGE® antioxidant (Invitrogen) added to the Running buffer in the upper chamber. Following electrophoresis, the gel was washed once with ddH_2O, stained with Coomassie (InstantBlue, Abcam) for 10 mins, and destained with ddH_2O overnight. The gel was imaged the following morning with the visible light setting on a SynGene G box.

2.6.4.2 Expression and purification

Glycerol stocks (2 x 10 µL) of each of the *E. coli* strains generated in Section 2.5.4 were used to inoculate 2 x 10 mL LB with 34 µg/mL chloramphenicol and 50 µg/mL kanamycin. These were grown overnight at 28ºC at 200 rpm. LB (2 x 1 L) containing 34 µg/mL chloramphenicol and 50 µg/mL kanamycin were each inoculated with 10 mL of overnight culture and grown at 28ºC at 200 rpm. The OD_{600} of these cultures were monitored every hour until an OD_{600} of 0.6-0.8 was reached. Protein expression was induced in all cultures with 1 mM IPTG they were left to grow at 18ºC at 200 rpm overnight. Cultures were centrifuged at 4,000*g* at 4ºC for 10 minutes, the supernatant was discarded, and the pellet frozen at -80ºC. Lysis buffer (20 mL, composition as in Section 2.6.4.1) was used to resuspend the frozen pellets, before sonication with the Soniprep 150 Plus (MSE) sonicator fitted with an exponential probe for 3 minutes with an amplitude of 15.0. The samples were incubated at 4ºC on an orbital shaker (Denley) for 60 minutes and then centrifuged at 20,000*g* at 4ºC for 10 minutes – the supernatant was reserved on ice.

The His-tagged proteins were purified using the Ni-NTA agarose resin (Qiagen). Before the resin was used, it was washed with 2 x 2 mL lysis buffer, spinning at 500*g* at 4ºC for 3 mins between washes. To the washed resin, the reserved supernatant was added and incubated on an orbital shaker (Denley) at 4ºC for 60 minutes. The suspension was centrifuged at 500*g* at 4ºC for 3 mins and the supernatant discarded.

The resin was washed with 5 x 5 mL Triton buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 40 mM Imidazole, 0.5% (*v/v*) Triton X-100 (Sigma)). Between each wash the samples were centrifuged at 500*g* at 4°C for 3 mins and the supernatant was removed. Following this, the resin was washed with 5 x 5 mL wash buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 40 mM Imidazole). Between each wash step, the samples were centrifuged at 500*g* at 4°C for 3 mins and the supernatant was removed. The proteins were eluted with 1 x 2 mL Elution Buffer 1 (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 250 mM Imidazole), 1 x 2 mL Elution Buffer 2 (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 100 mM Imidazole) and 2 x 2 mL Elution Buffer 3 (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 500 mM Imidazole). All elutions were combined such that there was only one tube of purified protein for each protein variant. Proteins were concentrated by passing through a 30 kDa cut-off centrifugal filter (Amicon) with spins conducted at 3124*g* at 4°C. The buffer of the concentrated proteins was exchanged to storage buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% (*v/v*) glycerol) using NAP5 Sephadex G-25 columns (Cytiva) at 4°C.

The concentrations of the purified proteins in storage buffer were measured by recording the absorbance at 280 nm (A_{280}) using a spectrophotometer (DeNovix DS-11 FX). The concentration was calculated as follows:

$$Concentration\ (\mu M) = A_{280} * 130680$$

Where the 130680 was the extinction coefficient as calculated using Geneious; this value was consistent for all protein variants.

$$Concentration\ (\mu M) = A_{280} * 1.28$$

Where 1.28 is the A_{280} of 1 µg/µl as calculated using Geneious; this value was consistent for all protein variants.

Purified proteins (1 µg) were resolved with SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis). This was run as previously described in Section 2.6.4.1, except that the volume of ddH_2O for each sample adjusted according to the protein concentration so that the total volume loaded was always 10 µL.

2.6.5 PullG6 assay

Purified protein (2 µg) in water in a total volume of 40 µL was pre-incubated at 40°C in a water bath for 5 minutes. In parallel, 50 µL of 4,6-O-Benzylidene-4-nitrophenyl-6^3-α-D-maltotriosyl-maltotriose (PullG6) with thermostable α- and β-glucosidase (Megazyme, K-PullG6 kit, Bottle 1) and 10 µL of 500 mM sodium acetate pH 5.0 was also pre-incubated at 40°C in a water bath for 5 minutes. The PullG6/Sodium acetate

solution was added to the purified protein solution and left in the 40ºC water for exactly, 5, 10 or 20 minutes, before addition of 1 mL of 2% (*v/v*) Tris-HCl, pH 9.0. The absorbance at 400 nm (A_{400}) was measured using a spectrophotometer (DeNovix DS-11 FX). The absorbance was converted to PullG6 units (mmoles of 4-nitrophenol produced per µg of LDA per min):

$$PullG6\ units = \frac{A_{400} * \ 0.0011\ L * 1000}{18100\ M^{-1} * 1\ cm * Time * 2\ \mu g}$$

Where 18100 M^{-1} is the molar attenuation coefficient (ε) of 4-nitrophenolate ions in 2% (*v/v*) Tris-HCl at 400 nm, 1 cm is the path length of the cuvette, 0.0011 L is the volume of the final solution, time is the time in minutes the PullG6 and protein reaction is conducted for, 2 µg is the amount of protein added, 1,000 is the conversion from moles to mmoles of 4-nitrophenol. For each time point and for each purified protein, three technical replicates were set up.

2.7 Genotyping for SNPs in *LDA*

2.7.1 DNA extraction

Extraction buffer (500 µL of 0.1 M Tris-HCl pH 7.5, 0.05 M EDTA, 1.25% (*v/v*) SDS (Severn Biotech Ltd)) was added to leaf tissue (~2 cm in length, 3-week-old plants) in 96-deep well plates (Fisher Scientific). Leaf tissue was ground using a ball mill (MM300, Retsch) at 25 Hz for 3 minutes with one 3 mm tungsten carbide ball. Samples were centrifuged at 3,000*g* for 5 minutes and incubated at 65ºC for 1 hour. Plates were cooled (4ºC for 15 minutes), 250 µL of 6 M sodium acetate was added to each sample and incubated at 4ºC for 15 minutes. Plates were centrifuged at 3,000*g* for 15 minutes and 350 µL of supernatant was added to 450 µL of ice-cold isopropanol and left on ice for 5 minutes. Samples were centrifuged at 3,000*g* for 15 minutes at 4ºC and the supernatant discarded. DNA pellets were washed with 500 µL of 70% (*v/v*) ethanol, centrifuged at 3,000*g* for 15 minutes at 4ºC and pellets were left to dry at room temperature for 30 minutes. Pellets were resuspended in 100 µL of ddH_2O and left overnight at 4ºC. DNA concentration was measured using a spectrophotometer (DeNovix DS-11 FX).

2.7.2 Kompetitive allele specific PCR

Primers were designed to genotype the SNPs in *LDA* (Table 2.13). Reactions (5 µL) were set up in FrameStar® 394 well plates (Scientific Laboratory Supplies Ltd) with 2.5 µL of 2X 2.7.2 Kompetitive allele specific PCR (KASP) V4 Mastermix (LGC Genomics Ltd) and 2.5 µL of DNA (1-100 ng/µL). KASP reactions were run on a thermocyler

(Mastercycler x50h, Eppendorf) with the following cycling conditions: 95°C for 15 minutes, 10 cycles: 94°C for 20 seconds, 65°C* for 1 minute (* decreases by 0.8°C each cycle), 40 cycles: 94°C for 20 seconds, 57°C for 1 minute. A PHERAstar plate reader (BMG Labtech) was used to measure fluorescence, and the KlusterCaller software was used to visualise results.

Table 2.13 KASP primers used to genotype the LDA variant produced by accessions in the *Ae. tauschii* diversity panel. Primer sequences are given in the 5'-3' direction. A VIC/HEX tail (5'- GAAGGTCGGAGTCAACGGATT) was added to the 5' end of the primers designed to identify the Ref alleles which encode for the Ref LDA variant. A FAM tail (5'- GAAGGTGACCAAGTTCATGCT) was added to the 5' end of primers designed to identify non-Ref alleles which encode for non-Ref LDA variant. Nucleotides represented in lower case letters are those which distinguish the variable base.

Nucleotide change (reference vs variant)	Amino acid change	Allele encoding for Ref LDA variant
G -> A	A140T	Ref allele: AGCTGCAACCGGAGAGCg Non-Ref allele: AGCTGCAACCGGAGAGCa Common: CGTCTTACGGTTTCCGGGA
G -> T	E248D	Ref allele: GAGACGGTTCAGCTCGAg Non-Ref allele: GAGACGGTTCAGCTCGAt Common: CTCTTGGTCCAGTAACACTCC
C -> T	A238V	Ref allele: GCTTCTTTGATGGTCCAGc Non-Ref allele: GCTTCTTTGATGGTCCAGt Common: TGAACCGTCTCCAGCACA
A -> C	N509T	Ref allele: CACTGTTCTCAATCTGGCCAt Non-Ref allele: CACTGTTCTCAATCTGGCCAg Common: ATCGTTCCTGGGTACTATGT

2.8 Protein structure

The barley LDA protein structure (4aio, (Møller *et al.*, 2012)) was obtained from the Protein Data Bank (PDB, https://www.rcsb.org/). Pymol was used to visualise the structure and highlight residues of interest.

2.9 Grain morphometrics

A MARViN seed analyser (Marvitech GmbH) was used to measure the total number of grains harvested per plant (59-218). All grains were used to calculate the mean values of grain length, width and areas.

2.10 Statistics and data visualisation

All statistical analyses were conducted with RStudio (version 2023.12.1, build 402). The statistical tests used, and any post hoc tests, are specified in the legend or text. Parametric tests were used, except in cases where the data did not follow a normal distribution, as determined by Shapiro-Wilk tests and visual inspection of quantile-quantile plots. In these cases, non-parametric tests were used instead. The 'ggplot2' package in RStudio was used to generate most of the graphs, with the exception of the graphs showing the Coulter counter fittings and Manhattan plots which were generated in Python 3.11.5, run on a Jupyter notebook (version 6.5.4), with the 'matplotlib' package. All cartons and diagrams were generated using Inkscape (version 1.3).

Chapter Three – Characterising variation in endosperm starch in the *Ae. tauschii* diversity panel

3.1 Introduction

Association genetics offers a powerful approach to utilise pre-existing variation in starch granule size and shape to discover new factors influencing granule formation. Many previous association genetic studies have used domesticated species and there is no published case of a genome wide association study for endosperm starch properties using a wild relative of a Triticeae species. Hence, these studies have not utilised the interesting variation in granule composition which is observed in wild Triticeae species (Chapter One; Table 1.1). In addition, starch related genes in wild species often have greater levels of diversity and could be a source of novel alleles for introduction into domesticated species to generate favourable starch characteristics. This is a benefit of utilising wild species as they are closely related to domesticated crops. Therefore, the knowledge that is gained from wild species, such as *Ae. tauschii*, can be directly applied to crops, such as hexaploid wheat.

3.1.1 Development of a diverse panel of Aegilops tauschii accessions

As outlined in Chapter One (Section 1.6.2), *Ae. tauschii* is the progenitor of the hexaploid wheat D genome. *Ae. tauschii* can be subdivided into three lineages: lineages one, two and three (Gaurav *et al.*, 2021). Genomic comparisons of individuals of different lineages with modern hexaploid wheats reveals that the wheat D genome is most related to lineage two *Ae. tauschii* accessions (Gaurav *et al.*, 2021). However, modern wheat also contains elements (~1-7%) from lineage three accessions – for example, a lineage three-derived allele of the glutenin *Glu-D1* gene is found in some modern wheat cultivars and is associated with higher quality (Delorean *et al.*, 2021). Meanwhile, lineage one accessions have made the smallest contribution (~0.7-1.1%) to the wheat D genome (Cavalet-Giorsa *et al.*, 2024). Hence it has been hypothesized that instead of a single hybridisation event (Figure 1.12) there were multiple hybridisation events between tetraploid wheat and *Ae. tauschii* (Cavalet-Giorsa *et al.*, 2024).

Although lineage two is the major contributor to the hexaploid wheat D genome, ~75% of the genetic diversity identified among lineage two accessions is not found in modern wheat cultivars (Gaurav *et al.*, 2021). This is a consequence of a bottleneck caused by extensive selection during domestication (Zhou *et al.*, 2020). This contrasts with the A and B genomes of modern hexaploid wheats which have more diversity than the D-genome. This is perhaps due to introgressions with tetraploid wild wheats, whereas

introgressions between hexaploid wheat and diploid *Ae. tauschii* are rare (He *et al.*, 2019). Therefore, there is much interest in exploring diversity in lineage two *Ae. tauschii* accessions and reintroducing this diversity into modern hexaploid wheats. Gaurav *et al.* (2021) assembled a diversity panel of 242 *Ae. tauschii* accessions which includes individuals from all three lineages (286 lineage one, 175 lineage two and 7 lineage three). These accessions originate from around the Fertile Crescent region (Figure 3.1). This is the domestication centre for modern day wheat; hence the authors argue that this panel covers almost all the genetic diversity in *Ae. tauschii* species worldwide. This diversity panel has extensive genetic and phenotypic variation, including in traits such as spike morphology, trichome number and disease resistance (Gaurav *et al.*, 2021). However, the variation in starch traits among this particular diversity panel remains unexplored.

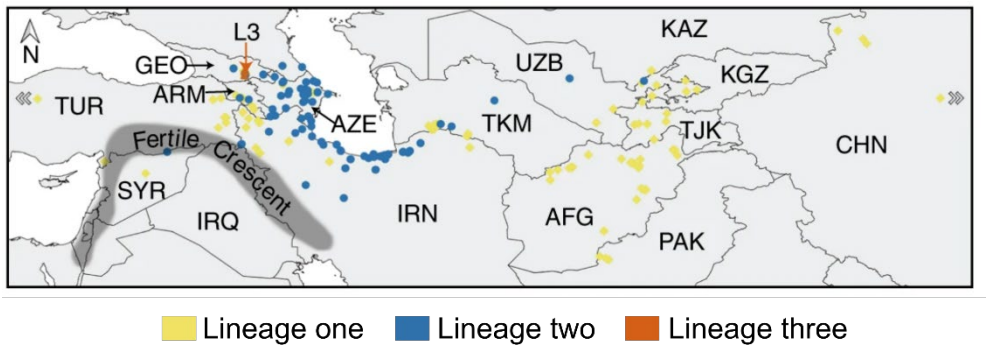


Figure 3.1 Distribution of *Ae. tauschii* accessions in the diversity panel. Locations of origin for 239 accessions in the *Ae. tauschii* diversity panel studied by Gaurav *et al.* (2021). Three accessions lie outside the map (one from Turkey and two from China) and these are represented with arrow heads. Country abbreviations are as follows – AFG = Afghanistan; ARM = Armenia; AZE = Azerbaijan; CHN = China; GEO = Georgia; IRN = Iran; IRQ = Iraq; KAZ = Kazakhstan; KGZ = Kyrgyzstan; PAK = Pakistan; SYR = Syria; TJK = Tajikistan; TUR = Turkey; TKM = Turkmenistan; UZB = Uzbekistan. This image has been adapted from Gaurav *et al.* (2021) under a CC-BY license.

3.1.2 Ae. tauschii has diversity in its starch granule parameters

Previous studies suggest that that *Aegilops tauschii* accessions vary in their starch granule parameters. The most extensive studies into *Ae. tauschii* starch granules was conducted by Stoddard (1999a). Stoddard describes variation in B-type granule content and granule size in 68 *Ae. tauschii* accessions. B-type granule content ranged from 15-38% in this population. When I compared this variation to that seen in modern cultivars (Introduction; Table 1.1), I suspected that *Ae. tauschii* might have unexploited

variation at the lower end of the B-type granule content range, as very few studies report B-type granule contents of less than 20% in modern wheat cultivars (Park *et al.*, 2009; Zhang *et al.*, 2016; Li *et al.*, 2017). Moreover, Stoddard (1999a) reports on granule size with modal diameters of 23.8 μm and 6.3 μm for A-type and B-type granules respectively. Interestingly, there were several accessions which had much larger A-type granules, with some granules as large as 57.5 μm observed. In addition to granule size and content, *Ae. tauschii* also has variation in amylose content (21-34%) (Chapter One; Table 1.1) (Mohammadkhani *et al.*, 1998; Stoddard, 1999a). Strikingly, no amylose-free *Ae. tauschii* accessions have ever been described. Beyond this, variation in other *Ae. tauschii* starch granule parameters, such as total starch content or granule number, have not been investigated.

Since these early studies on *Ae. tauschii* endosperm starch, there have been substantial technological advancements in the study of starch traits, allowing more quantitative characterisation of the variation. This is especially the case for A- and B-type granule distributions which historically were measured with laser scanning instruments. Laser scattering instruments generally underestimate granule size (Wilson *et al.*, 2006). Nowadays the preferred method is to use a Coulter counter, where the volume of granules is measured based on changes in electrical conductance. For accurate quantification of granule size of A- and B-type granules, mathematical curves should be fitted to size distribution traces (Tanaka *et al.*, 2017). The Tanaka *et al.* (2017) pipeline is not ideal, as it designed for laser scattering rather than Coulter counter data. It also fits three log-normal curves – one for A-type granules, one for B-type granules and one for C-type granules. C-type granules are a subpopulation of granules which are smaller than B-type granules; however these are only observed in a handful of studies (Wilson *et al.*, 2006). Other scripts, such as that used in Chen *et al.* (2024), fit two normal curves to Coulter counter data. A limitation of this is that it does not offer any flexibility if the size distribution traces do not fit a normal shape. Furthermore, not enough is known about the biology underlying starch granule size determination to assume a normal distribution. Overall, to accurately measure variation in granule size and content in *Ae. tauschii*, a method combining the Coulter counter and mathematical curves should be developed.

3.1.3 Ae. tauschii has greater diversity in its starch genes than domesticated wheat

In addition to the phenotypic variation, variation in the gene sequences of starch-related genes in *Ae. tauschii* has been described. For example, the nucleotide diversity of *GBSS* has been assessed in a panel of 59 *Aegilops* and *Triticum* species.

SNPs and InDels were found in every species when compared to published *T. aestivum GBSS* sequences. However, the majority of this variation occurred in intronic regions, or in the transit peptide, with only 17 amino acid changes in the mature protein sequence detected across the entire panel (Li *et al.*, 2016). Interestingly, the *Ae. tauschii GBSS* sequences cluster into two distinct phylogenetic groups, and only one of these clusters with *GBSS* from *T. aestivum* (Li *et al.*, 2016). Hence, there is unexploited diversity in *Ae. tauschii GBSS*, which has not been incorporated during hexaploid wheat evolution. Another example is *SS1*, where there is a loss of nucleotide diversity with increasing ploidy in wheat. In *Ae. tauschii* the SNP frequency across *SS1* is one SNP per 296 bases (Singh *et al.*, 2023), whereas in cultivated hexaploid wheat the frequency is one SNP per 824 bases (Singh *et al.*, 2023) – higher than the average SNP frequency of one SNP per 335 bases across the entire hexaploid wheat genome (Ravel *et al.*, 2006). Likewise, the nucleotide diversity in the genes encoding the small and large subunits of AGPase is reduced in tetraploid and hexaploid wheats compared to wild grass relatives (Hou *et al.*, 2017). Interestingly, there appears to have been stronger selection and greater reduction in nucleotide diversity in hexaploid, rather than tetraploid, wheats (Hou *et al.*, 2017). This reduction in genetic diversity of starch-related genes in domesticated species arises from selection during domestication. Hence, to discover novel alleles for known starch genes it is important to explore undomesticated wild relatives, such as *Ae. tauschii*, which have a greater nucleotide diversity.

In this chapter, I assess the diversity in endosperm starch from 117 lineage two members of the Gaurav *et al.* (2021) diversity panel. The lineage two accessions have been selected as they have the largest contribution to the D genome. The 117 accessions represent a subset selected from the entire collection and includes almost all lineage two diversity and reduces genomic redundancy. Firstly, I examine starch granule size distributions by developing a curve-fitting script in Python that can work on Coulter counter data. The values derived from this method revealed significant variation in granule size, content and number between the different accessions, which exceeds the variation that has previously been reported. Correlating these parameters with each other, with geographical parameters, or with phylogenetic relationships revealed no strong associations. Finally, I extend these analyses and survey the variation in a selection of modern wheats. Whilst these contain some variation, *Ae. tauschii* contains novel variation that could be useful for wheat breeding.

3.2 Results

3.2.1 There is variation in granule size distribution traces across the diversity panel

Endosperm starch was extracted from mature *Ae. tauschii* grains and starch granule size distribution traces were analysed with a Coulter counter. There was large variation in the size distribution traces produced. All accessions produced bimodal distributions, with a peak representing B-type granules occurring at diameters < 10 μm and a peak representing A-type granules occurring at diameters > 10 μm. There was no third peak, representing C-type granules, observed in any of the samples. A small subset of traces which are representative of some of the variation seen across the diversity panel are shown in Figure 3.2. Accession TOWWC021 (Figure 3.2 – yellow) is distinct from the other traces as it is has broad, less distinct peaks. This could be explained if there are large B-type granules and small A-type granules and the peaks have almost merged in the size distribution trace. From the remaining traces, I observed considerable variation in the height of both the A-type and B-type granule peaks and differences in diameter at which the peaks occur.

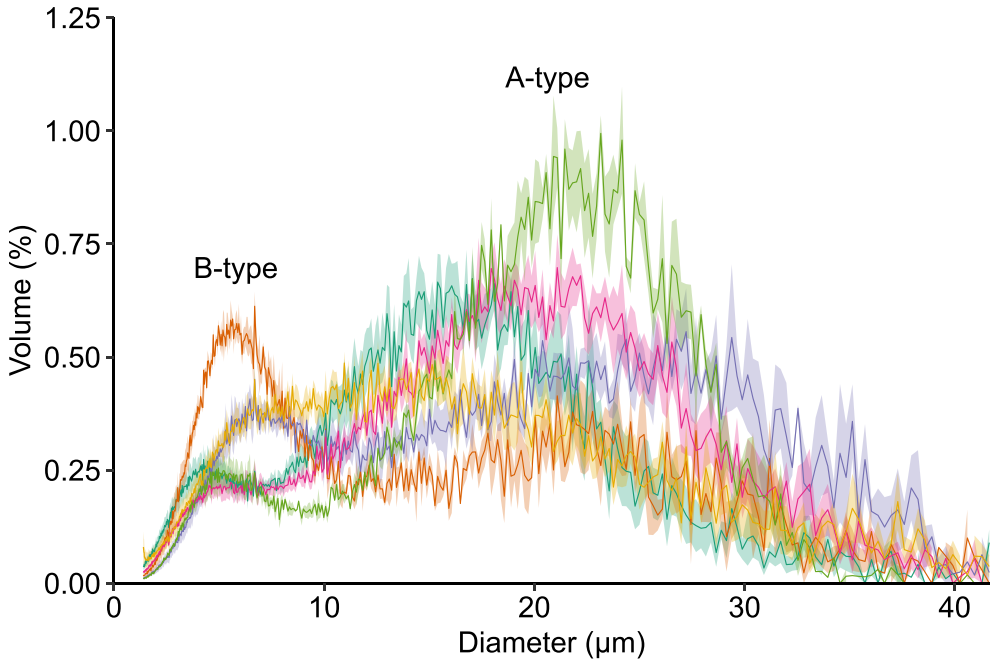


Figure 3.2 Size distribution traces from a subset of *Ae. tauschii* accessions are very diverse. A Coulter counter was used to measure size distribution traces. Six diverse traces from *Ae. tauschii* accessions representing some of the variation observed in the panel. The accessions represented are TOWWC016 (turquoise), TOWWC021 (yellow), TOWWC106 (purple), TOWWC141 (orange), TOWWC167 (pink), TOWWC193, which was used to generate the genome reference (green). The traces are displayed as means (solid lines) ± standard error of the mean (SE) (shading) and have been adjusted for representation on a linear *x* scale.

3.2.2 Curves were fitted to determine granule size and content

It is typical to fit curves to starch granule distribution traces to determine mean granule size and content by volume (Tanaka *et al.*, 2017). To overcome the limitations of the currently published scripts (Section 3.1.2) I wrote a Python pipeline for fitting bimodal mixed distributions to Coulter counter data. I designed this to fit three different distributions. First, it fits two normal curves (henceforth called N-N) which is consistent with the script used in Chen *et al.* (2024). However, I reasoned that a log-normal curve would be optimal for fitting the B-type granule peak, since the peak comes close to the *y*-axis and there can be no granules with a negative diameter value. Therefore, I tested fitting a log-normal curve to both the A-type and B-type peaks (henceforth called L-L), or a log-normal distribution to the B-type peak and a normal distribution to the A-type peak (henceforth called L-N) (Figure 3.3). The script was designed to report the total uncertainty of the reported parameters and the standard error of regression (S), which is a test of the goodness of the fit. An example of the three fittings, is shown in Figure 3.3a-d with a size distribution trace from accession TOWWC097, which falls in the middle of the diversity observed.

To determine which fitting worked best the standard error of regression was compared for each sample. For 79% of the samples the L-N fitting gave the lowest standard error of regression (Figure 3.3e). By contrast, the fitting which gave the lowest uncertainty value for most of samples was the N-N (Figure 3.3f). The uncertainty does not tell us about the goodness of the fit, only how much confidence there is in the parameters which describe the fit. It would be inappropriate to choose the N-N over the L-N fitting because there is little point having confidence in the parameters if the fitting is poor. Thus, the goodness of fit has been prioritised over the uncertainty and the L-N fitting has been used for all samples. By using the same fitting for all samples, it achieves consistency and means that errors in fitting are comparable across all samples. However, this does mean that for some of the samples, the fitting is not the most ideal. Generally, the values obtained by the fittings were consistent with each other in rank, such that samples having high values under one fit were also high in the other fits (Figure 3.4). There are some obvious outliers to this, and this decreases the reported Spearman rank correlation coefficients, yet all are above 0.7. The outliers often stem from the TOWWC021 curves (Figure 3.4 – pink) which, as already alluded to, have broad less distinct peaks which are difficult to accurately fit curves to (Figure 3.5). In Figure 3.5, although the combined curve follows the shape of the size distribution well, the constituent A- and B-type granule curves do not fit. Therefore, it is impossible to tell whether the proposed curves are correct. There are large differences in the constituent

peaks for the different fittings. This explains why there are inconsistencies between the parameters from the different fittings for these types of curves. The other outliers, for example the TOWWC087 curve which gives an extremely high B-type granule diameter (13.3 μm) and B-type granule content for the N-N fitting (86%), also has broad peaks which seem to merge into each other, such that it is not dissimilar from the curve in Figure 3.5. Crucially, this was not consistent across all three biological replicates for TOWWC087.

Even in cases where the fittings are good, the absolute values obtained for each parameter varied substantially between the different fittings. Mean A-type granule diameter was similar between the L-N and N-N fittings. The L-L fitting, however, resulted in higher A-type granule values. This trend was also seen in the B-type granule diameter. Both are likely a result of the long tail observed on the B-type granule peak under the L-L fitting, which also shifts the position of the A-type granule peak. The L-L fitting produced the highest B-type granule content values, which is again likely an overestimate produced by the long tail.

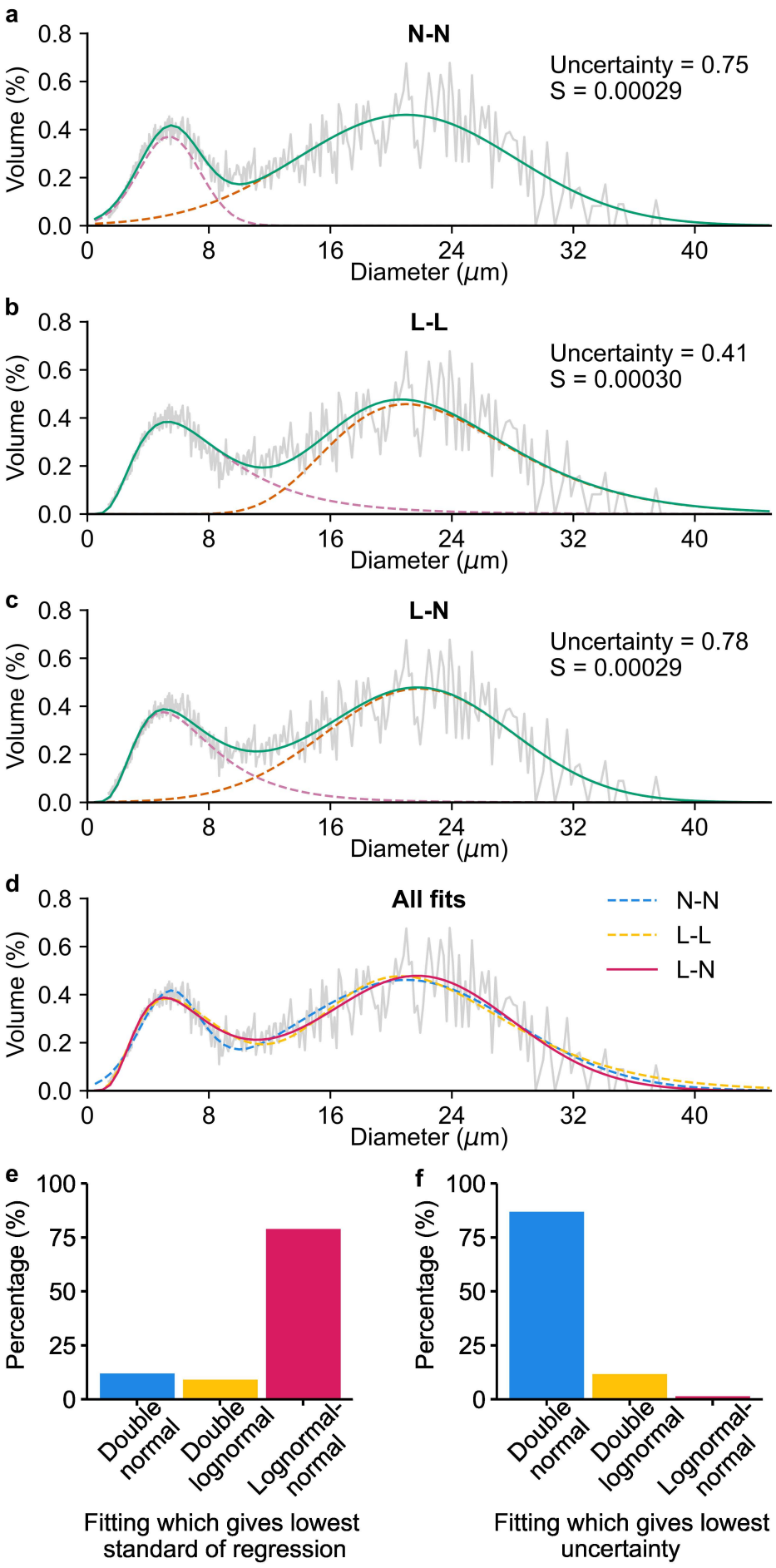

a
N-N
Uncertainty = 0.75
S = 0.00029
Volume (%)
Diameter (μm)
b
L-L
Uncertainty = 0.41
S = 0.00030
c
L-N
Uncertainty = 0.78
S = 0.00029
d
All fits
N-N
L-L
L-N
e
Percentage (%)
Double normal
Double lognormal
Lognormal-normal
Fitting which gives lowest standard of regression
f
Fitting which gives lowest uncertainty

Figure 3.3 Three different bimodal curve fittings to a volumetric distribution of *Ae. tauschii* endosperm starch. The bin size normalised volumetric size distribution was determined using a Coulter counter on endosperm starch from *Ae. tauschii* accession TOWWC097 (grey). Two normal (N-N) (a), two log-normal (L-L) (b), or a log-normal and a normal (L-N) (c) curves were fitted onto the size distribution. The dashed pink curve represents the A-type granule curve, the dashed orange curve represents B-type granule curve, and the solid turquoise curve is the sum of both curves. The uncertainty and standard error of regression (S) are reported in the top right. (d) All three fittings have been overlaid for comparison. 351 Coulter counter traces from *Ae. tauschii* accessions were ran through the fitting script, and the fitting with the lowest standard error of regression (e) and lowest uncertainty (f) were calculated and are shown as percentages.

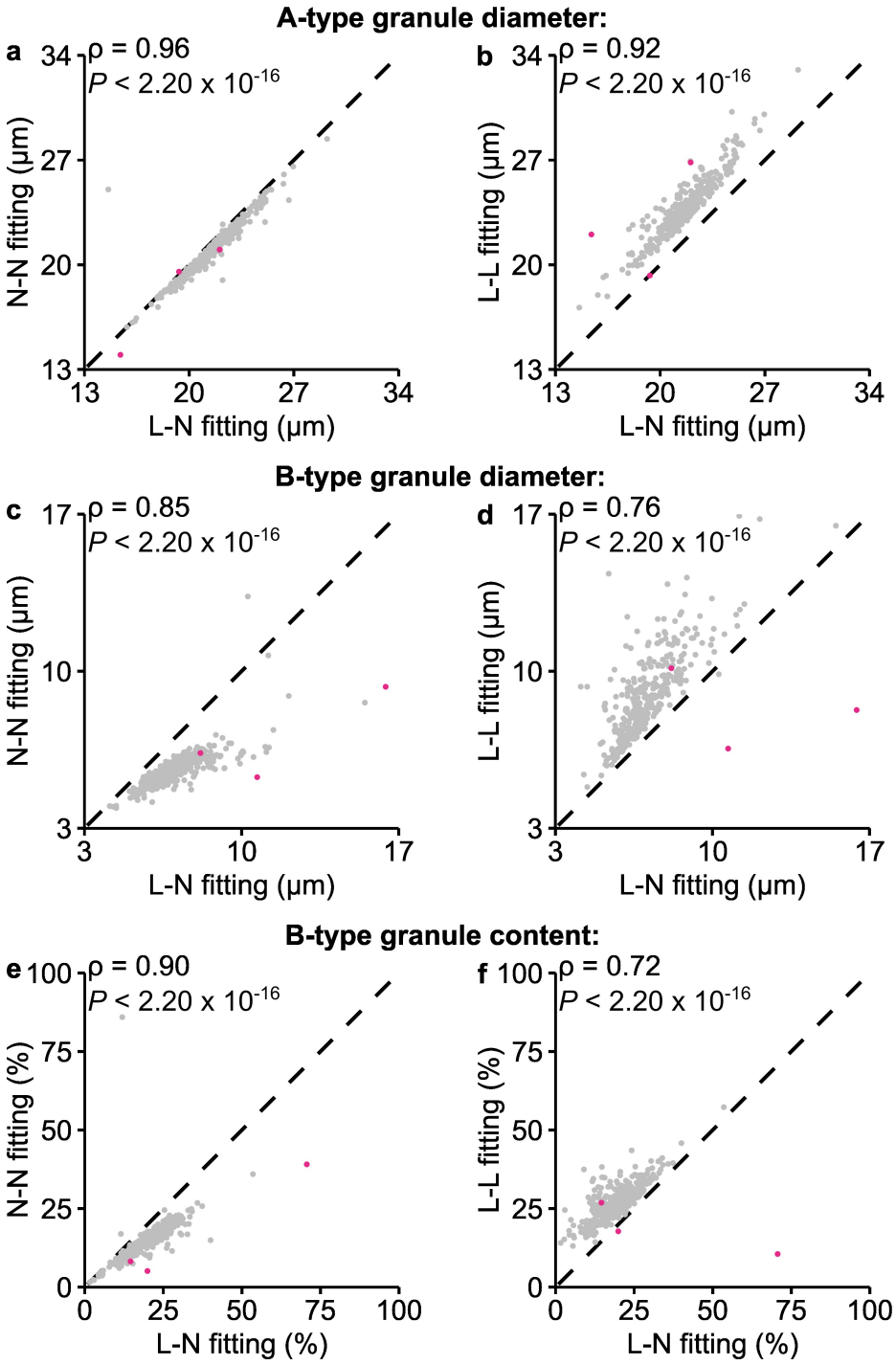


Figure 3.4 Comparison of parameters derived from curve fittings. A-type granule diameter (a-b), B-type granule diameter (c-d) and B-type granule content (e-f) were derived from the fitting two normal (N-N) (a, c, e), two log-normal (L-L) (b, d, f) and plotted against those derived from the log-normal-normal (L-N) fitting (a-f). Individual data points (n = 351) are shown as grey dots except for TOWWC021 which is highlighted in pink. The Spearman's rank coefficients (ρ) and corresponding P values shown in the top left of each graph. A dashed black line of $y=x$ is overlaid on each plot to show the expected relationship if the fittings gave the same values.

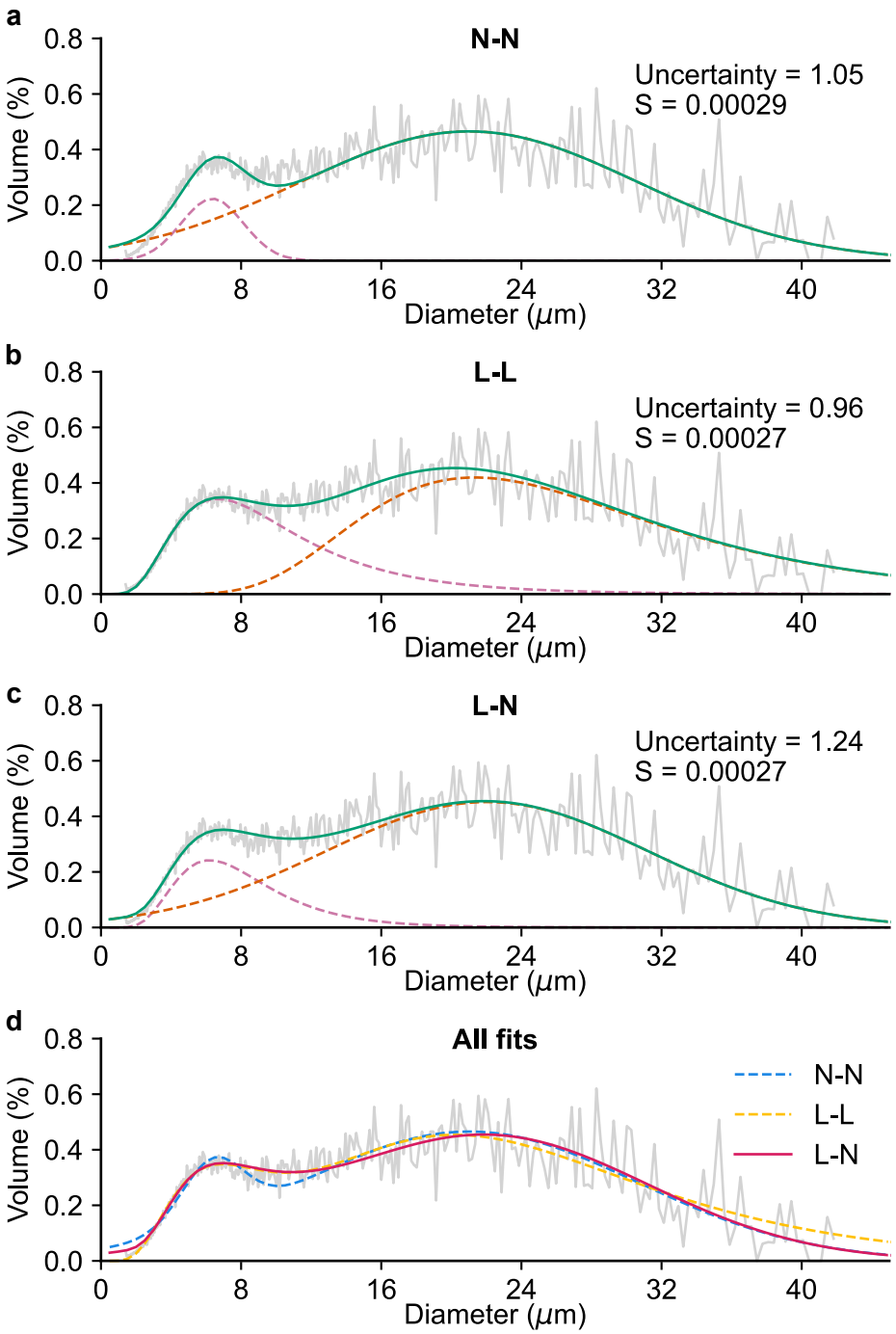


Figure 3.5 Curve fitting is not successful for non-bimodal size distributions. The bin size normalised volumetric size distribution was determined using a Coulter counter on endosperm starch from *Ae. tauschii* accession TOWWC021 (black). (a-c) Two normal (N-N) (a), two log-normal (L-L) (b), or a log-normal and a normal (L-N) (c) curves were fitted onto the size distribution. The dashed orange curve represents the A-type granule curve, the dashed pink curve represents B-type granule curve, and the solid turquoise curve is the sum of both curves. The uncertainty and standard error of regression (S) are reported in the top right. (d) All three fittings have been overlaid for comparison.

3.2.3 There is variation in granule parameters across the diversity panel

The mean values obtained from the curve fitting strategy were compared across the entire panel (Figure 3.6a-c). There was significant variation in the sizes of both A- and B-type granules within the accessions: A-type granule diameter ranged from 16.9 μm to 25.6 μm ($P < 0.05$, Kruskall-Wallis test) and B-type granule diameter ranged from 4.8 μm to 11.8 μm ($P < 0.05$, Kruskall-Wallis test). However, the most extreme variation was in B-type granule content which ranged from 6.4% to 36.6% ($P < 0.05$, Kruskall-Wallis test).

In addition to the values derived by curve fitting, I also estimated the percentage of B-type granules by number by setting a threshold at 10 μm (Chia *et al.*, 2020). Granules with a diameter < 10 μm were counted as B-type granules and those with a diameter ≥ 10 μm were counted as A-type granules. There was significant variation in the percentage number of B-type granules ranging from 86.9% to 96.7% ($P < 0.05$, Kruskall-Wallis test, Figure 3.6d).

The final starch parameters assessed across the entire panel were amylose content and total starch content (Figure 3.6e-f). Amylose content was measured via an iodine colourimetry method and ranged from 15.7% to 29.2% but did not vary significantly among the accessions ($P > 0.05$, Kruskall-Wallis test). The variation in total starch content was 31.7% to 60.9% ($P > 0.05$, Kruskall-Wallis test). However, there were no statistically significant differences across the panel, perhaps due to the fairly large variation within each accession. Hence variation in total starch content cannot underpin the observed variation in the starch granule parameters above.

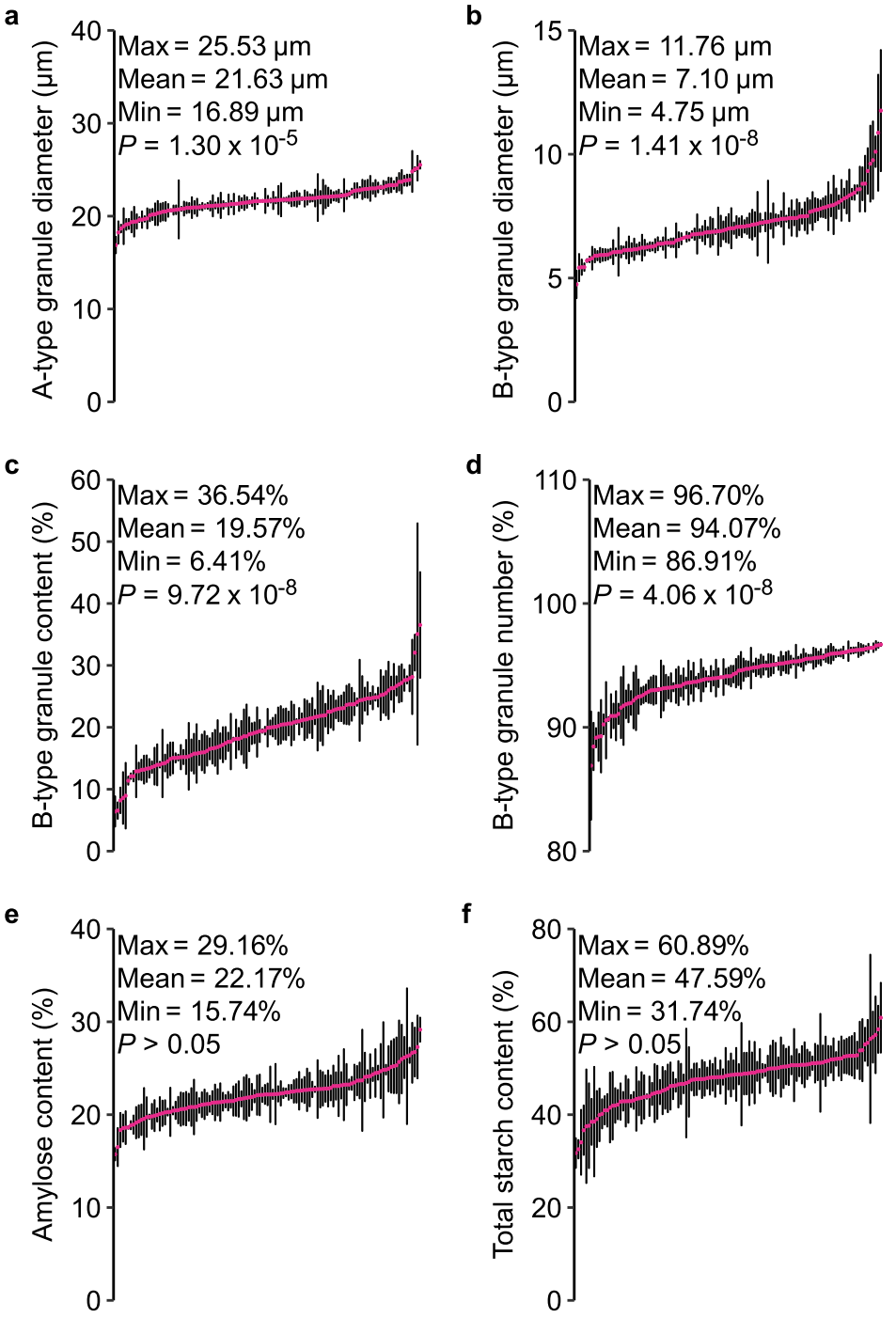


Figure 3.6 There is significant variation in granule size, content and number in the *Ae. tauschii* diversity panel. A Coulter counter was used to measure the size distribution of endosperm starch granules, a normal curve was fitted to model A-type granules and a log-normal curve was fitted to model B-type granules. The mean A-type granule diameter (a) and B-type granule diameter (b) were determined from these curves. The B-type granule content (c) is the relative proportion of B-type granules (by volume). (d) The B-type granule number is the relative number of granules with a diameter less than 10 µm. (e) The amylose content (% *w*/*w* of starch) of endosperm starch as determined by iodine colourimetry. (f) The total starch content (% *w*/*w* of

wholegrain flour) as determined by an enzymatic assay. For all graphs the data are shown as means (pink points) ± SE (black bars) with individual cultivars arranged from left to right in ascending order (*n* = 117, with 3 biological replicates per line). Statistically significant differences were calculated using Kruskal-Wallis tests.

3.2.4 Accessions with extreme starch granule parameters retain spherical and lenticular morphologies

As there was large variation in starch granule size, number and content (Figure 3.6), I investigated if this was linked to morphological variation in granule shape. To test this, I selected lines with extremely high, or extremely low values for A-type granule diameter, B-type granule diameter, B-type granule content and B-type granule number. These are henceforth referred to accessions with ‘extreme’ parameters (Table 3.1). Granule morphology was assessed using scanning electron microscopy (SEM) (Figure 3.7). All accessions had lenticular A-type granules and spherical B-type granules. Interestingly, the B-type granules sometimes had an indentation in their centres, giving almost a doughnut-like appearance, for example in the image of TOWWC141 in Figure 3.7. Likewise, some of the A-type granules had dips in their centres, this is exemplified in the images of TOWWC064 and TOWWC106 in Figure 3.7. Both these phenomena occurred at a low frequency in almost all accessions, and did not correlate with the starch granule phenotypes. The micrograph for TOWWC021 reveals that there are many intermediate granules. They appeared more spherical than lenticular in shape, so they are more likely to be large B-type granules rather than small A-type granules. This is consistent with the Coulter counter trace and gives confidence in the large B-type granule diameter calculated via the L-N curve fitting. Overall, the differences in granule size, content and number which are seen in the extreme accessions, does not stem from unusual A- or B-type granule shapes.

Table 3.1 *Ae tauschii* accessions with the most extreme starch granule parameter values. For each parameter, the three accessions with the highest and lowest values were identified. The table is arranged so that the accession with the lowest value is at the top of the low box, getting higher as you progress down the box, and the reverse for the high column so that the accession with the highest value is at the top of the high box.

Parameter	Low	High
A-type granule diameter	TOWWC016, TOWWC064, TOWWC168	TOWWC083, TOWWC106, TOWWC051
B-type granule diameter	TOWWC016, TOWWC046, TOWWC059	TOWWC021, TOWWC106, TOWWC187
B-type granule content	TOWWC167, TOWWC168, TOWWC064	TOWWC141, TOWWC021, TOWWC099
B-type granule number	TOWWC057, TOWWC187, TOWWC193	TOWWC155, TOWWC095, TOWWC003

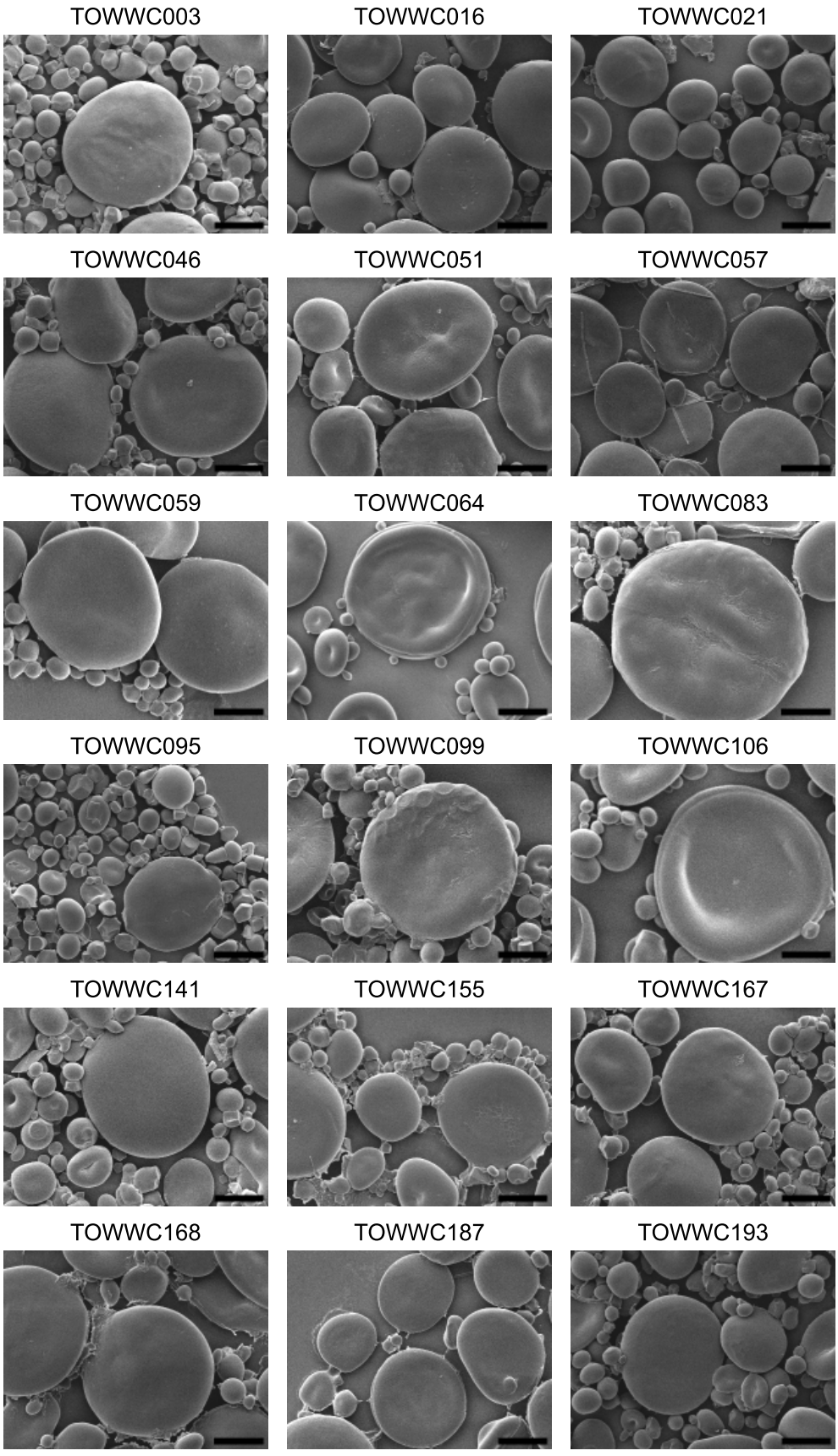


Figure 3.7 Starch from the 'extreme' *Aegilops tauschii* accessions has normal bimodal morphology. Representative scanning electron micrographs of purified endosperm starch. Accessions are arranged in numerical order, scale bar = 10 μm.

3.2.5 The starch granule parameters only weakly correlate with each other

By measuring numerous starch parameters on the same grains from many accessions, this study provided an opportunity to explore if there are any relationships between these parameters. All parameters that varied significantly across the panel (Figure 3.6) were correlated against each other (Figure 3.8). As the data were not normally distributed, the non-parametric Spearman's rank correlation test, was used to test the strength of the correlations between the parameters. The strongest significant correlation was between B-type granule content and B-type granule number (ρ = 0.61). This suggests that B-type granule content is partially influenced by number, but there must also be other factors that contribute. For all other correlations ρ was $\leq \pm 0.4$ yet $P \leq 0.05$, so these correlations are significant yet quite weak. Interestingly, it was only B-type granule diameter and B-type granule number which had a negative correlation coefficient (ρ = -0.35), and all other coefficients were positive. This could reflect a weak trend that when more B-type granules initiate the B-type granules are smaller. It is perhaps unsurprising that the correlation coefficients are weak when the graphs in Figure 3.81-f are considered. There are numerous examples where there are accessions which seem to be outliers. For example, Figure 3.8d shows a general trend that as B-type granule content increases, there is a corresponding increase in B-type granule diameter, however there is one accession (TOWWC141) has a high B-type granule content (36%) yet a relatively low B-type granule diameter (7.12 µm). These few outlier accessions may be decreasing the overall value of the correlation coefficients.

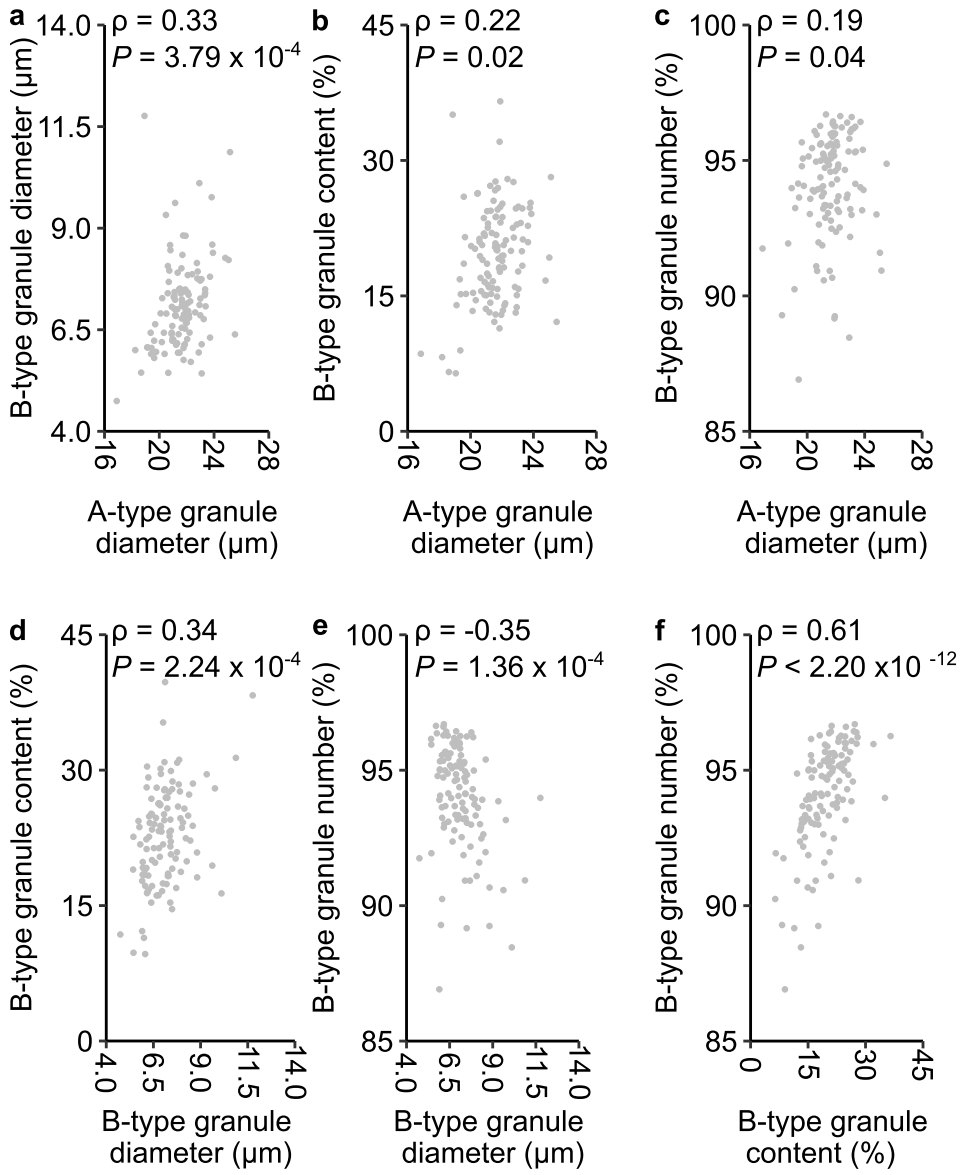


Figure 3.8 There are weak correlations between different starch granule parameters. (a-f) Starch granule parameters which varied significantly across the *Ae. tauschii* diversity panel were plotted against each other to reveal any correlations. The data are displayed as mean values (grey points, *n* = 117, with 3 biological replicates per accessions) and the axes range from the lowest to the highest values for each parameter. Spearman rank correlation tests were conducted the correlation coefficients (ρ), and corresponding *P* value are on the top left of each graph.

3.2.6 Correlations of starch granule parameters with geographical location and phylogeny are weak

It was hypothesized that variation in starch granule parameters might correlate to the geographical location of accession collection. This could be prominent if a particular starch granule phenotype has an adaptive advantage to specific environmental conditions. To test this, the locations from which the accessions were originally collected from were obtained from the JIC GRU (https://www.seedstor.ac.uk/search-browseaccessions.php?idCollection=38). Accurate location data were available for 93 accessions, and these were used to map the starch parameters that varied significantly across the entire diversity panel (Figure 3.9a-d). Visual representation of the parameters across a map revealed no obvious trends between any of the starch parameters and location. Longitude and latitude were separately correlated with the mean starch parameter values using Spearman's rank correlations tests (Figure 3.9e-l). When correlating with latitude, there was no relationship with either A-type granule diameter, B-type granule diameter or B-type granule content with the correlation coefficients being almost zero. With B-type granule number there was a significant, but weak, correlation ($\rho = 0.30$). Likewise, when the same analysis was repeated with longitude, no significant correlation was detected with any starch parameter.

In addition, it was hypothesized that there could be phylogenetically determined variation in starch granule parameters, such that closely related accessions might have more similar starch granule parameters than those that are more distantly related. This might be particularly prominent if the starch parameter has been selected for during *Ae. tauschii* evolution. To investigate this, a phylogenetic tree of lineage two *Ae. tauschii* accessions was generated using a script from Gaurav *et al.* (2021) (Figure 3.10). Lineage one accessions and wheat, which were not used in my project, were excluded from the tree. Unfortunately, 23 of the accessions included in this study were not included in the phylogenetic trees of Gaurav *et al.* (2021) and had to be excluded from this analysis. In terms of associations between starch granule parameters and phylogeny, there were no strong links. In fact, there were multiple cases where accessions within the same phylogenetic clade have contrasting starch granule parameters. For B-type granule number there is a clade (Figure 3.10d, clade towards the top left) which contains TOWWC187, TOWWC193, which both have low B-type granule numbers, and also TOWWC059, TOWC092, TOWWC119 which all have relatively high B-type granule number phenotypes. Likewise, the same clade has large variation in B-type granule diameter. Other clades also contain variation for instance TOWWC016, which has a low A-type granule diameter, is in a clade with accessions

with relatively high A-type granule diameter (Figure 3.10d, clade towards the bottom left). Moreover, TOWWC167 has an extremely low B-type granule content but the other accessions within its clade have much higher values for this parameter (Figure 3.10d, clade towards the top right). Therefore, there were no strong links between starch granule parameters and phylogeny. Furthermore, the distribution of the parameters across phylogeny would makes a genome wide association study a favourable strategy for discovering genetic factors underlying these parameters (Chapter Four).

The phylogenetic tree (Figure 3.10) also allows easy identification of accessions that have low or high values for multiple starch granule parameters. There are several accessions which stand out. Firstly TOWWC0064, which was identified as an accession with low A-type granule diameter and B-type granule content in Table 3.1, also has relatively low values for B-type granule diameter and B-type granule number compared to the other accessions studied. Secondly, TOWWC016 which has low A-type and B-type granule diameters (Table 3.1) also seems to have relatively low B-type granule content, yet B-type granule number is not as low when compared to other accessions. Some accessions, including TOWWC113 and TOWWC115, have relatively high values for all parameters measured.

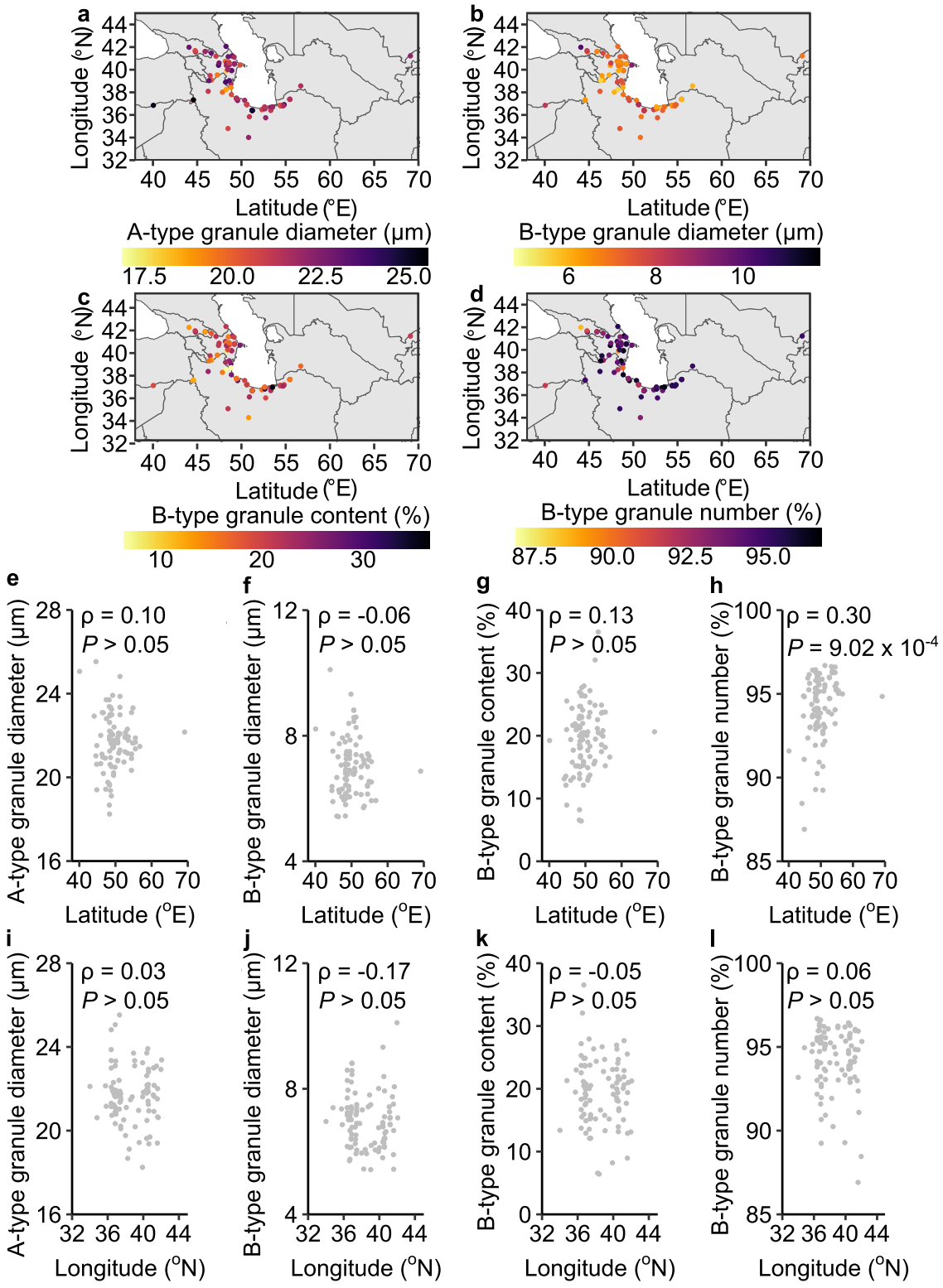


Figure 3.9 There was no clear trend between geographical collection location and starch parameters. Locations where the *Ae. tauschii* accessions were collected from are plotted on a map (a-d), with colours showing the values of the mean starch parameters (yellow = low, purple = high). The mean starch granule parameters were plotted individually against latitude (e-h) or longitude (i-l). Spearman's rank correlation coefficients (ρ) and the *P* values are displayed in the top left of each graph, *n* = 93.

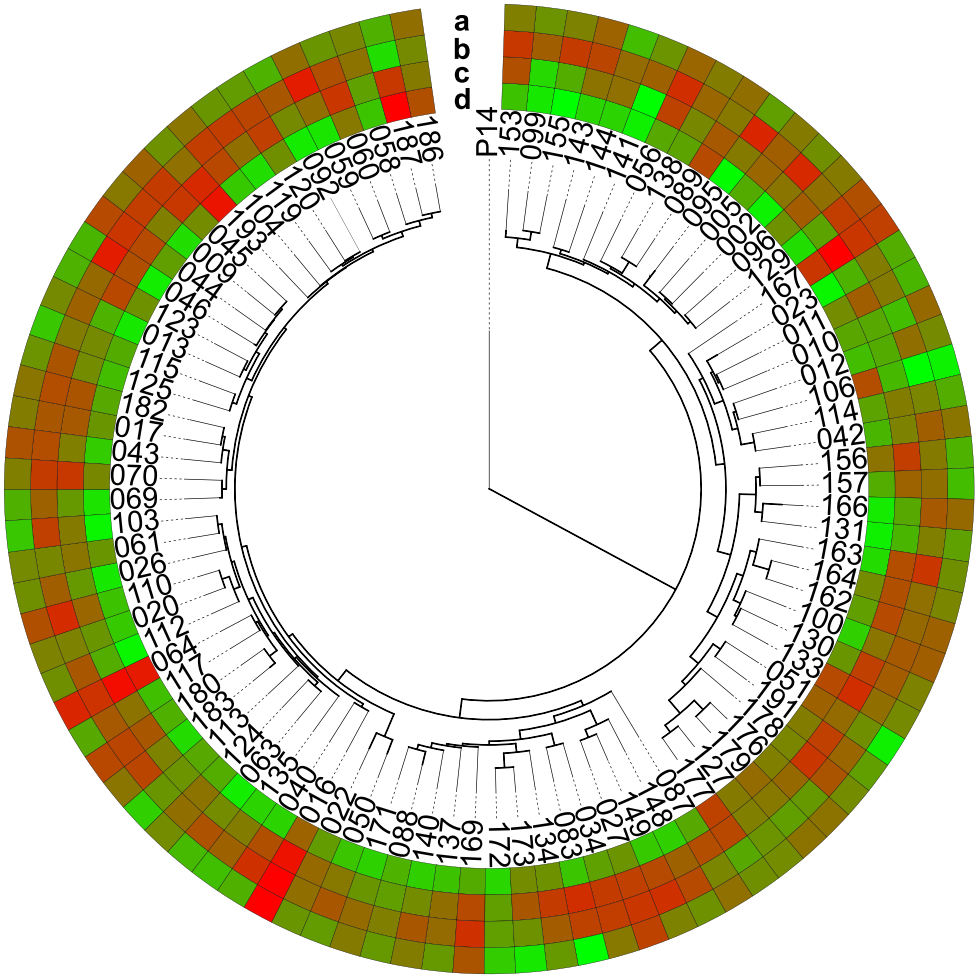


Figure 3.10 Phylogeny does not fully explain the variation in starch granule parameters. Phylogenetic tree of lineage two *Ae. tauschii* accessions based on k-mer phylogeny. The mean starch granule parameter which significantly vary across the entire panel are shown as boxes surrounding the tree, with a colour scale where dark red represents the lowest value for the parameter, and bright green represents the highest value for the parameter. (a) A-type granule diameter, (b) B-type granule diameter, (c) B-type granule content, (d) B-type granule number (*n* = 94). Accession names have been shortened for clarity, with the TOWWC prefix removed, P14 represents the tetraploid wheat outgroup Hoh-501-P14.

3.2.7 Ae. tauschii contains novel variation in endosperm starch compared to modern wheat

As *Ae. tauschii* is an undomesticated wild relative of modern bread wheat, I hypothesized that the variation seen in the endosperm starch of *Ae. tauschii* might not be present in hexaploid wheat due to potential loss during domestication. To test this, I selected 16 hexaploid wheats (Chapter Two; Table 2.2). These were diverse in growth

habitat and location of origin and chosen to capture maximum genetic diversity (Walkowiak *et al.*, 2020). The majority of these are modern cultivars, although there is one landrace – Chinese Spring. I obtained grains from these cultivars from Cristobal Uauy's lab (JIC), grown in at least triplicate under glasshouse conditions, although these were not grown at the same time as the *Ae. tauschii* panel.

Endosperm starch was purified from mature grains and the size distribution of starch granules was measured using a Coulter counter. There was significant variation found in the size, content and number of granules (Figure 3.11). Unlike the *Ae. tauschii* panel where no significant variation was found for amylose and total starch contents, for the wheat lines used here there was significant variation in both parameters ($P < 0.05$, one-way ANOVA (analysis of variance)). Surprisingly, the amylose values are all lower than those found in the *Ae. tauschii* panel, and lower than those previously reported in the literature (Chapter One; Table 1.1). Interestingly, the two lines with the lowest total starch content were Baj (42.09%) and Chinese Spring (46.54%). Chinese spring is a landrace, and although Baj is considered a cultivar, it is used by CIMMYT as a 'checker' line. This means it is a control used to compare experiments between years rather than being a cultivar for development. Hence the starch content may be lower in these lines as they have not undergone as intense selection pressure.

The variation in granule size, content and number was compared to those measured in *Ae. tauschii* (Figure 3.12). The starch content and amylose results were excluded from this analysis as there was no significant difference in these in *Ae. tauschii*. There was a significant difference in the mean A-type granule diameters with *Ae. tauschii* tending to have larger A-type granules compared to the wheat cultivars investigated here. Otherwise, there were no significant differences between the mean values for B-type granule diameter, B-type granule content, and B-type granule number between the *Aegilops tauschii* accessions and the wheat cultivars. However, when the total range is considered, there is variation in *Aegilops tauschii* that is not found in wheat. The wheat cultivars used here have less variation in B-type diameter (2.52 μm range in wheat and 7.01 μm in *Ae. tauschii*), A-type diameter (4.95 μm range in wheat and 8.64 μm in *Ae. tauschii*) and B-type granule number (4.97% range in wheat and 9.79% in *Ae. tauschii*). In contrast, the wheat cultivars had similar levels of variation in B-type granule content (27.71% range in wheat and 30.01% range in *Ae. tauschii*). The *Ae. tauschii* accessions with parameter values which lie outside those seen in wheat could be a source of useful variation.

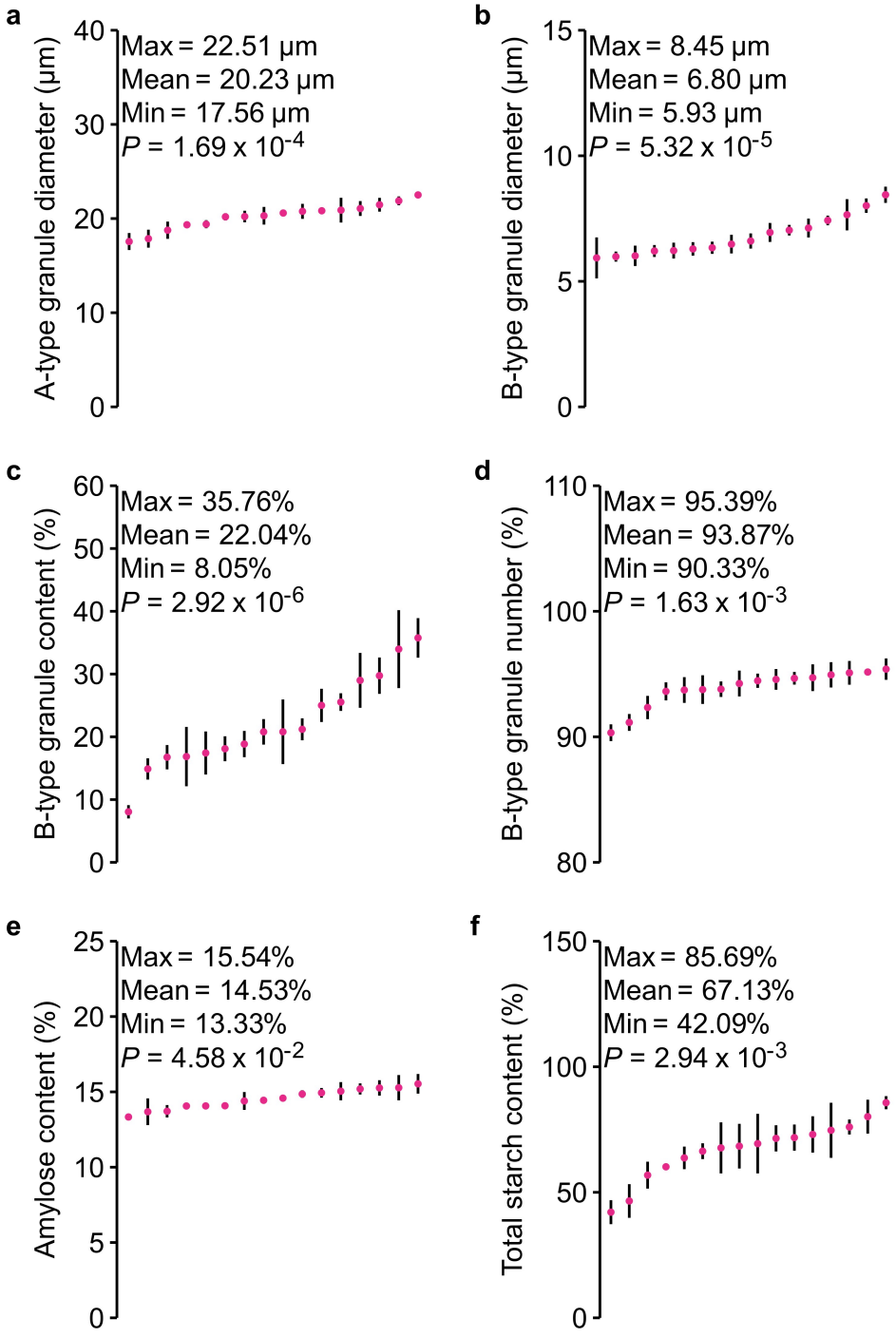


Figure 3.11 There is significant variation in granule size, content, number, amylose content and total starch content in selected *T. aestivum* cultivars. A Coulter counter was used to measure the size distribution of endosperm starch granules, a normal curve was fitted to model A-type granules and a log-normal curve was fitted to model B-type granules. The mean A-type granule diameter (a) and B-type granule diameter (b) were determined from these curves. The B-type granule content (c) is the relative proportion of B-type granules (by volume). (d) The B-type granule number is the relative number of granules with a diameter less than 10 µm. (e) The

amylose content (% *w*/*w* of starch) of endosperm starch as determined by a spectrophotometric assay. (f) The total starch content (% *w*/*w* of wholegrain flour) as determined by an enzymatic assay. For all graphs, the data are shown as means (red points) ± SE (black bars) with individual cultivars arranged from left to right in ascending order (*n* = 16, with 3-4 biological replicates per line). Statistically significant differences were calculated by a one-way ANOVA.

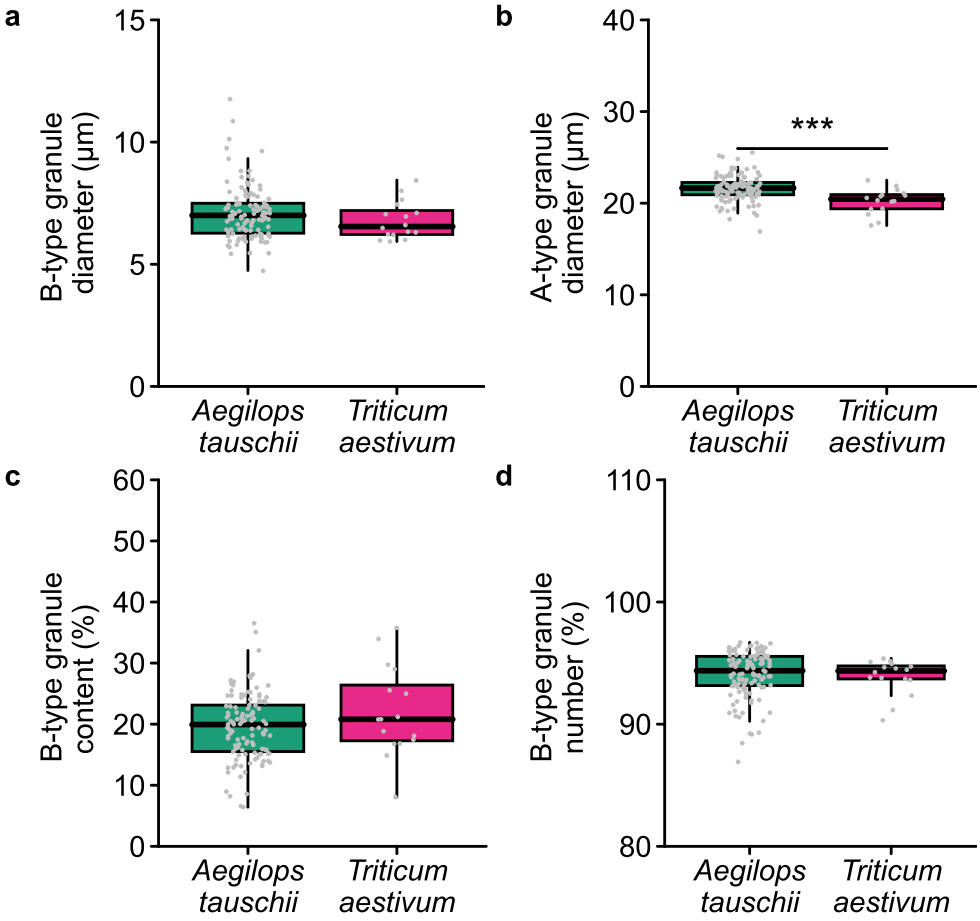


Figure 3.12 Comparing starch granule parameters of *Ae. tauschii* and *T. aestivum* reveals novel variation in *Ae. tauschii*. The granule properties for the *T. aestivum* cultivars were compared to those from the *Aegilops tauschii* accessions. (a) B-type granule diameter, (b) A-type granule diameter, (c) B-type granule content, (d) B-type granule number. Data are shown as box plots with the middle line representing the median, the box representing the interquartile range and the whiskers extending to 1.5x the interquartile range. The two-sample Wilcoxon signed rank test was used to determine if there was a significant difference between *Ae. tauschii* and *T. aestivum* values for each starch parameter (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Individual data points are shown as grey dots (*n* = 117 for *Ae. tauschii*, turquoise, and *n* = 16 for *T. aestivum*, pink).

3.3 Discussion

Ae. tauschii was a key species in the domestication of hexaploid bread wheat (Marcussen *et al.*, 2014). As it is a wild grass, it has greater genetic and phenotypic diversity than what is observed in domesticated wheat cultivars (Gaurav *et al.*, 2021). Previous studies have reported variation in granule size, B-type granule content and amylose content in *Ae. tauschii* (Mohammadkhani *et al.*, 1998; Stoddard, 1999a). However, these studies utilised a limited number of accessions and measured only a set number of characteristics. In this chapter, I performed an in-depth characterisation of endosperm starch in 117 different *Ae. tauschii* accessions, revealing novel variation and properties which appear not to be present in modern wheat cultivars.

3.3.1 A Python package allows accurate quantification of granule diameters and contents

I generated a Python package to analyse starch granule size distributions from *Ae. tauschii*. A curve fitting approach is considered more accurate compared to a binary threshold (Tanaka *et al.*, 2017) as it allows better resolution of the region where A-type and B-type granules overlap. Compared to pre-existing scripts (Tanaka *et al.*, 2017; Chen *et al.*, 2024), which all fit a single distribution, my script offers flexibility as it fits three different distributions – N-N, L-L and L-N. A goodness of fit test determined that the L-N fitting provided the best fit for most of the *Ae. tauschii* size distribution traces (Figure 3.3). This fitting has not previously been used in published literature. Not enough is known about the biology underlying starch granule size distribution to assume it should follow an L-N curve, and it is worth noting that the shape of the size distribution is also influenced by technical factors. For example, the Coulter counter assumes spherical particle size when it calculates a diameter value for the *x*-axis, which could influence the shape of the distribution. Overall, my script provides a new distribution (L-N) to accurately measure granule size and content in *Ae. tauschii*. In the future, it could be used to test whether starch from other Triticeae species also best fit a L-N distribution, or instead L-L or N-N distributions.

3.3.2 Variation in granule size and number exceeds that previously reported

Across the diversity panel, I saw significant variation in granule size, B-type granule number and B-type content (Figure 3.6). The greatest variation was observed was in B-type granule contents, 6-37%, which exceeds that previously reported by Stoddard (1999a) (15-38%) and Konik-Rose *et al.* (2009) (21-36%). The upper range of the B-type granule content is in line with previous literature, suggesting that these studies captured the upper limit of diversity in *Ae. tauschii* B-type granule content. It is curious

that the upper limit for B-type granule content in *Ae. tauschii* appears to be ~36-38%, whilst in other wild grasses much higher values are reported, such as 62% in *T. monococcum* (Stoddard, 1999a). Perhaps the high B-type granules observed in *T. monococcum* is due variation which is not present in *Ae. tauschii*. Alternatively, *T. monococcum* might have differences in amyloplast size affecting initiations and growth of B-type granules, or differences in the length of the grain filling period. At the other end of the spectrum, I have discovered accessions with low B-type granule content, as low as 6% (Table 3.1). Values like these are closer to published B-type granule contents of other *Aegilops* species, such as *Aegilops uniaristata*, 6-13%, or *Aegilops kotschyi*, 5-13% (Stoddard and Sarker, 2000). In other *Aegilops* species, such as *Ae. peregrina*, endosperm starch with no B-type granules have been reported. However, the SEM imaging performed here (Figure 3.7) reveal that even in the accessions with low B-type granule contents, B-type granules are still present. Thus, even though I see large variation in B-type granule content in *Ae. tauschii* endosperm starch, it is possible that there is even more interesting variation outside of *Ae. tauschii*, or alternatively in accessions not studied here.

3.3.3 Variation in amylose content was similar to previous reports

The other parameter I measured for which there is much published literature is amylose content. I measured a 16-30% difference in amylose content across the diversity panel. This is similar to the 21-34% reported by Mohammadkhani *et al.* (1998) and the 24-30% reported by Konik-Rose *et al.* (2009), but I have identified new variation in amylose content at the lower end of the scale. It is worth noting that I used an iodine colourimetry method for assaying amylose content (Washington *et al.*, 2000). This was desirable as it can be performed at small scales and at high throughput, hence enabling me to obtain apparent amylose contents for three biological replicates of all 117 *Ae. tauschii* accessions. This is a widely published method, although values can sometimes be inaccurate due to factors including inconsistencies in I_2/KI concentration and temperature (Vilaplana *et al.*, 2012). As all measurements here were conducted together and with the same set up, it was concluded that the values are comparable to each other. They are also consistent with the previous literature, once again giving confidence that the values measured here are reliable. The gold standard method for amylose content determination is size exclusion chromatography. However, this cannot be performed in a high throughput manner, making it unrealistic for this work. Importantly, Vilaplana *et al.* (2012) conclude that iodine colourimetry produces similar results to that of size exclusion chromatography.

Interestingly, 16% was the lowest amylose content recorded and there were no accessions with a waxy phenotype, where amylose content is 0%. In some other species, naturally occurring low amylose plants have been observed. For example, wild *Arabidopsis thaliana* accessions have been uncovered with <0.05% amylose contents (Seung *et al.*, 2020). The fact that no amylose-free *Ae. tauschii* accessions were identified suggests that this diversity panel has no loss-of-function mutations in the amylose synthesis gene *GBSSI*. It is interesting to wonder why no *waxy Ae. tauschii* accessions have been discovered, either here or in previous studies. It could be that, whilst *waxy* cereals can survive and thrive in agricultural settings, maybe it offers a disadvantage in the wild and is selected against.

For the other parameters measured, comparing the values to the literature is challenging due to the lack of previous studies. The mean value reported here for B-type granule diameter, 7.10 µm, is slightly higher than modal diameter Stoddard reports, 6.3 µm. By contrast, the mean value for A-type granule diameter, 21.63 µm, is lower than the modal diameter reported by Stoddard (1999a), 23.8 µm. No A-type granules as large as those reported by Stoddard (1999a) (57.5 µm) were observed. One reason could be that the particles observed by Stoddard were multiple granules instead of single granules.

One note of caution when interpreting these data, are that they are a result of a single round of growth in glasshouse conditions. However, each accession had at least three replicates grown in a randomised block design. Due to the amount of time and work it would have taken to repeat the growth and phenotyping, it was decided that another round of growth was not the best approach to take for this study. Instead, time was dedicated to the association mapping (Chapter Four) and characterisation of candidate genes influencing these parameters (Chapter Five). Hence, the robustness of these measured parameters across different periods of growth has not been assessed. Many factors can affect starch granule development, especially that drought or heat stress, which when applied across the entire growing period is associated with a lower B-type granule content in wheat (Zhang *et al.*, 2017). The plants were grown during early 2019 to avoid the height of the summer heat to reduce the impact of extreme drought or heat stress. Despite this, they may still have been affected by the environmental conditions. Therefore, it would be interesting in the future to repeat the growth and phenotyping of the *Ae. tauschii* diversity panel in the same and different conditions to test how robust the starch granule phenotypes are to environmental factors.

Whilst here I have only investigated the lineage two accessions of the *Ae. tauschii* diversity panel, the full panel also includes accessions from lineages one and three. These other lineages are proposed to be less important for the evolution of hexaploid wheat (Gaurav *et al.*, 2021) but could potentially contain interesting variation in starch traits. This has already been demonstrated for other parameters, for instance resistance to curl mite provided by *CURL MITE COLONISATION 4* (*CMC4*) is found almost entirely in lineage one (Gaurav *et al.*, 2021). Future work could extend my analysis and characterise endosperm starch in these additional *Ae. tauschii* lineages.

3.3.4 Starch granule size and number are complex parameters affected by genetic and environmental parameters

This study provided an opportunity to explore the relationships between the measured starch granule parameters using correlation analysis. The strongest correlation was between B-type granule number and B-type granule content. This suggests that B-granule content is more affected by granule number than granule size. Conversely, only weak correlations were found between the other starch parameters. This was surprising as it is easy to hypothesize that accessions with larger B-type granules might have smaller A-type granules if starch has been allocated to B-type granules instead. Another correlation I might have expected to see is between B-type granule number and diameter. For example, accessions with a larger number of B-type granules might have had smaller B-type granules. This would be expected from a resource allocation scenario, whereby the total starch is allocated into many more B-type granules, meaning that for each granule there is less substrate available for growth, resulting in smaller B-type granules. However, in both cases, only weak correlations were found, suggesting that other factors are influencing these parameters. To my knowledge, this is the first example of a comprehensive correlation analysis on granule size, content and number in any Triticeae species. Previous studies have typically correlated amylose content to the B-type granule content. It was decided that this type of analysis would be inappropriate here given that there was no significance difference in amylose content across the diversity panel (Figure 3.6e). For wheat, Li *et al.* (2008) found a negative relationship between amylose content and B-type granule content in wheat starch. This was consistent with the results of Peng *et al.* (1999) and Tao *et al.* (2016), who both concluded that A-type granules have a greater amylose content than B-type granules. Whilst in barley, Bathgate and Palmer (1972) found the opposite correlation with B-type granules having a greater amylose content. Interestingly, a previous study on *Ae. tauschii* starch found no correlation between amylose content and B-type granule content (Stoddard and Sarker, 2000).

In addition to exploring correlations between the different starch parameters, I hypothesized whether the observed variation in granule parameters could be explained, in part, by location of collection. The diversity panel was collected from a large area spanning 8º of longitude and 29º of latitude (Figure 3.9). Whilst there is no data about the specific environmental conditions where the accessions were collected, Brunazzi *et al.* (2018) sampled *Triticum urartu* accessions from similar locations and saw variation in mean temperature, annual precipitation and isothermality. Therefore, it is likely that the *Ae. tauschii* accessions I have investigated originate from areas with distinct climatic conditions. It is reasonable to assume that the climatic and environmental conditions could have applied selection pressure on the *Ae. tauschii* accessions, and this could be reflected in differing endosperm starch properties. However, to my knowledge, no example of this has been reported. When I tested this, there were no obvious correlations between location of origin and the starch properties (Figure 3.9). Whilst this analysis is insightful, it has only included two geographical parameters, longitude and latitude. I could not consider other important parameters including altitude, and environmental factors such as: distance from water, soil type, level of vegetation, as this information was not available. Thus, my analysis does not give a complete view of how the environment at the collection location influences the starch parameters. However, even if this information was available, the validity of such an approach should be questioned, because the starch parameter analysis was conducted on glasshouse grown plants, and the starch parameters might not exactly match the parameters in their natural environment. Hence even if trends between starch granules and geography or environmental conditions exist, it might not be possible to detect in this glasshouse grown population.

Combining the results of the environmental analysis and the starch parameter correlation analysis together suggests that these are complex parameters and could be affected by a multitude of parameters, both environmental and genetic.

3.3.5 Ae. tauschii contains novel diversity not found in wheat cultivars

Increasing starch granule diversity in wheat is important as the size distribution has a large impact on the quality and the end uses of the starch. Therefore, I compared the diversity which I had observed in *Ae. tauschii* to that in modern wheat. There were several *Ae. tauschii* accessions with more extreme A- or B-type granule diameters or B-type granule numbers than observed in modern wheat. In the future, *Ae. tauschii* could be used to increase variation in these traits. For B-type granule content, I identified numerous *Ae. tauschii* accessions with low B-type granule content, there

was only one wheat cultivar – the CIMMYT 'checker' line Baj (8.05%) – which had a B-type granule content of less than 10%. This is consistent with previous literature with fewer than ten wheat cultivars reported as having B-type granule contents of < 10%, with the lowest value being 7% (Zhang *et al.*, 2016). Therefore, *Ae. tauschii* could also be a source of novel diversity for low B-type granule content. One note of caution which must be exerted when drawing any conclusions from these data is that the *Ae. tauschii* and wheat cultivars were not grown in parallel, hence the environmental conditions experienced during the growth of the *Ae. tauschii* vs the wheat will have been different. Therefore, from these data, I can only suggest that *Ae. tauschii* contains novel variation compared to the wheat cultivars surveyed here. Ideally these plants would be grown at the same time for a thorough comparison. Growing in parallel does not come without issues as these plants have different growth conditions and life cycles, in particular differing vernalisation requirements.

Incorporating variation from *Ae. tauschii* into modern wheat is desirable for industry, such as breadmaking where B-type granules are less favourable than A-type granules. One approach to transfer the variation into wheat is using synthetic hexaploid wheats (SHWs). SHWs can be generated by crossing *Aegilops tauschii* with a tetraploid wheat and to produce haploid F1 plants followed by artificial chromosome doubling (Figure 3.13) (Rosyara *et al.*, 2019). I decided to investigate the NIAB (National Institute of Agricultural Botany) SHW collection to see if there were any lines which had been generated with *Ae. tauschii* accessions with extreme starch granule phenotypes (from Table 3.1). This would act as a proof of concept to see if the starch granule phenotype from *Ae. tauschii* would be dominant in a hexaploid background. Unfortunately, there were no suitable SHW lines. Therefore, future work could attempt to generate novel SHWs using the accessions in Table 3.1.

In this chapter I have discovered significant variation in granule size, content and number in *Ae. tauschii* which exceeds that previously reported. As the accessions were grown in controlled environments, it is hypothesized that this variation is mainly underpinned by genetic factors. Thus, I have used a genome wide association approach to uncover the genomic loci which underlie this variation in Chapter Four.

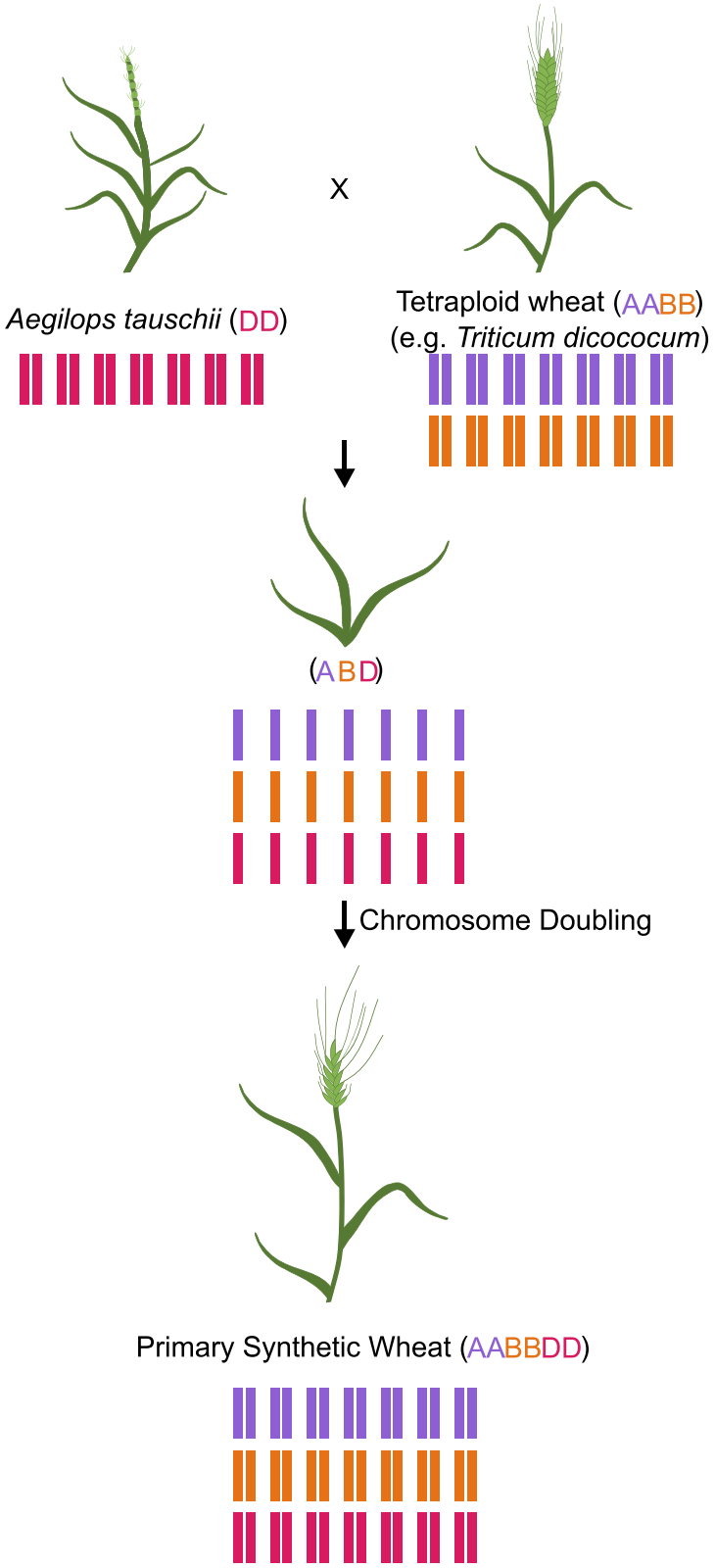


Figure 3.13 Generation of synthetic hexaploid wheat. Diagrammatic representation of the process of SHW generation. *Ae. tauschii* is crossed with a tetraploid wheat to produce haploid F_1 plants. Chromosome doubling is induced, for example with colchicine, to generate a primary synthetic wheat which can be backcrossed with hexaploid wheats.

Chapter Four – Using association genetics to identify loci associated with granule size and number

4.1 Introduction

In the previous chapter, I uncovered substantial variation in granule size, shape and number across the *Aegilops tauschii* diversity panel. In this chapter, I investigated the genetic factors that underpin this variation by conducting a Genome Wide Association Study (GWAS), and I identified candidate genes which may be involved in B-type granule formation for further characterisation.

4.1.1 There are extensive genomic resources available for Ae. tauschii

Association genetics relies on genomic resources, most notably markers and genome sequences. The first indication of the genetic architecture of *Ae. tauschii* came from a genetic map generated by Luo *et al.* (2009). This provided a sequence of markers along the chromosomes, but did not provide any physical distances. The first physical chromosome map of *Ae. tauschii* was generated from BACs (bacterial artificial chromosomes), consisting of just over 3,000 contigs which totalled to 4.02 Gb length (Luo *et al.*, 2013). Subsequently, there have been two high-quality genome assemblies for the lineage two accession TOWWC193, also called AL8/78, (Luo *et al.*, 2017; Wang *et al.*, 2021a). The later (2021) assembly is often referred to as v5. It is ~4.2 Gb in size with just under 33,000 high confidence annotated genes and 53,000 gaps, this is around 39,000 fewer gaps than the 2017 version (referred to as v4) (Wang *et al.*, 2021a). However, having a single reference genome comes with its limitations, most notably sequences that are not present in this accession are 'missing' and not known. To overcome this, genomes of three lineage one and a further three lineage two accessions have been published (Gaurav *et al.*, 2021; Zhou *et al.*, 2021). However, the most extensive effort comes from an international consortium which sequenced the *Ae. tauschii* pangenome containing 46 *Ae. tauschii* accessions, including 11 from lineage one, 34 from lineage two and one lineage three accession (Cavalet-Giorsa *et al.*, 2024). The accessions selected for the pangenome were carefully chosen and capture 99.3% of the diversity present in the *Ae. tauschii* panel I have been using in my research.

In addition to publication of genome sequences, there has been concurrent rapid development of genotyping markers available for *Ae. tauschii*. In 2013, there was release of the 10K *Ae. tauschii* Infinium SNP array, which contains 9485 functional SNPs which can be used in association mapping studies (Luo *et al.*, 2013). More recently, Gaurav *et al.* (2021) sequenced 242 diverse *Ae. tauschii* accessions to a low

coverage, between 7.5-30x, and used this to develop a presence/absence matrix of k-mers. These can act as genotyping markers during association mapping studies and offer the benefit that they can identify indels which cannot be deduced from using SNPs alone. They also reduce the reliance on a reference genome as k-mers can be produced and analysed for all accessions, meaning that variation in regions absent from the reference genome can be identified (Karikari *et al.*, 2023). Despite the advantages they still come with their limitations, most notably that the Gaurav *et al.* (2021) presence/absence matrix does not allow k-mers to be counted, and thus copy number variation (CNV) cannot be established.

4.1.2 GWAS can identify loci associated with phenotypic variation in Ae. tauschii

The plethora of genomic resources now available for *Ae. tauschii* means that it is timely to perform GWAS with this species. Over the last decade, many of these types of studies have been conducted and identified genomic loci associated with a range of different phenotypic traits (Table 4.1). Most of these studies concern either grain/kernel or pathogen traits, perhaps because these traits are easy to phenotype and are agronomically important. Furthermore, the majority of these have utilised SNP markers, potentially based on the availability of the 10K Infinium SNP array, and the lack of standardised methods for conducting a GWAS with k-mer markers (Karikari *et al.*, 2023). Regardless, this demonstrates that GWAS is a powerful approach to identify genetic variation underpinning phenotypic traits in *Ae. tauschii*.

Table 4.1 Published genome wide association studies conducted with *Aegilops tauschii*. A literature search identified GWAS conducted using *Ae. tauschii,* the type of marker used, whether SNP or k-mer, is reported. The studies are reported in order of publication and only studies where significant associations were found are included.

Phenotype	Description of *Ae. tauschii* population	Marker used	Reference
Phosphorous deficiency tolerance traits	380 accessions from the Triticeae Research Institute of Sichuan Agricultural University	SNPs	Liu *et al.* (2015b)
Plant morphology traits, including plant height, flag leaf length, lemma length, lemma width, etc…	322 accessions from the Triticeae Research Institute of Sichuan Agricultural University	SNPs	Liu *et al.* (2015a)
Tolerance to cadmium	235 accessions from the Triticeae Research Institute of Sichuan Agricultural University	SNPs	Qin *et al.* (2015)
Root and shoot weight, lengths and diameter under drought and non-drought conditions	373 accessions from the Triticeae Research Institute of Sichuan Agricultural University	SNPs	Qin *et al.* (2016)
Grain width, length and weight	177 accessions from the Wheat Germplasm Collection at Punjab Agricultural University	SNPs	Arora *et al.* (2017)
Amount of iron, zinc, copper and manganese in grains	114 accessions from Wheat Germplasm Collection at Punjab Agricultural University	SNPs	Arora *et al.* (2019a)
Stem rust resistance	195 accessions originating from diverse areas surrounding the Caspian Sea	k-mers in NLR (Nucleotide binding/leucine-rich repeat) genes	Arora *et al.* (2019b)

Table 4.1 Published genome wide association studies conducted with *Aegilops tauschii*. (continued)

Phenotype	Description of *Ae. tauschii* population	Marker used	Reference
Grain shape	221 accessions from China, the Middle East and Central Asia	SNPs	Zhao *et al.* (2021)
Kernel length, width, volume, surface area and width:length	223 accessions from the Triticeae Research Institute of Sichuan Agricultural University	SNPs	Wang *et al.* (2021b)
Stem rust resistance, flowering time, trichome number, spikelet number, powdery mildew resistance and wheat curl mite resistance	Between 147 and 210 accessions from an *Ae. tauschii* diversity panel collected from the Fertile Crescent region	k-mers	Gaurav *et al.* (2021)
Fusarium crown rot resistance	286 accessions	SNPs	Lin *et al.* (2022)
Fusarium head blight and deoxynivalenol resistance	147 accessions from an *Ae. tauschii* diversity panel collected from the Fertile Crescent region	k-mers	Kirana *et al.* (2023)
Leaf hair density	293 accessions originating from northern Syria and Turkey to western China	SNPs	Mahjoob *et al.* (2022)

4.1.3 GWAS in Ae. tauschii for starch related parameters

So far, there have been no GWAS studies in *Aegilops tauschii* that investigated endosperm starch properties, or even quality traits more generally (Table 4.1). Even when looking across the Triticeae, there are only a handful of GWAS studies concerning endosperm starch, and most focus on total starch content or amylose and amylopectin (Chapter One; Table 1.3). To my knowledge there is only one report of a GWAS in a Triticeae species, *Triticum aestivum*, linked to granule size parameters (Li *et al.*, 2017). Despite this, GWAS for granule size and content are still promising strategies and have been successful in other species. In a population of 266 maize lines, a GWAS strategy identified seven QTLs, across chromosomes 3, 6 and 7, associated with starch granule length and width (Liu *et al.*, 2018). Similarly, a GWAS for granule morphology in potato identified QTLs across chromosomes 1, 7, 11 and 12 (Khlestkin *et al.*, 2020).

Beyond GWAS, *Aegilops* species have been used in QTL mapping to discover factors influencing endosperm starch granule formation (Chapter One; Table 1.2). *Aegilops peregrina* is unique as it has almost no B-type granules (Stoddard, 1999a) because the second wave of granule initiation that produces the B-type granules fails to occur (Howard *et al.*, 2011). Hybridisation of *Ae. peregrina* with two synthetic tetraploid lines and subsequent phenotyping of the F_3 generation revealed that the B-less trait was segregating. This population was used in a QTL mapping experiment and the causative gene was identified as *BGC1* (Howard *et al.*, 2011; Chia *et al.*, 2020). This demonstrates the power of *Aegilops* species in association mapping techniques.

Therefore, there is good rationale for utilising the variation I observed in *Ae. tauschii* endosperm starch (Chapter Three) in an association mapping study. The *Ae. tauschii* k-mer GWAS pipeline by Gaurav *et al.* (2021) is an ideal tool, given the benefits over SNP based GWAS, and that it has been designed to use the same *Ae. tauschii* diversity panel I characterised in Chapter Three. In this chapter, I used GWAS to identify five loci associated with B-type granule diameter, and a further five loci associated with B-type granule number. Curiously these regions do not overlap, suggesting that they are under distinct genetic control. I identified the starch debranching enzyme LIMIT DEXTRINASE (*LDA*) as a major candidate gene within a region associated with B-type granule diameter, which will be explored further in Chapter Five. I also produced a list of 13 candidate genes by examining expression patterns of putative orthologs in wheat and predicted localisation, which could be investigated in future work. Overall, this chapter genetically dissects the control of B-

type granule formation in *Ae. tauschii* and identifies candidate genes which should be investigated in the future.

4.2 Results

4.2.1 GWAS identifies loci associated with B-type granule size

To understand the genetic variation underpinning the diversity in the starch granule parameters, a k-mer based association mapping strategy was employed. Firstly, the variation in mean B-type granule diameter (Chapter Three; Figure 3.6b) was analysed with the association mapping pipeline (Gaurav *et al.*, 2021). This was conducted three times, and each time the mapping was conducted against different a genome – TOWWC193 (AL8/78 v5) (Figure 4.1a), TOWWC106 (Figure 4.1b) or TOWWC112 (Figure 4.1c). The TOWWC106 and TOWWC112 genomes were generated by Gaurav *et al.* (2021). The Manhattan plots showed a striking difference in the correlations when mapped to the different assemblies, most notably that the significant peaks contain k-mers which have a negative correlation with B-type granule diameter when mapped to TOWWC193 or TOWWC112. Yet the k-mers had a positive correlation with B-type granule diameter when mapped to TOWWC106. This is likely because TOWWC106 has one of the largest B-type granule diameters (10.87 µm), whereas TOWWC112 had a lower value (6.88 µm), as does TOWWC193 (7.49 µm). This suggests that the genetic variation identified here by k-mers is mainly associated with larger B-type granule diameters.

There were multiple regions in the Manhattan plots that contained significant k-mers. Therefore, to focus the analyses, I decided to only include regions which contain significant k-mers when mapped to all three assemblies. The limitation of this approach is that if a region is absent from one of the assemblies, it would prevent k-mer mapping and mean that there are no significant hits in this region. There were five peaks which meet this criterion. These were located on chromosomes three (two peaks, called A and B), five (one peak, called C) and seven (two peaks, called D and E). Peak E (chromosome seven) contained many k-mer hits. When mapped to TOWWC106 there were more than 120,000 k-mers in this peak, although most of these fell below the significance threshold (black dashed line) so did not have significant associations with the B-type granule diameter phenotype. Similarly, for peak B (chromosome three) more than 35,000 k-mers mapped to TOWWC106 at this region, but again mostly with non-significant associations. The peaks varied in size, peak A (chromosome three) covering 1.99 Mb, peak B (chromosome three) being

13.44 Mb, peak C (chromosome five) being 0.58 Mb and peaks D and E (chromosome seven) being 4.81 Mb and 0.01 Mb respectively (Appendix One; Table S2). As there are several peaks, some of which are quite large, it was impossible to analyse all of them in depth within the scope of this thesis. Therefore, I identified three peaks that were of greater priority. These were peaks B, C and D - as they contain the k-mers with the highest association scores when mapped to TOWWC193, and when mapped to TOWWC106 and TOWWC112 there were many associating k-mers. These are prioritised for further analysis later in this chapter.

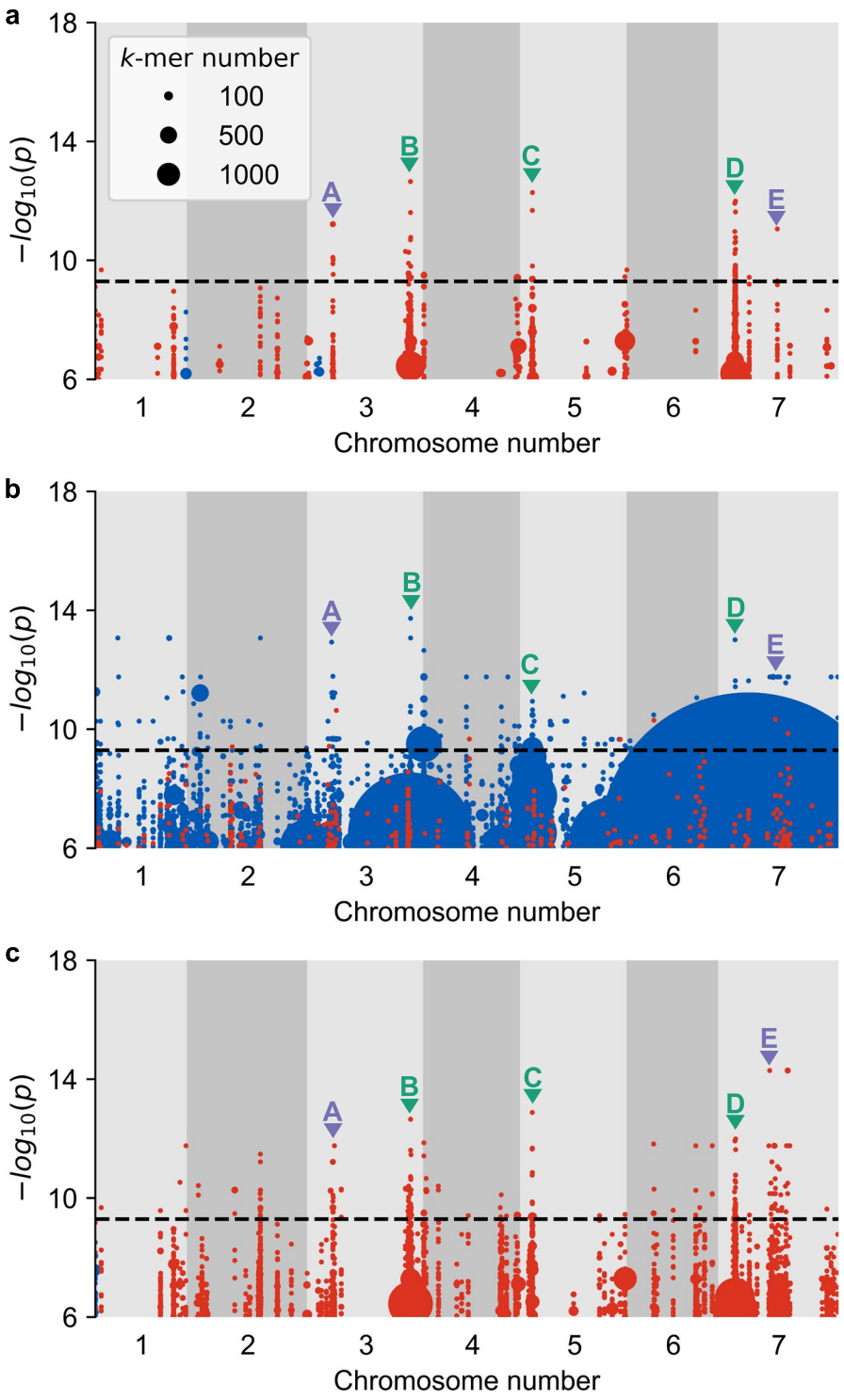


Figure 4.1 B-type granule diameter is associated with five genomic loci. Mean B-type granule diameter data were analysed with an association mapping pipeline (*n* = 117, where each value is the average of three biological replicates). K-mers associated with the phenotype were mapped to the TOWWC193/AL8/78 v5 (a), TOWWC106 (b) or TOWWC112 (c) *Aegilops tauschii* assemblies. Points above the black dashed line represent k-mers with significant association. Blue points represent positive correlations and red points represent negative correlations. The size of the points is

proportional to the number of k-mers associated and the key shown in (a) applies to all Manhattan plots here. Peaks identified as having a significant association with B-type granule diameter are identified with arrows and labelled with letters, turquoise arrows identify priority peaks and purple arrows identify non-priority peaks.

4.2.2 GWAS identifies loci associated with B-type granule number

Similar GWAS analyses as for B-type granule size above were conducted for B-type granule number (Figure 4.2 and Chapter Three; Figure 3.6d). Like with B-type granule diameter, there were also differences in the correlations of the k-mer hits between the different assemblies. When mapped to TOWWC193, most of the k-mers had a negative correlation, yet with TOWWC112 most had a positive correlation, whereas when mapped to TOWWC106 there was a mixture of positive and negative. Again, this aligns with the phenotype data, with TOWWC193 and TOWWC106 having low B-type granule numbers (89.16% and 90.93% respectively), and TOWWC112 has one of the highest B-type granule numbers (96.60%).

There were five peaks which were consistently significantly associated across all three Manhattan plots. These were all on chromosome three – for simplicity these are henceforth called V, W, X, Y and Z (Figure 4.2). The high number of k-mer hits in peak V, makes identification of peak W, and sometimes peak X, difficult. Therefore, the entirety of chromosome three been magnified (Figure 4.3). This magnification of the peaks on chromosome three also reveals peaks X and Y cover large regions (105 Mb and 23.3 Mb respectively) (Appendix One; Table S2).

I checked whether the peaks associated with B-type granule number were consistent with different binary thresholds. This is important as where to place the threshold for determining B-type granules is arguable. To test this, the threshold was set to different diameter values, 8 µm, 9 µm, 11 µm and 12 µm, and B-type granule number was recalculated (Figure 4.4a). There seemed to be similar trends regardless of the threshold used, with the same accessions having high or low B-type granule number values and very few outliers. To confirm this, Spearman's rank correlation tests were used, with the new data produced here correlated against the data with the 10 µm threshold (Chapter Three). All correlations were extremely strong with $r \geq 0.98$ ($r = 0.98$ with a threshold of 8 µm, $r = 1.00$ with a threshold of 9 µm and 11 µm and $r = 0.99$ with a threshold of 12 µm). Hence regardless of the threshold, the relative B-type granule numbers between accessions is similar.

These data were analysed with the association mapping pipeline and mapped against TOWWC193/AL8/78 v5 (Figure 4.4b-e and Figure 4.5). The Manhattan plots all have different levels of background k-mer hits, although peaks V and W are present in all plots. Peak X does not appear when a threshold of 8 µm or 9 µm is used. Peaks Y and Z are present in all plots, most visibly when chromosome three is magnified in Figure 4.5. However, in the plot with a threshold of 8 µm (Figure 4.5a), peaks Y and Z fall below the significance line. From this analysis, only peaks V and W are consistently associated with B-type granule number regardless of threshold is used for defining B-type granules. These are priority peaks for further analysis, whilst peaks X, Y and Z are classed as non-priority peaks.

The precise locations and sizes of peaks V and W were compared for all five thresholds for B-type granule number (Table 4.2). For the peak V there were differences in where the peak started with a threshold of 8 µm and 9 µm, although the differences were small, only 20 kb. As this 20 kb region is missing from the peak in these two examples, but the rest of the peak is still prominent, it was concluded that this region is less important for influencing B-type granule number. Similarly for peak W there are also differences in where the peak started with a threshold of 8 µm, 9 µm and 12 µm. Interestingly, with thresholds of 8 µm and 12 µm, the peak started 250 kb upstream of the peak with a threshold of 10 µm. Whereas, with a threshold of 9 µm the peak started 490 kb downstream of the peak at 10 µm. For further analyses, I will only use the regions which are consistent for all thresholds (Table 4.2).

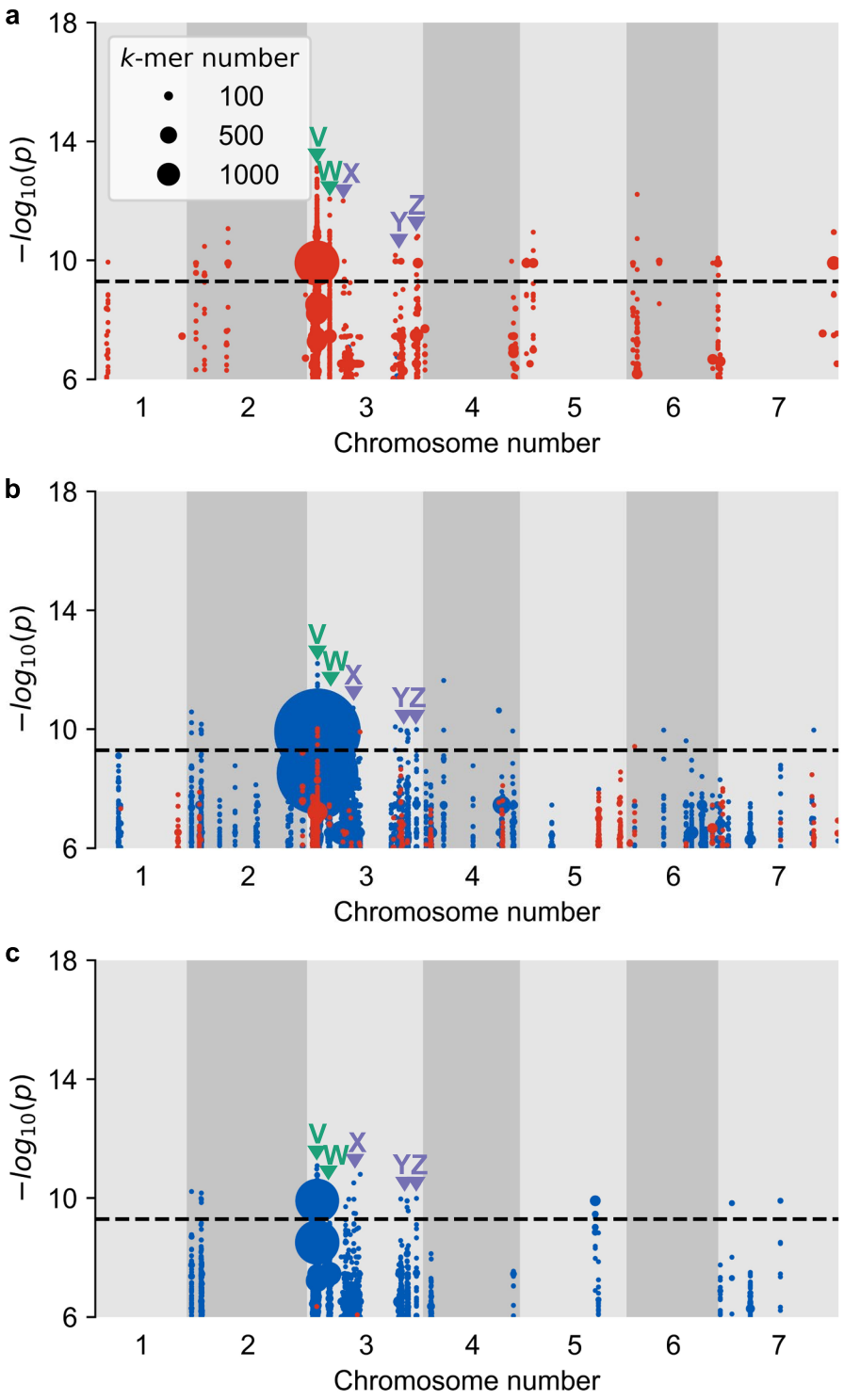


Figure 4.2 B-type granule number is associated with five genomic loci. Mean B-type granule number data were analysed with an association mapping pipeline (*n* = 117, where each value is the average of three biological replicates). K-mers associated with the phenotype were mapped to the TOWWC193/AL8/78 v5 (a), TOWWC106 (b) or TOWWC112 (c) *Aegilops tauschii* assemblies. Points above the black dashed line represent k-mers with significant association. Blue points represent positive correlations and red points represent negative correlations. The size of the points is

proportional to the number of k-mers associated and the key shown in (a) applies to all Manhattan plots here. Peaks identified as having a significant association with B-type granule number are identified with arrows and labelled with letters, turquoise arrows identify priority peaks and purple arrows identify non-priority peaks.

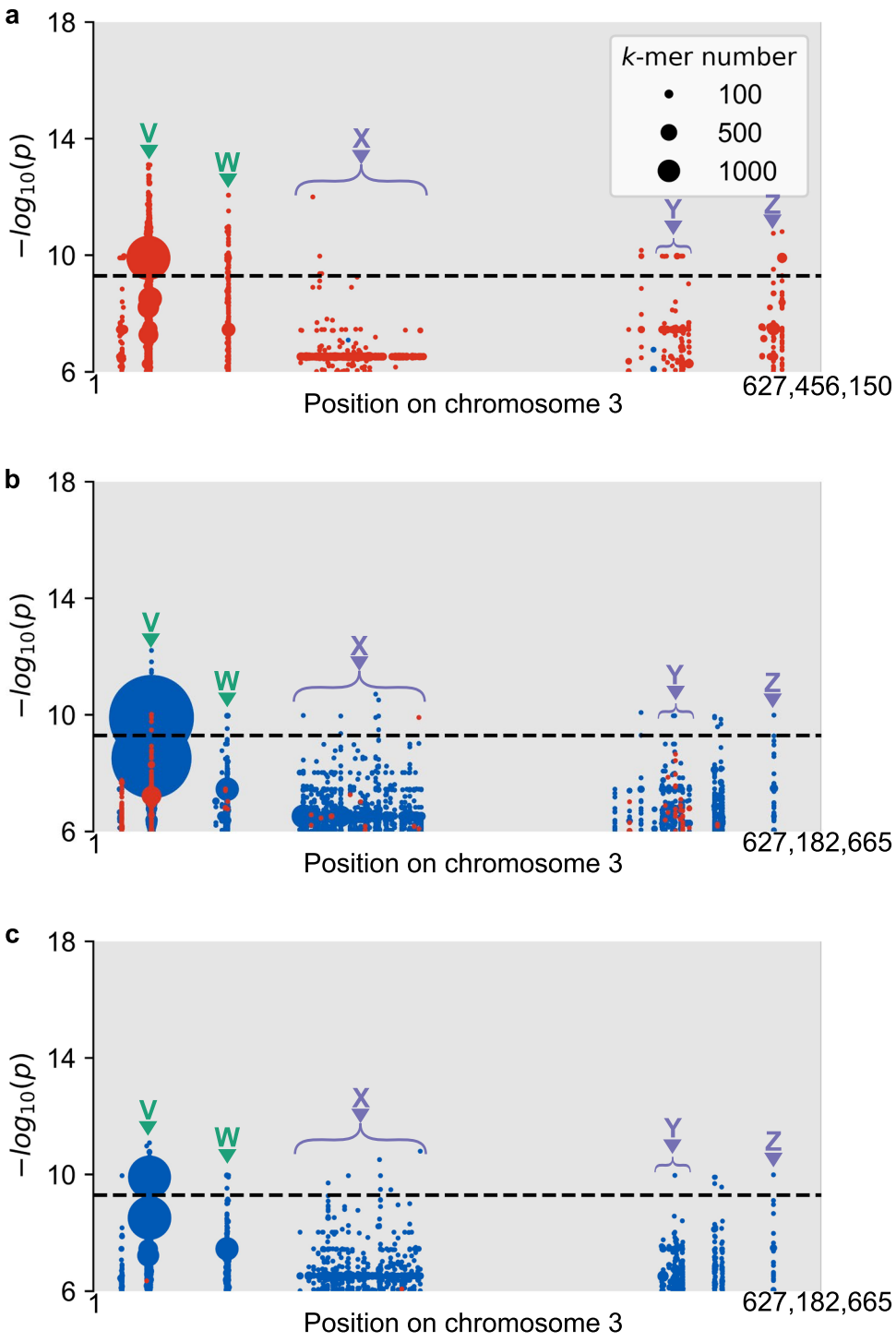


Figure 4.3 Two of the peaks on chromosome three are broad regions. Mean B-type granule number data were analysed with an association mapping pipeline (*n* = 117, where each value is the average of three biological replicates). The entire chromosome three from Figure 4.2 has been enlarged to show the distinct peaks, with the start and end positions of the chromosome given on the *x* axis. Peaks labelled with curly brackets arrows are broad regions.

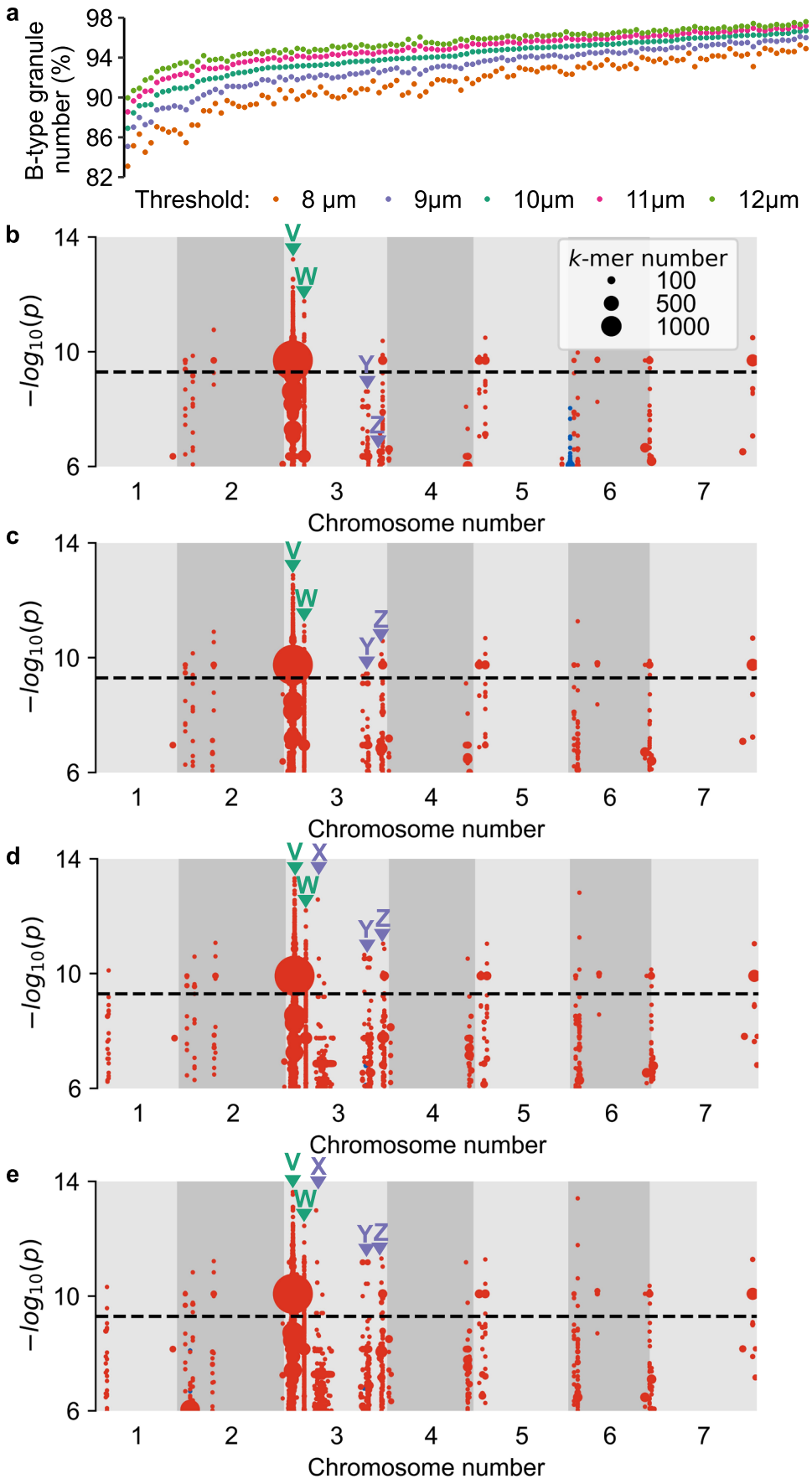

a
B-type granule number (%)
98
94
90
86
82
Threshold:
8 μm
9μm
10μm
11μm
12μm
b
c
d
e
−log10(p)
14
10
6
1
2
3
4
5
6
7
Chromosome number
k-mer number
100
500
1000
V
W
X
Y
Z
YZ

Figure 4.4 The V and W peaks on chromosome three are consistent regardless of the threshold used to calculate B-type granule number. The mean B-type granule number data were recalculated (a) with thresholds of 8 µm (orange), 9 µm (purple), 11 µm (pink) and 12 µm (green). Data are shown as means (points) with accessions arranged in ascending order based on the mean B-type granule number when calculated with a threshold of 10 µm (*n* = 117, where each value is the average of three biological replicates). These data were rerun through the association mapping pipeline, (b) 8 µm, (c) 9 µm, (d) 11 µm, (e) 12 µm. K-mers significantly associated with B-type granule number were mapped to the *Aegilops tauschii* reference genome AL8/78 v5. Points above the black dashed line represent k-mers with significant association. Blue points represent positive correlations and red points represent negative correlations. The size of the points is proportional to the number of k-mers associated and the key shown in (b) applies to all Manhattan plots here. Peaks identified as having a significant association with B-type granule number are identified with arrows and labelled with letters, turquoise arrows identify priority peaks and purple arrows identify non-priority peaks.

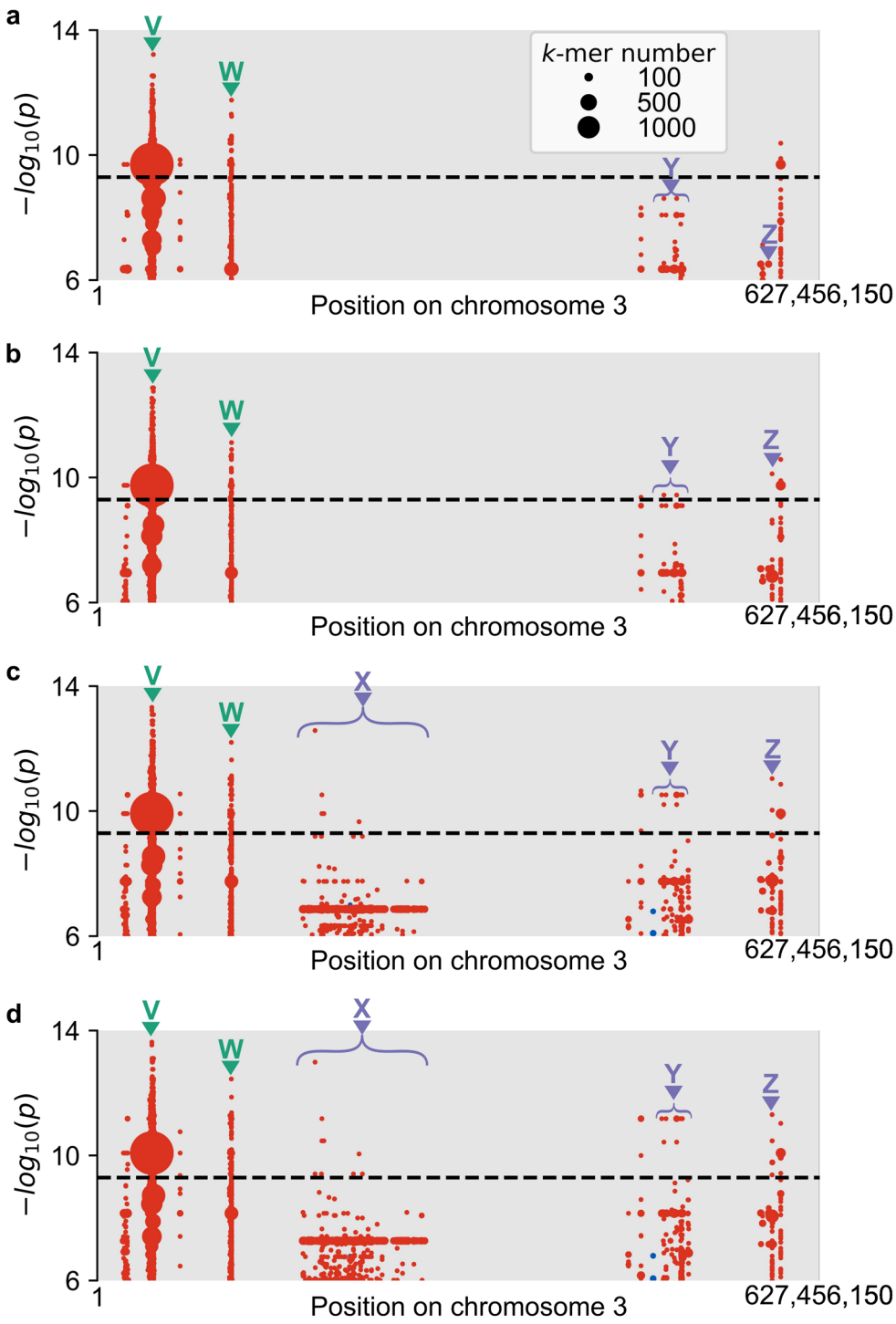


Figure 4.5 The V and W peaks on chromosome three are consistent regardless of the threshold used to calculate B-type granule number. The mean B-type granule number data were recalculated 8 µm (a), 9 µm (b), 11 µm (c) and 12 µm (d) (*n* = 117, where each value is the average of three biological replicates) and these data were rerun through the association mapping pipeline. The entire chromosome three from Figure 4.4b-e has been enlarged to show the distinct peaks, with the start and end positions of the chromosome given on the *x* axis. Peaks labelled with curly brackets arrows are broad regions.

Table 4.2 Locations of the peaks V and W on chromosome three when B-type granule number is calculated with different thresholds. The start and end of peaks were determined by finding the regions with the first and last k-mer hits, and the size of the peaks were calculated from these values.

Threshold (µm)	Peak V			Peak W		
	Start of peak	End of peak	Peak size (Mb)	Start of peak	End of peak	Peak size (Mb)
8	46,140,000	49,070,000	2.93	114,830,000	116,770,000	1.94
9	46,140,000	49,070,000	2.93	115,570,000	116,770,000	1.20
10	46,120,000	49,070,000	2.95	115,080,000	116,770,000	1.69
11	46,120,000	49,070,000	2.95	115,080,000	116,770,000	1.69
12	46,120,000	49,070,000	2.95	114,830,000	116,770,000	1.94
Consistent across all thresholds	46,140,000	49,070,000	2.93	115,570,000	116,770,000	1.20

4.2.3 The peaks associated with B-type granule number and diameter are distinct

To test whether any loci were significantly associated with both B-type granule number and B-type granule diameter, the Manhattan plots for these parameters were overlaid (Figure 4.6, Appendix One; Table S2). There are some loci where k-mers are found to be associated with both phenotypes, although in most cases these are below the significance threshold. Even when found above the significance threshold, they were not consistent when mapping back to all three assemblies. As an example when mapped to TOWWC106 (Figure 4.6b) peak E (chromosome seven – B-type granule diameter) also has some k-mer hits with B-type granule number, although these are absent when mapped back to TOWWC193 (Figure 4.6a). The same can be said for the peak C (chromosome five – B-type granule diameter) which when mapped to TOWWC193 (Figure 4.6a,b) also has significant k-mer hits with B-type granule number; although these are absent when mapped to TOWWC106 or TOWWC112 (Figure 4.6b,c). From this analysis, I concluded that the peaks associated with B-type granule number and diameter are distinct. Interestingly, both phenotypes have a lot of associating k-mers towards the ends of chromosome three, with peak B close to but not overlapping peaks Y and Z. Hence, these regions may contain many genes that influence B-type granules.

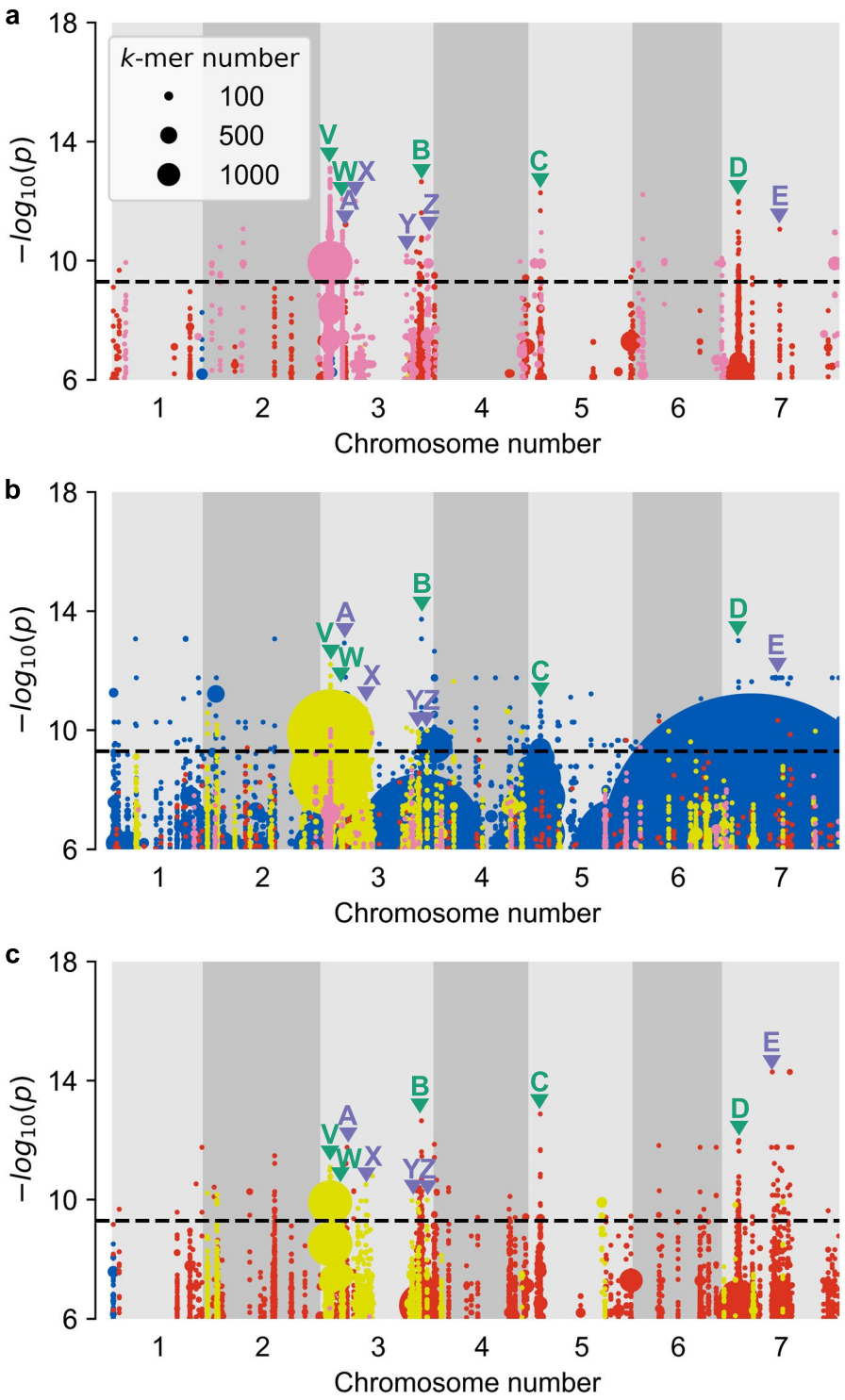


Figure 4.6 B-type granule number and diameter are under distinct genetic control. Mean B-type granule diameter and B-type granule number were independently analysed with an association mapping pipeline (*n* = 117, where each value is the average of three biological replicates). K-mers associated with the phenotype were mapped to the TOWWC193/AL8/78 v5 (a), TOWWC106 (b) or TOWWC112 (c) *Aegilops tauschii* assemblies and the Manhattan plots have been overlaid. Points above the black dashed line represent k-mers with significant

association. Blue and red points represent positive and negative associations with B-type granule diameter respectively. Yellow and pink points represent positive and negative associations with B-type granule number respectively. The size of the points is proportional to the number of k-mers associated and the key shown in (a) applies to all Manhattan plots here. Peaks identified as having a significant association with B-type granule diameter or B-type granule number are identified with arrows and labelled with letters, turquoise arrows identify priority peaks and purple arrows identify non-priority peaks. Peaks labelled A, B, C, D and E are associated with B-type granule diameter, whilst V, W, X, Y and Z are associated with B-type granule number.

4.2.4 No genomic loci were found to be associated with B-type granule content or A-type granule size

As B-type granule diameter and number had distinct peak patterns and both contribute to B-type granule content, I wanted to test whether any of these loci were also influencing B-type granule content. When the B-type granule content data (Chapter Three; Figure 3.6c) were analysed with the association mapping pipeline, although some k-mers were found to be associated with the phenotype, none of these reached the significance level used here (Figure 4.7). Hence it was concluded that there were no genomic loci significantly associated with B-type granule content. When run with A-type granule diameter (Chapter Three; Figure 3.6a), after filtering (Section 2.4.3), there were no k-mer hits with the phenotype and the Manhattan plot was empty. Unsurprisingly, the same was true when the mean data for total starch content and amylose content was analysed with the pipeline. This was expected given that there were no significant differences in these values across the diversity panel (Chapter Three; Figure 3.6e,f).

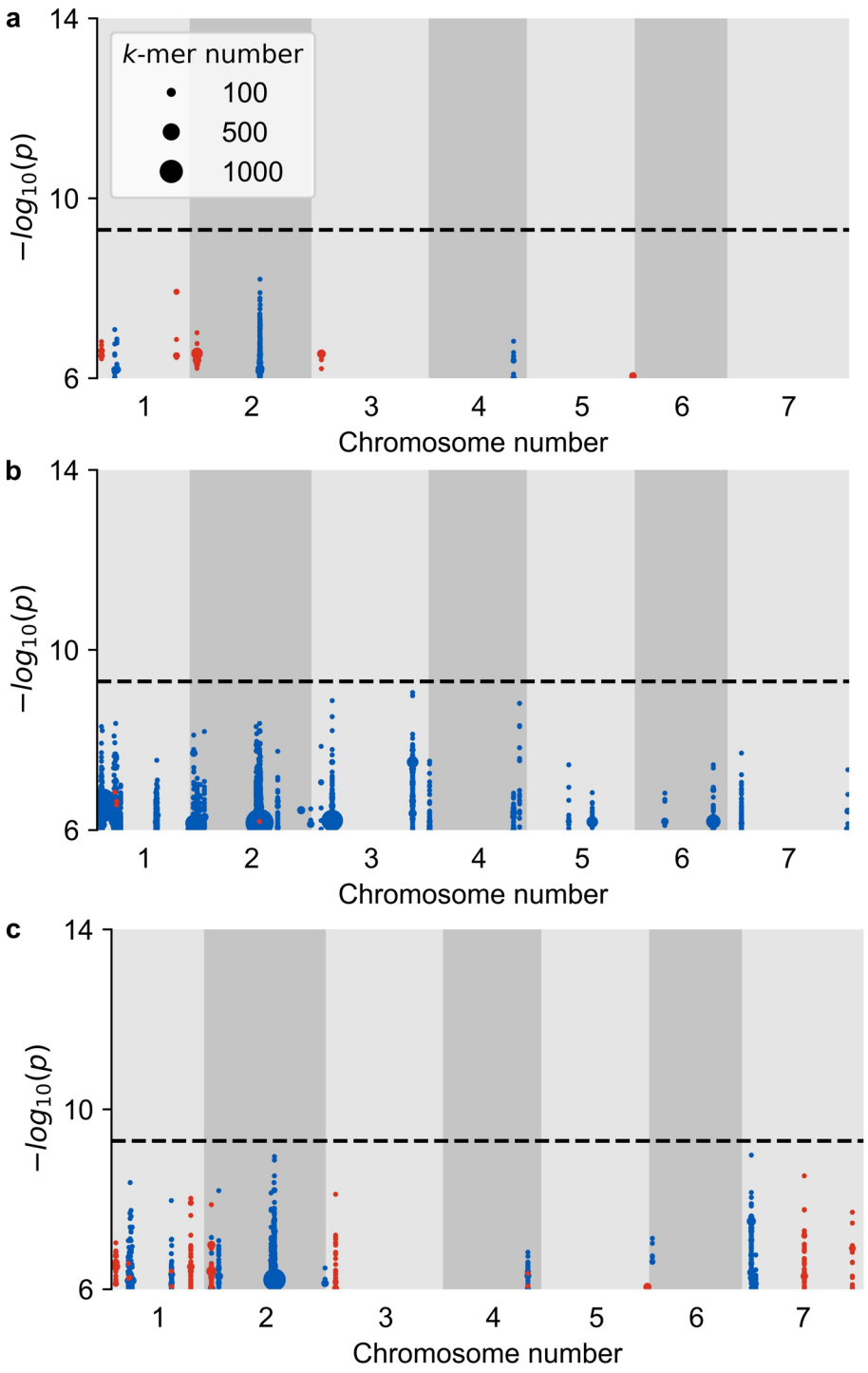


Figure 4.7 There were no significant peaks associated with B-type granule content. Mean B-type granule content data were analysed with an association mapping pipeline (*n* = 117, where each value is the average of three biological replicates). K-mers associated with the phenotype were mapped to the TOWWC193/AL8/78 v5 (a), TOWWC106 (b) or TOWWC112 (c) *Aegilops tauschii* assemblies. Points above the black dashed line represent k-mers with significant association. Blue points represent positive correlations and red points represent negative correlations. The size of the points is proportional to the number of k-mers associated and the key shown in (a) applies to all Manhattan plots here.

4.2.5 Limit Dextrinase is beneath a peak associated with B-type granule diameter

Next, I wanted to identify if any of the peaks contained genes already known to be associated with starch synthesis or granule initiation. A list of starch-related genes was obtained from Chen *et al.* (2023a), and the protein sequences encoded by each gene was used in BLASTp searches against the *Ae. tauschii* v4 genome to obtain the *Ae. tauschii* name and sequences in this version of the genome (Appendix One; Table S3). No putative orthologs could be identified for *SBE1.2* and *SBE1.3*, whilst for all other genes, putative orthologs were successfully identified. As the GWAS has been mapped back to v5 of the genome and there could be small differences in the positions of genes between the different assemblies, the sequence for each individual *Ae. tauschii* starch gene from v4 was blasted against the v5 genome. The locations of the starch genes in v5 of the genome was compared to the locations of the peaks identified in the Manhattan plots (Figure 4.8). Peak D, associated with B-type granule diameter, contains a known starch gene *LIMIT DEXTRINASE*, a starch debranching enzyme, which will henceforth be referred to as *LDA*. None of the other peaks contained a known starch gene, hence suggesting that they could contain novel genes involved in B-type granule formation.

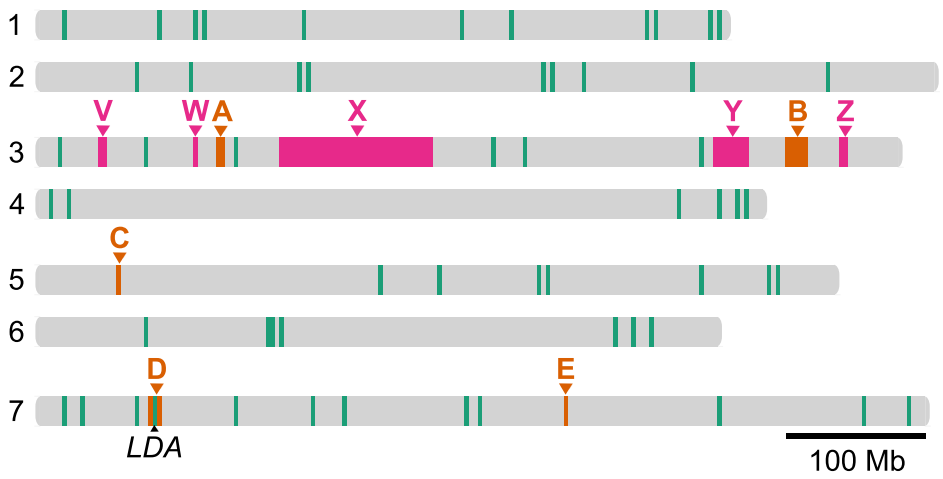


Figure 4.8 The B-type granule diameter peak on chromosome 7 overlaps with *LDA*. The locations of genes involved in starch granule initiation and starch biosynthesis (turquoise) were plotted on *Ae. tauschii* chromosomes (grey). The peaks (both priority and non-priority) significantly associated with B-type granule diameter (orange) and B-type granule number (pink) were also plotted and labelled. The scale bar represents 100 Mb.

4.2.6 Genes under the other peaks are varied

Genes underneath the priority peaks were identified using gene annotations of the AL8/78 v5 reference genome (Wang *et al.*, 2021a). There were some discrepancies between the gene annotations of the Wang *et al.* (2021a) v5 genome uploaded to NCBI and JBrowse, with some genes only being annotated in one upload. Hence genes which were annotated in both, and genes which were annotated in one plus v4 of the reference genome (Luo *et al.*, 2017), were classed as genes here. Gene models annotated as low confidence were not included for this analysis. In total, there were 221 genes under the priority peaks for B-type granule diameter and number. Specifically, for B-type granule diameter there were: 124 genes under the peak B, 4 genes under the peak C and 56 genes under the peak D, including *LDA* (Appendix One; Table S4). For B-type granule number there were: 25 genes under the peak V and 12 genes under peak W (Appendix One; Table S5). The gene descriptions were based on the descriptions in the aforementioned genome databases or based on domain annotations predicted using the InterPro tool. The genes beneath the peaks are quite varied, with no genes predicted to encode carbohydrate binding modules (CBMs) which are common in proteins involved in starch granule formation (Chapter One).

4.2.7 Prediction of plastid localised proteins

As endosperm starch granule formation occurs inside plastids, any proteins directly partaking in this process will be plastid localised. There are bioinformatic tools available to predict intracellular localisation based on the protein's amino acid sequence (Horton *et al.*, 2007). I employed WoLF pSORT to predict whether the protein encoded by the genes under the GWAS peaks were plastid localised in *Ae. tauschii* (Figure 4.9, Appendix One; Tables S6 and S7). Many proteins were predicted to be in the plastid (46 proteins, 21% of total proteins) or the nucleus (40 proteins, 18% of total proteins). Fourteen percent of the total proteins (31 proteins) were predicted to be cytosolic, with the rest of the cellular, and extracellular, locations having few proteins predicted to localised there (Figure 4.9). One protein, encoded by *AET3Gv20951100*, was predicted to be an extracellular protein, or in the plastid, with equal probability. For the remaining 66 *Ae. tauschii* genes (30%) I was unable to predict the localisation of the encoded protein. This was because the gene models were poor – there were often more than twenty predicted transcripts per gene and these transcripts sometimes lacked apparent start and stop codons.

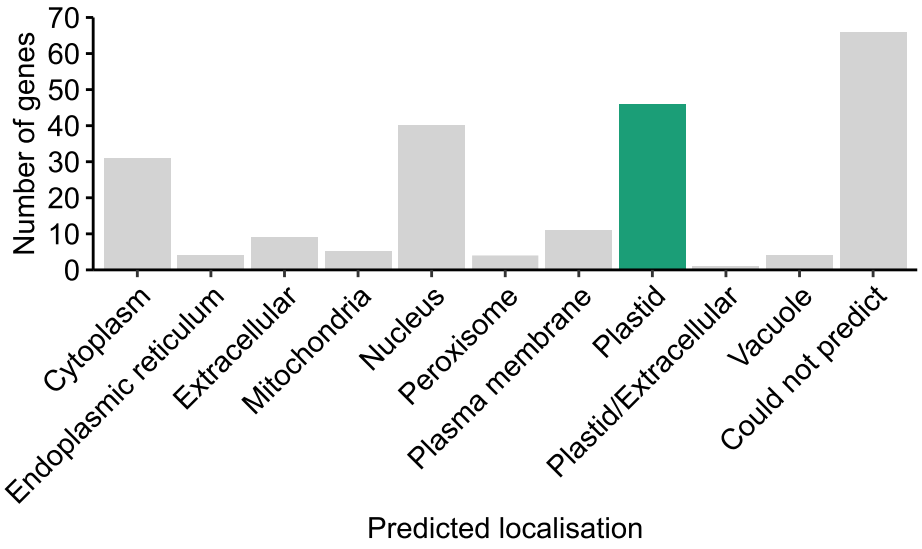


Figure 4.9 Predicted localisation of the 221 genes under the GWAS peaks in *Ae. tauschii*. WoLF PSORT was used to predict protein localisation of the protein encoded by the transcript which was labelled in Ensembl Plants as the canonical transcript (*n* = 221). For some genes it was not possible to predict the canonical transcript or protein localisation – these are shown as 'could not predict'. Bars are arranged in alphabetical order, with the exception of 'could not predict' which is on the far right. The plastidial bar is highlighted in turquoise, as these are most likely to be directly involved in starch granule formation.

4.2.8 Some genes have a similar expression pattern to starch granule initiation genes

I investigated expression patterns to identify which of the genes (Appendix One; Tables S4 and S5) are strong candidates for influencing B-type granule formation. Expression data is often used to identify putative candidate genes following a GWAS experiment, for example in Liu *et al.* (2023). This is especially valuable in this situation as it would be expected that the best candidates will be endosperm expressed. Moreover, there is already good knowledge of the timing of A- and B-type granule initiation and when genes involved in granule formation are expressed during grain development. There is no endosperm expression dataset for *Ae. tauschii*, so instead, endosperm expression datasets from *T. turgidum* and *T. aestivum* were used. This relied on identification of *T. aestivum* and *T. turgidum* putative orthologs for the genes identified in the GWAS, which was achieved using BLASTp searches (Appendix One; Tables S8-S12). Twenty percent of the genes under the GWAS peaks had no putative ortholog in the *T. aestivum* D genome, which is the most closely related to *Ae. tauschii*.

It could be that these genes have been lost in the hexaploid. Alternatively, presence/absence variation could mean that they were not present in the *Ae. tauschii* accessions which were important in the evolution of hexaploid wheat, and thus were not incorporated into wheat. Interestingly, some of the genes had more than one putative ortholog in wheat, suggesting there has been gene duplication. Genes without putative wheat orthologs had to be excluded from the expression analyses.

I used the wheat expression browser (Borrill *et al.*, 2016), to investigate expression of the putative *T. aestivum* orthologs, as this contains information for the D genome. Endosperm expressed genes were defined as those with a mean TPM (transcripts per million) of ≥ 1. For genes with an A genome ortholog, 67 (42%) were endosperm expressed, for genes with a B genome ortholog, 59 (38%) were endosperm expressed and for genes with a D genome ortholog, 64 (36%) were endosperm expressed (Appendix One; Tables S8-S12). Following this I investigated expression across developmental time – the highest quality endosperm expression time course dataset was produced using *T. turgidum* (Chen *et al.*, 2023a). Whilst this does not include information about the D genome, it has previously been used to investigate starch metabolism and thus is suitable for my analyses here. Firstly, I filtered the *T. turgidum* putative orthologs to identify those which were endosperm expressed (mean TPM ≥ 1 for at least one time point). There were 94 *Ae. tauschii* genes with endosperm expression of at least one *T. turgidum* putative ortholog (Figure 4.10). Many of the putative orthologs, for example the A and B orthologs of *AET3Gv20963400*, *AET3Gv20964800* and *AET3Gv20979000*, had constant low levels of expression throughout grain development. These are less likely to be involved in starch granule formation. Other putative orthologs had more dynamic expression pattern across development, for instance, the A and B orthologs of *AET3Gv20957000* had decreasing expression from 6-30 DPA, whilst the A and B putative orthologs of *AET5Gv20149000* tended to be more expressed as grain development progressed. Some orthologs, for example the A and B putative orthologs of *AET3Gv20980300* and *AET7Gv20336800*, increase in expression around 8-13 DPA, and then experience a fall in expression, to low levels, by 30 DPA.

To identify if any of the expression patterns portrayed in Figure 4.10 were like those of known starch granule initiation genes I used a clustering approach. I retrieved endosperm expression profiles of the key players in starch granule initiation: *BGC1* (*TRITD4Av1G198830*, *TRITD0Uv1G034540*), *PHS1* (*TRITD5Av1G205670*, *TRITD5Bv1G201740*), *MRC* (*TRITD6Av1G081580*), *MFP1.1* (*TRITD1Av1G054690*,

TRITD1Bv1G062760), *MFP1.2* (*TRITD3Av1G038460*, *TRITD3Bv1G047250*) and *SS4* (*TRITD1Av1G202430* and *TRITD1Bv1G193680*) from Chen *et al.* (2023a). These were analysed using the *Clust* package (Abu-Jamous and Kelly, 2018) alongside the 152 endosperm expressed putative *T. turgidum* orthologs. *Clust* was chosen as it has been successfully used on the entire endosperm expression dataset from Chen *et al.* (2023a). I used a minimum cluster size of two as I was interested in which putative orthologs had a similar expression pattern to the starch granule initiation genes, instead of how many. This program has an additional parameter called 'tightness' which controls how tight the clusters are, with larger values generally giving smaller tighter clusters. I therefore tested different tightness values and chose the values which led to clustering of the most starch genes (Table 4.3). There was no tightness value where all the starch granule initiation genes were placed in clusters. Therefore, I decided to use a tightness where the highest number of starch granule initiation genes clustered. Tightness values of both 0.2 and 0.4 placed six of the starch genes in clusters which included most genes known to be involved in B-type granule formation (Table 4.3). I decided to proceed with a tightness value of 0.4 as this produced a greater number of clusters, allowing me to investigate more diverse patterns of expression across grain development. Using this approach, 84 of the 163 inputted genes (52%) could be grouped into five clusters with distinct expression patterns (Figure 4.11, Table S13). Nearly half the genes did not cluster, this is similar to what was achieved in Chen *et al.* (2023a), where 53% of their differentially expressed genes did not cluster. Clusters one, three and four contained known starch granule initiation genes. Surprisingly, the A and B orthologs of the genes did not always cluster together. This is most notable for *MFP1.1* where the A ortholog is in cluster three, yet the B ortholog is in cluster one. Likewise, *PHS1-A*, was detected in cluster four, yet the B ortholog did not fall into any cluster. This could be a result of slight differences in the expression of the two orthologs, or a consequence of the parameters used for the clustering here. Thus, these clustering results should be analysed cautiously. There were 32 genes which clustered with *MFP1.1-B* and *MRC-A* (cluster one), eight genes which clustered with *MFP1.1-A* (cluster three) and 27 genes which clustered with *BGC1-A, BGC1-B* and *PHS1-A* (cluster four). Interestingly, both the A and B genome putative orthologs of *LDA* fall within cluster four. Hence in *T. turgidum LDA* has a similar pattern of expression to *BGC1* and *PHS1*, which is a further line of evidence that it could be involved in B-type granule formation. None of the other genes which fall within clusters one, three, and four, have known links to starch. As they have a similar expression pattern to the starch granule initiation genes, it means that they should be high priority candidates for being involved in starch granule formation.

Combining these results with my predicted localisation analysis (Section 4.2.8), I composed a list of *Ae. tauschii* genes which were predicted to be plastid localised, and whose putative *T. turgidum* orthologs had a similar expression pattern to starch granule initiation genes (Table 4.4). It is noted that for most of these genes, only one of the putative *T. turgidum* orthologs falls within clusters one, three or four, which contained the starch granule initiation genes. These are still relevant because I did not always retrieve both the A and B copies of the starch granule initiation genes together in clusters either, thus the presence of only one putative ortholog cannot rule it out from participating in granule initiation and formation. This analysis has reduced the 221 genes identified in the GWAS to 13 which are most likely to be directly involved in starch granule formation. These should be prioritised for characterisation in future work.

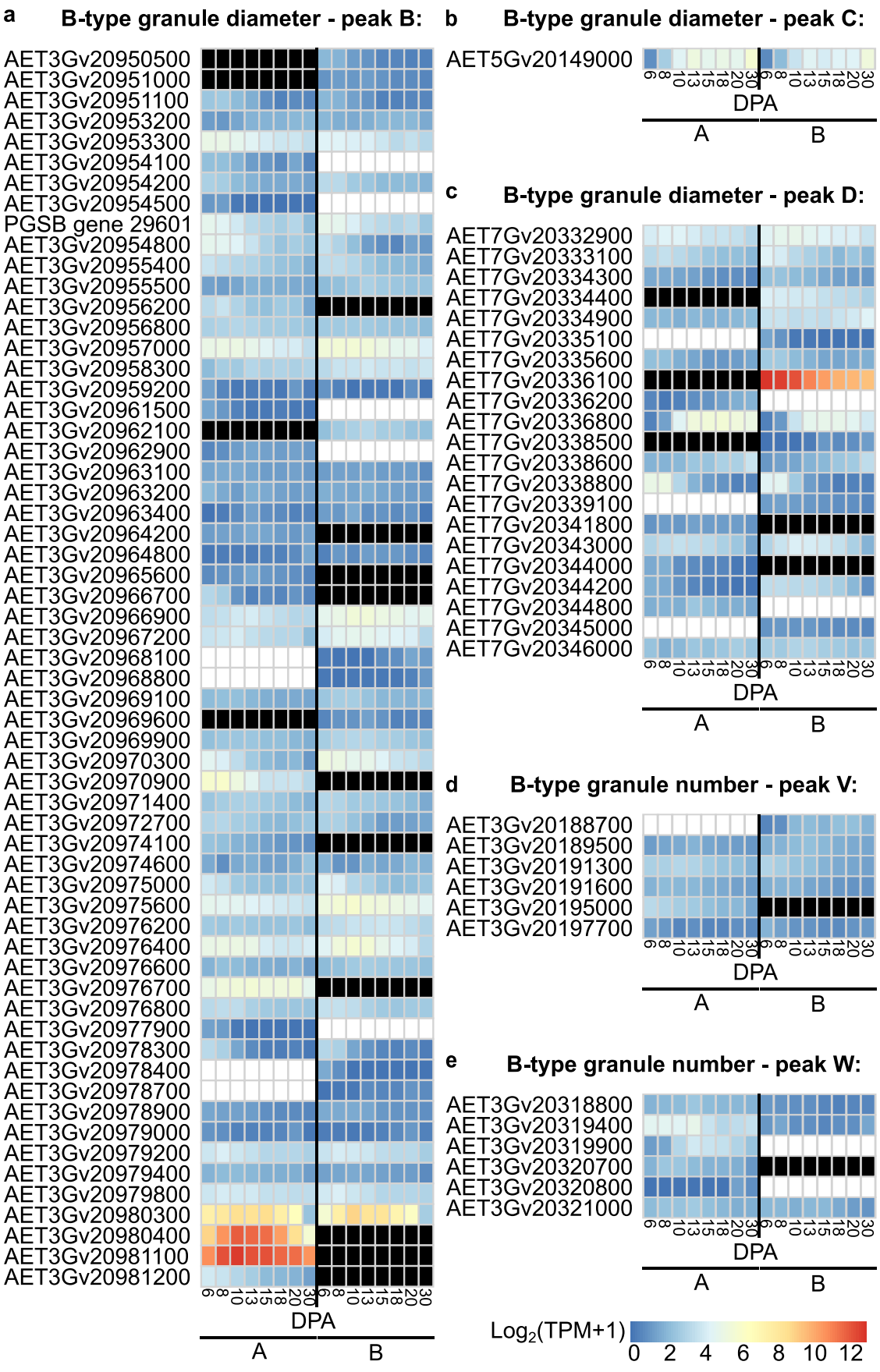

a
B-type granule diameter - peak B:
AET3Gv20950500
AET3Gv20951000
AET3Gv20951100
AET3Gv20953200
AET3Gv20953300
AET3Gv20954100
AET3Gv20954200
AET3Gv20954500
PGSB gene 29601
AET3Gv20954800
AET3Gv20955400
AET3Gv20955500
AET3Gv20956200
AET3Gv20956800
AET3Gv20957000
AET3Gv20958300
AET3Gv20959200
AET3Gv20961500
AET3Gv20962100
AET3Gv20962900
AET3Gv20963100
AET3Gv20963200
AET3Gv20963400
AET3Gv20964200
AET3Gv20964800
AET3Gv20965600
AET3Gv20966700
AET3Gv20966900
AET3Gv20967200
AET3Gv20968100
AET3Gv20968800
AET3Gv20969100
AET3Gv20969600
AET3Gv20969900
AET3Gv20970300
AET3Gv20970900
AET3Gv20971400
AET3Gv20972700
AET3Gv20974100
AET3Gv20974600
AET3Gv20975000
AET3Gv20975600
AET3Gv20976200
AET3Gv20976400
AET3Gv20976600
AET3Gv20976700
AET3Gv20976800
AET3Gv20977900
AET3Gv20978300
AET3Gv20978400
AET3Gv20978700
AET3Gv20978900
AET3Gv20979000
AET3Gv20979200
AET3Gv20979400
AET3Gv20979800
AET3Gv20980300
AET3Gv20980400
AET3Gv20981100
AET3Gv20981200
6 8 10 13 15 18 20 30 6 8 10 13 15 18 20 30
DPA
A B
b
B-type granule diameter - peak C:
AET5Gv20149000
6 8 10 13 15 18 20 30 6 8 10 13 15 18 20 30
DPA
A B
c
B-type granule diameter - peak D:
AET7Gv20332900
AET7Gv20333100
AET7Gv20334300
AET7Gv20334400
AET7Gv20334900
AET7Gv20335100
AET7Gv20335600
AET7Gv20336100
AET7Gv20336200
AET7Gv20336800
AET7Gv20338500
AET7Gv20338600
AET7Gv20338800
AET7Gv20339100
AET7Gv20341800
AET7Gv20343000
AET7Gv20344000
AET7Gv20344200
AET7Gv20344800
AET7Gv20345000
AET7Gv20346000
6 8 10 13 15 18 20 30 6 8 10 13 15 18 20 30
DPA
A B
d
B-type granule number - peak V:
AET3Gv20188700
AET3Gv20189500
AET3Gv20191300
AET3Gv20191600
AET3Gv20195000
AET3Gv20197700
6 8 10 13 15 18 20 30 6 8 10 13 15 18 20 30
DPA
A B
e
B-type granule number - peak W:
AET3Gv20318800
AET3Gv20319400
AET3Gv20319900
AET3Gv20320700
AET3Gv20320800
AET3Gv20321000
6 8 10 13 15 18 20 30 6 8 10 13 15 18 20 30
DPA
A B
Log2(TPM+1)
0 2 4 6 8 10 12

Figure 4.10 Endosperm expression pattern of putative *T. turgidum* orthologs of genes under GWAS peaks. Putative *T. turgidum* orthologs of genes under the priority peaks associated with B-type granule diameter (a-c) and B-type granule number (d-e) were identified via BLASTp searches. Their endosperm pattern of expression was retrieved from Chen *et al.* (2023a) and is represented as heatmaps. Genes shown are those with at least one putative ortholog which is endosperm expressed, mean TPM ≥ 1 for at least one time point. For simplicity, where there is more than one ortholog in either the A or B genome, the average expression for all putative orthologs is shown. Genes with no putative ortholog or no endosperm expression are represented with black and white respectively. Average $\log_2$(TPM +1), as some values were zero, are used to represent expression level, with dark blue representing no or very low expression, and red representing the highest expression. TPM = transcripts per million, DPA = Days Post Anthesis, A = A genome putative ortholog, B = B genome putative ortholog.

Table 4.3 Effect of varying the tightness value in the *Clust* program on cluster number and the number of starch genes within these clusters. The *Clust* program was ran with a minimum cluster size of 2 and varying tightness values, tightness values ran are shown in this table in numerical order. The number of clusters and the starch granule initiation genes within the clusters are listed, and there were 11 starch granule initiation genes included in this analysis in total. Starch genes that are not present in the clusters did not cluster.

Tightness value	Number of clusters	Number and names of starch granule initiation genes in clusters
0	1	2 *(MRC, MFP1.1-B)*
0.2	2	6 (*BGC1-A, BGC1-B, PHS1-A, MRC, MFP1.1-B, MFP1.2-A)*
0.4	5	6 (*BGC1-A, BGC1-B, PHS1-A, MRC, MFP1.1-A, MFP1.1-B)*
0.6	1	1 *(MFP1.1-B)*
0.8	1	1 *(MFP1.1-B)*
1.0	1	1 *(MFP1.1-B)*
2.0	3	3 *(BGC1-A, BGC1-B, MFP1.1-B)*
4.0	3	3 *(BGC1-A, BGC1-B, MFP1.1-B)*
6.0	3	3 *(BGC1-A, BGC1-B, MFP1.1-B)*
8.0	2	2 *(BGC1-A, PHS1-A)*
10.0	2	2 *(BGC1-A, PHS1-A)*

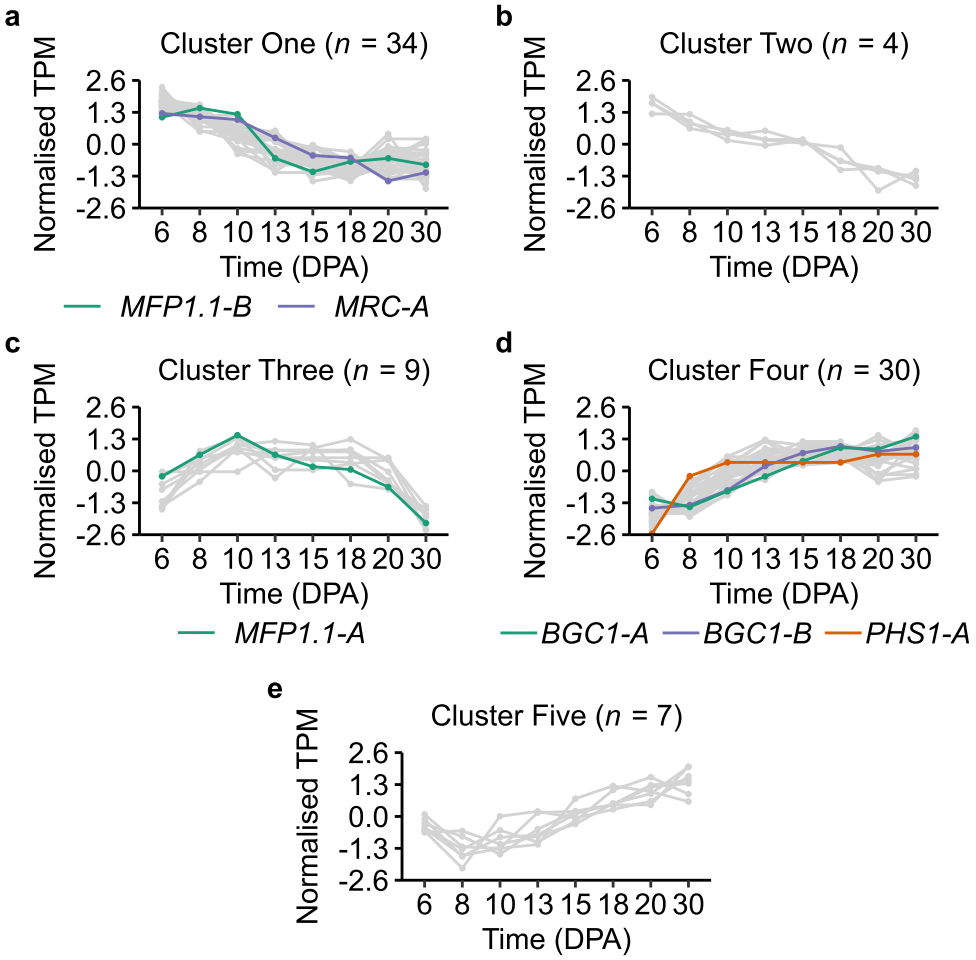


Figure 4.11 Some of the putative *T. turgidum* orthologs have similar expression patterns to genes involved in starch granule initiation. The putative *T. turgidum* orthologs of the genes identified in the GWAS and the orthologs of *SS4, BGC1, PHS1, MRC and MFP1*, were clustered using the *Clust* package, producing five clusters (a-e). Genes in each cluster are plotted using the inbuilt normalisation values of *Clust*, and the number of genes (*n*) is shown in the title above each plot. The starch granule initiation genes are highlighted in turquoise, purple and orange, with genes identified in the GWAS shown in grey. TPM = transcripts per million, DPA = Days Post Anthesis.

Table 4.4 High priority gene candidates from the GWAS for future investigation. Genes whose proteins were predicted to be plastid encoded, and whose putative *T. turgidum* orthologs clustered with starch granule initiation genes were collated. The *Ae. tauschii* gene name and description of the predicted protein are listed. The clusters in which the putative *T. turgidum* orthologs were found are followed by whether this was the A or B genome ortholog (in brackets). Note that not all putative *T. turgidum* orthologs were grouped into a cluster or fell into a cluster which did not contain a starch granule initiation gene (italics). The table is separated into genes found under the different GWAS peaks; peaks excluded from this table had no genes which fit the criteria.

Gene	Predicted protein	Cluster(s) for putative *T. turgidum* ortholog(s)
Peak B - B-type granule diameter:		
AET3Gv20953300	Transcription and mRNA export factor ENY2	1 (A)
AET3Gv20954100	Amidohydrolase-related protein	1 (A)
AET3Gv20955400	Mannosyl-oligosaccharide glucosidase with GH63 domain	1 (A), *2 (B)*
AET3Gv20955500	Exostosin GT47 domain containing protein	4 (A, B)
AET3Gv20966700	Stress enhanced protein 1	1 (A)
AET3Gv20967200	50S ribosomal protein L27	3 (A)
AET3Gv20975600	E2 ubiquitin-conjugating enzyme	1 (B)
AET3Gv20979400	Aspartokinase	3 (B)
Peak D - B-type granule diameter:		
AET7Gv20335100	DUF761 domain containing protein	1 (B)
AET7Gv20336200	HNH endonuclease	4 (A)
AET7Gv20336800	LDA (Limit dextrinase)	4 (A, B)
AET7Gv20341800	FAR1 DNA binding domain protein	4 (A)
Peak V - B-type granule number:		
AET3Gv20188700	RAP domain-containing protein	4 (B)

4.3 Discussion

The release of the k-mer association mapping pipeline for the *Ae. tauschii* diversity panel characterised in Chapter Three (Gaurav *et al.*, 2021), made it timely to undertake a GWAS for starch granule parameters. To my knowledge, there has only been one GWAS conducted on starch granule properties in the Triticeae (Li *et al.*, 2017), and this study did not utilise undomesticated wild relatives, which tend to have greater genetic and phenotypic diversity. In this chapter, I have provided the first example of a GWAS for starch granule parameters using *Ae. tauschii*. I have successfully identified multiple loci associated with B-type granule size and number and high priority candidate genes which could be influencing these traits.

4.3.1 Identification of genomic loci associated with B-type granules

I identified five peaks associated with B-type granule diameter (Figure 4.1, peaks A-E) and a further five peaks associated with B-type granule number (Figure 4.2, peaks V-Z). These range in size from 0.01-105 Mb. Curiously, they tend to be larger than the GWAS peaks identified by Gaurav *et al.* (2021), which were mainly between 50-520 kb, when the same pipeline was used for phenotypes including stem rust resistance, trichome number and spikelet number. This could be due to the nature of the traits investigated and how prevalent the 'extreme' phenotypes are, both in terms of number and phylogenetic distribution. Other studies which have used the same association mapping pipeline obtained peaks that are more similar in size to the peaks I report here. For instance, a 15.4 Mb peak associated with deoxynivalenol resistance (Kirana *et al.*, 2023). This 2023 publication also gives confidence that my approach of excluding the broad peaks (X and Y) in my downstream analysis is acceptable, as Kirana *et al.* (2023) excluded a 130 Mb peak on chromosome 7 as it contained more than 3,000 genes associated with fusarium head blight resistance.

There are no published association genetics studies for B-type granule number, and the only previous study for B-type granule size was a QTL study on a *T. aestivum* RIL population, identifying four QTLs on chromosome 4D (Igrejas *et al.*, 2002). Interestingly, I did not identify any peaks on chromosome four which were consistent when mapped to AL8/78 /TOWWC193, TOWWC106 and TOWWC112 (Figure 4.1). Likewise, Igrejas *et al.* (2002) did not find loci which overlapped with the loci I have identified here. There were some traits where I saw no significantly associated peaks in my study, such as A-type granule size and B-type granule content (Figure 4.7) which have reported QTLs in the literature. Igrejas *et al.* (2002) identified one QTL each of the long arm of chromosome 1B and the short arm of chromosome 7A, and two QTLs

on the short arm of chromosome 4D associated with A-type granule diameter. Whilst Borém *et al.* (1999) found one QTL on chromosome 2 and one on chromosome 7 associated with mean maximum A-type granule diameter. For B-type granule content, as previously explained, Howard *et al.* (2011) identified a QTL on the short arm of chromosome 4S in an *Aegilops* mapping population, this QTL was subsequently found to contain *BGC1*. Li *et al.* (2017) conducted a GWAS with an advanced intercross line *T. aestivum* population from the Yellow and Huai Valley Facultative Wheat Region and identified 23 loci associated with B-type granule content across chromosomes 2A, 2B, 2D, 3A, 3D, 4A, 4B, 4D, 5A, 5D, 6A, 6B, 7A, 7B and 7D. It is widely accepted that QTLs identified in one study are often not detected in subsequent studies (Mohammadi *et al.*, 2020), and this can come down to many factors including the diversity of the population used, marker density, and the environmental conditions the plants were grown in, which can greatly affect starch granule formation (as discussed in Chapter Three; Section 3.3.3).

4.3.2 B-type granule number and diameter are under distinct genetic control

One of the most interesting findings of the GWAS was that the peaks significantly associated with B-type granule number and diameter did not overlap (Figure 4.6). This agrees with the correlation analysis conducted (Chapter Three; Figure 3.8e) where the correlation coefficient between B-type granule number and B-type granule diameter was -0.35. This was surprising, as it was expected that fewer B-type granules would cause less competition for glucan substrates and more space and substrate for each B-type granule to grow larger. Such as in the *phs1 T. turgidum* mutant where there are fewer, yet larger B-type granules (Kamble *et al.*, 2023). However, this trade-off between number and size would only be seen in situations where the total B-type granule content is the same. In my GWAS this does not seem to be the case as the significant peaks associated with B-type granule number and size do not overlap, and no peaks were found for B-type granule content (Figure 4.7). One possible explanation for this is that in the *Ae. tauschii* diversity panel the variation in B-type granule content is not a simple trade-off between granule number and size and there is additional variation. One contributing factor could be due to differences in grain filling period. Differences in flowering time of this panel have already been documented (Gaurav *et al.*, 2021). If there are also differences in the timing of senescence, then it is reasonable to assume that there will be differences in the grain filling periods between accessions. Differences in grain filling period are linked to differences in B-type granule content (Stoddard, 1999b), although no one has investigated the specific effects of grain filling on granule number or size. It would be interesting in the future to

measure grain filling period in these *Ae. tauschii* accessions and conduct a GWAS with these data to see if this produces any loci which overlap with the peaks identified here.

The distinct peaks for B-type granule number and size could suggest that mechanistic factors influencing these traits are different. For instance, the key determinant of granule size could be available space in the amyloplast or number of stromules, since larger amyloplasts facilitate the growth of larger granules (Esch *et al.*, 2023). Whereas the major determinant of B-type granule number might be the abundance of protein initiation complexes (Chapter One; Section 1.5), which is affected by gene expression and protein stability. Regardless of the mechanism, the finding that B-type granule size and number are under distinct genetic control could prove useful in wheat breeding, as it offers the potential to adjust these traits independently of each other. The industrial benefits of this could be substantial, for instance, B-type granule size, but not number, is important for coating paper (Shevkani *et al.*, 2017).

4.3.3 Challenges in finding causative genes

All the peaks identified in my GWAS analysis contain multiple genes, which makes identification of a causative gene challenging. Even those that only contain a small number of genes, such as peak C which had only four genes, there is no strong candidate; for instance, there are no genes with an annotated CBM. One approach to identify causative genes would be to retrieve the sequences of the k-mers which have a significant association and map these back to the genes. However, this is not possible with the pipeline released by Gaurav *et al.* (2021). Regardless, this approach would still be difficult, as the most highly associated k-mers may not necessarily be causing the phenotype. Instead, they might be linked to the causative allele within the region of linkage disequilibrium (LD). Likewise, I cannot rely purely on the number of k-mers which have a significant association, as indels will result in a larger number of mismatched k-mers (length of k-mer x number of nucleotides in insertion/deletion) then singular SNPs (length of k-mer x 1). Hence, even though peak V contains a significant number of highly associated k-mers, identifying the causative variation is difficult.

4.3.4 Identification of LDA as a strong candidate

Peak D, associated with B-type granule diameter, contained the gene encoding the starch debranching enzyme LDA (Figure 4.8). LDA is α-glucan debranching enzyme capable of cleaving amylopectin's α-1,6-linkages. Previous studies in cereals have revealed that LDA has roles in germination (Iwaki and Fuwa, 1981; Dinges *et al.*, 2003) and transitory starch metabolism (Dinges *et al.*, 2003) (Chapter Five; Section

5.1). There is a report of a single amino acid mutation in the barley LDA which results in a greater B-type granule content (Sparla *et al.*, 2014). However, this has not been explored any further, so how the mutation is affecting LDA's mechanistic role is unclear. Due to its established links to starch, and its similar expression pattern to known starch granule initiation genes (Figure 4.11), I decided that *LDA* was the strongest candidate identified in the GWAS and will investigate this in more detail in Chapter Five.

In total, I identified ten peaks in my GWAS, which were categorised into priority and non-priority. This approach of only analysing the most promising peaks is a common approach taken in GWAS studies, whereby all peaks are reported but only a few investigated in more detail (Kirana *et al.*, 2023). Except for peak D, none of the peaks contained any genes with known links to starch metabolism. The non-priority peaks are not explored any further, however from this analysis, I have demonstrated that I have not missed any known starch genes under these peaks. In the future, it would be valuable to retrieve the gene candidates from underneath these peaks to see if they contain any promising candidates, such as genes encoding proteins with CBMs.

4.3.5 Thirteen priority gene candidates identified using expression data and localisation identified

I chose to prioritise gene candidates under my priority peaks based on predicted localisation and similar expression pattern to known starch genes. This strategy reduced the 221 candidates under my five priority peaks to 13 candidates that are most likely to have direct roles in starch granule formation. My strategy was appropriate to my aims, which was to discover new components directly involved in starch synthesis. However, this approach does not come without its limitations. I could not predict localisation for 66 (30%) of the genes (Figure 4.9), hence all of these were excluded from the final shortlist (Table 4.4), despite some of them potentially being localised in the plastid. Even where I could predict localisation, the data must be analysed cautiously as these are merely predictions based off the amino acid sequence. The accuracy of WoLF PSORT is estimated to be 55-70% (Horton *et al.*, 2007; Kaundal and Raghava, 2009). Its precision is greater than other prediction tools, such as TargetP (31%) and Plant-Ploc (33%) (Chou and Shen, 2007, 2008; Kaundal and Raghava, 2009; Almagro Armenteros *et al.*, 2019), which is why I selected it for my analyses here. However, there is still a large percentage error in the protein localisation predictions, meaning that I could have both false positives and false negatives in my dataset. To confirm the localisation, I would need to conduct *in planta*

experiments, for example tagging the proteins with a fluorescent tag and transiently expressing them in *Nicotiana benthamiana* to assess localisation *in planta*. Despite being more accurate than the bioinformatic predictions, there are potentials for false positives and negatives, especially given that the proteins are being expressed in leaf tissue of *N. benthamiana* rather than the endosperm of *Ae. tauschii*. Therefore, the gold standard approach would be to check subcellular localisation in the endosperm by generating transgenic plants expressing the fluorescently tagged proteins under their native promoters, alongside an amyloplast reporter and see if the two localise in endosperm tissue. However, this would be a resource intensive strategy and unfeasible for 221 proteins, hence why the bioinformatic strategy used here is a reasonable compromise.

Another method of prioritisation involved investigating gene expression using the high-quality *T. turgidum* endosperm dataset from Chen *et al.* (2023a). Again, this approach is not without weaknesses given that *Ae. tauschii* is the D genome progenitor, yet *T. turgidum* only has the A and B genomes. Thus, there were many instances (50, 23%) where no putative *T. turgidum* orthologs could be identified, and it is possible that these genes are unique to the D genome. This was particularly pertinent for the start of peak B, as none of the first five genes within the peak had putative *T. turgidum* orthologs. These genes had to be excluded from the analysis, so could not be included in the high priority shortlist (Table 4.4). Of the remaining 171 genes, 94 (55%) had endosperm expression (Figure 4.10). To my knowledge, no study has accurately assessed the percentage of genes which are endosperm expressed in *T. turgidum*, however some estimate it may be approximately 32% in *T. aestivum* (Clarke *et al.*, 2000). If true, then I have many more genes which are endosperm expressed than would be expected, perhaps suggesting that endosperm expressed genes are enriched in these regions of the genome.

Peaks C and W contain no high priority gene candidates (Table 4.4). This could be due to the aforementioned limitations of the prioritisation methods. It is possible that the genes under these peaks are not directly involved in starch granule formation – for example, any gene involved in transcriptional regulation in the nucleus would be excluded from this analysis. Likewise, genes that are expressed primarily in leaves and influence resource allocation to the endosperm (e.g., sugar transporters important for phloem loading) would also be excluded. Hence not containing any high priority candidates does not necessarily mean that these peaks are false positives. This is a limitation of my analysis - it is biased towards genes which are directly involved in

granule formation, since I designed my analysis to find candidates that act directly in our mechanistic model of granule formation, which is one of the key objectives overriding this work.

4.3.6 Some of the high priority candidates have links to starch and sugar metabolism

Of the 13 high priority candidates in Table 4.4, *AET3Gv20955400* is particularly interesting as it is predicted to encode a mannosyl-oligosaccharide glucosidase with a glycoside hydrolase (GH) 63 domain. In plants, there are 29 different GH families, with many members in each individual family. All are associated with carbohydrate degradation, although their specific roles are diverse ranging from cell wall biogenesis and remodelling, to chitin cleavage during fungal pathogen attack, and hormone metabolism (Minic, 2008). Several enzymes involved in starch metabolism have GH domains, including GH13, GH14 and GH31 domains (Minic, 2008; Smith and Zeeman, 2020). To my knowledge, there is no protein involved in starch metabolism with a GH63 domain. In fact, there are very few characterised proteins containing a GH63 domain in plants, with the CAZY database containing just one example from each of: *Arabidopsis thaliana*, rice and the lycophyte *Selaginella moellendorffii* (Lombard *et al.*, 2013). Phylogenetic analysis reveals that *AET3Gv20955400* is the direct ortholog of the rice protein, called OsMOGS (MANNOSYL-OLIGOSACCHARIDE GLUCOSIDASE). OsMOGs is localised in the endoplasmic reticulum (ER) and is expressed throughout the plant, with the strongest expression in the root (Wang *et al.*, 2014). It catalyses the removal of a glucose residue from an ER-derived precursor, $Glc_3Man_9GlcNAc_2$ (Glc = glucose, Man = mannose, GlcNAc = N-Acetylglucosamine), to produce $Glc_2Man_9GlcNAc_2$, which is subsequently used to produce N-glycosylated glycoproteins. Rice *mogs* plants have impaired cellulose synthesis resulting in thinner cell walls in root epidermal cells and altered auxin distributions, which disrupts cell elongation resulting in short roots (Wang *et al.*, 2014). A similar role is found for the characterised *Arabidopsis thaliana* GH63-containing protein, called GCS1 (Generative Cell-Specific Protein 1). When knocked out, there is an accumulation of the ER-derived precursor and reduced production of N-glycosylated glycoproteins, leading to altered seed development and shrunken seeds (Boisson *et al.*, 2001). It would be interesting to investigate whether it is involved in the production of N-glycosylated proteins in *Ae. tauschii* or whether it has been co-opted to participate in starch granule formation.

There was only one high priority candidate beneath peak V - *AET3Gv20188700* which encodes a protein with a RAP (RNA-binding domain abundant in Apicomplexans)

domain. In apicomplexans, RAP domains are important for RNA binding and appear to have a restriction endonuclease-like fold (Hollin *et al.*, 2021). However, no RAP domain-containing proteins have been characterised in plants. It is interesting to speculate whether this RAP domain protein could have been co-opted for a role in starch synthesis. This is relevant as the putative *Arabidopsis thaliana* ortholog has punctate localisation in the plastid (Kleinknecht *et al.*, 2014) which is characteristic of starch initiation proteins (Seung *et al.*, 2018). Transient localisation experiments in *N. benthamiana* could be conducted to see if the puncta of the RAP domain protein co-localise to starch granule initiation puncta.

In this chapter, I have identified independent genomic loci associated with B-type granule number and B-type granule diameter. Utilisation of publicly available resources and datasets allowed me to prioritise which genes underneath these peaks which are most likely to be involved in starch granule formation. This can form the basis for future work to explore the role of these genes in Triticeae species. I have chosen the highest priority candidate, *LDA*, and will explore the role of this further in Chapter Five.

Chapter Five – Exploring the role of Limit Dextrinase in endosperm starch formation

5.1 Introduction

From the Genome Wide Association Study in the Chapter Four, the starch debranching enzyme Limit Dextrinase (LDA) emerged as a strong candidate to be involved in B-type granule formation and will be explored in more detail here. LDA is an α-glucan debranching enzyme which can hydrolyse the α-1,6-linkages of amylopectin, releasing linear malto-oligosaccharides (MOS). It is also often referred to as pullulanase, as it can degrade the α-1,6-linkages joining maltotriose units in the polymer pullulan produced by *Aureobasidium* fungi (Pfister and Zeeman, 2016). This ability is unique to LDA and cannot be carried out by other plant starch debranching enzymes.

5.1.1 Role of LDA in starch degradation

LDA is most commonly associated with starch degradation, but its exact physiological role is not well understood. There is evidence that it is involved in seed germination and degrading transitory starch in leaves. Most of the work of the role of LDA in germination has been conducted in barley due to the importance of starch degradation in malting. The production of gibberellic acid, one of the initial steps in germination, leads to high aleurone expression of *LDA* within 12 hours of the initiation of germination, although its activity does not dramatically increase until 4 days later (Burton *et al.*, 1999). This disconnect between expression and activity can be partially explained by the presence of an endogenous protein inhibitor, LDI (LIMIT DEXTRINASE INHIBITOR) (MacGregor *et al.*, 1994). LDI binds to LDA in a 1:1 ratio at the LDA catalytic site (Møller *et al.*, 2015a). Within 3-4 days of the start of germination the amount of LDI declines, allowing LDA activity to increase (Ross *et al.*, 2003). *LDA* expression is also high in germinating rice (Iwaki and Fuwa, 1981) and maize (Dinges *et al.*, 2003). Maize *lda* mutants germinate much slower than the wild-type (Dinges *et al.*, 2003). To my knowledge, there are no examples of a *lda* mutation completely abolishing germination. This could be because of the presence of other starch degradation enzymes, most notably α-amylase which is the major player in endosperm starch degradation (Zeeman *et al.*, 2010). However, neither α-amylase, nor the other degradation enzymes active during germination (β-amylase and α-glucosidase) can hydrolyse α-1,6-linkages. Therefore, in the absence of LDA, whether an as yet unknown debranching enzyme carries out this activity, or whether hydrolysis of α-1,4-linkages alone is sufficient for germination, still needs addressing.

Regarding transitory starch degradation, maize *lda* mutants fail to degrade all their starch during the dark period (Dinges *et al.*, 2003). In contrast, the *Arabidopsis thaliana lda* mutant has no starch degradation phenotype unless combined with other mutations (Wattebled *et al.*, 2005; Delatte *et al.*, 2006). When combined with *isa3* (i.e., in the *lda isa3* double mutant), plants have higher starch contents across a light-dark cycle due to impaired nighttime degradation (Delatte *et al.*, 2006). Furthermore, introducing an *amy3* mutation into *lda isa3* plants completely blocks degradation of starch in the dark (Streb *et al.*, 2012). It is interesting that a phenotype is seen in maize *lda* mutants yet, in *Arabidopsis thaliana* a phenotype is only observed when *lda* is combined with other mutations. This could suggest there is greater redundancy of LDA in *Arabidopsis thaliana*.

5.1.2 Role of LDA in starch synthesis

Whilst there is still some mystery surrounding LDA's precise role in starch degradation, its role in endosperm starch synthesis is even less clear. In barley it is expressed during grain development, from approximately 5 DPA and peaks during B-type granule formation at approximately 21 DPA (Sissons *et al.*, 1993). There is no full *lda* mutant reported in a Triticeae species. A barley *lda* TILLING mutant with a V270I mutation, which falls within the CBM48 domain (Chapter One; Section 5.1.3), has a higher B-type granule content (Sparla *et al.*, 2014). However it is not clear whether this is a gain or loss of function mutation. Intriguingly an RNAi strategy against *LDI* in barley leads to increased LDA activity and fewer B-type granules (Stahl *et al.*, 2004). This needs to be analysed cautiously since there is no evidence for LDI being localised in the amyloplast, and if it is not, how it would exert an inhibitory effect on LDA in the amyloplast is unclear (Nakamura *et al.*, 1996). Cao *et al.* (2015) compared *LDA* expression between a wheat and an *Aegilops crassa* species which has no B-type granules. They found differences in the expression pattern of *LDA*, in addition to other starch genes, and suggest that LDA could be contributing to differences in B-type granule size. A major step forward in enhancing our understanding of the role of LDA in B-type granule formation would be the generation of a *lda* mutant in a Triticeae species, so its role in the process can be directly assessed. In endosperm starch formation outside the Triticeae, maize and rice *lda* mutants have no differences in granule size or shape compared to wild-type plants (Dinges *et al.*, 2003; Fujita *et al.*, 2009), although defects are seen in MOS levels in maize *lda* plants.

As introduced in Chapter One (Figure 1.11), the current working model of granule initiation in the Triticeae is that A-type granules are initiated at around 6-8 DPA,

followed by B-type granule formation approximately 10 days later. Initiation of B-type granules is reliant on, amongst others, PHS1 which can elongate MOS primers (Kamble *et al.*, 2023). These MOS primers are proposed to be released from existing A-type granules, but the enzyme(s) that partake in this process have not been identified. I hypothesize that LDA, which can produce MOS through hydrolysing α-1,6-linkages, could contribute to this process. There are multiple lines of evidence to support this as *LDA* expression peaks in B-type granule development (Chen *et al.*, 2023a) and has known links to B-type granules, as explained in earlier in this section.

5.1.3 The structure of LDA is well characterised

The structure of the barley LDA has been elucidated (Vester-Christensen *et al.*, 2010a; Møller *et al.*, 2012) (Figure 5.1). It is a member of the GH13 family and has four distinct domains: a CBM21-like domain at its N-terminal, a CBM48 domain, a catalytic domain and a C-terminal domain (Møller *et al.*, 2012). CBMs are non-catalytic domains that usually bind directly to glucan substrates and can be separated into 99 conserved families based on their amino acid sequences (You *et al.*, 2024). In LDA, the CBM21-like domain is named based on structural homology to the starch binding CBM21 domain in glucoamylase from the filamentous fungus *Rhizopus oryzae.* However, the key starch binding residues are not conserved so it is unclear whether this domain can bind substrates. This is relevant because the structure of this domain bound to a substrate has not yet been obtained (Tung *et al.*, 2008; Møller *et al.*, 2012). Despite this, the CBM21-like domain is important for determining substrate specificity and enzyme function. Two residues on the exposed surface of the CBM21-like domain, Serine-14 (serine = Ser) and Histidine-108 (histidine = His), are important for overall catalytic activity of the barley LDA (Andersen *et al.*, 2020). Mutants for either of these residues have a 60-95% reduction in activity in *in vitro* assays, despite being 40 Å away from the active site (Andersen *et al.*, 2020). Ser-14 and His-108 are important in multiple species (Gilding *et al.*, 2013; Chen and Bao, 2016). Alignment of the CBM21-like domain from LDA across the plant kingdom reveals that the residues surrounding Serine-14 and Histidine-108 are more highly conserved (Andersen *et al.*, 2020). Thus, perhaps there are even more essential residues in the CBM21-like domain which have not yet been explored.

In other proteins, CBM48's are important for glucan and substrate binding. For instance, the CBM48 of PTST2 interacts with MOS (Seung *et al.*, 2017). However, there are no studies demonstrating substrate binding to the CBM48 of LDA (Møller *et al.*, 2016). Substrate binding was also not observed in the barley LDA structure when

co-crystallised with cyclodextrin substrates (Vester-Christensen *et al.*, 2010a). This does not rule out substrate binding as the interaction may be difficult to detect if it is transient or weak. Regardless, it has been proposed that this may not be a canonical CBM48, but instead be an intermediate between a CBM48 and CBM20 domain. This is because one of its conserved Tryptophan residues is shared with CBM20's (Janeček *et al.*, 2011). Within the catalytic domain, there are two key catalytic residues - an aspartic acid that acts as a nucleophile, and a glutamic acid residue that can act as a general acid/base. Together, these perform a double displacement reaction leading to the cleavage of α-1,6-linkages (Møller *et al.*, 2015b). The role of the C-terminal domain has not yet been characterised.

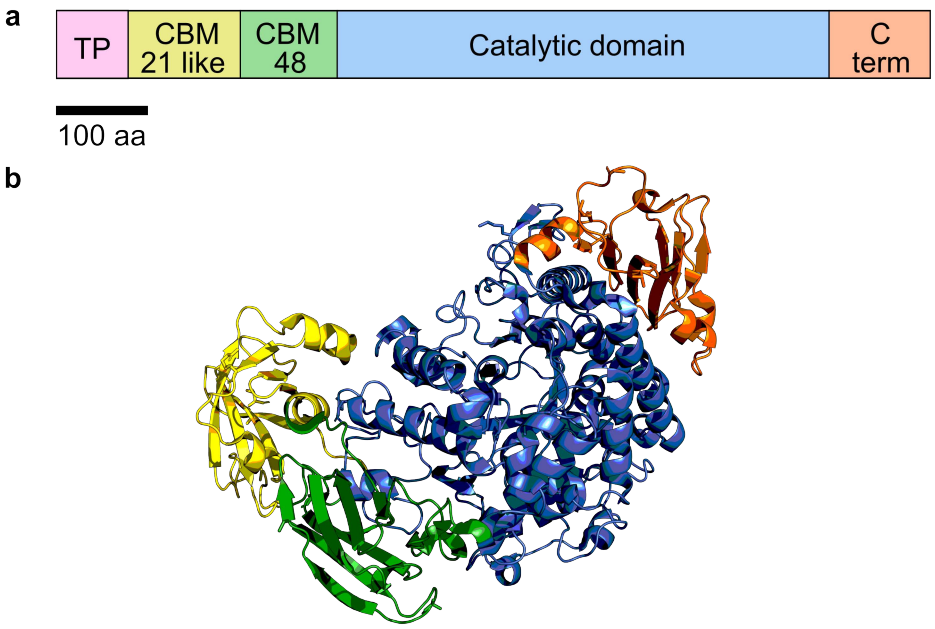


Figure 5.1 The structure of barley LDA. Diagrammatic representation of the LDA protein. Domains are shown as coloured boxes: TP = transit peptide, C term = C-terminal domain, scale bar = 100 amino acids and total length of the protein = 884 amino acids. (b) The barley LDA crystal structure produced by Møller *et al.* (2012) (pdb accession: 4aio). The domains have been colour coded and correspond to that in (a). Note that the transit peptide is absent in the structure.

Elucidating the function of LDA in Triticeae crops is of high priority and is the overall goal of this chapter. I have adopted genetic and biochemical approaches to understand LDA role in the Triticeae. To test the role of *LDA in planta*, I examined complete *lda* knockouts from the *T. turgidum* TILLING population. These had small reductions in B-type granule number. In parallel, I used cDNA sequencing to discover two novel LDA variants encoded by *Ae. tauschii* accessions with large B-type granule diameters. I hypothesized that these variants have different catalytic activities that could be influencing B-type granule formation. To test this, the different variants were expressed recombinantly and purified. *In vitro* activity assays revealed increased activity against a simple carbohydrate substrate. Combining the genetic and biochemical data suggest that it could be increased LDA activity, rather than a loss of activity, which is associated with large B-type granules in *Ae. tauschii*.

5.2 Results

5.2.1 LDA is active in the endosperm

In Chapter Four, I showed that *LDA* is highly expressed in the *T. turgidum* endosperm and has a similar expression pattern to *BGC1* and *PHS1* (Chapter Four; Figure 4.11 and Chen *et al.* (2023a)). To assess whether LDA activity follows a similar pattern I collected developing grains from the *Ae. tauschii* genome reference accession (TOWWC193) at 10, 15, 19 and 22 DPA (Figure 5.2a). The endosperm was dissected and protein extracted. LDA activity can be assayed by running extracted protein on a native gel containing red pullulan (Delatte *et al.*, 2006). LDA is the only enzyme capable of degrading the red pullulan, resulting in clearing of the red colour. Figure 5.2b shows that LDA activity is low in early grain development (10 DPA). It then rapidly increases and has much greater activity when B-type granules are initiated (from ~15 DPA), with the strongest activity at 22 DPA. To see whether this pattern was consistent in wheat, endosperm protein extracts from developing *T. turgidum* (8 to 28 DPA) were obtained from Nitin Uttam Kamble and analysed on red pullulan native gels (Figure 5.2c). Even though the precise time points are different, the activity follows a similar pattern to that of *Ae. tauschii*, increasing throughout grain development. The fact that LDA is highly active in the endosperm of both *Ae. tauschii* and *T. turgidum* during grain development provides strong rationale for investigating it further in this chapter.

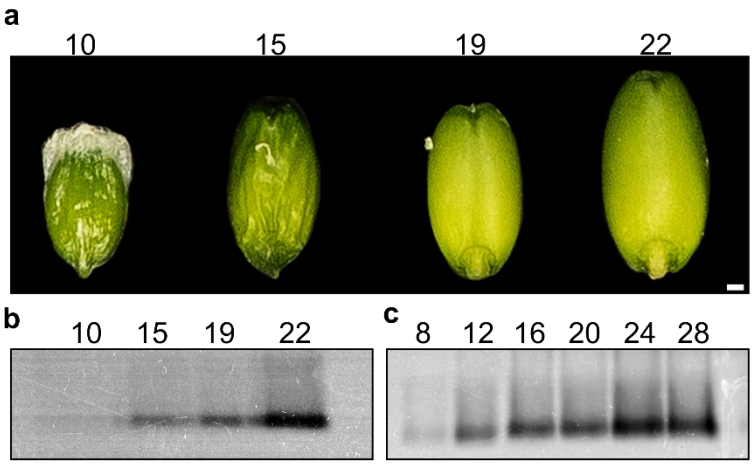


Figure 5.2 LDA is active in the endosperm. (a) Representative images of TOWWC193 *Ae. tauschii* grains across development, scale bar = 1 mm. Endosperm protein extracts (35 µg) from *Ae. tauschii* TOWWC193 (b) or *T. turgidum* (c) were analysed on a native gel containing red pullulan, bands represent LDA activity with darker bands indicating more activity. For (a-c) numbers represent days post anthesis (DPA).

5.2.2 Identification of T. turgidum plants with no LDA activity

As LDA is active in the endosperm (Figure 5.2) and has links to starch granule formation (Section 5.1.2) I wanted to explore the effect of completely eliminating LDA activity in a Triticeae species. I decided to pursue this in wheat due to the availability of mutants. To study LDA in wheat, I first identified gene models of *LDA* in *T. aestivum* using Ensembl plants. It was decided to use the *T. aestivum* genome as the gene models were poorly predicted for the *T. turgidum* genome. All three *T. aestivum* homeologs had 27 exons (Figure 5.3a). Noticeably the D genome homeolog has a large intron, ~5 kb, between exons 22 and 23. Overlaying these gene models onto the *T. turgidum* genome sequence revealed that the predicted amino acid sequences for LDA between *T. aestivum* and *T. turgidum* were 99.9% and 99.8% identical for *LDA-A1* and *LDA-B1*, respectively.

To test the effect of complete loss of LDA activity on endosperm starch granule formation, *T. turgidum cv* Kronos mutants were generated by Brendan Fahy. The *T. turgidum* TILLING resource was used to identify three mutants: one in *TaLDA-A1* - K2607 – and two in *TaLDA-B1* - K983 and K3912 (Figure 5.3a). K2607 and K983 cause premature stop codons at amino acids 127 of LDA-A1 and 591 of LDA-B1. K3912 has a mutation at the splice site following the 18th exon, likely resulting in mis-

splicing. Crucially, all three mutations are likely to be loss of function alleles. The *lda-1* line was generated by crossing K2607 with K3912, and the *lda-2* line was generated by crossing K2607 with K983. Genotyping was carried out by the JIC genotyping service with KASP markers (Chapter Two; Section 2.1.3). Plants that were homozygous for only the A genome mutation (*aa*BB), only the B genome mutation (AA*bb*) or both mutations (*aabb*) were isolated. As a control, wild-type (WT) segregants (AABB) were obtained from the cross; these do not contain the mutations in *LDA* but will possess some of the other background mutations. In the following analyses, I have also included a true Kronos wild-type (WT) that does not contain any EMS-induced mutations.

To assess whether the mutations cause complete loss of LDA activity, I extracted protein from leaves so as not to affect grain development and any grain phenotype. The extracted proteins were run on a native gel containing red pullulan (Figure 5.3b). For *lda-1*, there was a complete loss of activity in the double homozygous mutant (*aabb*) as no bands are observed at all. The bands in the single homozygous mutants (*aa*BB and AA*bb*) are slightly narrower compared to the WT segregant (AABB) suggesting reduced LDA activity; this would have to be confirmed in a quantitative assay. Unexpectedly, *lda-2 aa*BB has no detectable LDA activity, despite having the same A allele as *lda-1*. It could be that this plant is actually a double mutant (*aabb*) and the genotyping for the K983 mutation is inaccurate, despite forming distinct clusters in the KASP genotyping (Figure 5.3c). Primer re-design did not resolve this, producing identical results. Consequently, due to uncertainty in genotyping the wild-type B allele, I excluded *lda-2* AABB and *lda-2 aa*BB plants from further experiments. However, from Figure 5.3, the *lda-2 aabb* plants are full *LDA* knockouts. These plants have been included in subsequent work as they act as an additional *lda* mutant, which is important given that the TILLING mutants have background mutations. As the single homozygous mutants (*lda-1 aa*BB, *lda-1* AA*bb*, *lda-2* AA*bb*) had detectable LDA activity, I decided that they were of lower priority to analyse, and due to time constraints, I did not have time to analyse them here. Therefore, for the remainder of this thesis I only include Kronos WT, *lda-1* AABB, *lda-1 aabb* and *lda-2 aabb*.

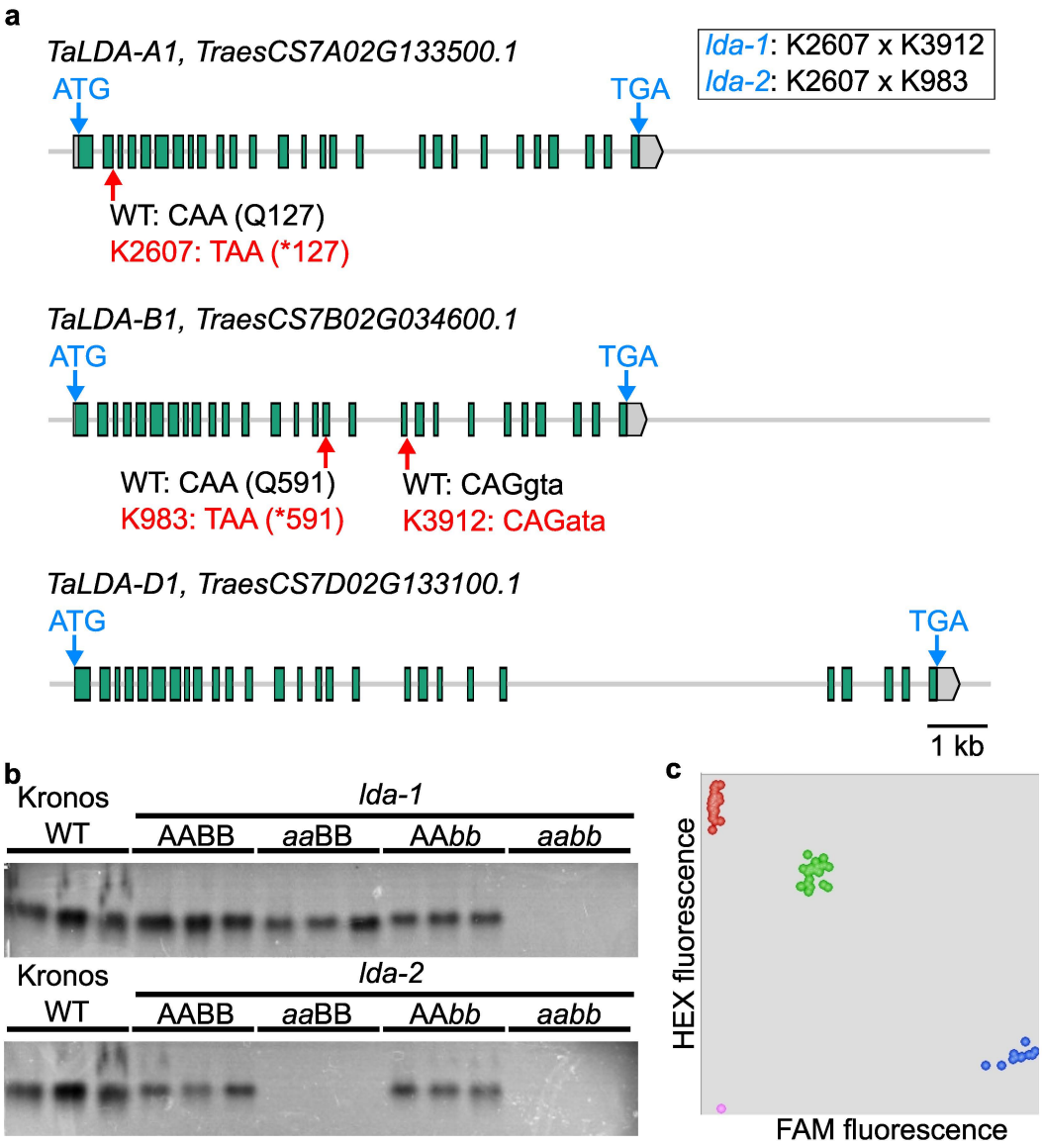


Figure 5.3 Identification of *T. turgidum* mutants with no LDA activity. (a) Diagrammatic representation of the *LDA* gene models in *T. aestivum* Chinese Spring. Exons are represented as green boxes and the 5' and 3' untranslated regions (UTRs) as grey boxes. The locations of the start (ATG) and stop codons (TGA) are indicated with blue arrows. The positions and resulting changes in the TILLING mutants are represented by red arrows, scale bar = 1 kb. (b) Leaf protein (100 µg) was run on a native gel containing red pullulan. Bands represent LDA activity with each individual band corresponding to protein extracted from a different plant (hence *n* = 3 for each genotype). (c) KASP genotyping for the K983 mutation. Red dots represent plants called as wild-type (BB) for the K983, green dots represent plants called as heterozygous (B*b*) for K983, blue dots represent plants called as homozygous (*bb*) for K983, pink dots represent cases where the KASP has failed.

5.2.3 T. turgidum lda plants have a small reduction in B-type granule number

I investigated whether the *lda* plants had any growth or grain phenotype. There was no detectable difference in growth and appearance between the mutants and the Kronos WT or *lda-1* AABB (Figure 5.4a). There was no significant difference in the height of mature plants (Figure 5.4b). For tiller number, there was no difference between the *lda-1 aabb* plants and the controls (Figure 5.4c). Whilst the *lda-2 aabb* plants had significantly more tillers than the Kronos WT; this could be a result of background mutations as I have no *lda-2* WT segregant to compare with. The morphology of the *lda* grains were indistinguishable from the controls (Figure 5.5a). There was also no difference in grain yield, thousand grain weight (TGW), or grain area, length or width (Figure 5.5b-g).

With assistance from Becca Testa, I extracted starch from mature grains and analysed granule morphology. There was no observable difference in granule shape and normal A-type and B-type morphologies were observed (Figure 5.6a). Size distribution traces from the Coulter counter were similar between the mutants and the controls (Figure 5.6b). The only noticeable difference was in the *lda-2 aabb* trace, where, the B-type peak had a lower maximum height, the A-type peak had a greater maximum height, and both peaks were shifted towards smaller diameters compared to the controls. To quantify if these differences were significant, log-normal - normal (L-N) bimodal distributions were fitted to the size distribution curves, as conducted in Chapter Three (Figure 5.6c-d). Mean A-type granule size (Figure 5.6c) was significantly smaller in *lda-2 aabb* than the Kronos WT (19.2 µm and 21.2 µm respectively). However, in the absence of a true WT segregant, it is not possible to conclude that this is a result of the *lda-2* mutations. The trend does not apply for *lda-1 aabb*, where no difference is observed when compared to the WT segregant (*lda-1* AABB). Thus, I concluded that knocking out *LDA* has no effect on A-type granule size. There was no significant difference in B-type granule diameter or content between *lda-1 aabb*, *lda-2 aabb* and the controls. Interestingly, there was a difference in B-type granule number, with *lda-1 aabb* and *lda-2 aabb* having significantly lower values compared to the controls (Figure 5.6f). It is worth noting that this difference is small ~0.5%, so it needs to be confirmed in a separate experiment with larger sample numbers. This difference in B-type granule number does not stem from differences in total starch content, as there were no differences in total starch content between the mutants and controls. There were also no significant differences in amylose content (Figure 5.6g-h).

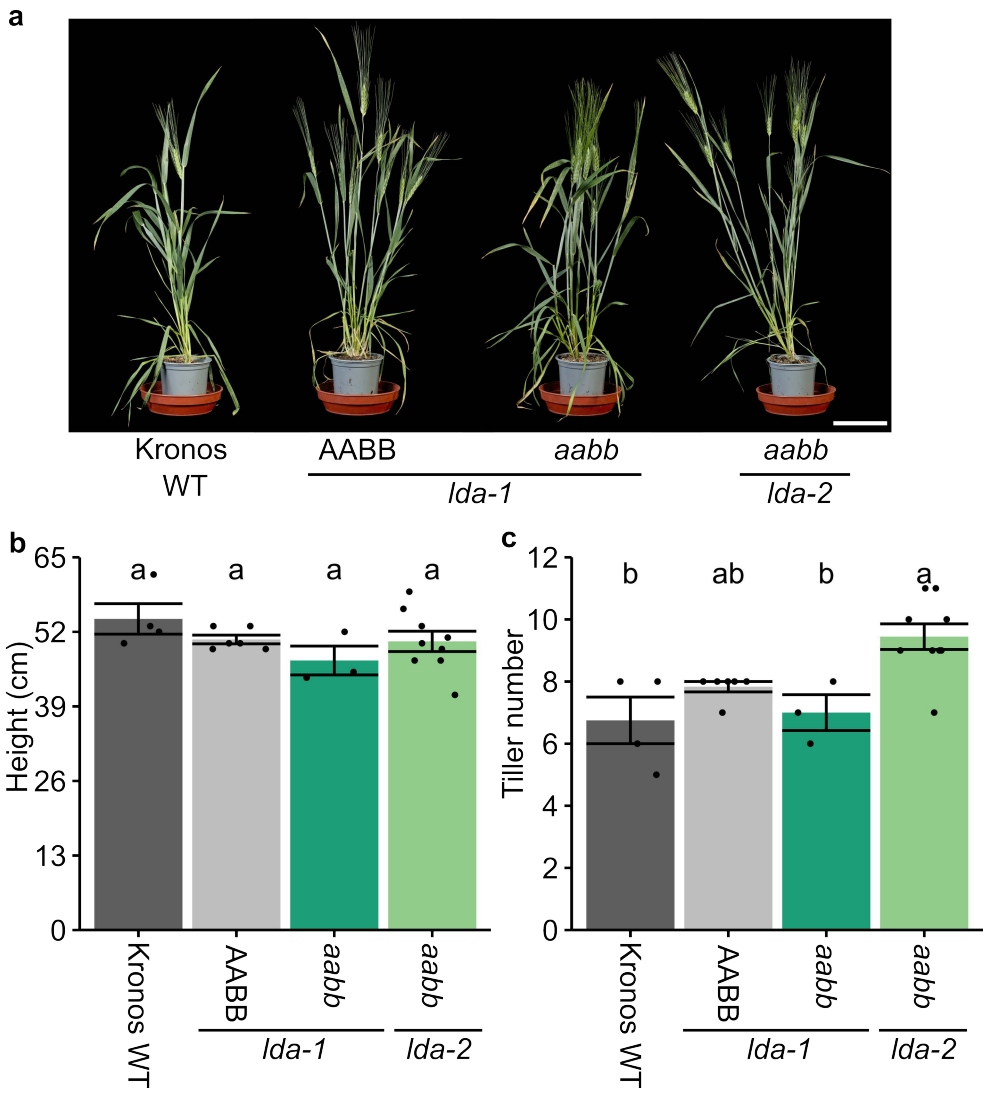


Figure 5.4 The *lda* plants have no defects in growth.(a) Photographs of 8-week-old *lda-1* (*aabb*) and *lda-2* (*aabb*) plants with the corresponding WT sibling control from the *lda-1* cross (*lda-1* AABB) and a Kronos WT, scale bar = 10 cm. (b) Plant height at maturity. (c) Tiller number at maturity. For (b-c) data are displayed as means ± SE with individual data points (black dots) representing the mean value from each individual plant. Values with different letters were significantly different under a one-way ANOVA with Tukey's post hoc test ($P < 0.05$). $n = 4$ for Kronos WT, $n = 6$ for *lda-1* AABB, $n = 3$ for *lda-1 aabb* and $n = 9$ for *lda-2 aabb*.

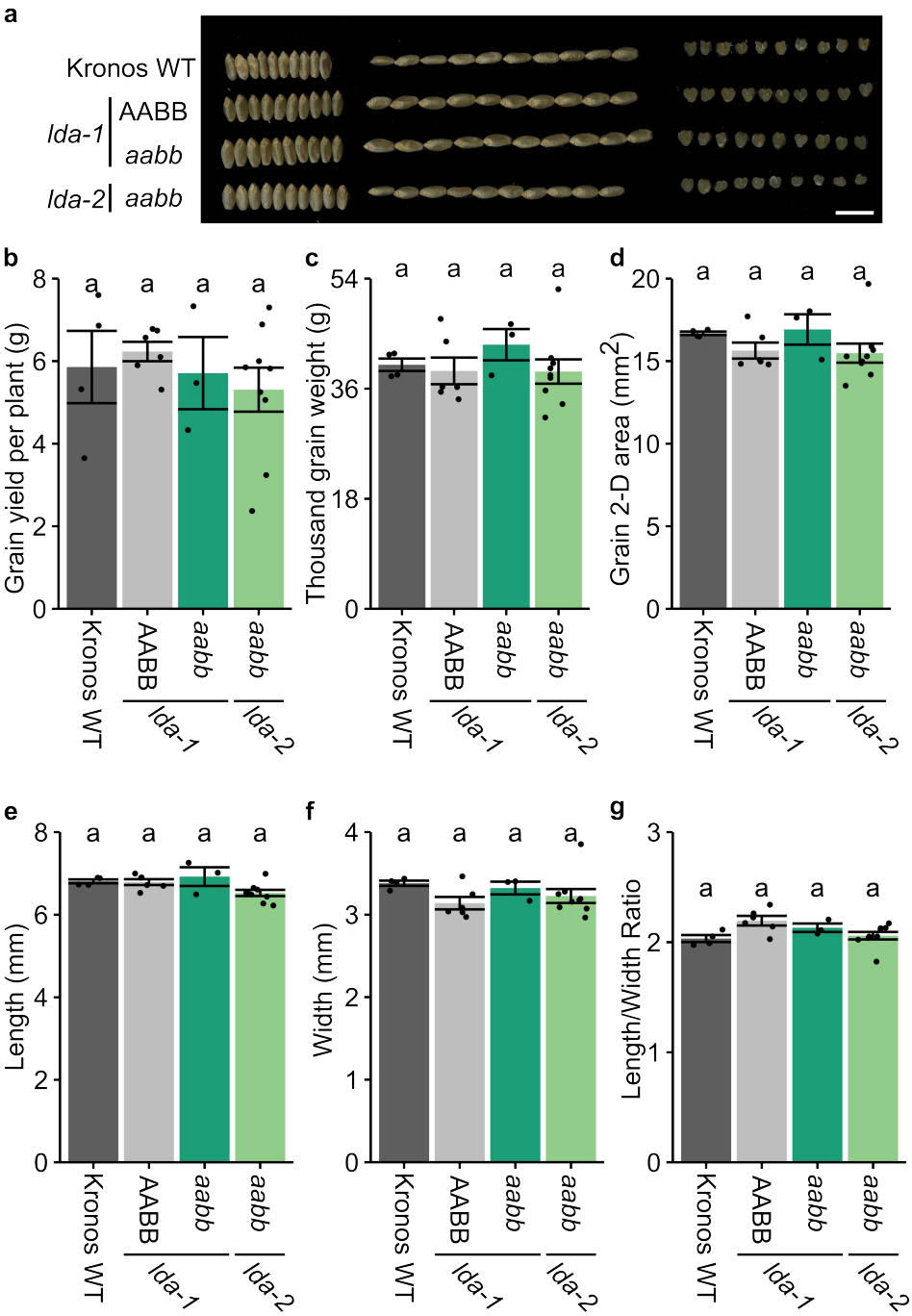


Figure 5.5 There is no effect on grain yield per plant or morphology in *lda* mutants. (a) Photographs of representative mature grains from each genotype. Grains are shown from the dorsal side and are arranged upright (left), or sideways (centre) with cross sections through the middle of the grain on the right, scale bar = 1 cm. (b) Grain yield per plant, (c) Thousand grain weight, (d) Grain 2-D area, (e) Grain length, (f) Grain width, (g) Grain length/width ratio. For (b-g), data are displayed as means ± SE with individual data points (black dots) representing the mean value from each individual plant. Values with different letters were significantly different under a one-way ANOVA with Tukey's post hoc test ($P < 0.05$). $n = 4$ for Kronos WT, $n = 6$ for *lda-1* AABB, $n = 3$ for *lda-1 aabb* and $n = 9$ for *lda-2 aabb*.

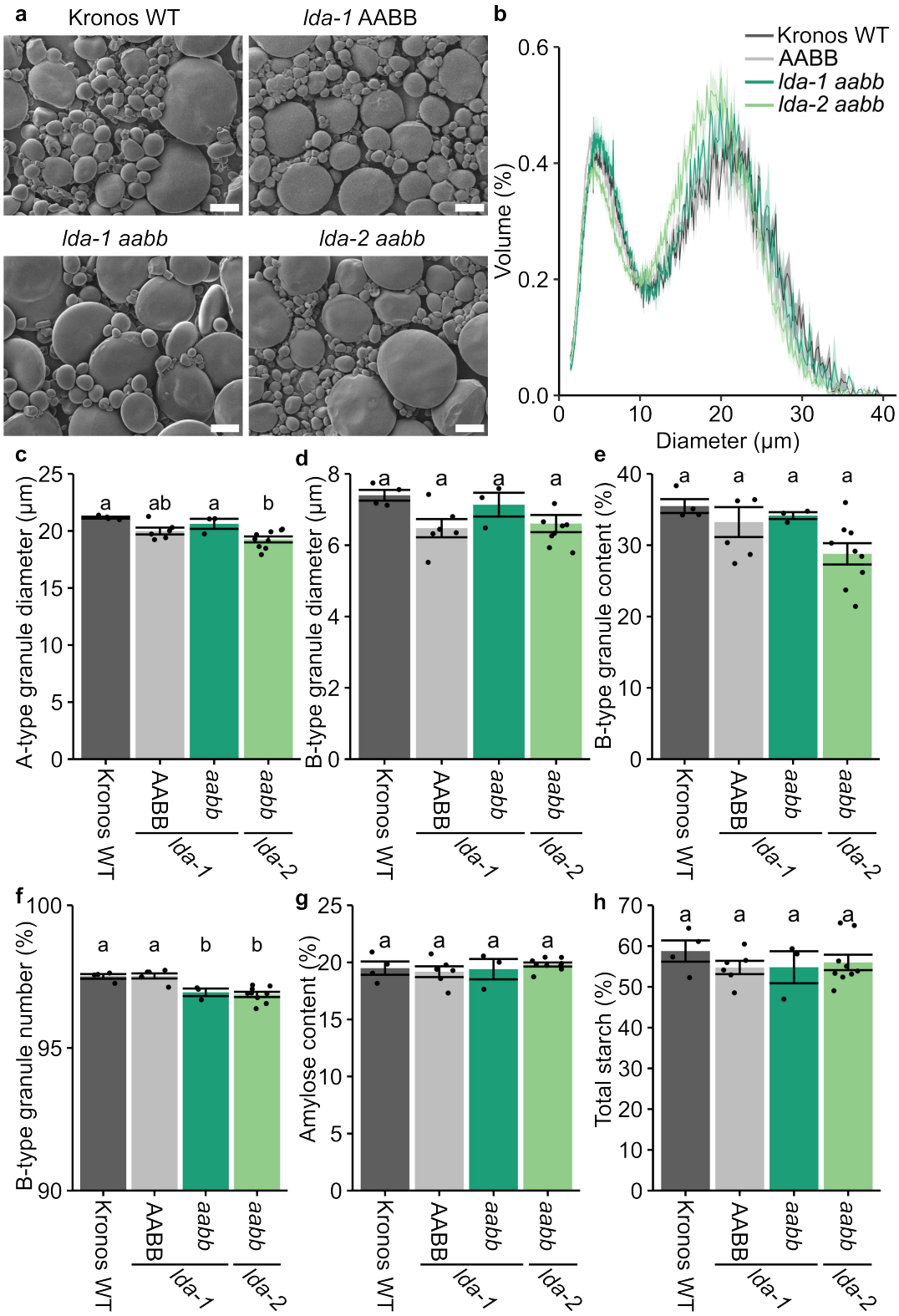

a
Kronos WT
lda-1 AABB
lda-1 aabb
lda-2 aabb
b
Kronos WT
AABB
lda-1 aabb
lda-2 aabb
Volume (%)
Diameter (μm)
c
A-type granule diameter (μm)
d
B-type granule diameter (μm)
e
B-type granule content (%)
f
B-type granule number (%)
g
Amylose content (%)
h
Total starch (%)
Kronos WT
AABB
aabb
aabb
lda-1
lda-2

Figure 5.6 There is a small reduction in B-type granule number in *lda* grains.(a) Scanning electron micrographs of purified endosperm starch, scale bar = 10 μm. (b) Granule size distributions were determined with a Coulter counter, volume (%) vs diameter traces are displayed as means (solid lines) ± SE (shading) and have been adjusted for representation on a linear *x* scale. Mean A-type granule diameter (c), B-type granule diameter (d) and B-type granule content (e) were derived from the size distribution curves by fitting log-normal - normal distributions. (f) The number of B-type granules (< 10 μm) was calculated directly form the Coulter counter data, please note that this *x* axis starts at 90%. (g) Amylose content, (h) Total starch content. For (c-h) data are displayed as means ± SE with individual data points shown as black dots. Values with different letters were significantly different under a one-way ANOVA with Tukey's post hoc test ($P < 0.05$). $n = 4$ for Kronos WT, $n = 6$ for *lda-1* AABB, $n = 3$ for *lda-1 aabb* and $n = 9$ for *lda-2 aabb*.

5.2.4 Identification of two LDA variants associated with large B-type granules

The phenotype of my *T. turgidum lda* mutant were weak, so next I wanted to test whether a loss of LDA activity is seen in the *Ae. tauschii* diversity panel. To investigate this, I selected ten *Ae. tauschii* accessions with high and ten accessions with low B-type granule diameters, in addition to the genome reference TOWWC193 (Table 5.1). I extracted protein from the leaves of these plants and analysed LDA activity on red pullulan native gels (Figure 5.7). I used leaf tissue rather than endosperm tissue as it is easier to collect a larger quantity of tissue. This allows more protein to be extracted which produces higher quality native gels. From a practical side, leaf tissue can be collected early in plant development (~2 weeks post sowing) whereas it takes much longer (~8 months) before developing grains can be harvested. It is worth noting that LDA activity in the leaf might not be representative of LDA activity in the endosperm. However, if there is a loss of function mutation in *LDA*, resulting in little or no LDA activity in the leaf, it is likely that activity will be lost in the endosperm as there is only one known isoform of LDA. I detected bands from the protein extracts of all the selected accessions (Figure 5.7), suggesting that there is functional LDA activity in all accessions tested here. It is hard to draw any conclusions about the intensity of the bands as they vary slightly with each biological replicate, and native PAGE is not a quantitative way of assessing LDA activity. However, there were differences in the band migration pattern, particularly in accessions with large B-type granules. Notably, the bands from TOWWC177 and TOWWC179 were shifted to slightly lower mobility. These differences in migration could stem from many factors including differences in amino acid sequence, protein folding, protein-protein interactions.

Next, I explored whether the different migration patterns stemmed from differences in amino acid sequence. For the 21 accessions shown in Figure 5.7, RNA was extracted from leaves and used to synthesize cDNA. From this, I amplified the cDNA region of LDA encoding for amino acids 79-963 (Figure 5.8), which includes all the main domains and only excludes the transit peptide region. It would not be expected that variation in the transit peptide would have major effects on the protein's activity or interactions, as it is cleaved upon plastid import. The cDNA was sequenced and was converted into the amino acid sequence. Amino acid sequences were aligned with the ClustalW algorithm and selected parts of the alignment are shown in Figure 5.9, with complete alignments provided in Appendix Three; Figures S1-S5. Fourteen of the accessions produced LDA with amino acid sequences identical to that produced by TOWWC193, the genome reference. This protein sequence is henceforth referred to as Ref. Three of the accessions (TOWWC021, TOWWC010, TOWWC139) with large

B-type granules carried an A140T substitution (caused by a guanine to adenine substitution) in the CBM21-like domain and an E248D substitution (caused by a guanine to thymine substitution) in the CBM48 domain (Figure 5.9a). These non-synonymous substitutions were caused by the same SNPs in all three accessions. Furthermore, these accessions had no other non-synonymous substitutions across the entire protein sequence when compared to the protein sequence of the TOWWC193 genome reference. Curiously, these substitutions were not seen in any of the accessions with small B-type granules. Therefore, I classed this as an LDA variant of interest for further investigation – this is henceforth referred to as the "TD" variant. Another interesting variant was detected in the TOWWC177 accession (Figure 5.9b). This had 16 non-synonymous substitutions in LDA that were not found in any of the other LDA sequences reported here (Appendix Three; Figures S1-S5). Although substitutions were found in every protein domain, they were not in the key catalytic residues (D551 and E588). The number of substituted residues made this an interesting variant of LDA to investigate further and is henceforth referred to as "177-like". One substitution, M417K, was found in both the low B-type granule and high B-type granule diameter lines, so this was unlikely to be the SNP affecting B-type granule size. For the substitutions in the TD and 177-like variants, I identified the corresponding residues in the barley LDA protein and highlighted these on the barley structure (Figure 5.10). For both variants, the substitutions almost exclusively occur in exposed residues and appear in linker regions rather than in residues that are important for the secondary structures.

To test for the presence of these different LDA variants across the entire *Ae. tauschii* diversity panel, I designed KASP markers which detected the SNPs causing the A140T, E248D, A238V and N509T substitutions (Figure 5.11a-d). This allowed me to genotype all but two members of the diversity panel for which the DNA was of too poor quality for the KASP. These accessions were TOWWC129 and TOWWC138, which had mean B-type granule diameters of 6.23 μm and 7.24 μm respectively. There were 15 accessions with the TD variant. I compared the B-type granule diameter phenotype of accessions producing the Ref and the TD variants, and found that the TD variant was significantly associated with larger B-type granules and not found in accessions with very small (< 6.9 μm) B-type granules (Figure 5.11e). In contrast, the genotyping for the 177-like variant was more complex. Surprisingly, the TOWWC179 accession consistently grouped with the TOWWC177 for both KASP markers, despite the cDNA sequencing showing that the TOWWC179 produced the Ref, and not the 177-like, LDA variant. Otherwise, there were no inconsistencies between the KASP and cDNA

sequencing results. Usually, more emphasis would be placed on the cDNA sequencing than the KASP. However, the KASP and cDNA sequencing results should be evaluated alongside the native PAGE results in Figure 5.7. Here, the bands for TOWWC179 and TOWWC177 have an almost identical migration pattern, suggesting the presence of the same protein variant. This provides additional evidence that TOWWC179 might indeed encode a 177-like LDA variant, which was not detected during cDNA sequencing, which needs to be explored in more detail. Hence, due to uncertainty around the true TOWWC179 genotype, I have excluded it from the subsequent analysis. In addition to TOWWC177 one accession (TOWWC176) also had the 177-like variant. Again, this accession had larger B-type granules (8.45 μm), suggesting that the 177-like variant is more likely to be found in accessions with larger B-type granules (Figure 5.11e). Hence both variants tend to be associated with larger B-type granules. There are, however, some accessions with large B-type granules which produce the Ref variant of LDA. This could be due to the presence of other mutations in other parts of the genome. This is unsurprising given that B-type granule diameter produced multiple peaks in the GWAS (Figure 4.1), so this one locus is not controlling the entire phenotype. Regardless, this provides additional evidence that these altered LDA variants might be influencing the B-type granule phenotype.

To test whether the residues which are substituted in the TD and 177-like variants are conserved, I compared 31 LDA protein sequences from across the plant kingdom. These sequences were identified by Andersen *et al.* (2020) as orthologs of LDA. The protein sequences were aligned with ClustalW and rendered in ESPript (Figure 5.12). Neither A140 nor E248 were well conserved across LDAs across the plant kingdom. Even within the Triticeae there were differences, for instance, the D-genome ortholog of LDA in *T. aestivum* has T140 and D248 (the TD variant). Many of the substituted residues in the 177-like variant were not well conserved when compared across the plant kingdom (Appendix Three; Figures S6-S10). The exception was A178 which was conserved in everything except *Asparagus officinalis* (Figure 5.12b). A178 occurs within a highly conserved region of the protein, which could suggest it is important for structure or function. Beyond this, there were five residues which, although not strongly conserved across the plant kingdom, were strictly conserved across the Triticeae - V187, A238, A279, A320 and N509 (Figure 5.12b).

Table 5.1 *Ae. tauschii* accessions for which proteins were extracted and analysed on red pullulan native gels. Accessions are listed from smallest (bottom) to largest (top) mean B-type granule diameter. The accession used to generate the original genome reference sequence is marked with a *.

Accession	Mean B-type granule diameter (μm)
TOWWC016	4.75
TOWWC046	5.42
TOWWC059	5.44
TOWWC168	5.44
TOWWC095	5.71
TOWWC002	5.76
TOWWC110	5.80
TOWWC057	5.90
TOWWC056	5.90
TOWWC090	5.93
TOWWC193*	7.49
TOWWC179	8.58
TOWWC172	8.59
TOWWC139	8.81
TOWWC177	8.82
TOWWC010	9.33
TOWWC028	9.62
TOWWC025	9.76
TOWWC187	10.11
TOWWC106	10.87
TOWWC021	11.76

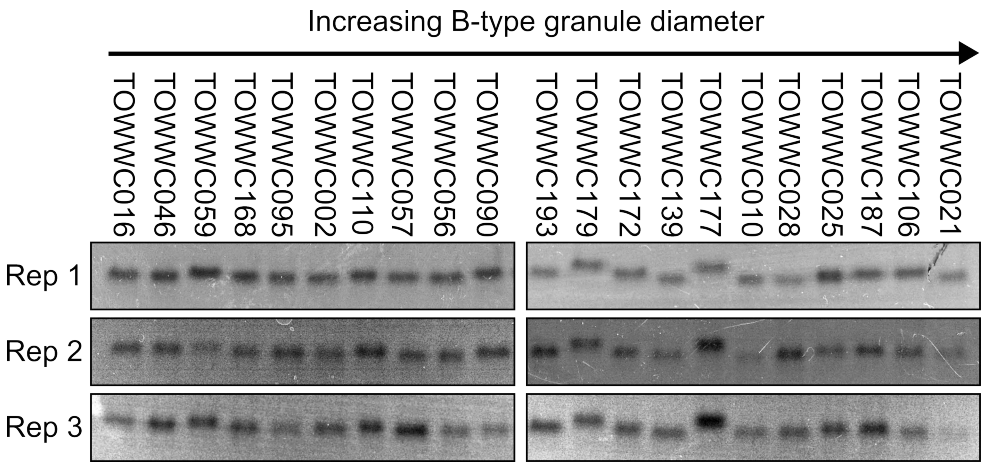


Figure 5.7 LDA from *Ae. tauschii* accessions migrates differently on native gels. Leaf protein (100 µg) was run on a native gel containing red pullulan. Bands represent LDA activity and each replicate corresponds to proteins extracted from a different plant.

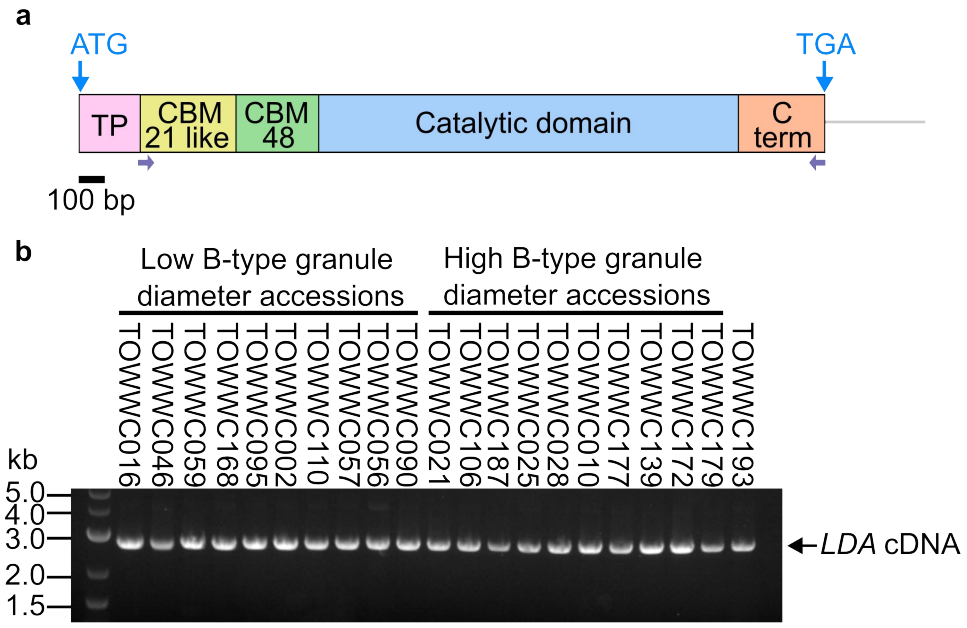


Figure 5.8 Successful amplification of *LDA* cDNA. (a) Diagrammatic representation of the *LDA* cDNA. The regions which encode distinct domains or regions are shown as coloured boxes and the 3'UTR is represented as a grey line. The locations of the start codons (ATG) and stop codons (TGA) are indicated with blue arrows. The binding sites of the primers used to amplify the cDNA in (b) are represented as purple arrows. Scale bar = 100 bp, TP = transit peptide, C term = C-terminal domain. (b) *LDA* cDNA was amplified and resolved on a 1% (w/v) agarose gel, the band just below 3 kb (2.7 kb) represents LDA.

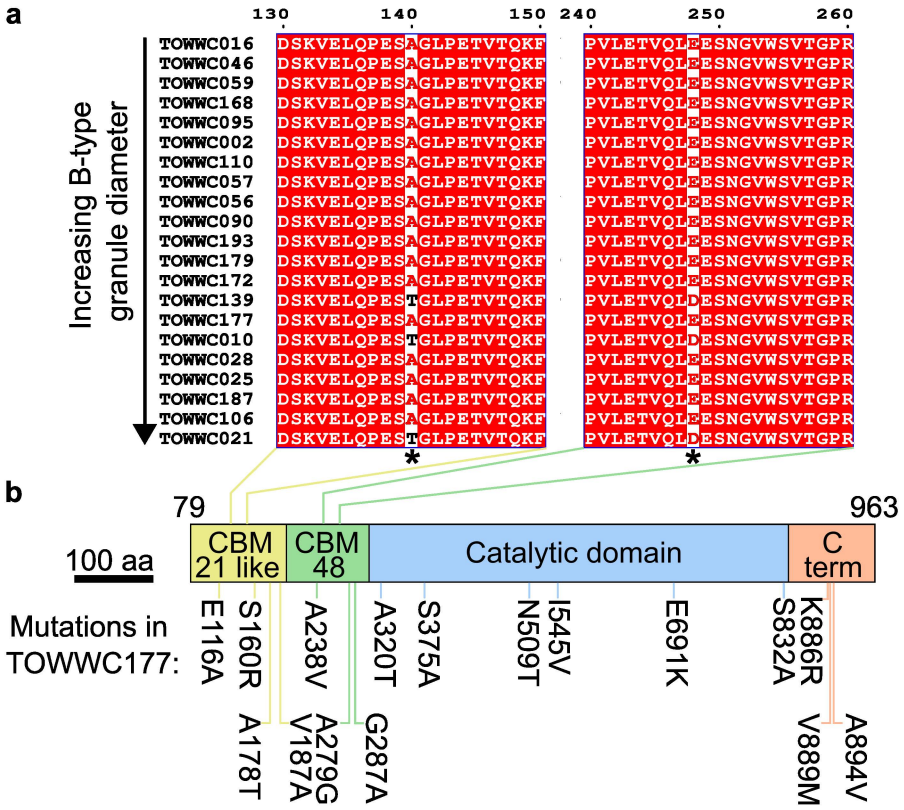


Figure 5.9 Identification of two LDA variants which are associated with accessions with large B-type granules. (a) Multiple sequence alignment of parts of the CBM21-like (left) and CBM48 (right) domains of LDA. Strictly (100%) conserved residues are represented by white letters with a red background; residues with >70% conservation, based on, physicochemical properties of the residues, have red text on a white background. All conserved residues are framed in blue. Residues with <70% conservation, again based on physicochemical properties of the residues, are shown in black on a white background. The A140T and E248D substitutions are marked with a *. (b) Diagrammatic representation of the region of the LDA protein for which the sequence has been determined in this study. Domains are shown as coloured boxes. The regions containing the A140T and E248D substitutions are indicated above the protein with coloured lines. The non-synonymous substitutions exclusively found in TOWWC177 are shown beneath. TP = transit peptide, C term = C-terminal domain, 79 and 963 represent amino acid numbering, scale bar = 100 amino acids.

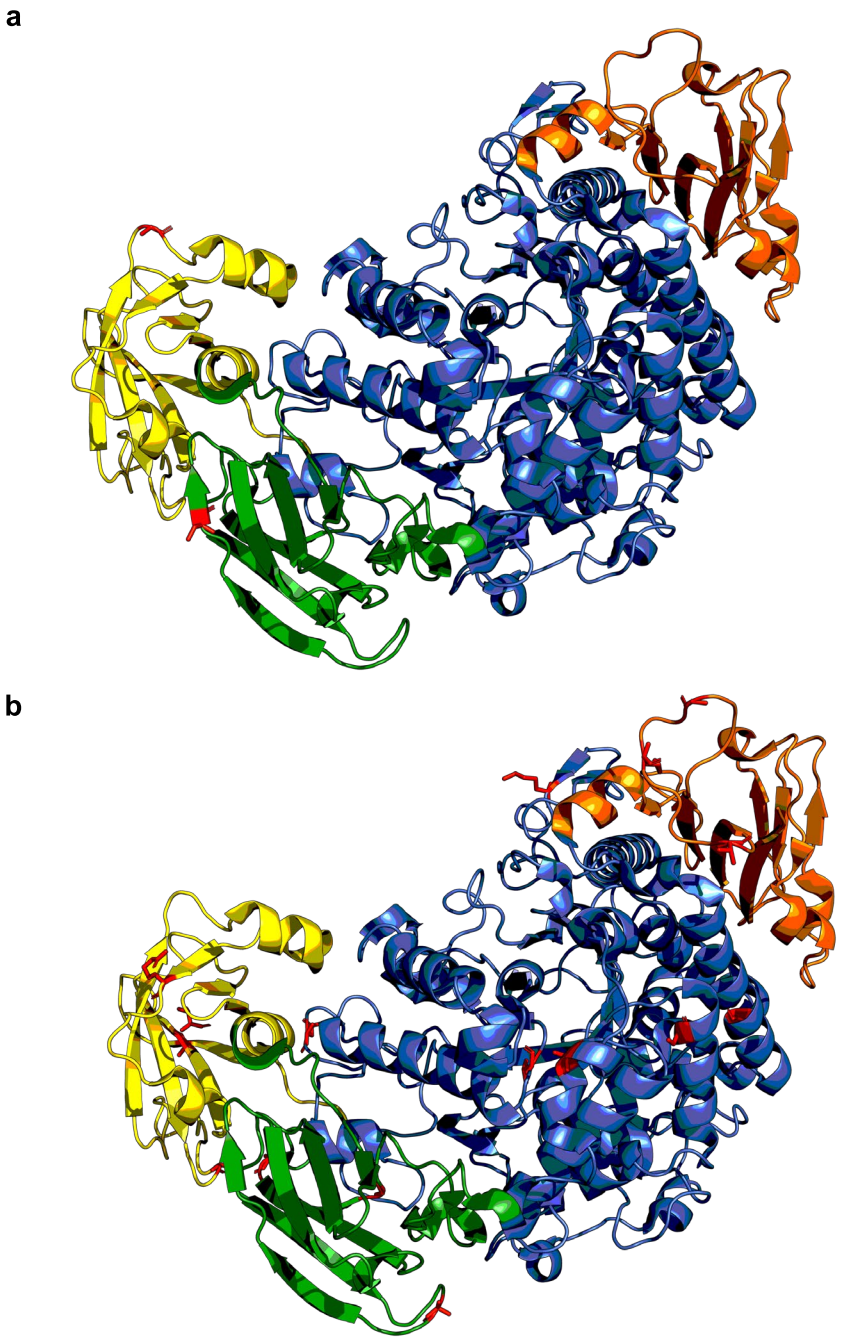


Figure 5.10 The substitutions found in the LDA variants are in exposed residues. Positions of the non-synonymous substitutions, A140T and E248D (a) and the TOWWC177 substitutions (b), were mapped onto the barley structure (PDB = 4aio, (Møller *et al.*, 2012)). Substitutions are in red, and protein domains are coloured with: yellow = CBM21-like, green = CBM48, blue = catalytic domain and orange = C-terminal domain.

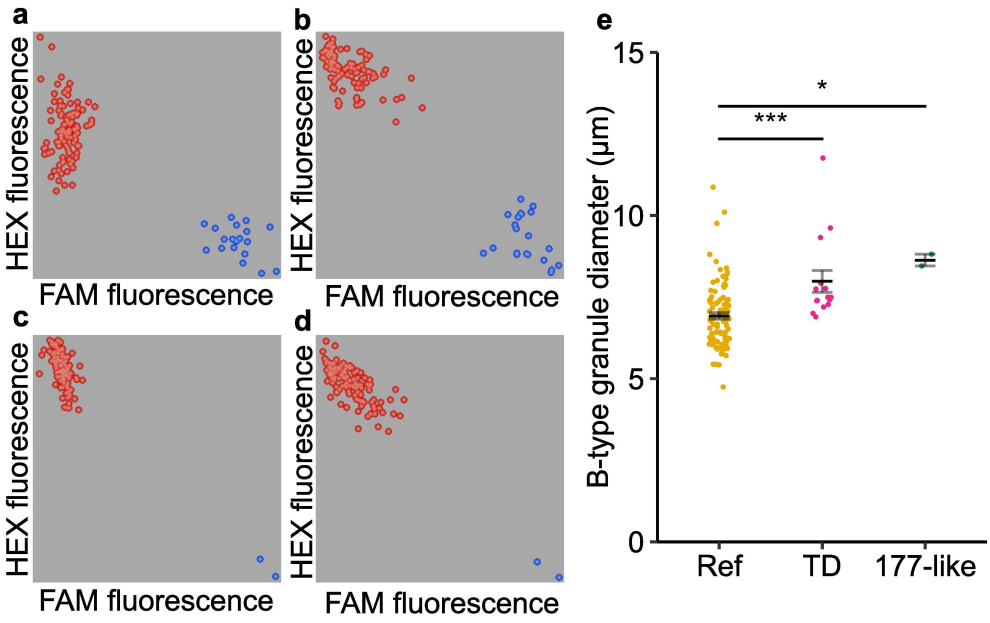


Figure 5.11 The TD and 177-like variants are significantly associated with large B-type granules. (a-d) Examples of KASP results for the SNPs which cause the A140T (a), E248D (b), A238V (c) and N509T (d) substitutions. Red dots represent plants with the Ref allele, blue dots represent plants with the substituted allele. *n* = 127 in a and b, 125 in c and 128 in d. (e) The mean B-type granule diameter from Chapter Three (Figure 3.6b) has been replotted according to the LDA variant encoded by the accession. The mean (solid black line) ± SE (grey lines) are overlaid on the individual data points. The Kruskal Walis test followed by the post hoc Wilcoxon rank sum test was used to determine if there was a significant difference between the groups (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). *n* = 97 for Ref, 15 for TD and 2 for 177-like.

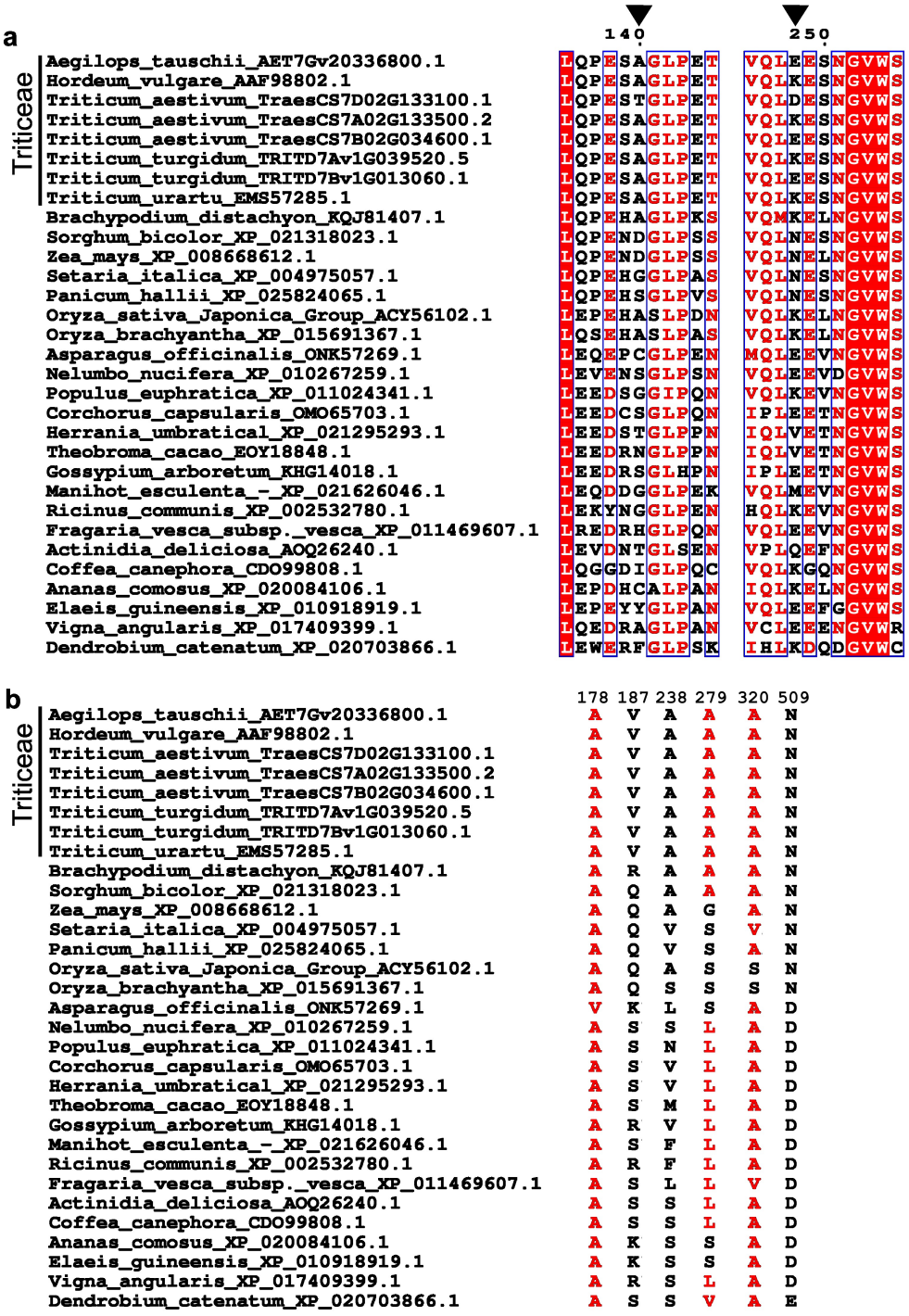


Figure 5.12 The A178 residue of LDA is highly conserved. Multiple sequence alignment of (a) the regions surrounding the A140 and E248 residues (residues highlighted with black arrows) and (b) residues that are altered in the 177-like LDA variant yet, are conserved across the Triticeae. Accessions are arranged in the same order as in Andersen *et al.* (2020). Strictly (100%) conserved residues are represented by white letters with a red background; residues with >70% conservation have red text on a white background and all conserved residues are framed in blue. Residues with <70% conservation are shown in black on a white background. Numbering is relative to the *Ae. tauschii* sequence.

5.2.5 The different LDA variants appeared to have altered in vitro activity

I hypothesized whether the different LDA variants might have altered enzymatic activities, that could be influencing B-type granule formation. To test this, I conducted *in vitro* activity assays with recombinant LDA protein expressed in *Escherichia coli*. For this, the *Ae. tauschii LDA* sequence (encoding amino acids 57-973, which includes all the main domains and only excludes some of the transit peptide region) was expressed in a pET-28a vector in *E. coli* with an N-terminal 6xHis tag, which was used for purification. This method was employed given its success with barley LDA (Møller *et al.*, 2015b). Recombinant protein was produced for the three LDA variants identified in Section 5.2.4 (Ref, TD, 177-like). To dissect the roles of the two substitutions within the TD variant, recombinant protein was also produced for LDA with a single A140T or a single E248D mutation. Furthermore, I produced the *Ae. tauschii* LDA protein carrying the same mutation (V271I) that causes increased B-type granule content in barley (Sparla *et al.* (2014)).

I conducted a small-scale solubility test, to assess whether the Ref LDA could be produced as a soluble protein (Figure 5.13a). Comparing the uninduced soluble fraction with the induced soluble fractions (lanes one and three, respectively) showed the presence of a band just under 130 kDa, which corresponds to LDA as it is only present in the induced sample. The recombinant protein ran slightly higher than its calculated molecular weight of 102 kDa. There was sufficient LDA in this soluble fraction to proceed with large-scale recombinant protein expression and purification for all six LDA variants. It is likely that the large band between 95-130 kDa in the induced soluble fraction is also LDA, suggesting that most of it is insoluble. Unfortunately, my uninduced insoluble fraction was too viscous to load onto the gel for a valid comparison.

Recombinant LDA was successfully purified for all six variants and protein concentration was quantified (Table 5.2). All recombinant proteins were run on an SDS-PAGE gel to assess the quality of the purification. There are few detectable bands in the gel, suggesting the purification has worked well and there is little contamination by other proteins (Figure 5.13b). There were small differences in how the proteins migrated. The Ref, A140T, E248D and V271I run at a slightly higher molecular weight (purple arrows in Figure 5.13b-c) than TD and 177-like (turquoise arrows in Figure 5.13b-c). Likewise, the Ref, A140T, E248D and V271I proteins appear to have a second, faint band which is at a slightly smaller molecular weight than the protein in TD and 177-like (orange arrow in Figure 5.13b-c). This second faint band is

not observed in the TD and 177-like lanes, although may be difficult to detect given its proximity to the main band. From this gel, it is not possible to conclude whether this second band is LDA, which is perhaps slightly degraded, or contamination from bacterial proteins. The major band for all purifications is most likely the full-length LDA protein, hence the main protein activity will be LDA with only minor contributions from any contaminants.

Activity was tested against the commercial PullG6 (4,6-O-Benzylidene-4-nitrophenyl-6^3-α-Dmaltotriosyl-maltotriose) substrate in a colourimetric assay (Figure 5.14a). All LDA variant recombinant proteins had activity against PullG6 (Figure 1.7b). This gives confidence that the bands seen on the Coomassie gel (Figure 5.13b-c) represent LDA that was folded correctly and not denatured in solution. There were significant differences in the LDA activity against PullG6 between the different variants (Table 5.3). Most notably, the 177-like and TD variants had significantly greater activity at all time points, whereas, the V271I variant had significantly lower activity than the Ref at all time points. The A140T and E248D single mutants had no difference in activity compared to the Ref protein, except for the 20-minute timepoint when the E248D mutant which had greater activity, although this increase was small compared to the increase in activity seen in the 177-like and TD variants. When combined with the results of the TD allele, this suggests that both substitutions are, most likely, needed to cause an impact on activity. Overall, I have shown that the V271I, TD and 177-like variants have substitutions that impact LDA activity.

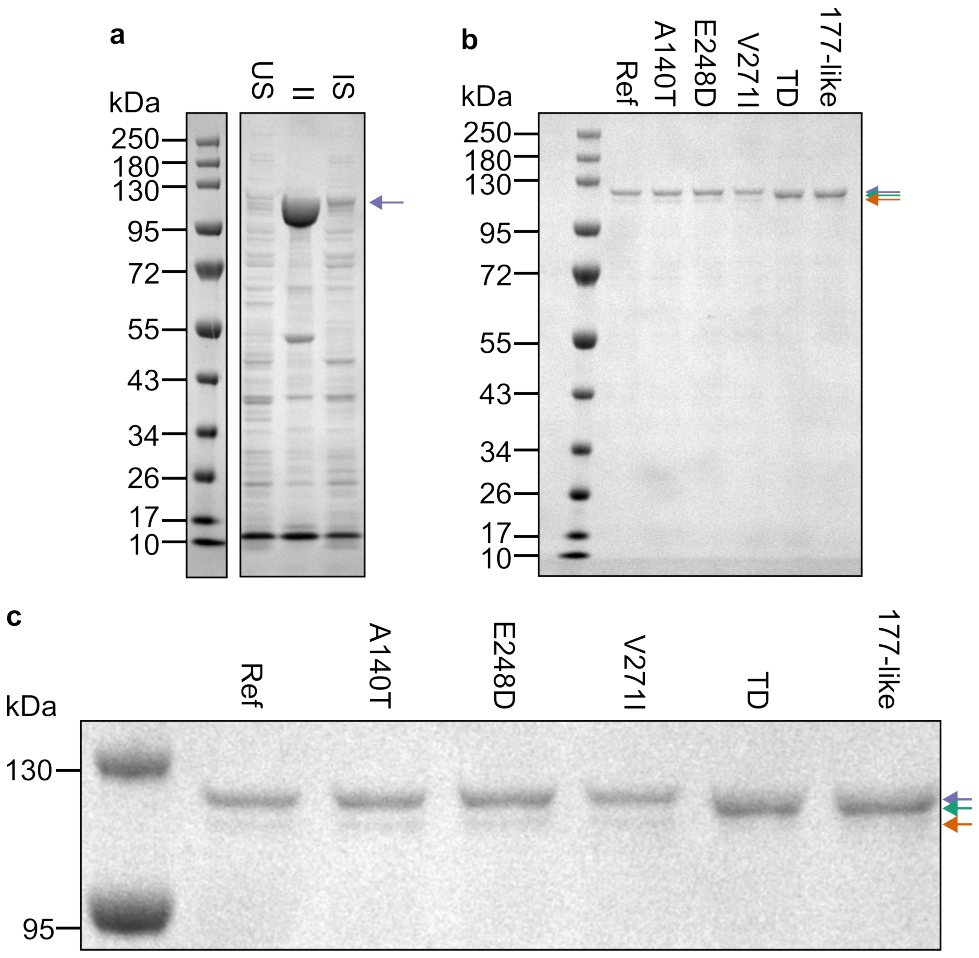
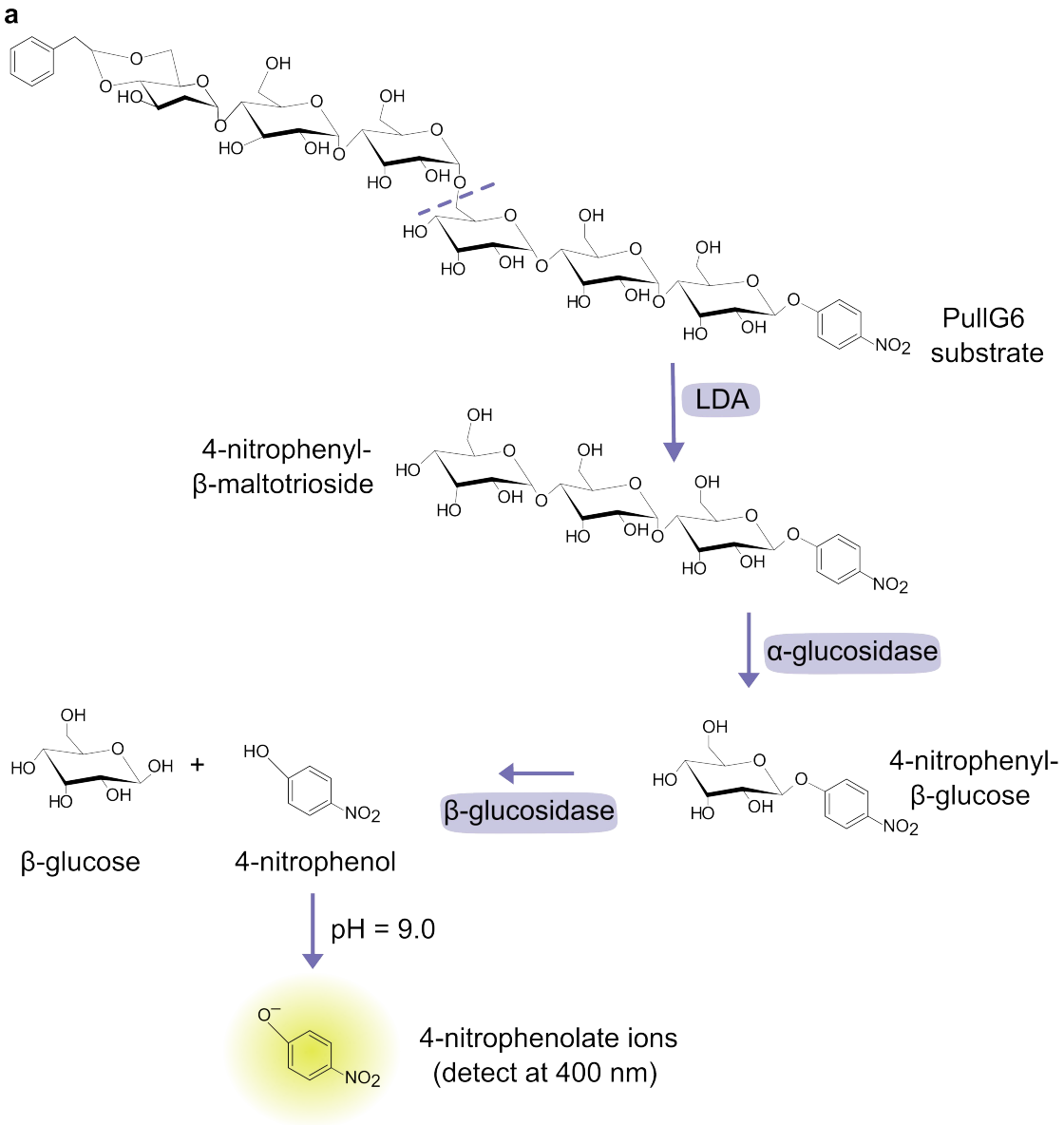


Figure 5.13 Recombinant LDA has high purity. (a) Coomassie-stained SDS-PAGE gel for a small-scale solubility test with the LDA Ref protein (3 µL loaded per lane). The band representing LDA is highlighted with a purple arrow. US = uninduced soluble fraction, II = induced insoluble fraction, IS = induced soluble fraction, the uninduced insoluble fraction could not be loaded onto the gel as it was too viscous. (b) Coomassie-stained SDS-PAGE gel for purified recombinant LDA (1 µg per lane). The bands representing LDA in Ref, A140T, E248D and V271I are highlighted with a purple arrow, the bands representing LDA in TD and 177-like are highlighted with a turquoise arrow, a faint band is detected in some lanes and is highlighted with an orange arrow. (c) The 95-130 kDa region of the gel in (b) has been enlarged for easier identification of the bands. For (a-c), the NEB broad range protein standard was used, and molecular weights are shown in kDa.

Table 5.2 Concentrations of purified recombinant LDA. Concentrations were determined based on absorbance at 280 nm are reported in both μg/μL and μM.

Protein	Concentration (μg/μL)	Concentration (μM)
Ref	0.93	5.54
A140T	1.25	7.45
E248D	0.89	5.33
V271I	0.87	5.21
TD	0.90	5.39
177-like	0.89	5.31


a
PullG6 substrate
LDA
4-nitrophenyl-
β-maltotrioside
α-glucosidase
4-nitrophenyl-
β-glucose
β-glucosidase
β-glucose
4-nitrophenol
pH = 9.0
4-nitrophenolate ions
(detect at 400 nm)

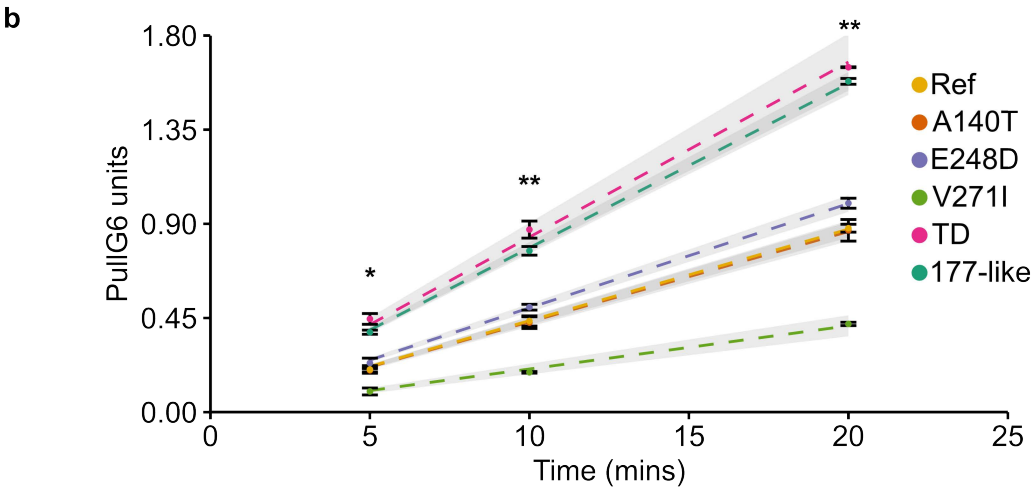

b
PullG6 units
1.80
1.35
0.90
0.45
0.00
0
5
10
15
20
25
Time (mins)
*
**
**
Ref
A140T
E248D
V271I
TD
177-like

Figure 5.14 The different LDA variants have different activities against PullG6. (a) Principle of the PullG6 *in vitro* assay for testing LDA activity. The PullG6 substrate is cleaved by LDA, the position of cleavage is indicated by the dashed purple line. This produces 4-nitrophenyl-β-maltotrioside which is cleaved by an excess of α-glucosidase to produce 4-nitrophenyl-β-glucose. An excess of β-glucosidase converts 4-nitrophenyl-β-glucose into glucose and 4-nitrophenol, which in basic conditions (pH = 9.0) leads to generation of 4-nitrophenolate ions and production of a yellow colour which can be detected at a wavelength of 400 nm. (b) Activity of the different LDA variants against the PullG6 substrate. Data are represented as means ± SE (n = 3) with a trendline (dashed line) with an intercept of zero. Kruskal-Wallis tests were used to assess if there was a significant difference between the activity of the different variants at each time point (* = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$), post-hoc comparisons are given in Table 5.3. One PullG6 unit = mmoles of 4-nitrophenol produced per µg of LDA per min).

Table 5.3 The LDA variants have significantly different activity against PulIG6. Activity of the different LDA variants against PulIG6 was measured (Figure 5.14). Kruskal-Wallis tests followed by the Dunnett post-hoc test were used to assess if there was a significant difference between the activity of the different variants compared to the Ref at each time point. The Dunnett test was conducted in R with the DescTools package, statistically significant comparisons are highlighted with asterisks, where * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$. CI = confidence interval.

Comparison	Difference	Lower CI	Upper CI	*P*
5 minutes				
177-like-Ref	0.17807	0.10367	0.25246	4.20×10^{-5}***
A140T-Ref	-0.0041	-0.0784	0.07034	0.9999
E248D-Ref	0.03079	-0.0436	0.10518	0.6459
TD-Ref	0.24147	0.16708	0.31586	1.70×10^{-6}***
V271I-Ref	-0.1052	-0.1796	-0.0308	0.0058**
10 minutes				
177-like-Ref	0.33537	0.23371	0.43703	1.90×10^{-7}***
A140T-Ref	-0.0078	-0.1095	0.09386	0.9995
E248D-Ref	0.06604	-0.0356	0.1677	0.2688
TD-Ref	0.43696	0.3353	0.53862	5.60×10^{-10}***
V271I-Ref	-0.2453	-0.347	-0.1437	5.90×10^{-5}***
20 minutes				
177-like-Ref	0.70183	0.5978	0.80587	4.10×10^{-15}***
A140T-Ref	-0.01	-0.1141	0.09401	0.9985
E248D-Ref	0.11942	0.01538	0.22346	0.0232*
TD-Ref	0.76929	0.66525	0.87333	$< 2 \times 10^{-16}$***
V271I-Ref	-0.4575	-0.5616	-0.3535	1.20×10^{-8}***

5.3 Discussion

A GWAS in Chapter Four identified *LDA* as a potential candidate for influencing B-type granule formation. Up until now, there has been limited work on the role of *LDA* in the developing Triticeae endosperm. One single point mutant in *LDA* has been published in barley, and this is associated with increased B-type granule content (Sparla *et al.*, 2014). However, no full *lda* knockout has been produced and characterised. Elucidating the role and function of LDA in the Triticeae endosperm is therefore important. In this chapter, I approached this from a genetics angle by characterising *T. turgidum lda* plants and a biochemical angle by assessing the activity of different LDA variants encoded by the *Ae. tauschii* diversity panel.

5.3.1 Complete elimination of LDA in T. turgidum causes a small reduction in B-type granule number

Noone has previously characterised a full *lda* knockout in the Triticeae. Here I identified *T. turgidum lda* mutants and characterised them, showing that there is a ~0.5% reduction in B-type granule number. Besides this, I saw no difference in any starch granule parameter phenotypes including granule diameter or B-type granule content (Figure 5.6). The small difference in B-type granule number could be due to reduced MOS formation, and hence there are less primers to initiate B-type granules. This agrees with the hypothesis I proposed in Section 5.1, where a loss of LDA would lead to less MOS formation, thus affecting B-type granule formation. Total MOS levels in the endosperm would need to be measured in the *lda* mutants to confirm or refute this. Curiously, there was no difference in B-type granule diameter in the *lda* plants, even though I identified *LDA* in my B-type granule diameter GWAS (Chapter Four). Despite the mild effects observed on starch granule number in the *lda* mutants, there are many other aspects still to explore. For example, whether the chain length distribution of the amylopectin is altered, such is observed in rice *lda* mutants (Fujita *et al.*, 2009). Also whether there is an effect on germination rates as observed in maize (Dinges *et al.*, 2003).

It was surprising that there was no difference in B-type granule content in the *lda* plants, given that the barley *lda* V210I mutant characterised by Sparla *et al.* (2014) exhibited a higher B-type granule content. Moreover, purified *Ae. tauschii* LDA with the corresponding mutation (V271I) had lower activity than the Ref variant (Figure 5.14b). From combining these results, it would be expected that the higher B-type granule content of the barley *lda* mutants is caused by a reduction in LDA activity. Hence it might be expected that the *T. turgidum* mutants, which have no detectable LDA

activity, would have a similar phenotype. One reason for this apparent discrepancy is that when LDA activity is completely eliminated, there may be compensation by other debranching enzymes. This level of compensation may not occur if activity is only increased or decreased by small amounts, as seems to be the case in the barley *lda* V270I lines. Alternatively, the mechanisms of compensation between barley and wheat may be different. Compensation by another enzyme is a plausible scenario given the presence of other debranching enzymes in the wheat endosperm, which can also produce MOS. One debranching enzyme is *ISA1*, which is expressed almost uniformly across grain development (Chen *et al.*, 2023a). To test whether ISA1 activity is increased in the *lda* plants, activity assays for ISA activity could be conducted. This has the advantage over qPCR as it allows the ISA activity in the endosperm, rather than gene expression to be measured. If found to be true, it would be interesting to generate a *lda isa1* mutant and characterise its phenotype. Sestili *et al.* (2016) generated RNAi *T. turgidum* lines which had a significant reduction in the expression of both *ISA1* and *LDA* and saw a slight increase in B-type granule content, hence it would be interesting to see if the same phenotype was present in *lda isa1* knockouts.

5.3.2 Identification of two novel LDA variants

Sequencing cDNA and translating it to protein sequence allowed me to identify two LDA variants of interest in the *Ae. tauschii* diversity panel. The TD variant had two substitutions, one in each CBM domain, whilst the 177-like variant had 16 substitutions scattered across the length of the protein. LDA is also known to have variation in other species – Gilding *et al.* (2013) identified two substitutions (G32R and D105A), in the CBM21-like domain, in a natural occurring accession of *S. bicolor.* These substitutions cause a 67% increase in catalytic activity against pullulan. Moreover, Chen and Bao (2016) identified nine non-synonymous substitutions (H100R, L135I, F316L, S473N, L739S, D770E, K802E, G835V, and A854T) in a rice cultivar. Like the 177-like variant, these mutations were scattered across the entire protein sequence with at least one in every domain. However, the authors did not quantify the effect of these on LDA activity. None of the mutated residues identified in *S. bicolor* and rice correspond to the residues I have identified in the TD or 177-like LDA variants, nor do they correspond to S14 or H108, which were identified in the barley LDA to be important for substrate specificity (Andersen *et al.*, 2020). Therefore, the LDA variants I have identified in this study are novel and could be specific to *Ae. tauschii*. It would be interesting to expand this and look at LDA variants in different *Ae. tauschii* lineages to see if more variation could be uncovered.

I identified one residue in the 177-like variant, A178, which is highly conserved in the LDA protein sequences across the plant kingdom (Figure 5.12). This could suggest that it is essential for protein function. In LDA the A178 residue is located towards the end of a β-sheet, hence if this residue is mutated it could prevent proper folding of the β-sheet structure. It would be interesting to generate a mutant for this specific residue and investigate how it affects LDA activity and starch formation. In addition to this, the five other residues that are usually conserved in the Triticeae but are mutated in the 177-like variant are of interest (Figure 5.12b). These residues could adopt a particular role specifically during the formation of the A-type and B-type granules in the endosperm. More work is needed to address this as no study has looked at these residues specifically. It would be interesting to explore all these mutations further, perhaps first by generating recombinant proteins with single mutations to dissect the role of each residue on overall protein function and protein-protein interactions.

My approach successfully identified variation in the LDA protein which correlates with B-type granule diameter. However, by sequencing cDNA, I am missing variation in the introns and the promoter, both of which can influence gene expression (Shaul, 2017). This is important due to the temporal nature of A-type and B-type granule formation, which means that the timing of gene expression affects granule formation and starch accumulation. For instance, Kang *et al.* (2010) identified a *T. aestivum* cultivar which had greater expression of the large cytosolic subunit of AGPase from 15 DPA, and this correlated with a higher total starch content. One way of assessing variation in promoters and introns in *LDA* in the *Ae. tauschii* diversity panel would be to sequence genomic DNA. A limitation of this approach is that it does not provide information about precise gene expression and protein accumulation. Regardless, this strategy was successful for Schwarte *et al.* (2013) who identified variation in the *SS4* promoter of different *Arabidopsis thaliana* accessions. This variation correlated with *SS4* expression. Here, it might be more insightful to directly assess endosperm expression of *LDA*, and the corresponding protein abundance, in *Ae. tauschii* accessions with high and low B-type granule diameter. This could be achieved using endosperm dissected from grains across developmental time (Figure 5.2a) and performing qPCR and quantitative assays for LDA activity, like the PullG6 assay (Figure 5.14a).

5.3.3 LDA activity can be influenced by residues in the CBM21-like and CBM48 domains

Both the TD and 177-like variants had increased activity against the simple substrate PullG6 (Figure 5.14b). Whilst the 177-like variant has many substitutions across the

whole protein, including the catalytic domain, the TD variant only had substitutions in the CBMs (Figure 5.9b). This suggests that these CBMs are affecting enzymatic activity. Interestingly, at the early time points of my assay, the increase in activity was only observed when both substitutions were combined and not in single mutants. This is suggestive of an important role of both CBMs in LDA function. The CBM21-like domain is known to be important for activity and mutations in conserved residues cause a 60-95% reduction in activity (Andersen *et al.*, 2020). How it influences activity is not clear as there is no evidence that the domain can bind starch (Tung *et al.*, 2008; Møller *et al.*, 2012). In glucoamylase from *R. oryzae,* the CBM21 domain rapidly binds a large range of MOS ligands (Cabral *et al.*, 2024). The substrate size is important in this interaction and influences precisely where the substrate binds (Cabral *et al.*, 2024). Usually, CBM21-substrate interactions are transient, although the timing will depend on the specific substrate, and bound substrates can be directed to the catalytic site (Cabral *et al.*, 2024). It is tempting to speculate that a similar mechanism could operate in LDA, whereby the CBM21-like is recognising appropriate substrates and directing them to the LDA active site. Future work needs to focus on identifying if the CBM21-like domain can bind to glucan chains, and if so, what preferences it has. Once these are determined it will be easier to confirm a mechanism.

As explained in Section 5.1.3, binding of carbohydrate substrates to the CBM48 domain of LDA has also not been demonstrated (Wilkens *et al.*, 2018). In other starch granule formation proteins, CBM48s can bind to various glucan substrates. For example, the CBM48 in PTST2 binds MOS primers for delivery to SS4 (Seung *et al.*, 2017). It is interesting to speculate whether both the CBMs might be involved in recognition, perhaps of slightly different substrates, and delivery of these to the LDA active site. This could explain why when only one CBM contains a substitution, the second can compensate for substrate recognition and delivery to the active site, and hence why an effect is only observed with the TD variant and not with the A140T or E248D single mutants. This could explain why individually these residues are not well conserved because as long as one CBM has activity, the enzyme will be functional.

However, the results of the LDA activity assay (Figure 5.14b) should be analysed cautiously, as the experimental set up and subsequent calculations take into account the total amount of protein added. Total protein quantification can be affected by background contamination, and I observed on the SDS-PAGE gel that there seems to be some contamination and, the LDA bands were not of equal abundance (Figure 5.13). The V271I appears narrower and fainter than the Ref, whilst the TD and 177-like

bands appear larger. Hence there could be slightly more TD and 177-like LDA, and slightly less V271I LDA in the activity assays when compared to the Ref. This corresponds with the activity results where V271I gave the lowest activity, whilst the TD and 177-like variants had the highest activity. Consequently, some of the differences in the activity could stem from slight differences in the amount of LDA added to each reaction. A more precise way would be to quantitively determine the LDA concentration rather than total protein concentration, perhaps using antibodies specific to the His-tag of the recombinantly produced LDA. Once more accurate protein concentrations are determined, the assay could then be repeated with a more complex substrate, such as purified starch, and catalytic parameters including K_m and k_{cat} could be calculated. This is important as PullG6 is not a physiological substrate, so activity against this substrate might not reflect the activity on the more complex substrates LDA acts on *in planta*.

As the TD and 177-like LDA variants had higher activity in the activity assay (Figure 5.14b), it is reasonable to assume that the larger B-type granule diameter phenotype in *Ae. tauschii* might be caused by an increase in activity rather than a loss of activity. This could be another reason why there is no strong phenotype in the *lda* mutants. Stahl *et al.* (2004) argue that increased LDA activity results in reduced B-type granules in *LDI* RNAi barley lines. However, this must be analysed cautiously as the total LDA activity in developing endosperm extracts was comparable to their wild-type controls. Beyond this, there are no other examples of increased *LDA* expression affecting starch granules in plants. Consequently, an interesting extension to the work conducted here would be to generate a wheat line overexpressing *LDA*. Or alternatively recreate the substitutions observed in the TD and 177-like *LDA* alleles in wheat and observe the phenotype.

Overall, in this chapter I have employed a genetics strategy and characterised *T. turgidum lda* mutants, which is the first example of an LDA knockout mutant in the Triticeae. In parallel, I have investigated LDA at a biochemical level by identifying two interesting variants and measured that they have increased activity against a simple substrate. By combining these results together, I suggest that the large B-type granule observed in some *Ae. tauschii* accessions may stem from increased LDA activity rather than a loss of activity. This provides greater insights into the roles LDA might be playing in endosperm starch formation.

Chapter Six – General Discussion

The aims of my thesis were to characterise natural diversity in *Ae. tauschii* endosperm starch and use association mapping to identify novel genes and variation that influence endosperm starch synthesis in the Triticeae. The importance of this cannot be understated as there are gaps in our understanding of A-type and B-type granule formation, yet the presence of these two granule types greatly influences the physiochemical and digestive properties of starch and therefore the end uses of wheat. Hence understanding granule formation could allow starch with improved quality to be bred and developed.

To date, most studies which have utilised association genetics to investigate endosperm starch properties have used domesticated species rather than wild relatives. However, wild relatives often harbour more variation. This is true for *Ae. tauschii* endosperm starch as in Chapter Three I discovered significant variation in granule size, content and number. Interestingly some of this variation does not appear to be present in modern wheat cultivars (Chapter Three). Therefore, *Ae. tauschii* is a source of untapped variation for improving the quality of Triticeae starch. This gives confidence that conducting an association study with *Ae. tauschii*, rather than a domesticated Triticeae species, is a promising strategy for identifying novel genes and variation important for starch granule properties. Indeed, I successfully employed this strategy to identify multiple, novel significant loci associated with both B-type granule size and number (Chapter Four). These loci could be useful for breeding wheat with altered B-type granules for use in food and industrial processes. In addition to these novel loci, I identified the gene encoding the starch debranching enzyme *LDA* under a peak associated with B-type granule diameter. Interestingly, I discovered two variants of LDA present in *Ae. tauschii* accessions with large B-type granules which appear to have increased activity against a simple glucan substrate (Chapter Five). This suggests a potential role of LDA in granule formation in the Triticeae. Complete elimination of *LDA* in tetraploid wheat caused only minor effects on endosperm starch granule formation. This suggests that a gain-of-function mutation may be required to affect granule morphology. Here I will discuss my work in a broader context, focusing on how it advances our fundamental understanding of starch synthesis and its direct applications on wheat breeding and quality improvement.

6.1.1 Insights into the role of LDA in starch granule formation in the Triticeae

Our current model of B-type granule formation is that MOS are released from A-type granules and subsequently used to prime the initiation of B-type granules. This relies

on PHS1 and BGC1 (Chapter One; Figure 1.11) (Chia *et al.*, 2020; Kamble *et al.*, 2023). The enzyme which releases MOS from the A-type granules had not been identified. When *LDA* was found beneath a peak associated with B-type granule size in my GWAS (Chapter Four) I hypothesized that it could be fulfilling this role. Analysis of *T. turgidum lda* mutants (Chapter Five) revealed only a small defect in B-type granule number. When considered more broadly, the lack of phenotype in the *lda* mutants is not surprising as many *Ae. tauschii* accessions with large B-type granules appear to encode for LDA variants with increased catalytic activity. Therefore, it is possible that *LDA* overexpression or gain-of-function mutants may have larger effects on B-type granules. How increased LDA catalytic activity might result in larger B-type granules is difficult to rationalise with our current mechanistic understanding of starch granule formation. Under this model, one would predict that simply increasing LDA activity would lead to more MOS release from A-type granules. This would result in more primers for B-type granule initiation causing an increase in the number of B-type granules. In situations where total starch content is equal, a greater number of B-type granules would mean that there would be less carbon and space allocated to each growing granule. It would therefore be expected that each granule would be smaller. This goes against what I have observed in *Ae. tauschii* where increased LDA activity is associated with larger B-type granules.

However, the model may not be as simple as this. Here, I have only measured LDA catalytic activity. Yet, it is possible that the substitutions in the LDA variants also affect protein-protein interactions as most of the substitutions are on the protein surface (Chapter Five; Figure 5.10). Differences in protein-protein interactions could be more important than the altered catalytic activity. If the LDA variants cannot interact with key interaction partners, it could prevent the efficient transfer of MOS from LDA to the proteins involved in B-type granule initiation. Such effect could result in decreased B-type granule initiation, resulting in more amyloplast space and substrates for the B-type granules which are initiated to grow larger. To test this hypothesis, the protein-protein interactions of the reference LDA should be determined. It would then be possible to assess if there are differences with the LDA variants (TD and 177-like) identified here.

Other factors, such as the length of the MOS primer, may influence the efficiency of B-type granule initiation. The proteins involved in granule initiation in other species show specific substrate specificity for MOS length. *Arabidopsis thaliana* PTST2 (the ortholog of BGC1) preferentially binds longer MOS rather than shorter MOS such as

maltoheptaose (7 glucose units) (Seung *et al.*, 2017). Moreover, rice PHS1 has almost no synthesis activity on MOS with less than four glucose units, but is efficient with longer MOS (≥ 4 glucose units) (Hwang *et al.*, 2010). The precise selectivity of PHS1 also depends on the physiological conditions as Pi, which is produced as a byproduct of the synthesis reaction and competitively inhibits the reaction. This inhibition is less severe for short chain MOS (Hwang *et al.*, 2010). So far, all experiments into PHS1 and BGC1 selectivity has been conducted in isolation. However, PHS1 and BGC1 interact *in planta* (Kamble *et al.*, 2023), and whether this BGC1-PHS1 interaction affects specificity of the proteins for MOS has not been investigated. Furthermore, the MOS selectivities of BGC1 and PHS1 in the Triticeae has not been studied. If the MOS profile affects the efficiency of the granule initiation proteins, then the length of MOS produced by LDA could be important for determining the amount of granule initiation. The true MOS release profile of LDA is unknown as previous studies have quantified it using pullulan which is not a physiological substrate in plants (Wu *et al.*, 2002; Wangpaiboon *et al.*, 2023). Thus, future work should determine whether there are differences in the MOS release pattern from the LDA variants versus the reference. This could help explain how the variation in LDA affects the initiation of B-type granules, and consequently their size. If the different LDA variants release MOS which is of a less favourable length for PHS1 and BGC1, the overall amount of B-type granule initiation might decrease. Less B-type granule initiations may result in more space in the amyloplast for the B-type granules which are initiated to grow larger. This could explain why LDA variants were associated with larger B-type granules in some *Ae. tauschii* accessions.

If LDA is important for MOS metabolism and the generation of primers for B-type granule initiation it is surprising that eliminating *LDA* only has a small effect. Perhaps a complete loss can be compensated by a different debranching enzyme – as discussed in Chapter Five; Section 5.3.1. Therefore, generation of lines with more than one debranching enzyme mutation (e.g. *lda isa1* or *lda isa3*) will be important for elucidating this. One factor not yet considered is possible interactions between different debranching enzymes. For example, in cassava co-incubation of LDA and ISA3 with pullulan revealed that they work synergistically to affect the overall MOS profile (Wangpaiboon *et al.*, 2023). Therefore, analysis of the entire debranching profile of the endosperm may be important to assess which enzyme or enzymes are important for producing MOS for B-type granule initiation. Overall, both genetic approaches, such as generation of overexpression lines and lines with more than one debranching enzyme mutation (e.g. *lda isa1* or *lda isa3*), and biochemical approaches,

such as characterising the MOS released from the reference LDA and the variants identified here, will be important to fully understand how LDA is influencing B-type granule formation in the *Ae. tauschii* diversity panel and beyond.

I also found that B-type granule parameters are complex traits controlled by multiple genomic loci, as there were many peaks in the Manhattan plots for B-type granule size and number (Chapter Four). Therefore, the interaction of *LDA* with these other genomic loci could be influencing the overall B-type granule phenotype. This is possible as the *Ae. tauschii* accessions in the diversity panel are highly homozygous, so even if loci are on different chromosomes, they can be inherited together. Within the loci associated with B-type granule diameter I identified eleven other genes apart from *LDA*, which are predicted to be plastid localised and have *T. turgidum* orthologs with similar expression patterns to known starch granule initiation genes. It would be interesting to see if any of these directly interact with LDA through protein-protein interactions, or how mutating these genes in the *lda* background influences the observed phenotype.

6.1.2 Insights into starch synthesis beyond the Triticeae

More broadly, my work could have a large impact on our fundamental understanding of starch synthesis both in the Triticeae and beyond. Firstly, LDA is usually believed to be involved in starch degradation, for example in germination (Burton *et al.*, 1999) and transitory starch metabolism (Pfister and Zeeman, 2016). However, here I suggest that this degradation activity may be important for MOS metabolism and influencing B-type granule initiation – which is a biosynthetic process. Whether LDA has been specifically co-opted for this role in bimodal starch granule formation in the Triticeae is not yet known. In simple granule producing maize and compound granule producing rice, there is no starch granule size or shape phenotype in *lda* plants (Dinges *et al.*, 2003; Fujita *et al.*, 2009). This is consistent with my work here where *T. turgidum lda* plants had only a small B-type granule phenotype. Future work in maize and rice could focus on increasing LDA activity, for example through overexpression studies. This will help to elucidate whether LDA also plays a role in starch synthesis in species which produce non-bimodal granules, and consequently, help to answer a key question in the starch field of how different starch granule morphologies are achieved.

Furthermore, I identified new avenues for research for granule formation, including 13 genes with no previous links to granule formation (Chapter Four). It would be interesting to assess the starch granule phenotype for mutants of these genes in the

Triticeae, and if present outside the Triticeae. This could help increase our understanding of how bimodal starch granule initiation arose. Each gene in our current model for bimodal starch granule formation in the Triticeae endosperm has at least some link to other starch synthesis processes, for instance transitory starch metabolism. It will be interesting to investigate whether any of the genes I have identified have specific roles in A- or B-type granule formation. Alternatively, if these genes also have roles in starch formation in other tissues and species it adds weight to the idea that the generation of distinct starch granule distributions relies on the expression pattern and abundance of the same small set of proteins involved in starch metabolism.

More generally, my work raises interesting discussions on the adaptive significance of starch granules parameters and how these have been influenced by domestication. *Ae. tauschii* had variation in its starch properties which were not observed in modern wheat cultivars (Chapter Three). It is interesting to speculate whether starch granule parameters might provide an adaptive significance for *Ae. tauschii* accessions. For instance, one could imagine that having more A-type granules, which tend to be more easily degraded during germination (Li *et al.*, 2012), might promote faster germination and seedling establishment. Earlier establishment could help *Ae. tauschii* plants to outcompete others which are slower to germinate and develop. It seems likely that some of the starch granule parameters may have been selected for during domestication. For example, I saw a decrease in mean A-type granule diameter in modern wheat cultivars compared to *Ae. tauschii*. Most likely this would have been indirect selection as a consequence of selecting for other properties such as responses to stress and increased grain yield (Kiszonas and Morris, 2018). Hence, exclusively studying domesticated species could mean that interesting starch granule properties are missed. Future research should therefore evaluate how wild species can be incorporated into research programmes on starch biosynthesis, which could provide greater insights into starch granule formation.

6.1.3 The genomic loci associated with granule size and number could be useful to the breeding industry

Identifying QTLs associated with starch granule quality parameters and producing markers to genotype these, as I have achieved here, is important for breeding wheat with high functional and nutritional quality. Historically the major focus of the wheat breeding industry has been increasing yield, resulting in ~0.6% increase in potential yield per annum globally (Fischer, 2022). Quality is not ignored in modern breeding

pipelines; however it is often negatively correlated with yield (Yao *et al.*, 2018), and yield is prioritised. Accurate quality selection can be challenging, often requiring low-throughput, expensive tests in advanced breeding generations (Kiszonas and Morris, 2018). One quality test involves using a Rapid Visco Analyser (RVA) which examines viscosity by mimicking processing steps (mixing, heating and cooling). However, this is too time consuming to utilise during early breeding generations (Cozzolino, 2016). Hence, positive quality traits in early breeding generations may be unknowingly discarded. The development of genomic markers could assist in high-throughput selection of grain quality at earlier generations, to avoid situations where poor quality grains are maintained and progressed through to advanced generations, which would be a loss of investment and resources.

Several markers are already available for starch quality traits, although they have not been heavily utilised (Guzmán *et al.*, 2022). One example is a marker being used to detect loss of function of *GBSSI* on chromosome 4A, which results in lower amylose content for producing high-quality udon noodles (McLauchlan *et al.*, 2001). Moreover, markers are used to detect deletions and SNPs in *SBE2a* and *SBE2b*, which result in high amylose wheat with increased resistant starch, which is an important nutritional quality trait (see Section 6.1.4) (Regina *et al.*, 2015). Other markers have been developed for *SS1*, *SS2a*, *SS3* and *SBE1*, although the extent to which these has been implemented in breeding has not been assessed (Guzmán et al., 2022). Regardless, none of these markers allow selection for starch granule size, content or number, despite these traits influencing end uses (Section 6.1.4). Therefore, development of markers for these traits is an area which could be exploited to help breeders increase starch quality in commercial wheat lines.

The lack of markers for starch quality traits is surprising given that numerous QTLs have been reported for starch granule properties in the Triticeae (Chapter One; Table 1.2 and Table 1.3). Consequently, there is a disconnect between identification of QTLs and the generation of markers to make these QTLs useful for the breeding industry. To overcome this, for the chromosome seven QTL associated with B-type granule size (Chapter Four, peak D), I have identified specific mutations in *LDA* which are correlated with larger B-type granules (Chapter Five). Development of markers for genotyping in *Ae. tauschii* was highly successful. The next step is to deploy these in wheat pre-breeding pipelines to identify material containing these specific alleles. There is a possibility that the specific alleles from *Ae. tauschii* may not be present in wheat varieties or landraces. In these cases, SHWs could be made to transfer these

loci into wheat (Chapter Three; Figure 3.13) (Li *et al.*, 2018a). Further work needs to be conducted on the other loci identified in Chapter Four to design markers which can genotype these regions and be used in wheat breeding programmes, which could be impactful as the first examples of genetic markers for B-type granule size and number.

Before a large-scale effort is deployed to use these loci and develop markers for the breeding industry, an important step is to test the robustness of the loci in different environmental conditions. This is important as loci are more useful in breeding if they are not heavily influenced by the environment. This is particularly relevant as starch granule initiation is impacted by environmental conditions, including drought and heat stress (Zhang *et al.*, 2017). The gold standard approach for tackling this would be to grow the entire diversity panel, re-phenotype and repeat the GWAS. By re-conducting the GWAS, the reproducibility of the peaks could be assessed – this could be important for some of the smaller, minor peaks. However, this approach would be time and labour intensive. An alternative strategy would be to grow a small subset of accessions which have different genotypes for the loci identified in this study in contrasting environmental conditions. For each genotype the effect of the environment on the phenotype could be measured and the robustness assessed. Regardless of the approach taken, the link between genotype and phenotype does not necessarily have to be 100% reproducible. Many of the wheat quality traits and markers already in use only have accuracies of 40-60% (Guzmán *et al.*, 2022). Even if the loci identified here only have a 40% link to the B-type granule phenotype this still provides some knowledge on breeding for these phenotypes.

6.1.4 There are multiple benefits of altering B-type granule size and number

The variation in B-type granules (Chapter Three) and the corresponding genomic loci identified in the GWAS (Chapter Four), might mean it is possible to produce wheat with more or fewer B-type granules. This could be achieved through breeding (Section 1.1.2) or genetic engineering. One particularly desirable phenotype would be fewer but larger B-type granules. This phenotype could be useful for milling as some studies report better milling efficiencies with grains with fewer B-type granules (Edwards *et al.*, 2008). For end users, this phenotype could be beneficial for brewing, where B-type granules are highly undesirable as they ferment inefficiently and gelatinise resulting in a hazy beer (Bathgate and Palmer, 1972; Tillett *et al.*, 1993). This phenotype may also be beneficial for the food and health industries more generally, as in the gut, starch digestion results in the release of sugars. Starch which is more difficult to digest, commonly known as resistant starch, causes slower release of sugars and reduces the overall peak of the blood glucose spike (DeMartino and Cockburn, 2020). Granule

size has a large influence on digestibility in the gut – A-type granules are more difficult to digest than B-type granules (Li *et al.*, 2023). Therefore, starch with a small number of large B-type granules should have more resistant starch, as there are fewer B-type granules, and those which remain are larger and should be slightly more resistant to digestion. Consequently, my work could help to develop wheat with higher resistant starch, which could be beneficial for the food and health industries.

More generally, the variation I have discovered in B-type granule content could be useful as different B-type granule contents are desirable for different process. For example, high B-type granule contents are desirable for the pasta industry as B-type granules are associated with increased quality (Soh *et al.*, 2006). Furthermore, a higher number of B-type granules, but not necessarily large B-type granules, would be useful for the paper making industry where B-type granules fill the pores on the paper's surface (Shevkani *et al.*, 2017). Low B-type granule content is desirable for the production of starch films (Montaño-Leyva *et al.*, 2008).

Consequently, my work has the potential to impact many industries, from milling to food to paper production. It is therefore essential that there is engagement with breeders and end users so that the genomic loci I have identified can be used to create the most useful phenotypes for industrial purposes.

References

Abt, M.R., Pfister, B., Sharma, M., Eicke, S., Bürgy, L., Neale, I., Seung, D., and Zeeman, S.C. (2020). STARCH SYNTHASE5, a noncanonical starch synthase-like protein, promotes starch granule initiation in *Arabidopsis*. Plant Cell **32,** 2543-2565.

Abu-Jamous, B., and Kelly, S. (2018). *Clust*: automatic extraction of optimal co-expressed gene clusters from gene expression data. Genome biology **19,** 172.

Ahmed, H.I., Heuberger, M., Schoen, A., Koo, D.-H., Quiroz-Chavez, J., Adhikari, L., Raupp, J., Cauet, S., Rodde, N., Cravero, C., Callot, C., Lazo, G.R., Kathiresan, N., Sharma, P.K., Moot, I., Yadav, I.S., Singh, L., Saripalli, G., Rawat, N., Datla, R., Athiyannan, N., Ramirez-Gonzalez, R.H., Uauy, C., Wicker, T., Tiwari, V.K., Abrouk, M., Poland, J., and Krattinger, S.G. (2023). Einkorn genomics sheds light on history of the oldest domesticated wheat. Nature **620,** 830-838.

Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. Life Science Alliance **2,** e201900429.

Andersen, S., Svensson, B., and Møller, M.S. (2020). Roles of the N-terminal domain and remote substrate binding subsites in activity of the debranching barley LIMIT DEXTRINASE. Biochimica et Biophysica Acta (BBA) Proteins Proteomics **1868,** 140294.

Aranzana, M.J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., and Tang, C. (2005). Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genetics **1,** e60.

Arora, S., Cheema, J., Poland, J., Uauy, C., and Chhuneja, P. (2019a). Genome-wide association mapping of grain micronutrients concentration in *Aegilops tauschii*. Frontiers in Plant Science **10,** 54.

Arora, S., Singh, N., Kaur, S., Bains, N.S., Uauy, C., Poland, J., and Chhuneja, P. (2017). Genome-wide association study of grain architecture in wild wheat *Aegilops tauschii*. Frontiers in Plant Science **8,** 886.

Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y., Matny, O., Johnson, R., Enk, J., Periyannan, S., Singh, N., Asyraf Md Hatta, M., Athiyannan, N., Cheema, J., Yu, G., Kangara, N., Ghosh, S., Szabo, L.J., Poland, J., Bariana, H., Jones, J.D.G., Bentley, A.R., Ayliffe, M., Olson, E., Xu, S.S., Steffenson, B.J., Lagudah, E., and Wulff, B.B.H. (2019b).

Resistance gene cloning from a wild crop relative by sequence capture and association genetics. Nature Biotechnolgy **37,** 139-143.

Asare, E.K., Båga, M., Rossnagel, B.G., and Chibbar, R.N. (2012). Polymorphism in the barley *GRANULE BOUND STARCH SYNTHASE 1* (*GBSS1*) gene associated with grain starch variant amylose concentration. Journal of Agricultural and Food Chemistry **60,** 10082-10092.

Badenhuizen, N.P. (1963). Formation and distribution of amylose and amylopectin in the starch granule. Nature **197,** 464-467.

Ballicora, M.A., Frueauf, J.B., Fu, Y., Schürmann, P., and Preiss, J. (2000). Activation of the potato tuber ADP-GLUCOSE PYROPHOSPHORYLASE by thioredoxin. Journal of Biological Chemistry **275,** 1315-1320.

Ballicora, M.A., Laughlin, M.J., Fu, Y., Okita, T.W., Barry, G.F., and Preiss, J. (1995). ADENOSINE 5'-DIPHOSPHATE-GLUCOSE PYROPHOSPHORYLASE from potato tuber (significance of the N terminus of the small subunit for catalytic properties and heat stability). Plant Physiology **109,** 245-251.

Banks, W., Greenwood, C.T., and Walker, J.T. (1970). Studies on the starches of barley genotypes: the waxy starch. Starch - Stärke **22,** 149-152.

Batey, I.L., Hayden, M.J., Cai, S., Sharp, P.J., Cornish, G.B., Morell, M.K., and Appels, R. (2001). Genetic mapping of commercially significant starch characteristics in wheat crosses. Australian Journal of Agricultural Research **52,** 1287-1296.

Bathgate, G.N., and Palmer, G.H. (1972). A reassessment of the chemical structure of barley and wheat starch granules. Starch - Stärke **24,** 336-341.

Bechtel, D.B., Zayas, I., Kaleikau, L., and Pomeranz, Y. (1990). Size-distribution of wheat starch granules during endosperm development. Cereal chemistry **67,** 59-63.

Beló, A., Zheng, P., Luck, S., Shen, B., Meyer, D.J., Li, B., Tingey, S., and Rafalski, A. (2008). Whole genome scan detects an allelic variant of *FAD2* associated with increased oleic acid levels in maize. Molecular Genetics and Genomics **279,** 1-10.

BeMiller, J.N., and Whistler, R.L. (2009). Starch: chemistry and technology. (Academic press).

Bertoft, E. (2013). On the building block and backbone concepts of amylopectin structure. Cereal Chemistry **90,** 294-311.

Bertoft, E. (2015). Fine structure of amylopectin. In Starch: Metabolism and Structure, Y. Nakamura, ed (Tokyo: Springer Japan), pp. 3-40.

Bertolini, A. (2009). Starches: characterization, properties and applications. (Boca Raton, Florida, USA: CRC Press).

Bhattacharyya, M.K., Smith, A.M., Ellis, T.H.N., Hedley, C., and Martin, C. (1990). The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding *STARCH-BRANCHING ENZYME*. Cell **60,** 115-122.

Black, C.K., Panozzo, J.F., Wright, C.L., and Lim, P.C. (2000). Survey of white salted noodle quality characteristics in wheat landraces. Cereal Chemistry **77,** 468-472.

Bogracheva, T.Y., Morris, V.J., Ring, S.G., and Hedley, C.L. (1998). The granular structure of C-type pea starch and its role in gelatinization. Biopolymers **45,** 323-332.

Boisson, M., Gomord, V., Audran, C., Berger, N., Dubreucq, B., Granier, F., Lerouge, P., Faye, L., Caboche, M., and Lepiniec, L. (2001). *Arabidopsis GLUCOSIDASE I* mutants reveal a critical role of *N*-glycan trimming in seed development. The EMBO Journal **20,** 1010-1019.

Bond, J.L., Rogols, S., and Salter, J.W. (1975). Size classified cereal starch granules (Google Patents).

Borém, A., Mather, D.E., Rasmusson, D.C., Fulcher, R.G., and Hayes, P.M. (1999). Mapping quantitative trait loci for starch granule traits in barley. Journal of Cereal Science **29,** 153-160.

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: a customizable RNA-seq data analysis and visualization platform. Plant Physiology **170,** 2172-2186.

Brandolini, A., Hidalgo, A., Gabriele, S., and Heun, M. (2015). Chemical composition of wild and feral diploid wheats and their bearing on domesticated wheats. Journal of Cereal Science **63,** 122-127.

Brunazzi, A., Scaglione, D., Talini, R.F., Miculan, M., Magni, F., Poland, J., Enrico Pè, M., Brandolini, A., and Dell'Acqua, M. (2018). Molecular diversity and landscape genomics of the crop wild relative *Triticum urartu* across the fertile crescent. The Plant Journal **94,** 670-684.

Bürgy, L., Eicke, S., Kopp, C., Jenny, C., Lu, K.J., Escrig, S., Meibom, A., and Zeeman, S.C. (2021). Coalescence and directed anisotropic growth of starch granule initials in subdomains of *Arabidopsis thaliana* chloroplasts. Nature Communications **12,** 6944.

Burton, R.A., Zhang, X.-Q., Hrmova, M., and Fincher, G.B. (1999). A single *LIMIT DEXTRINASE* gene is expressed both in the developing endosperm and in germinated grains of barley. Plant Physiology **119,** 859-872.

Burton, R.A., Jenner, H., Carrangis, L., Fahy, B., Fincher, G.B., Hylton, C., Laurie, D.A., Parker, M., Waite, D., Van Wegen, S., Verhoeven, T., and Denyer, K. (2002). Starch granule initiation and growth are altered in barley mutants that lack *ISOAMYLASE* activity. The Plant Journal **31,** 97-112.

Busi, M.V., Palopoli, N., Valdez, H.A., Fornasari, M.S., Wayllace, N.Z., Gomez-Casati, D.F., Parisi, G., and Ugalde, R.A. (2008). Functional and structural characterization of the catalytic domain of the STARCH SYNTHASE III from *Arabidopsis thaliana*. Proteins: Structure, Function, and Bioinformatics **70,** 31-40.

Buttrose, M.S. (1960). Submicroscopic development and structure of starch granules in cereal endosperms. Journal of Ultrastructure Research **4,** 231-257.

Cabral, V.Á., Govoni, B., and Verli, H. (2024). Unravelling carbohydrate binding module 21 (CBM21) dynamics of interaction with amylose. Carbohydrate Polymers **330,** 121792.

Cao, H., Imparl-Radosevich, J., Guan, H., Keeling, P.L., James, M.G., and Myers, A.M. (1999). Identification of the soluble starch synthase activities of maize endosperm. Plant Physiology **120,** 205-216.

Cao, H., Yan, X., Chen, G., Zhou, J., Li, X., Ma, W., and Yan, Y. (2015). Comparative proteome analysis of A- and B-type starch granule-associated proteins in bread wheat (*Triticum aestivum* L.) and *Aegilops crassa*. Journal of Proteomics **112,** 95-112.

Cavalet-Giorsa, E., González-Muñoz, A., Athiyannan, N., Holden, S., Salhi, A., Gardener, C., Quiroz-Chávez, J., Rustamova, S.M., Elkot, A.F., Patpour, M., Rasheed, A., Mao, L., Lagudah, E.S., Periyannan, S.K., Sharon, A., Himmelbach, A., Reif, J.C., Knauft, M., Mascher, M., Stein, N., Chayut, N., Ghosh, S., Perovic, D., Putra, A., Perera, A.B., Hu, C.-Y., Yu, G., Ahmed, H.I., Laquai, K.D., Rivera, L.F., Chen, R., Wang, Y., Gao, X., Liu, S., Raupp, W.J., Olson, E.L., Lee, J.-Y., Chhuneja, P., Kaur, S., Zhang, P., Park, R.F., Ding, Y., Liu, D.-C., Li, W., Nasyrova, F.Y., Dvorak, J., Abbasi, M., Li, M., Kumar, N., Meyer, W.B., Boshoff, W.H.P., Steffenson, B.J., Matny, O., Sharma, P.K., Tiwari, V.K., Grewal, S., Pozniak, C.J., Chawla, H.S., Ens, J., Dunning, L.T., Kolmer, J.A., Lazo, G.R., Xu, S.S., Gu, Y.Q., Xu, X., Uauy, C., Abrouk, M., Bougouffa, S., Brar, G.S., Wulff, B.B.H., and Krattinger, S.G.

(2024). Origin and evolution of the bread wheat D genome. Nature **633**, 848-855.

Ceballos, H., Sánchez, T., Morante, N., Fregene, M., Dufour, D., Smith, A.M., Denyer, K., Pérez, J.C., Calle, F., and Mestres, C. (2007). Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). Journal of Agricultural and Food Chemistry **55,** 7469-7476.

Chen, J. (2022). Characterisation of SS4 and MRC in starch granule initiation (John Innes Centre. University of East Anglia, Norwich, UK.).

Chen, J., Hawkins, E., and Seung, D. (2021). Towards targeted starch modification in plants. Current Opinion in Plant Biology **60,** 102013.

Chen, J., Watson-Lazowski, A., Kamble, N.U., Vickers, M., and Seung, D. (2023a). Gene expression profile of the developing endosperm in durum wheat provides insight into starch biosynthesis. BMC Plant Biology **23,** 363.

Chen, J., Chen, Y., Watson-Lazowski, A., Hawkins, E., Barclay, J.E., Fahy, B., Denley Bowers, R., Corbin, K., Warren, F.J., Blennow, A., Uauy, C., and Seung, D. (2024). Wheat *MYOSIN-RESEMBLING CHLOROPLAST PROTEIN* controls B-type starch granule initiation timing during endosperm development. Plant Physiology **196**, 1980-1996.

Chen, J., Yu, R., Li, N., Deng, Z., Zhang, X., Zhao, Y., Qu, C., Yuan, Y., Pan, Z., Zhou, Y., Li, K., Wang, J., Chen, Z., Wang, X., Wang, X., He, S.-N., Dong, J., Deng, X.W., and Chen, H. (2023b). Amyloplast sedimentation repolarizes LAZYs to achieve gravity sensing in plants. Cell **186,** 4788-4802..

Chen, X., Fang, W., Ji, M., Xu, S., Jiang, Y., Song, S., Chen, G., Tian, J., and Deng, Z. (2019). Genome-wide association study of total starch and its components in common wheat. Euphytica **215,** 1-13.

Chen, Y., and Bao, J. (2016). Underlying mechanisms of zymographic diversity in STARCH SYNTHASE I and PULLULANASE in rice-developing endosperm. Journal of Agricultural and Food Chemistry **64,** 2030-2037.

Cheng, J., Khan, M.A., Qiu, W.M., Li, J., Zhou, H., Zhang, Q., Guo, W., Zhu, T., Peng, J., Sun, F., Li, S., Korban, S.S., and Han, Y. (2012). Diversification of genes encoding *GRANULE-BOUND STARCH SYNTHASE* in monocots and dicots is marked by multiple genome-wide duplication events. PLoS One **7,** e30088.

Chia, T., Adamski, N.M., Saccomanno, B., Greenland, A., Nash, A., Uauy, C., and Trafford, K. (2017). Transfer of a starch phenotype from wild wheat to bread wheat by deletion of a locus controlling B-type starch granule content. Journal of Experimental Botany **68,** 5497-5509.

Chia, T., Chirico, M., King, R., Ramirez-Gonzalez, R., Saccomanno, B., Seung, D., Simmonds, J., Trick, M., Uauy, C., Verhoeven, T., and Trafford, K. (2020). A carbohydrate-binding protein, B-GRANULE CONTENT 1, influences starch granule size distribution in a dose-dependent manner in polyploid wheat. Journal of Experimental Botany **71,** 105-115.

Chou, K.C., and Shen, H.B. (2007). Large-scale plant protein subcellular location prediction. Journal of Cellular Biochemistry **100,** 665-678.

Chou, K.C., and Shen, H.B. (2008). Cell-PLoc: a package of web servers for predicting subcellular localization of proteins in various organisms. Nature Protocols **3,** 153-162.

Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A., Chen, H., Frazer, K.A., Huson, D.H., Schölkopf, B., Nordborg, M., Rätsch, G., Ecker, J.R., and Weigel, D. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science **317,** 338-342.

Clarke, B.C., Hobbs, M., Skylas, D., and Appels, R. (2000). Genes active in developing wheat endosperm. Functional & Integrative Genomics **1,** 44-55.

Cozzolino, D. (2016). The use of the rapid visco analyser (RVA) in breeding and selection of cereals. Journal of Cereal Science **70,** 282-290.

Crumpton-Taylor, M., Grandison, S., Png, K.M., Bushby, A.J., and Smith, A.M. (2012). Control of starch granule numbers in *Arabidopsis* chloroplasts. Plant Physiology **158,** 905-916.

Crumpton-Taylor, M., Pike, M., Lu, K.J., Hylton, C.M., Feil, R., Eicke, S., Lunn, J.E., Zeeman, S.C., and Smith, A.M. (2013). *STARCH SYNTHASE 4* is essential for coordination of starch granule formation with chloroplast division during *Arabidopsis* leaf expansion. New Phytologist **200,** 1064-1075.

Cuesta-Seijo, J.A., Ruzanski, C., Krucewicz, K., Meier, S., Hägglund, P., Svensson, B., and Palcic, M.M. (2017). Functional and structural characterization of plastidic starch *PHOSPHORYLASE* during barley endosperm development. PLoS One **12,** e0175488.

Dai, Z., Yin, Y., and Wang, Z. (2009). Starch granule size distribution from seven wheat cultivars under different water regimes. Cereal Chemistry **86,** 82-87.

Darvasi, A., and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. Genetics **141,** 1199-1207.

Delatte, T., Trevisan, M., Parker, M.L., and Zeeman, S.C. (2005). *Arabidopsis* mutants *Atisa1* and *Atisa2* have identical phenotypes and lack the same

multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. The Plant Journal **41,** 815-830.

Delatte, T., Umhang, M., Trevisan, M., Eicke, S., Thorneycroft, D., Smith, S.M., and Zeeman, S.C. (2006). Evidence for distinct mechanisms of starch granule breakdown in plants. Journal of Biological Chemistry **281,** 12050-12059.

Delorean, E., Gao, L., Lopez, J.F.C., Mehrabi, A., Bentley, A., Sharon, A., Keller, B., Wulff, B., Steffenson, B., Steuernagel, B., Sansaloni, C.P., Liu, D.-C., Lagudah, E., Nasyrova, F., Brown-Guedira, G., Sela, H., Dvorak, J., Poland, J., Mayer, K., Krasileva, K., Gaurav, K., Mao, L., Caccamo, M., Mascher, M., Luo, M., Chhuneja, P., Davey, R., Faris, J., Xu, S., Nicholson, P., Chayut, N., Ambrose, M., Rawat, N., Tiwari, V.K., Wulff, B.B.H., Ibba, M.I., Poland, J., and Open Wild Wheat, C. (2021). *HIGH MOLECULAR WEIGHT GLUTENIN* gene diversity in *Aegilops tauschii* demonstrates unique origin of superior wheat quality. Communications Biology **4,** 1242.

Delvallé, D., Dumez, S., Wattebled, F., Roldán, I., Planchot, V., Berbezy, P., Colonna, P., Vyas, D., Chatterjee, M., Ball, S., Mérida, A., and D'Hulst, C. (2005). *SOLUBLE STARCH SYNTHASE I*: a major determinant for the synthesis of amylopectin in *Arabidopsis thaliana* leaves. The Plant Journal **43,** 398-412.

DeMartino, P., and Cockburn, D.W. (2020). Resistant starch: impact on the gut microbiome and health. Current Opinion in Biotechnology **61,** 66-71.

Deng, Z., Fang, W., Guo, X., Zhao, X., Guo, H., Hu, S., and Tian, J. (2018). Genetic dissection of interactions between wheat flour starch and its components in two populations using two QTL mapping methods. Molecular Breeding **38,** 41.

Deng, Z., Hu, S., Chen, F., Li, W., Chen, J., Sun, C., Zhang, Y., Wang, S., Song, X., and Tian, J. (2015). Genetic dissection of interaction between wheat protein and starch using three mapping populations. Molecular Breeding **35,** 12.

Dengate, H., and Meredith, P. (1984). Variation in size distribution of starch granules from wheat grain. Journal of Cereal Science **2,** 83-90.

Denver, K., Barber, L.M., Burton, R., Hedley, C.L., Hylton, C.M., Johnson, S., Jones, D.A., Marshall, J., Smith, A.M., Tatge, H., Tomlinson, K., and Wang, T.L. (1995). The isolation and characterization of novel low-amylose mutants of *Pisum sativum* L. Plant, Cell & Environment **18,** 1019-1026.

Denyer, K., Dunlap, F., Thorbjornsen, T., Keeling, P., and Smith, A.M. (1996). The major form of ADP-GLUCOSE PYROPHOSPHORYLASE in maize endosperm is extra-plastidial. Plant Physiology **112,** 779-785.

Dian, W., Jiang, H., and Wu, P. (2005). Evolution and expression analysis of *STARCH SYNTHASE III* and *IV* in rice. Journal of Experimental Botany **56,** 623-632.

Dinges, J.R., Colleoni, C., James, M.G., and Myers, A.M. (2003). Mutational analysis of the *PULLULANASE*-type debranching enzyme of maize indicates multiple functions in starch metabolism. The Plant Cell **15,** 666-680.

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2021). The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research **50,** D571-D577.

Drummond, G.S., Smith, E.E., and Whelan, W.J. (1972). Purification and properties of potato -1,4-glucan. -1,4-Glucan 6-glycosyltransferase (Q-enzyme). European Journal of Biochemistry **26,** 168-176.

Duan, H., Li, J., Sun, L., Xiong, X., Xu, S., Sun, Y., Ju, X., Xue, Z., Gao, J., Wang, Y., Xie, H., Ding, D., Zhang, X., and Tang, J. (2023). Identification of novel loci associated with starch content in maize kernels by a genome-wide association study using an enlarged SNP panel. Molecular Breeding **43,** 91.

Edwards, M.A., Osborne, B.G., and Henry, R.J. (2008). Effect of endosperm starch granule size distribution on milling yield in hard wheat. Journal of Cereal Science **48,** 180-192.

Esch, L., Ngai, Q.Y., Barclay, J.E., McNelly, R., Hayta, S., Smedley, M.A., Smith, A.M., and Seung, D. (2023). Increasing amyloplast size in wheat endosperm through mutation of *PARC6* affects starch granule morphology. New Phytologist **240,** 224-241.

Fahy, B., Gonzalez, O., Savva, G.M., Ahn-Jarvis, J.H., Warren, F.J., Dunn, J., Lovegrove, A., and Hazard, B.A. (2022). Loss of *STARCH SYNTHASE IIIA* changes starch molecular structure and granule morphology in grains of hexaploid bread wheat. Scientific Reports **12,** 10806.

Feike, D., Seung, D., Graf, A., Bischof, S., Ellick, T., Coiro, M., Soyk, S., Eicke, S., Mettler-Altmann, T., Lu, K.J., Trick, M., Zeeman, S.C., and Smith, A.M. (2016). The starch granule-associated protein EARLY STARVATION 1 is required for the control of starch degradation in *Arabidopsis thaliana* leaves. Plant Cell **28,** 1472-1489.

Feldman, M., and Levy, A.A. (2015). Origin and evolution of wheat and related triticeae species. In Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics, M. Molnár-Láng, C. Ceoloni, and J. Doležel, eds (Cham: Springer International Publishing), pp. 21-76.

Feng, N., He, Z., Zhang, Y., Xia, X., and Zhang, Y. (2013). QTL mapping of starch granule size in common wheat using recombinant inbred lines derived from a PH82-2/Neixiang 188 cross. The Crop Journal **1,** 166–171.

Figueroa, C.M., Asencion Diez, M.D., Ballicora, M.A., and Iglesias, A.A. (2022). Structure, function, and evolution of plant *ADP-GLUCOSE PYROPHOSPHORYLASE*. Plant Molecular Biology **108,** 307-323.

Fischer, R. (2022). History of wheat breeding: a personal view. Wheat Improvement: Food Security in a Changing Climate. Springer Nature: Switzerland**,** 17-30.

Flores-Castellanos, J., and Fettke, J. (2022). The plastidial glucan PHOSPHORYLASE affects the maltooligosaccharide metabolism in parenchyma cells of potato (*Solanum tuberosum* L.) tuber discs. Plant and Cell Physiology **64,** 422-432.

French, D. (1972). Fine structure of starch and its relationship to the organization of starch granules. Journal of the Japanese Society of Starch Science **19,** 8-25.

Fujita, N., Toyosawa, Y., Utsumi, Y., Higuchi, T., Hanashiro, I., Ikegami, A., Akuzawa, S., Yoshida, M., Mori, A., Inomata, K., Itoh, R., Miyao, A., Hirochika, H., Satoh, H., and Nakamura, Y. (2009). Characterization of *pullulanase* (*pul*)-deficient mutants of rice (*Oryza sativa* L.) and the function of *PUL* on starch biosynthesis in the developing rice endosperm. Journal of Experimental Botany **60,** 1009-1023.

Fuwa, H., Sugimoto, Y., Tanaka, M., and Glover, D. (1978). Susceptibility of various starch granules to amylases as seen by scanning electron microscope. Starch-Stärke **30,** 186-191.

Gahlaut, V., Zinta, G., Jaiswal, V., and Kumar, S. (2020). Quantitative epigenetics: a new avenue for crop improvement. Epigenomes **4,** 25.

Gámez-Arjona, F.M., and Mérida, Á. (2021). Interplay between the N-terminal domains of A*arabidopsis* STARCH SYNTHASE 3 determines the interaction of the enzyme with the starch granule. Frontiers in Plant Science **12,** 704161.

Gámez-Arjona, F.M., Raynaud, S., Ragel, P., and Mérida, A. (2014). STARCH SYNTHASE 4 is located in the thylakoid membrane and interacts with plastoglobule-associated proteins in *Arabidopsis*. The Plant Journal **80,** 305-316.

Gámez-Arjona, F.M., Li, J., Raynaud, S., Baroja-Fernández, E., Muñoz, F.J., Ovecka, M., Ragel, P., Bahaji, A., Pozueta-Romero, J., and Mérida, Á. (2011). Enhancing the expression of *STARCH SYNTHASE CLASS IV* results in increased levels of both transitory and long-term storage starch. Plant Biotechnology Journal **9,** 1049-1060.

Ganal, M.W., Wieseke, R., Luerssen, H., Durstewitz, G., Graner, E.-M., Plieske, J., and Polley, A. (2014). High-throughput SNP profiling of genetic resources in crop plants using genotyping arrays. In Genomics of Plant Genetic Resources: Volume 1. Managing, sequencing and mining genetic resources, R. Tuberosa, A. Graner, and E. Frison, eds (Dordrecht: Springer Netherlands), pp. 113-130.

Gao, M., Fisher, D.K., Kim, K.N., Shannon, J.C., and Guiltinan, M.J. (1997). Independent genetic control of maize *STARCH-BRANCHING ENZYMES IIA* and *IIB* (isolation and characterization of a *SBE2A* cDNA). Plant Physiology **114,** 69-78.

Gaurav, K., Arora, S., Silva, P., Sánchez-Martín, J., Horsnell, R., Gao, L., Brar, G.S., Widrig, V., John Raupp, W., Singh, N., Wu, S., Kale, S.M., Chinoy, C., Nicholson, P., Quiroz-Chávez, J., Simmonds, J., Hayta, S., Smedley, M.A., Harwood, W., Pearce, S., Gilbert, D., Kangara, N., Gardener, C., Forner-Martínez, M., Liu, J., Yu, G., Boden, S.A., Pascucci, A., Ghosh, S., Hafeez, A.N., O'Hara, T., Waites, J., Cheema, J., Steuernagel, B., Patpour, M., Justesen, A.F., Liu, S., Rudd, J.C., Avni, R., Sharon, A., Steiner, B., Kirana, R.P., Buerstmayr, H., Mehrabi, A.A., Nasyrova, F.Y., Chayut, N., Matny, O., Steffenson, B.J., Sandhu, N., Chhuneja, P., Lagudah, E., Elkot, A.F., Tyrrell, S., Bian, X., Davey, R.P., Simonsen, M., Schauser, L., Tiwari, V.K., Randy Kutcher, H., Hucl, P., Li, A., Liu, D.-C., Mao, L., Xu, S., Brown-Guedira, G., Faris, J., Dvorak, J., Luo, M.-C., Krasileva, K., Lux, T., Artmeier, S., Mayer, K.F.X., Uauy, C., Mascher, M., Bentley, A.R., Keller, B., Poland, J., and Wulff, B.B.H. (2021). Population genomic analysis of *Aegilops tauschii i*dentifies targets for bread wheat improvement. Nature Biotechnology **40,** 422-431.

Geddes, K., and Greenwood, C.T. (1969). Observations on the biosynthesis of the starch granule. Starch - Stärke **21,** 148-153.

Ghosh, H.P., and Preiss, J. (1966). ADENOSINE DIPHOSPHATE GLUCOSE PYROPHOSPHORYLASE: A regulatory enzyme in the biosynthesis of starch in spinach leaf chloroplasts. Journal of Biological Chemistry **241,** 4491-4504.

Gilding, E.K., Frere, C.H., Cruickshank, A., Rada, A.K., Prentis, P.J., Mudge, A.M., Mace, E.S., Jordan, D.R., and Godwin, I.D. (2013). Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. Nature communications **4,** 1483.

Glynn, J.M., Yang, Y., Vitha, S., Schmitz, A.J., Hemmes, M., Miyagishima, S.-y., and Osteryoung, K.W. (2009). PARC6, a novel chloroplast division factor,

influences FtsZ assembly and is required for recruitment of PDV1 during chloroplast division in *Arabidopsis*. The Plant Journal **59,** 700-711.

Goering, K.J., Fritts, D.H., and Eslick, R.F. (1973). A study of starch granule size and distribution in 29 barley varieties. Starch - Stärke **25,** 297-302.

Goldstein, A., Annor, G., Vamadevan, V., Tetlow, I., Kirkensgaard, J.J.K., Mortensen, K., Blennow, A., Hebelstrup, K.H., and Bertoft, E. (2017). Influence of diurnal photosynthetic activity on the morphology, structure, and thermal properties of normal and *waxy* barley starch. International Journal of Biological Macromolecules **98,** 188-200.

Gómez-Casati, D.F., and Iglesias, A.A. (2002). ADP-GLUCOSE PYROPHOSPHORYLASE from wheat endosperm. Purification and characterization of an enzyme with novel regulatory properties. Planta **214,** 428-434.

Goren, A., Ashlock, D., and Tetlow, I.J. (2018). Starch formation inside plastids of higher plants. Protoplasma **255,** 1855-1876.

Graf, A., Schlereth, A., Stitt, M., and Smith, A.M. (2010). Circadian control of carbohydrate availability for growth in *Arabidopsis* plants at night. Proceedings of the National Academy of Sciences **107,** 9458-9463.

Guan, H.P., and Preiss, J. (1993). Differentiation of the properties of the branching isozymes from maize (*Zea mays*). Plant Physiology **102,** 1269-1273.

Guérardel, Y., Leleu, D., Coppin, A., Liénard, L., Slomianny, C., Strecker, G., Ball, S., and Tomavo, S. (2005). Amylopectin biogenesis and characterization in the protozoan parasite *Toxoplasma gondii*, the intracellular development of which is restricted in the HepG2 cell line. Microbes and Infection **7,** 41-48.

Guo, H., Liu, Y., Li, X., Yan, Z., Xie, Y., Xiong, H., Zhao, L., Gu, J., Zhao, S., and Liu, L. (2017). Novel mutant alleles of the starch synthesis gene *TaSSIVb-D* result in the reduction of starch granule number per chloroplast in wheat. BMC Genomics **18,** 1-10.

Guo, Q., He, Z., Xia, X., Qu, Y., and Zhang, Y. (2014). Effects of wheat starch granule size distribution on qualities of chinese steamed bread and raw white noodles. Cereal Chemistry **91,** 623-630.

Guo, Y., Tian, J., Chen, J., Bao, Y., and Deng, Z. (2023). Genetic dissection of protein and starch during wheat grain development using QTL mapping and GWAS. Frontiers in Plant Science **14,** 1189887.

Gupta, P.K. (2021). Quantitative genetics: pan-genomes, SVs, and k-mers for GWAS. Trends in Genetics **37,** 868-871.

Guzmán, C., Ibba, M.I., Álvarez, J.B., Sissons, M., and Morris, C. (2022). Wheat quality. In Wheat Improvement: Food Security in a Changing Climate, M.P. Reynolds and H.-J. Braun, eds (Cham: Springer International Publishing), pp. 177-193.

Hädrich, N., Hendriks, J.H., Kötting, O., Arrivault, S., Feil, R., Zeeman, S.C., Gibon, Y., Schulze, W.X., Stitt, M., and Lunn, J.E. (2012). Mutagenesis of cysteine 81 prevents dimerization of the APS1 subunit of ADP-GLUCOSE PYROPHOSPHORYLASE and alters diurnal starch turnover in *Arabidopsis thaliana* leaves. The Plant Journal **70,** 231-242.

Hannah, L.C. (2007). Starch formation in the cereal endosperm. In Endosperm: Developmental and Molecular Biology, O.-A. Olsen, ed (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 179-193.

Haque, E., Shirasawa, K., Suematsu, K., Tabuchi, H., Isobe, S., and Tanaka, M. (2023). Polyploid GWAS reveals the basis of molecular marker development for complex breeding traits including starch content in the storage roots of sweet potato. Frontiers in Plant Science **14,** 1181909.

Hawkins, E., Chen, J., Watson-Lazowski, A., Ahn-Jarvis, J., Barclay, J.E., Fahy, B., Hartley, M., Warren, F.J., and Seung, D. (2021). *STARCH SYNTHASE 4* is required for normal starch granule initiation in amyloplasts of wheat endosperm. New Phytologist **230,** 2371-2386.

Hayakawa, K., Tanaka, K., Nakamura, T., Endo, S., and Hoshino, T. (1997). Quality characteristics of waxy hexaploid wheat (*Triticum aestivum* L.): properties of starch gelatinization and retrogradation. Cereal Chemistry **74,** 576-580.

Hazard, B., Zhang, X., Colasuonno, P., Uauy, C., Beckles, D.M., and Dubcovsky, J. (2012). Induced mutations in the *STARCH BRANCHING ENZYME II (SBEII)* genes increase amylose and resistant starch content in durum wheat. Crop Science **52,** 1754-1766.

He, F., Pasam, R., Shi, F., Kant, S., Keeble-Gagnere, G., Kay, P., Forrest, K., Fritz, A., Hucl, P., Wiebe, K., Knox, R., Cuthbert, R., Pozniak, C., Akhunova, A., Morrell, P.L., Davies, J.P., Webb, S.R., Spangenberg, G., Hayes, B., Daetwyler, H., Tibbits, J., Hayden, M., and Akhunov, E. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nature Genetics **51,** 896-904.

He, S., Hao, X., Wang, S., Zhou, W., Ma, Q., Lu, X., Chen, L., and Zhang, P. (2022). *STARCH SYNTHASE II* plays a crucial role in starch biosynthesis and the formation of multienzyme complexes in cassava storage roots. Journal of Experimental Botany **73,** 2540-2557.

Hendriks, J.H., Kolbe, A., Gibon, Y., Stitt, M., and Geigenberger, P. (2003). ADP-GLUCOSE PYROPHOSPHORYLASE is activated by posttranslational redox-modification in response to light and to sugars in leaves of *Arabidopsis* and other plant species. Plant Physiology **133,** 838-849.

Hennen-Bierwagen, T.A., Liu, F., Marsh, R.S., Kim, S., Gan, Q., Tetlow, I.J., Emes, M.J., James, M.G., and Myers, A.M. (2008). Starch biosynthetic enzymes from developing maize endosperm associate in multisubunit complexes. Plant Physiology **146,** 1892-1908.

Hollin, T., Jaroszewski, L., Stajich, J.E., Godzik, A., and Le Roch, K.G. (2021). Identification and phylogenetic analysis of RNA binding domain abundant in apicomplexans or RAP proteins. Microbial Genomics **7,** 000541.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor Nucleic Acids Research **35,** W585-W587.

Hou, J., Li, T., Wang, Y., Hao, C., Liu, H., and Zhang, X. (2017). *ADP-GLUCOSE PYROPHOSPHORYLASE* genes, associated with kernel weight, underwent selection during wheat domestication and breeding. Plant Biotechnology Journal **15,** 1533-1543.

Hovenkamp-Hermelink, J.H.M., Jacobsen, E., Ponstein, A.S., Visser, R.G.F., Vos-Scheperkeuter, G.H., Bijmolt, E.W., de Vries, J.N., Witholt, B., and Feenstra, W.J. (1987). Isolation of an amylose-free starch mutant of the potato (*Solanum tuberosum* L.). Theoretical and Applied Genetics **75,** 217-221.

Howard, T., Rejab, N.A., Griffiths, S., Leigh, F., Leverington-Waite, M., Simmonds, J., Uauy, C., and Trafford, K. (2011). Identification of a major QTL controlling the content of B-type starch granules in *Aegilops*. Journal of Experimental Botany **62,** 2217-2228.

Hu, S., Wang, M., Zhang, X., Chen, W., Song, X., Fu, X., Fang, H., Xu, J., Xiao, Y., Li, Y., Bai, G., Li, J., and Yang, X. (2021). Genetic basis of kernel starch content decoded in a maize multi-parent population. Plant Biotechnology Journal **19,** 2192-2205.

Huang, L., Tan, H., Zhang, C., Li, Q., and Liu, Q. (2021). Starch biosynthesis in cereal endosperms: An updated review over the last decade. Plant Communications **2,** 100237.

Huang, S., Sirikhachornkit, A., Faris, J.D., Su, X., Gill, B.S., Haselkorn, R., and Gornicki, P. (2002). Phylogenetic analysis of the *ACETYL-COA CARBOXYLASE* and *3-PHOSPHOGLYCERATE KINASE* loci in wheat and other grasses. Plant Molecular Biology **48,** 805-820.

Hussain, H., Mant, A., Seale, R., Zeeman, S., Hinchliffe, E., Edwards, A., Hylton, C., Bornemann, S., Smith, A.M., Martin, C., and Bustos, R. (2003). Three isoforms of ISOAMYLASE contribute different catalytic properties for the debranching of potato glucans. The Plant Cell **15,** 133-149.

Hwang, S.-K., Nishi, A., Satoh, H., and Okita, T.W. (2010). Rice endosperm-specific plastidial α-glucan PHOSPHORYLASE is important for synthesis of short-chain malto-oligosaccharides. Archives of Biochemistry and Biophysics **495,** 82-92.

Hylton, C., and Smith, A.M. (1992). The *rb* mutation of peas causes structural and regulatory changes in adp glucose pyrophosphorylase from developing embryos. Plant Physiology **99,** 1626-1634.

Igrejas, G., Faucher, B., Bertrand, D., Guibert, D., Leroy, P., and Branlard, G. (2002). Genetic analysis of the size of endosperm starch granules in a mapped segregating wheat population. Journal of Cereal Science **35,** 103-107.

Iwaki, K., and Fuwa, H. (1981). Purification and some properties of debranching enzyme of germinating rice endosperm. Agricultural and Biological Chemistry **45,** 2683-2688.

Jacobsen, E., Hovenkamp-Hermelink, J.H.M., Krijgsheld, H.T., Nijdam, H., Pijnacker, L.P., Witholt, B., and Feenstra, W.J. (1989). Phenotypic and genotypic characterization of an amylose-free starch mutant of the potato. Euphytica **44,** 43-48.

Jaiswal, S., Båga, M., Ahuja, G., Rossnagel, B.G., and Chibbar, R.N. (2014). Development of barley (*Hordeum vulgare* L.) lines with altered starch granule size distribution. Journal of Agricultural and Food Chemistry **62,** 2289-2296.

Jamann, T.M., Balint-Kurti, P.J., and Holland, J.B. (2015). QTL mapping using high-throughput sequencing. Plant Functional Genomics: Methods and Protocols **1284,** 257-285.

James, M.G., Robertson, D.S., and Myers, A.M. (1995). Characterization of the maize gene *sugary1*, a determinant of starch composition in kernels. Plant Cell **7,** 417-429.

Jane, J.-l., and Shen, J.J. (1993). Internal structure of the potato starch granule revealed by chemical gelatinization. Carbohydrate Research **247,** 279-290.

Jane, J.-L., Kasemsuwan, T., Leas, S., Zobel, H., and Robyt, J.F. (1994). Anthology of starch granule morphology by scanning electron microscopy. Starch - Stärke **46,** 121-129.

Jane, J., Xu, A., Radosavljevic, M., and Seib, P. (1992). Location of amylose in normal starch granules. I. Susceptibility of Amylose and Amylopectin to Cross-Linking Reagents **69,** 405-409.

Janeček, Š., Svensson, B., and MacGregor, E.A. (2011). Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. Enzyme and Microbial Technology **49,** 429-440.

Jenkins, P.J., and Donald, A.M. (1995). The influence of amylose on starch granule structure. International Journal of Biological Macromolecules **17,** 315-321.

Jeong, S.Y., Rose, A., and Meier, I. (2003). MFP1 is a thylakoid-associated, nucleoid-binding protein with a coiled-coil structure. Nucleic Acids Research **31,** 5175-5185.

Johnson, P.E., Patron, N.J., Bottrill, A.R., Dinges, J.R., Fahy, B.F., Parker, M.L., Waite, D.N., and Denyer, K. (2003). A low-starch barley mutant, *Risø 16*, lacking the cytosolic small subunit of ADP-GLUCOSE PYROPHOSPHORYLASE, reveals the importance of the cytosolic isoform and the identity of the plastidial small subunit. Plant Physiology **131,** 684-696.

Jung, Y.-J., Nogoy, F.M., Lee, S.-K., Cho, Y.-G., and Kang, K.-K. (2018). Application of ZFN for site directed mutagenesis of rice *SSIVa* gene. Biotechnology and Bioprocess Engineering **23,** 108-115.

Kamble, N.U., Makhamadjonov, F., Fahy, B., Martins, C., Saalbach, G., and Seung, D. (2023). Initiation of B-type starch granules in wheat endosperm requires the plastidial α-glucan phosphorylase *PHS1*. The Plant Cell **35,** 4091-4110.

Kang, G.Z., Wang, Y.H., Liu, C., Shen, B.Q., Zheng, B.B., Feng, W., and Guo, T.C. (2010). Difference in AGPase subunits could be associated with starch accumulation in grains between two wheat cultivars. Plant Growth Regulation **61,** 61-66.

Karikari, B., Lemay, M.-A., and Belzile, F. (2023). K-mer-based genome-wide association studies in plants: advances, challenges, and perspectives. Genes **14,** 1439.

Kaundal, R., and Raghava, G.P.S. (2009). RSLpred: an integrative system for predicting subcellular localization of rice proteins combining compositional and evolutionary information. Proteomics **9,** 2324-2342.

Kavakli, I.H., Park, J.S., Slattery, C.J., Salamone, P.R., Frohlick, J., and Okita, T.W. (2001). Analysis of allosteric effector binding sites of potato ADP-GLUCOSE PYROPHOSPHORYLASE through reverse genetics. Journal of Biological Chemistry **276,** 40834-40840.

Kawagoe, Y. (2013). The characteristic polyhedral, sharp-edged shape of compound-type starch granules in rice endosperm is achieved via the septum-like structure of the amyloplast. Journal of Applied Glycoscience **60,** 29-36.

Khlestkin, V.K., Erst, T.V., Rozanova, I.V., Efimov, V.M., and Khlestkina, E.K. (2020). Genetic loci determining potato starch yield and granule morphology revealed by genome-wide association study (GWAS). PeerJ **8,** e10286.

Kirana, R.P., Gaurav, K., Arora, S., Wiesenberger, G., Doppler, M., Michel, S., Zimmerl, S., Matic, M., Eze, C.E., Kumar, M., Topuz, A., Lemmens, M., Schuhmacher, R., Adam, G., Wulff, B.B.H., Buerstmayr, H., and Steiner, B. (2023). Identification of a UDP-GLUCOSYLTRANSFERASE conferring deoxynivalenol resistance in *Aegilops tauschii* and wheat. Plant Biotechnology Journal **21,** 109-121.

Kiszonas, A.M., and Morris, C.F. (2018). Wheat breeding for quality: A historical review. Cereal Chemistry **95,** 17-34.

Kleczkowski, L.A., Villand, P., Luthi, E., Olsen, O.A., and Preiss, J. (1993). Insensitivity of barley endosperm ADP-GLUCOSE PYROPHOSPHORYLASE to 3-phosphoglycerate and orthophosphate regulation. Plant Physiology **101,** 179-186.

Kleinknecht, L., Wang, F., Stübe, R., Philippar, K., Nickelsen, J., and Bohne, A.-V. (2014). RAP, the sole octotricopeptide repeat protein in *Arabidopsis*, is required for chloroplast 16S rRNA maturation. The Plant Cell **26,** 777-787.

Knight, M.E., Harn, C., Lilley, C.E., Guan, H., Singletary, G.W., Mu-Forster, C., Wasserman, B.P., and Keeling, P.L. (1998). Molecular cloning of *STARCH SYNTHASE I* from maize (W64) endosperm and expression in *Escherichia coli*. The Plant Journal **14,** 613-622.

Konik-Rose, C.M., Rahman, S., Appels, R., Moss, R., McMaster, G., Marshall, D.R., and Stoddard, F.L. (2009). Starch characterisation and variability in *GBSS* loci of synthetic hexaploid wheats and their durum and *Aegilops tauschii* parents. Euphytica **167,** 203-216.

Kossmann, J., Abel, G.J., Springer, F., Lloyd, J.R., and Willmitzer, L. (1999). Cloning and functional analysis of a cDNA encoding a starch synthase from potato (*Solanum tuberosum* L.) that is predominantly expressed in leaf tissue. Planta **208,** 503-511.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X., Borrill, P., Fosker, C., Ayling, S., Phillips, A.L., Uauy, C., and Dubcovsky, J. (2017). Uncovering hidden variation in polyploid wheat. Proceedings of the National Academy of Sciences **114,** E913-E921.

Kremling, K.A.G., Diepenbrock, C.H., Gore, M.A., Buckler, E.S., and Bandillo, N.B. (2019). Transcriptome-wide association supplements genome-wide association in *Zea mays*. G3 Genes|Genomes|Genetics **9,** 3023-3033.

Kubo, A., Colleoni, C., Dinges, J.R., Lin, Q., Lappe, R.R., Rivenbark, J.G., Meyer, A.J., Ball, S.G., James, M.G., Hennen-Bierwagen, T.A., and Myers, A.M. (2010). Functions of heteromeric and homomeric isoamylase-type starch-debranching enzymes in developing maize endosperm. Plant Physiology **153,** 956-969.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution **35,** 1547-1549.

Langeveld, S.M.J., van Wijk, R., Stuurman, N., Kijne, J.W., and de Pater, S. (2000). B-type granule containing protrusions and interconnections between amyloplasts in developing wheat endosperm revealed by transmission electron microscopy and *GFP* expression. Journal of Experimental Botany **51,** 1357-1361.

Lee, S.K., Eom, J.S., Hwang, S.K., Shin, D., An, G., Okita, T.W., and Jeon, J.S. (2016). Plastidic *phosphoglucomutase* and *adp-glucose pyrophosphorylase* mutants impair starch synthesis in rice pollen grains and cause male sterility. Journal of Experimental Botany **67,** 5557-5569.

Li, A., Liu, D., Yang, W., Kishii, M., and Mao, L. (2018a). Synthetic hexaploid wheat: yesterday, today, and tomorrow. Engineering **4,** 552-558.

Li, C., Huang, Y., Huang, R., Wu, Y., and Wang, W. (2018b). The genetic architecture of amylose biosynthesis in maize kernel. Plant Biotechnology Journal **16,** 688-695.

Li, C., Zhou, D., Fan, T., Wang, M., Zhu, M., Ding, J., Zhu, X., Guo, W., and Shi, Y.-C. (2020). Structure and physicochemical properties of two waxy wheat starches. Food Chemistry **318,** 126492.

Li, C.Y., Li, C., Lu, Z.X., Li, W.H., and Cao, L.P. (2012). Morphological changes of starch granules during grain filling and seed germination in wheat. Starch - Stärke **64,** 166-170.

Li, J., Rasheed, A., Guo, Q., Dong, Y., Liu, J., Xia, X., Zhang, Y., and He, Z. (2017). Genome-wide association mapping of starch granule size distribution in common wheat. Journal of Cereal Science **77,** 211-218.

Li, M., Daygon, V.D., Solah, V., and Dhital, S. (2023). Starch granule size: Does it matter? Critical Reviews in Food Science and Nutrition **63,** 3683-3703.

Li, M., Geng, L., Xie, S., Wu, D., Ye, L., and Zhang, G. (2021). Genome-wide association study on total starch, amylose and amylopectin in barley grain reveals novel putative alleles. International Journal of Molecular Sciences **22,** 553.

Li, W.-y., Yan, S.-h., Yin, Y.-p., Li, Y., Liang, T.-b., Gu, F., Dai, Z.-m., and Wang, Z.-l. (2008). Comparison of starch granule size distribution between hard and soft wheat cultivars in eastern China. Agricultural Sciences in China **7,** 907-914.

Li, W., Fu, B.-B., Li, Z., Liu, Y.-X., Pu, Z.-E., Qi, P.-F., Jiang, Q.-T., Chen, G.-Y., Wang, J.-R., Wei, Y.-M., and Zheng, Y.-L. (2016). Characterization of the *waxy* gene in diploid *Triticum* L. and *Aegilops* L. species and its geographic distribution. Genetic Resources and Crop Evolution **63,** 987-1002.

Li, Z., Mouille, G., Kosar-Hashemi, B., Rahman, S., Clarke, B., Gale, K.R., Appels, R., and Morell, M.K. (2000). The structure and expression of the wheat *STARCH SYNTHASE III* gene: motifs in the expressed gene define the lineage of the *STARCH SYNTHASE III* gene family. Plant Physiology **123,** 613-624.

Lin, Y., Wang, Q., Chen, H., Yan, N., Wu, F., Wang, Z., Li, C., and Liu, Y. (2022). Genome-wide association mapping of *Fusarium* crown rot resistance in *Aegilops tauschii*. Frontiers in Plant Science **13,** 998622.

Liu, H., Si, X., Wang, Z., Cao, L., Gao, L., Zhou, X., Wang, W., Wang, K., Jiao, C., Zhuang, L., Liu, Y., Hou, J., Li, T., Hao, C., Guo, W., Liu, J., and Zhang, X. (2023). TaTPP-7A positively feedback regulates grain filling and wheat grain yield through T6P-SnRK1 signalling pathway and sugar–ABA interaction. Plant Biotechnology Journal **21,** 1159-1175.

Liu, N., Xue, Y., Guo, Z., Li, W., and Tang, J. (2016). Genome-wide association study identifies candidate genes for starch content regulation in maize kernels. Frontiers in Plant Science **7,** 1046.

Liu, N., Zhang, Z., Xue, Y., Meng, S., Huang, Y., Li, W., Huang, J., and Tang, J. (2018). Identification of quantitative trait loci and candidate genes for maize starch granule size through association mapping. Scientific Reports **8,** 14236.

Liu, Q., Zhou, Y., and Fettke, J. (2021a). Starch granule size and morphology of *Arabidopsis thaliana* starch-related mutants analyzed during diurnal rhythm and development. Molecules **26,** 5859.

Liu, Q., Li, X., and Fettke, J. (2021b). Starch granules in *Arabidopsis thaliana* mesophyll and guard cells show similar morphology but differences in size and number. International Journal of Molecular Sciences **22,** 5666.

Liu, Y., Wang, L., Mao, S., Liu, K., Lu, Y., Wang, J., Wei, Y., and Zheng, Y. (2015a). Genome-wide association study of 29 morphological traits in *Aegilops tauschii*. Scientific Reports **5,** 15562.

Liu, Y., Wang, L., Deng, M., Li, Z., Lu, Y., Wang, J., Wei, Y., and Zheng, Y. (2015b). Genome-wide association study of phosphorus-deficiency-tolerance traits in *Aegilops tauschii*. Theoretical and Applied Genetics **128,** 2203-2212.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research **42,** D490-D495.

Lopez-Rubio, A., Flanagan, B.M., Gilbert, E.P., and Gidley, M.J. (2008). A novel approach for calculating starch crystallinity and its correlation with double helix content: A combined XRD and NMR study. Biopolymers: Original Research on Biomolecules **89,** 761-768.

Lou, H., Zhang, R., Liu, Y., Guo, D., Zhai, S., Chen, A., Zhang, Y., Xie, C., You, M., Peng, H., Liang, R., Ni, Z., Sun, Q., and Li, B. (2021). Genome-wide association study of six quality-related traits in common wheat (*Triticum aestivum* L.) under two sowing conditions. Theoretical and Applied Genetics **134,** 399-418.

Lu, K.-J., Pfister, B., Jenny, C., Eicke, S., and Zeeman, S.C. (2017). Distinct functions of STARCH SYNTHASE 4 domains in starch granule formation. Plant Physiology **176,** 566-581.

Luo, M.-C., Gu, Y.Q., You, F.M., Deal, K.R., Ma, Y., Hu, Y., Huo, N., Wang, Y., Wang, J., Chen, S., Jorgensen, C.M., Zhang, Y., McGuire, P.E., Pasternak, S., Stein, J.C., Ware, D., Kramer, M., McCombie, W.R., Kianian, S.F., Martis, M.M., Mayer, K.F.X., Sehgal, S.K., Li, W., Gill, B.S., Bevan, M.W., Šimková, H., Doležel, J., Weining, S., Lazo, G.R., Anderson, O.D., and Dvorak, J. (2013). A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. Proceedings of the National Academy of Sciences **110,** 7940-7945.

Luo, M.-C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T., Wang, L., Wang, Y., McGuire, P.E., Liu, S., Long, H., Ramasamy, R.K., Rodriguez, J.C., Van, S.L., Yuan, L., Wang, Z., Xia, Z., Xiao, L., Anderson, O.D., Ouyang, S., Liang, Y., Zimin, A.V., Pertea, G., Qi, P., Bennetzen, J.L., Dai, X., Dawson, M.W., Müller, H.-G., Kugler, K., Rivarola-Duarte, L., Spannagl, M., Mayer, K.F.X., Lu, F.-H., Bevan, M.W., Leroy, P., Li, P., You, F.M., Sun, Q., Liu, Z., Lyons, E., Wicker, T., Salzberg, S.L., Devos, K.M.,

and Dvořák, J. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature **551,** 498-502.

Luo, M., Deal, K., Akhunov, E., Akhunova, A., Anderson, O., Anderson, J., Blake, N., Clegg, M., Coleman-Derr, D., and Conley, E. (2009). Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. Proceedings of the National Academy of Sciences **106,** 15780-15785.

MacGregor, A.W., Macri, L.J., Schroeder, S.W., and Bazin, S.L. (1994). Purification and characterisation of LIMIT DEXTRINASE INHIBITORS from barley. Journal of Cereal Science **20,** 33-41.

Mahjoob, M.M.M., Kamal, N.M., Gorafi, Y.S.A., and Tsujimoto, H. (2022). Genome-wide association study reveals distinct genetic associations related to leaf hair density in two lineages of wheat-wild relative *Aegilops tauschii*. Scientific Reports **12,** 17486.

Mai, K.K.K., Yeung, W.-T., Han, S.-Y., Cai, X., Hwang, I., and Kang, B.-H. (2019). Electron tomography analysis of thylakoid assembly and fission in chloroplasts of a single-cell C4 plant, *Bienertia sinuspersici*. Scientific Reports **9,** 19640.

Malinova, I., and Fettke, J. (2017). Reduced starch granule number per chloroplast in the *dpe2/phs1* mutant is dependent on initiation of starch degradation. PLoS One **12,** e0187985.

Malinova, I., Qasim, H.M., Brust, H., and Fettke, J. (2018). Parameters of starch granule genesis in chloroplasts of *Arabidopsis thaliana*. Frontiers in Plant Science **9,** 761.

Malinova, I., Mahlow, S., Alseekh, S., Orawetz, T., Fernie, A.R., Baumann, O., Steup, M., and Fettke, J. (2013). Double knockout mutants of *Arabidopsis* grown under normal conditions reveal that the plastidial *PHOSPHORYLASE* isozyme participates in transitory starch metabolism. Plant Physiology **164,** 907-921.

Malinova, I., Alseekh, S., Feil, R., Fernie, A.R., Baumann, O., Schöttler, M.A., Lunn, J.E., and Fettke, J. (2017). *STARCH SYNTHASE 4* and *PLASTIDAL PHOSPHORYLASE* differentially affect starch granule number and morphology. Plant Physiology **174,** 73-85.

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

Matsushima, R., and Hisano, H. (2019). Imaging amyloplasts in the developing endosperm of barley and rice. Scientific Reports **9,** 3745.

Matsushima, R., Maekawa, M., and Sakamoto, W. (2015). Geometrical formation of compound starch grains in rice implements Voronoi diagram. Plant and Cell Physiology **56,** 2150-2157.

Matsushima, R., Yamashita, J., Kariyama, S., Enomoto, T., and Sakamoto, W. (2013). A phylogenetic re-evaluation of morphological variations of starch grains among Poaceae species. Journal of Applied Glycoscience **60,** 37-44.

McLauchlan, A., Ogbonnaya, F.C., Hollingsworth, B., Carter, M., Gale, K.R., Henry, R.J., Holton, T.A., Morell, M.K., Rampling, L.R., Sharp, P.J., Shariflou, M.R., Jones, M.G.K., and Appels, R. (2001). Development of robust PCR-based DNA markers for each homoeo-allele of *GRANULE-BOUND STARCH SYNTHASE* and their application in wheat breeding programs. Australian Journal of Agricultural Research **52,** 1409-1416.

Mérida, A., and Fettke, J. (2021). Starch granule initiation in *Arabidopsis thaliana* chloroplasts. The Plant Journal **107,** 688-697.

Mingo-Castel, A.M., Pelacho, A.M., and de Felipe, M.R. (1991). Amyloplast division in kinetin induced potato tubers. Plant Science **73,** 211-217.

Minic, Z. (2008). Physiological roles of plant glycoside hydrolases. Planta **227,** 723-740.

Mishra, A., Sharma, V., Rahim, M.S., Sonah, H., Pal, D., Mantri, S., Sharma, T.R., and Roy, J. (2021). Genotyping-by-sequencing based QTL mapping identified a novel *waxy* allele contributing to high amylose starch in wheat. Euphytica **217,** 131.

Mohammadi, M., Xavier, A., Beckett, T., Beyer, S., Chen, L., Chikssa, H., Cross, V., Freitas Moreira, F., French, E., Gaire, R., Griebel, S., Lopez, M.A., Prather, S., Russell, B., and Wang, W. (2020). Identification, deployment, and transferability of quantitative trait loci from genome-wide association studies in plants. Current Plant Biology **24,** 100145.

Mohammadkhani, A., Stoddard, F.L., and Marshall, D.R. (1998). Survey of amylose content in Secale cereale, *Triticum monococcum*, *T. turgidum* and *T. tauschii.* Journal of Cereal Science **28,** 273-280.

Møller, M.S., Henriksen, A., and Svensson, B. (2016). Structure and function of α-glucan debranching enzymes. Cellular and Molecular Life Sciences **73,** 2619-2641.

Møller, M.S., Abou Hachem, M., Svensson, B., and Henriksen, A. (2012). Structure of the starch-debranching enzyme barley LIMIT DEXTRINASE reveals

homology of the N-terminal domain to CBM21. Acta Crystallographica Section F: Structural Biology and Crystallization Communications **68,** 1008-1012.

Møller, M.S., Vester-Christensen, M.B., Jensen, J.M., Hachem, M.A., Henriksen, A., and Svensson, B. (2015a). Crystal structure of barley LIMIT DEXTRINASE-LIMIT DEXTRINASE INHIBITOR (LD-LDI) complex reveals insights into mechanism and diversity of cereal type inhibitors. Journal of Biological Chemistry **290,** 12614-12629.

Møller, M.S., Windahl, M.S., Sim, L., Bøjstrup, M., Abou Hachem, M., Hindsgaul, O., Palcic, M., Svensson, B., and Henriksen, A. (2015b). Oligosaccharide and substrate binding in the starch debranching enzyme barley LIMIT DEXTRINASE. Journal of Molecular Biology **427,** 1263-1277.

Montaño-Leyva, B., Torres-Chávez, P., Ramírez-Wong, B., Plascencia-Jatomea, M., and Brown-Bojórquez, F. (2008). Physical and mechanical properties of durum wheat (*Triticum durum*) starch films prepared with A- and B-type granules. Starch - Stärke **60,** 559-567.

Morell, M.K., Blennow, A., Kosar-Hashemi, B., and Samuel, M.S. (1997). Differential expression and properties of starch branching enzyme isoforms in developing wheat endosperm. Plant Physiology **113,** 201-208.

Morell, M.K., Kosar-Hashemi, B., Cmiel, M., Samuel, M.S., Chandler, P., Rahman, S., Buleon, A., Batey, I.L., and Li, Z. (2003). Barley *sex6* mutants lack STARCH SYNTHASE IIA activity and contain a starch with novel properties. The Plant Journal **34,** 173-185.

Morrison, W.R., and Gadan, H. (1987). The amylose and lipid contents of starch granules in developing wheat endosperm. Journal of Cereal Science **5,** 263-275.

Muehlbauer, G.J., Specht, J.E., Thomas-Compton, M.A., Staswick, P.E., and Bernard, R.L. (1988). Near-isogenic lines—a potential resource in the integration of conventional and molecular marker linkage maps. Crop Science **28,** 729-735.

Muqaddasi, Q.H., Brassac, J., Ebmeyer, E., Kollers, S., Korzun, V., Argillier, O., Stiewe, G., Plieske, J., Ganal, M.W., and Röder, M.S. (2020). Prospects of GWAS and predictive breeding for European winter wheat's grain protein content, grain starch content, and grain hardness. Scientific Reports **10,** 12541.

Nakamura, T., Yamamori, M., Hirano, H., Hidaka, S., and Nagamine, T. (1995). Production of waxy (amylose-free) wheats. Molecular and General Genetics **248,** 253-259.

Nakamura, Y., and Kainuma, K. (2022). On the cluster structure of amylopectin. Plant Molecular Biology **108,** 291-306.

Nakamura, Y., Umemoto, T., Ogata, N., Kuboki, Y., Yano, M., and Sasaki, T. (1996). *STARCH DEBRANCHING ENZYME* (*R-ENZYME* or *PULLULANASE*) from developing rice endosperm: purification, cDNA and chromosomal localization of the gene. Planta **199,** 209-218.

Németh, R., and Tömösközi, S. (2021). Rye: Current state and future trends in research and applications. Acta Alimentaria **50,** 620-640.

Nie, H., Park, H., Kim, S., Kim, D., Kim, S., Kwon, S.Y., and Kim, S.H. (2023). Genetic diversity assessment and genome-wide association study reveal candidate genes associated with component traits in sweet potato (*Ipomoea batatas* (L.) Lam). Molecular Genetics and Genomics **298,** 653-667.

Nielsen, T.H., Krapp, A., Röper-Schwarz, U., and Stitt, M. (1998). The sugar-mediated regulation of genes encoding the small subunit of Rubisco and the regulatory subunit of ADP-glucose pyrophosphorylase is modified by phosphate and nitrogen. Plant, Cell & Environment **21,** 443-454.

Nikuni, Z. (1969). Starch and cooking. Journal of Cookery Science of Japan **2,** 6-14.

Nishi, A., Nakamura, Y., Tanaka, N., and Satoh, H. (2001). Biochemical and genetic analysis of the effects of *amylose-extender* mutation in rice endosperm. Plant Physiology **127,** 459-472.

Oliveira, A.B., Rasmusson, D.C., and Fulcher, R.G. (1994). Genetic aspects of starch granule traits in barley. Crop Science **34,** 1176-1180.

Orawetz, T., Malinova, I., Orzechowski, S., and Fettke, J. (2016). Reduction of the *PLASTIDIAL PHOSPHORYLASE* in potato (*Solanum tuberosum* L.) reveals impact on storage starch structure during growth at low temperature. Plant Physiology and Biochemistry **100,** 141-149.

Pan, D.D., and Jane, J.I. (2000). Internal structure of normal maize starch granules revealed by chemical surface gelatinization. Biomacromolecules **1,** 126-132.

Park, S.-H., Chung, O.K., and Seib, P.A. (2005). Effects of varying weight ratios of large and small wheat starch granules on experimental straight-dough bread. Cereal Chemistry **82,** 166-172.

Park, S.-H., Wilson, J.D., and Seabourn, B.W. (2009). Starch granule size distribution of hard red winter and hard red spring wheat: Its effects on mixing and breadmaking quality. Journal of Cereal Science **49,** 98-105.

Pasam, R.K., Sharma, R., Malosetti, M., van Eeuwijk, F.A., Haseneyer, G., Kilian, B., and Graner, A. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC plant biology **12,** 16.

Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E., and Tanksley, S.D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature **335,** 721-726.

Patron, N.J., Greber, B., Fahy, B.F., Laurie, D.A., Parker, M.L., and Denyer, K. (2004). The *lys5* mutations of barley reveal the nature and importance of plastidial ADP-Glc transporters for starch synthesis in cereal endosperm. Plant Physiology **135,** 2088-2097.

Patron, N.J., Smith, A.M., Fahy, B.F., Hylton, C.M., Naldrett, M.J., Rossnagel, B.G., and Denyer, K. (2002). The altered pattern of amylose accumulation in the endosperm of low-amylose barley cultivars is attributable to a single mutant allele of *GRANULE-BOUND STARCH SYNTHASE I* with a deletion in the 5′-non-coding region. Plant Physiology **130,** 190-198.

Peng, C., Wang, Y., Liu, F., Ren, Y., Zhou, K., Lv, J., Zheng, M., Zhao, S., Zhang, L., Wang, C., Jiang, L., Zhang, X., Guo, X., Bao, Y., and Wan, J. (2014). *FLOURY ENDOSPERM6* encodes a CBM48 domain-containing protein involved in compound granule formation and starch synthesis in rice endosperm. The Plant Journal **77,** 917-930.

Peng, M., Gao, M., Abdel-Aal, E.S., Hucl, P., and Chibbar, R. (1999). Separation and characterization of A-and B-type starch granules in wheat endosperm. Cereal Chemistry **76,** 375-379.

Pérez, S., Baldwin, P.M., and Gallant, D.J. (2009). Chapter 5 - Structural features of starch granules I. In Starch (Third Edition), J. BeMiller and R. Whistler, eds (San Diego: Academic Press), pp. 149-192.

Pfister, B., and Zeeman, S.C. (2016). Formation of starch in plant cells. Cellular and molecular life sciences **73,** 2781-2807.

Pilling, E., and Smith, A.M. (2003). Growth ring formation in the starch granules of potato tubers. Plant Physiology **132,** 365-371.

Plaxton, W.C., and Preiss, J. (1987). Purification and properties of nonproteolytic degraded adpglucose pyrophosphorylase from maize endosperm. Plant Physiology **83,** 105-112.

Pollard, D.A. (2012). Design and construction of recombinant inbred lines. In Quantitative Trait Loci (QTL): Methods and Protocols, S.A. Rifkin, ed (Totowa, NJ: Humana Press), pp. 31-39.

Praphasanobol, P., Purnama, P.R., Junbuathong, S., Chotechuen, S., Moung-Ngam, P., Kasettranan, W., Paliyavuth, C., Comai, L., Pongpanich, M.,

Buaboocha, T., and Chadchawan, S. (2023). Genome-wide association study of starch properties in local Thai rice. Plants (Basel) **12**.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics **155,** 945-959.

Qin, P., Wang, L., Liu, K., Mao, S., Li, Z., Gao, S., Shi, H., and Liu, Y. (2015). Genome wide association study of *Aegilops tauschii* traits under seedling-stage cadmium stress. The Crop Journal **3,** 405-415.

Qin, P., Lin, Y., Hu, Y., Liu, K., Mao, S., Li, Z., Wang, J., Liu, Y., Wei, Y., and Zheng, Y. (2016). Genome-wide association study of drought-related resistance traits in *Aegilops tauschii*. Genetics and Molecular Biology **39,** 398-407.

Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S.J., Cory, A.T., Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout, C.J., Chalhoub, B., Mayer, K.F.X., Benhamed, M., Latrasse, D., Bendahmane, A., Wulff, B.B.H., Appels, R., Tiwari, V., Datla, R., Choulet, F., Pozniak, C.J., Provart, N.J., Sharpe, A.G., Paux, E., Spannagl, M., Bräutigam, A., and Uauy, C. (2018). The transcriptional landscape of polyploid wheat. Science **361,** eaar6089.

Ravel, C., Praud, S., Murigneux, A., Canaguier, A., Sapet, F., Samson, D., Balfourier, F., Dufour, P., Chalhoub, B., and Brunel, D. (2006). Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome **49,** 1131-1139.

Regina, A., Kosar-Hashemi, B., Ling, S., Li, Z., Rahman, S., and Morell, M. (2010). Control of starch branching in barley defined through differential RNAi suppression of *STARCH BRANCHING ENZYME IIA* and *IIB*. Journal of Experimental Botany **61,** 1469-1482.

Regina, A., Kosar-Hashemi, B., Li, Z., Pedler, A., Mukai, Y., Yamamoto, M., Gale, K., Sharp, P.J., Morell, M.K., and Rahman, S. (2005). *STARCH BRANCHING ENZYME IIB* in wheat is expressed at low levels in the endosperm compared to other cereals and encoded at a non-syntenic locus. Planta **222,** 899-909.

Regina, A., Berbezy, P., Kosar-Hashemi, B., Li, S., Cmiel, M., Larroque, O., Bird, A.R., Swain, S.M., Cavanagh, C., Jobling, S.A., Li, Z., and Morell, M. (2015). A genetic strategy generating wheat with very high amylose content. Plant Biotechnology Journal **13,** 1276-1286.

Reif, J.C., Gowda, M., Maurer, H.P., Longin, C.F.H., Korzun, V., Ebmeyer, E., Bothe, R., Pietsch, C., and Würschum, T. (2011). Association mapping for

quality traits in soft winter wheat. Theoretical and Applied Genetics **122,** 961-970.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research **42,** W320-W324.

Roldán, I., Wattebled, F., Mercedes Lucas, M., Delvallé, D., Planchot, V., Jiménez, S., Pérez, R., Ball, S., D'Hulst, C., and Mérida, A. (2007). The phenotype of soluble *starch synthase IV* defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. The Plant Journal **49,** 492-504.

Ross, H.A., Sungurtas, J., Ducreux, L., Swanston, J.S., Davies, H.V., and McDougall, G.J. (2003). LIMIT DEXTRINASE in barley cultivars of differing malting quality: activity, inhibitors and limit dextrin profiles. Journal of Cereal Science **38,** 325-334.

Rosyara, U., Kishii, M., Payne, T., Sansaloni, C.P., Singh, R.P., Braun, H.-J., and Dreisigacker, S. (2019). Genetic contribution of synthetic hexaploid wheat to CIMMYT's spring bread wheat breeding germplasm. Scientific Reports **9,** 12355.

Rydberg, U., Andersson, L., Andersson, R., Åman, P., and Larsson, H. (2001). Comparison of *STARCH BRANCHING ENZYME I* and *II* from potato. European Journal of Biochemistry **268,** 6140-6145.

Ryley, J.F., Bentley, M., Manners, D.J., and Stark, J.R. (1969). Amylopectin, the storage polysaccharide of the coccidia *Eimeria brunetti* and *E. tenella*. The Journal of Parasitology **55,** 839-845.

Saccomanno, B., Chambers, A.H., Hayes, A., Mackay, I., McWilliam, S.C., and Trafford, K. (2017). Starch granule morphology in oat endosperm. Journal of Cereal Science **73,** 46-54.

Saccomanno, B., Berbezy, P., Findlay, K., Shoesmith, J., Uauy, C., Viallis, B., and Trafford, K. (2022). Characterization of wheat lacking B-type starch granules. Journal of Cereal Science **104,** 103398.

Sadali, N.M., Sowden, R.G., Ling, Q., and Jarvis, R.P. (2019). Differentiation of chromoplasts and other plastids in plants. Plant Cell Reports **38,** 803-818.

Saito, M., Tanaka, T., Sato, K., Vrinten, P., and Nakamura, T. (2018). A single nucleotide polymorphism in the "*Fra*" gene results in fractured starch granules in barley. Theoretical and Applied Genetics **131,** 353-364.

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

Sapkota, S., Boatwright, J.L., Jordan, K., Boyles, R., and Kresovich, S. (2020). Identification of novel genomic associations and gene candidates for grain starch content in *Sorghum*. Genes (Basel) **11,** 1448.

Satoh, H., Shibahara, K., Tokunaga, T., Nishi, A., Tasaki, M., Hwang, S.K., Okita, T.W., Kaneko, N., Fujita, N., Yoshida, M., Hosaka, Y., Sato, A., Utsumi, Y., Ohdan, T., and Nakamura, Y. (2008). Mutation of the plastidial alpha-glucan *PHOSPHORYLASE* gene in rice affects the synthesis and structure of starch in the endosperm. Plant Cell **20,** 1833-1849.

Schönhals, E.M., Ding, J., Ritter, E., Paulo, M.J., Cara, N., Tacke, E., Hofferbert, H.R., Lübeck, J., Strahwald, J., and Gebhardt, C. (2017). Physical mapping of QTL for tuber yield, starch content and starch yield in tetraploid potato (*Solanum tuberosum* L.) by means of genome wide genotyping by sequencing and the 8.3 K SolCAP SNP array. BMC Genomics **18,** 642.

Schwarte, S., Brust, H., Steup, M., and Tiedemann, R. (2013). Intraspecific sequence variation and differential expression in starch synthase genes of *Arabidopsis thaliana*. BMC Research Notes **6,** 84.

Scott, M.F., Ladejobi, O., Amer, S., Bentley, A.R., Biernaskie, J., Boden, S.A., Clark, M., Dell'Acqua, M., Dixon, L.E., Filippi, C.V., Fradgley, N., Gardner, K.A., Mackay, I.J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R.K., Thudi, M., Varshney, R.K., Venturini, L., Whan, A., Cockram, J., and Mott, R. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity **125,** 396-416.

Sears, E.R., and Miller, T.E. (1985). The history of chinese spring wheat. Cereal Research Communications **13,** 261-263.

Sestili, F., Sparla, F., Botticella, E., Janni, M., D'Ovidio, R., Falini, G., Marri, L., Cuesta-Seijo, J.A., Moscatello, S., Battistelli, A., Trost, P., and Lafiandra, D. (2016). The down-regulation of the genes encoding *ISOAMYLASE 1* alters the starch composition of the durum wheat grain. Plant Science **252,** 230-238.

Seung, D. (2020). Amylose in starch: towards an understanding of biosynthesis, structure and function. New Phytologist **228,** 1490-1504.

Seung, D., and Smith, A.M. (2019). Starch granule initiation and morphogenesis-progress in *Arabidopsis* and cereals. Journal of Experimental Botany **70,** 771-784.

Seung, D., Echevarría-Poza, A., Steuernagel, B., and Smith, A.M. (2020). Natural polymorphisms in *Arabidopsis* result in wide variation or loss of the amylose component of starch. Plant Physiology **182,** 870-881.

Seung, D., Schreier, T.B., Bürgy, L., Eicke, S., and Zeeman, S.C. (2018). Two plastidial coiled-coil proteins are essential for normal starch granule initiation in *Arabidopsis*. Plant Cell **30,** 1523-1542.

Seung, D., Soyk, S., Coiro, M., Maier, B.A., Eicke, S., and Zeeman, S.C. (2015). PROTEIN TARGETING TO STARCH is required for localising GRANULE-BOUND STARCH SYNTHASE to starch granules and for normal amylose synthesis in *Arabidopsis*. PLoS Biology **13,** e1002080.

Seung, D., Boudet, J., Monroe, J., Schreier, T.B., David, L.C., Abt, M., Lu, K.J., Zanella, M., and Zeeman, S.C. (2017). Homologs of *PROTEIN TARGETING TO STARCH* control starch granule initiation in *Arabidopsis* Leaves. Plant Cell **29,** 1657-1677.

Shannon, J.C., Pien, F.M., Cao, H., and Liu, K.C. (1998). BRITTLE-1, an adenylate translocator, facilitates transfer of extraplastidial synthesized ADP-glucose into amyloplasts of maize endosperms. Plant Physiology **117,** 1235-1252.

Sharma, M., Abt, M.R., Eicke, S., Ilse, T.E., Liu, C., Lucas, M.S., Pfister, B., and Zeeman, S.C. (2024). MFP1 defines the subchloroplast location of starch granule initiation. Proceedings of the National Academy of Sciences **121,** e2309666121.

Sharma, S., Friberg, M., Vogel, P., Turesson, H., Olsson, N., Andersson, M., and Hofvander, P. (2023). *PHO1A* (*PLASTID STARCH PHOSPHORYLASE*) is duplicated and essential for normal starch granule phenotype in tubers of *Solanum tuberosum* L. Frontiers in Plant Science **14**.

Shaul, O. (2017). How introns enhance gene expression. The International Journal of Biochemistry & Cell Biology **91,** 145-155.

Shevkani, K., Singh, N., Bajaj, R., and Kaur, A. (2017). Wheat starch production, structure, functionality and applications—a review. International Journal of Food Science & Technology **52,** 38-58.

Shu, X., and Rasmussen, S.K. (2014). Quantification of amylose, amylopectin, and β-glucan in search for genes controlling the three major quality traits in barley by genome-wide association studies. Frontiers in Plant Science **5,** 197.

Siekmann, D., Jansen, G., Zaar, A., Kilian, A., Fromme, F.J., and Hackauf, B. (2021). A genome-wide association study pinpoints quantitative trait genes for plant height, heading date, grain quality, and yield in rye (*Secale cereale* L.). Frontiers in Plant Science **12,** 718081.

Sikka, V.K., Choi, S.-B., Kavakli, I.H., Sakulsingharoj, C., Gupta, S., Ito, H., and Okita, T.W. (2001). Subcellular compartmentation and allosteric regulation of

the rice endosperm ADP GLUCOSE PYROPHOSPHORYLASE. Plant Science **161,** 461-468.

Singh, A., Compart, J., AL-Rawi, S.A., Mahto, H., Ahmad, A.M., and Fettke, J. (2022). *LIKE EARLY STARVATION 1* alters the glucan structures at the starch granule surface and thereby influences the action of both starch-synthesizing and starch-degrading enzymes. The Plant Journal **111,** 819-835.

Singh, M., Kaur, S., Kaur, A., Yadav, I.S., Sharma, P., Chhuneja, P., and Singh, K. (2023). Nucleotide diversity and molecular characterization of *SOLUBLE STARCH SYNTHASE I* gene in wheat and its ancestral species. Journal of Plant Biochemistry and Biotechnology **32,** 92-105.

Sissons, M.J., Lance, R.C.M., and Sparrow, D.H.B. (1993). Studies on *LIMIT DEXTRINASE* in barley. 3. *LIMIT DEXTRINASE* in developing kernels. Journal of Cereal Science **17,** 19-24.

Smith, A.M., and Zeeman, S.C. (2020). Starch: a flexible, adaptable carbon store coupled to plant growth. Annual Review of Plant Biology **71,** 217-245.

Soh, H.N., Sissons, M.J., and Turner, M.A. (2006). Effect of starch granule size distribution and elevated amylose content on durum dough rheology and spaghetti cooking quality. Cereal Chemistry. **83,** 513-519.

Sokolov, L.N., Déjardin, A., and Kleczkowski, L.A. (1998). Sugars and light/dark exposure trigger differential regulation of *ADP-GLUCOSE PYROPHOSPHORYLASE* genes in *Arabidopsis thaliana* (thale cress). Biochemical Journal **336,** 681-687.

Soulaka, A.B., and Morrison, W.R. (1985). The amylose and lipid contents, dimensions, and gelatinisation characteristics of some wheat starches and their A- and B-granule fractions. Journal of the Science of Food and Agriculture **36,** 709-718.

Sparla, F., Falini, G., Botticella, E., Pirone, C., Talamè, V., Bovina, R., Salvi, S., Tuberosa, R., Sestili, F., and Trost, P. (2014). New Starch Phenotypes Produced by TILLING in Barley. PLoS One **9,** e107779.

Stahl, Y., Coates, S., Bryce, J.H., and Morris, P.C. (2004). Antisense downregulation of the barley *LIMIT DEXTRINASE INHIBITOR* modulates starch granule size distribution, starch composition and amylopectin structure. The Plant Journal **39,** 599-611.

Stoddard, F.L. (1999a). Survey of starch particle-size distribution in wheat and related species. Cereal Chemistry **76,** 145-149.

Stoddard, F.L. (1999b). Variation in grain mass, grain nitrogen, and starch B-granule content within wheat heads. Cereal Chemistry **76,** 139-144.

Stoddard, F.L., and Sarker, R. (2000). Characterization of starch in *Aegilops* species. Cereal Chemistry **77,** 445-447.

Streb, S., Eicke, S., and Zeeman, S.C. (2012). The simultaneous abolition of three starch hydrolases blocks transient starch breakdown in *Arabidopsis*. Journal of Biological Chemistry **287,** 41745-41756.

Stuber, C.W., Edwards, M., and Wendel, J.F. (1987). Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors Influencing Yield and Its Component Traits. Crop Science **27,** 639-648.

Sulpice, R., Flis, A., Ivakov, A.A., Apelt, F., Krohn, N., Encke, B., Abel, C., Feil, R., Lunn, J.E., and Stitt, M. (2014). *Arabidopsis* coordinates the diurnal regulation of carbon allocation and growth across a wide range of photoperiods. Molecular Plant **7,** 137-155.

Sun, C., Sathish, P., Ahlandsberg, S., and Jansson, C. (1998). The two genes encoding *STARCH-BRANCHING ENZYMES IIA* and *IIB* are differentially expressed in barley. Plant Physiology **118,** 37-49.

Sun, H., Lü, J., Fan, Y., Zhao, Y., Kong, F., Li, R., Wang, H., and Li, S. (2008). Quantitative trait loci (QTLs) for quality traits related to protein and starch in wheat. Progress in Natural Science **18,** 825-831.

Szydlowski, N., Ragel, P., Hennen-Bierwagen, T.A., Planchot, V., Myers, A.M., Mérida, A., d'Hulst, C., and Wattebled, F. (2011). Integrated functions among multiple starch synthases determine both amylopectin chain length and branch linkage location in *Arabidopsis* leaf starch. Journal of Experimental Botany **62,** 4547-4559.

Szydlowski, N., Ragel, P., Raynaud, S., Lucas, M.M., Roldán, I., Montero, M., Muñoz, F.J., Ovecka, M., Bahaji, A., Planchot, V., Pozueta-Romero, J., D'Hulst, C., and Mérida, A. (2009). Starch granule initiation in *Arabidopsis* requires the presence of either class IV or class III starch synthases. Plant Cell **21,** 2443-2457.

Tanaka, E., Ral, J.F., Li, S., Gaire, R., Cavanagh, C.R., Cullis, B.R., and Whan, A. (2017). Increased accuracy of starch granule type quantification using mixture distributions. Plant Methods **13,** 107.

Tang, X.J., Peng, C., Zhang, J., Cai, Y., You, X.M., Kong, F., Yan, H.G., Wang, G.X., Wang, L., Jin, J., Chen, W.W., Chen, X.G., Ma, J., Wang, P., Jiang, L., Zhang, W.W., and Wan, J.M. (2016). ADP-GLUCOSE PYROPHOSPHORYLASE LARGE SUBUNIT 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. Plant Science **249,** 70-83.

Tao, H., Wang, P., Wu, F., Jin, Z., and Xu, X. (2016). Particle size distribution of wheat starch granules in relation to baking properties of frozen dough. Carbohydrate Polymers **137,** 147-153.

Tateoka, T. (1962). Starch grains of endosperm in grass systematics. Shokubutsugaku Zasshi **75,** 377-383.

Tatge, H., Marshall, J., Martin, C., Edwards, E.A., and Smith, A.M. (1999). Evidence that amylose synthesis occurs within the matrix of the starch granule in potato tubers. Plant, Cell & Environment **22,** 543-550.

Tetlow, I.J. (2011). Starch biosynthesis in developing seeds. Seed Science Research **21,** 5-32.

Tetlow, I.J., and Emes, M.J. (2014). A review of starch-branching enzymes and their role in amylopectin biosynthesis. IUBMB Life **66,** 546-558.

Tetlow, I.J., and Bertoft, E. (2020). A review of starch biosynthesis in relation to the building block-backbone model. International Journal of Molecular Sciences **21,** 7011.

Tetlow, I.J., Morell, M.K., and Emes, M.J. (2004). Recent developments in understanding the regulation of starch metabolism in higher plants. Journal of Experimental Botany **55,** 2131-2145.

Thorbjørnsen, T., Villand, P., Denyer, K., Olsen, O.-A., and Smith, A.M. (1996). Distinct isoforms of ADP-GLUCOSE PYROPHOSPHORYLASE occur inside and outside the amyloplasts in barley endosperm. The Plant Journal **10,** 243-250.

Tian, B., Deng, Z., Xie, Q., and Tian, J. (2015). Genetic dissection of the developmental behaviour of total starch content and its components in wheat grain. Crop and Pasture Science **66,** 445-455.

Tian Bin, T.B., Liu Bin, L.B., Zhu Zhan-Ling, Z.Z., Xie QuanGang, X.Q., and Tian JiChun, T.J. (2011). Conditional and unconditional QTL mapping of grain starch accumulation in wheat. Scientia Agricultura Sinica **44,** 4551-4559.

Tibbs Cortes, L., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. The Plant Genome **14,** e20077.

Tiessen, A., Hendriks, J.H., Stitt, M., Branscheid, A., Gibon, Y., Farré, E.M., and Geigenberger, P. (2002). Starch synthesis in potato tubers is regulated by post-translational redox modification of ADP-GLUCOSE PYROPHOSPHORYLASE: a novel regulatory mechanism linking starch synthesis to the sucrose supply. Plant Cell **14,** 2191-2213.

Tillett, I.J.L., Bryce, J.H., and Narzillß, L. (1993). The regulation of starch granule size in endosperm of developing barley grains. In European Brewery

Convention: Proceedings of the 24th Congress Oslo 1993 (Oxford University Press), pp. 45-52.

Toyosawa, Y., Kawagoe, Y., Matsushima, R., Crofts, N., Ogawa, M., Fukuda, M., Kumamaru, T., Okazaki, Y., Kusano, M., Saito, K., Toyooka, K., Sato, M., Ai, Y., Jane, J.-L., Nakamura, Y., and Fujita, N. (2016). Deficiency of *STARCH SYNTHASE IIIA* and *IVB* alters starch granule morphology from polyhedral to spherical in rice endosperm. Plant Physiology **170,** 1255-1270.

Tsai, C.Y., and Nelson, O.E. (1966). Starch-deficient maize mutant lacking *ADENOSINE DEPHOSPHATE GLUCOSE PYROPHOSPHORYLASE* activity. Science **151,** 341-343.

Tsai, C.Y., and Nelson, O.E. (1969). Mutations at the *shrunken-4* locus in maize that produce three altered phosphorylases. Genetics **61,** 813-821.

Tsai, H.-Y., Janss, L.L., Andersen, J.R., Orabi, J., Jensen, J.D., Jahoor, A., and Jensen, J. (2020). Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. Scientific Reports **10,** 3347.

Tuncel, A., Cakir, B., Hwang, S.-K., and Okita, T.W. (2014). The role of the large subunit in redox regulation of the rice endosperm ADP-GLUCOSE PYROPHOSPHORYLASE. The FEBS Journal **281,** 4951-4963.

Tung, J.-Y., Chang, Margaret D.-T., Chou, W.-I., Liu, Y.-Y., Yeh, Y.-H., Chang, F.-Y., Lin, S.-C., Qiu, Z.-L., and Sun, Y.-J. (2008). Crystal structures of the starch-binding domain from *Rhizopus oryzae* GLUCOAMYLASE reveal a polysaccharide-binding path. Biochemical Journal **416,** 27-36.

Vandromme, C., Spriet, C., Dauvillée, D., Courseaux, A., Putaux, J.L., Wychowski, A., Krzewinski, F., Facon, M., D'Hulst, C., and Wattebled, F. (2019). PII1: a protein involved in starch initiation that determines granule number and size in *Arabidopsis* chloroplast. New Phytologist **221,** 356-370.

Vester-Christensen, M.B., Abou Hachem, M., Svensson, B., and Henriksen, A. (2010a). Crystal structure of an essential enzyme in seed starch degradation: barley LIMIT DEXTRINASE in complex with cyclodextrins. Journal of Molecular Biology **403,** 739-750.

Vester-Christensen, M.B., Hachem, M.A., Naested, H., and Svensson, B. (2010b). Secretory expression of functional barley LIMIT DEXTRINASE by *Pichia pastoris* using high cell-density fermentation. Protein Expression and Purification **69,** 112-119.

Vilaplana, F., Hasjim, J., and Gilbert, R.G. (2012). Amylose content in starches: Toward optimal definition and validating experimental methods. Carbohydrate Polymers **88,** 103-111.

Voichek, Y., and Weigel, D. (2020). Identifying genetic variants underlying phenotypic variation in plants without complete genomes. Nature Genetics **52,** 534-540.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez, R.H., Kolodziej, M.C., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B., Gundlach, H., Bandi, V., Siri, J.N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D., Ban, T., Venturini, L., Bevan, M., Clavijo, B., Koo, D.-H., Ens, J., Wiebe, K., N'Diaye, A., Fritz, A.K., Gutwin, C., Fiebig, A., Fosker, C., Fu, B.X., Accinelli, G.G., Gardner, K.A., Fradgley, N., Gutierrez-Gonzalez, J., Halstead-Nussloch, G., Hatakeyama, M., Koh, C.S., Deek, J., Costamagna, A.C., Fobert, P., Heavens, D., Kanamori, H., Kawaura, K., Kobayashi, F., Krasileva, K., Kuo, T., McKenzie, N., Murata, K., Nabeka, Y., Paape, T., Padmarasu, S., Percival-Alwyn, L., Kagale, S., Scholz, U., Sese, J., Juliana, P., Singh, R., Shimizu-Inatsugi, R., Swarbreck, D., Cockram, J., Budak, H., Tameshige, T., Tanaka, T., Tsuji, H., Wright, J., Wu, J., Steuernagel, B., Small, I., Cloutier, S., Keeble-Gagnère, G., Muehlbauer, G., Tibbets, J., Nasuda, S., Melonek, J., Hucl, P.J., Sharpe, A.G., Clark, M., Legg, E., Bharti, A., Langridge, P., Hall, A., Uauy, C., Mascher, M., Krattinger, S.G., Handa, H., Shimizu, K.K., Distelfeld, A., Chalmers, K., Keller, B., Mayer, K.F.X., Poland, J., Stein, N., McCartney, C.A., Spannagl, M., Wicker, T., and Pozniak, C.J. (2020). Multiple wheat genomes reveal global variation in modern breeding. Nature **588,** 277-283.

Wang, L., Zhu, T., Rodriguez, J.C., Deal, K.R., Dubcovsky, J., McGuire, P.E., Lux, T., Spannagl, M., Mayer, K.F.X., Baldrich, P., Meyers, B.C., Huo, N., Gu, Y.Q., Zhou, H., Devos, K.M., Bennetzen, J.L., Unver, T., Budak, H., Gulick, P.J., Galiba, G., Kalapos, B., Nelson, D.R., Li, P., You, F.M., Luo, M.-C., and Dvorak, J. (2021a). *Aegilops tauschii* genome assembly Aet v5.0 features greater sequence contiguity and improved annotation. G3 Genes|Genomes|Genetics **11,** jkab325.

Wang, Q., Yan, N., Chen, H., Li, S., Hu, H., Lin, Y., Shi, H., Zhou, K., Jiang, X., and Yu, S. (2021b). Genome-wide association study of kernel traits in *Aegilops tauschii*. Frontiers in Genetics **12,** 651785.

Wang, S., Wang, J., Zhang, W., Li, C., Yu, J., and Wang, S. (2015). Molecular order and functional properties of starches from three waxy wheat varieties grown in China. Food Chemistry **181,** 43-50.

Wang, S., Xu, Y., Li, Z., Zhang, S., Lim, J.-M., Lee, K.O., Li, C., Qian, Q., Jiang, D.A., and Qi, Y. (2014). *OsMOGS* is required for *N*-glycan formation and auxin-mediated root development in rice (*Oryza sativa* L.). The Plant Journal **78,** 632-645.

Wang, W., Wei, X., Jiao, G., Chen, W., Wu, Y., Sheng, Z., Hu, S., Xie, L., Wang, J., Tang, S., and Hu, P. (2020). *GBSS-BINDING PROTEIN*, encoding a CBM48 domain-containing protein, affects rice quality and yield. Journal of Integrative Plant Biology **62,** 948-966.

Wang, Y., White, P., Pollak, L., and Jane, J. (1993). Amylopectin and intermediate materials in starches from mutant genotypes. Cereal Chemistry **70,** 521-525.

Wangpaiboon, K., Charoenwongpaiboon, T., Klaewkla, M., Field, R.A., and Panpetch, P. (2023). Cassava *PULLULANASE* and its synergistic debranching action with *ISOAMYLASE 3* in starch catabolism. Frontiers in Plant Science **14,** 1114215.

Washington, J., Box, A., Karakousis, A., and Barr, A. (2000). Developing waxy barley cultivars for food, feed and malt. In Barley genetics VIII. Int. Barley Genet. Symp., 8th, Adelaide, Australia, pp. 22-27.

Watson-Lazowski, A., Raven, E., Feike, D., Hill, L., Barclay, J.E., Smith, A.M., and Seung, D. (2022). Loss of *PROTEIN TARGETING TO STARCH 2* has variable effects on starch synthesis across organs and species. Journal of Experimental Botany **73,** 6367-6379.

Wattebled, F., Dong, Y., Dumez, S., Delvallé, D., Planchot, V.r., Berbezy, P., Vyas, D., Colonna, P., Chatterjee, M., Ball, S., and D'Hulst, C. (2005). Mutants of *Arabidopsis* lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant Physiology **138,** 184-195.

Wei, X., Liu, G.-L., Jia, S.-L., Chi, Z., Hu, Z., and Chi, Z.-M. (2021). Pullulan biosynthesis and its regulation in *Aureobasidium* spp. Carbohydrate Polymers **251,** 117076.

Wei, X., Jiao, G., Lin, H., Sheng, Z., Shao, G., Xie, L., Tang, S., Xu, Q., and Hu, P. (2017). *GRAIN INCOMPLETE FILLING 2* regulates grain filling and starch synthesis during rice caryopsis development. Journal of Integrative Plant Biology **59,** 134-153.

Whitt, S.R., Wilson, L.M., Tenaillon, M.I., Gaut, B.S., and Buckler, E.S. (2002). Genetic diversity and selection in the maize starch pathway. Proceedings of the National Academy of Sciences **99,** 12959-12962.

Wilkens, C., Svensson, B., and Møller, M.S. (2018). Functional roles of starch binding domains and surface binding sites in enzymes involved in starch biosynthesis. Frontiers in Plant Science **9,** 1652.

Wilson, J.D., Bechtel, D.B., Todd, T.C., and Seib, P.A. (2006). Measurement of wheat starch granule size distribution using image analysis and laser diffraction technology. Cereal Chemistry **83,** 259-268.

Wu, C., Colleoni, C., Myers, A.M., and James, M.G. (2002). Enzymatic properties and regulation of ZPU1, the maize pullulanase-type starch debranching enzyme. Archives of Biochemistry and Biophysics **406,** 21-32.

Yan, H., Zhang, W., Wang, Y., Jin, J., Xu, H., Fu, Y., Shan, Z., Wang, X., Teng, X., Li, X., Wang, Y., Hu, X., Zhang, W., Zhu, C., Zhang, X., Zhang, Y., Wang, R., Zhang, J., Cai, Y., You, X., Chen, J., Ge, X., Wang, L., Xu, J., Jiang, L., Liu, S., Lei, C., Zhang, X., Wang, H., Ren, Y., and Wan, J. (2024). Rice *LIKE EARLY STARVATION1* cooperates with *FLOURY ENDOSPERM6* to modulate starch biosynthesis and endosperm development. The Plant Cell **36,** 1892-1912.

Yao, J., Zhao, D., Chen, X., Zhang, Y., and Wang, J. (2018). Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.). The Crop Journal **6,** 353-365.

Yazdani, M., Rouse, M.N., Steffenson, B.J., Bajgain, P., Patpour, M., Johansson, E., and Rahmatov, M. (2023). Developing adapted wheat lines with broad-spectrum resistance to stem rust: Introgression of *Sr59* through backcrossing and selections based on genotyping-by-sequencing data. PLoS One **18,** e0292724.

You, Y., Kong, H., Li, C., Gu, Z., Ban, X., and Li, Z. (2024). Carbohydrate binding modules: Compact yet potent accessories in the specific substrate binding and performance evolution of carbohydrate-active enzymes. Biotechnology Advances **73,** 108365.

Yun, M.-S., and Kawagoe, Y. (2009). Amyloplast division progresses simultaneously at multiple sites in the endosperm of rice. Plant and Cell Physiology **50,** 1617-1626.

Yun, M.-S., and Kawagoe, Y. (2010). Septum formation in amyloplasts produces compound granules in the rice endosperm and is regulated by plastid division proteins. Plant and Cell Physiology **51,** 1469-1479.

Yun, M.-S., Umemoto, T., and Kawagoe, Y. (2011). Rice debranching enzyme ISOAMYLASE3 facilitates starch metabolism and affects plastid morphogenesis. Plant and Cell Physiology **52,** 1068-1082.

Zeeman, S.C., Kossmann, J., and Smith, A.M. (2010). Starch: Its Metabolism, Evolution, and Biotechnological Modification in Plants. Annual Review of Plant Biology **61,** 209-234.

Zeeman, S.C., Tiessen, A., Pilling, E., Kato, K.L., Donald, A.M., and Smith, A.M. (2002). Starch synthesis in *Arabidopsis*. Granule synthesis, composition, and structure. Plant Physiology **129,** 516-529.

Zhang, L., Li, N., Zhang, J., Zhao, L., Qiu, J., and Wei, C. (2022). The CBM48 domain-containing protein FLO6 regulates starch synthesis by interacting with SSIVb and GBSS in rice. Plant Molecular Biology **108,** 343-361.

Zhang, M.-Z., Fang, J.-H., Yan, X., Liu, J., Bao, J.-S., Fransson, G., Andersson, R., Jansson, C., Åman, P., and Sun, C. (2012). Molecular insights into how a deficiency of amylose affects carbon allocation – carbohydrate and oil analyses and gene expression profiling in the seeds of a rice waxy mutant. BMC Plant Biology **12,** 230.

Zhang, W., Gu, J., Wang, Z., Wei, C., Yang, J., and Zhang, J. (2017). Comparison of structural and functional properties of wheat starch under different soil drought conditions. Scientific Reports **7,** 12312.

Zhang, X., Myers, A.M., and James, M.G. (2005). Mutations affecting *STARCH SYNTHASE III* in *Arabidopsis* alter leaf starch structure and increase the rate of starch synthesis. Plant Physiology **138,** 663-674.

Zhang, X., Szydlowski, N., Delvallé, D., D'Hulst, C., James, M.G., and Myers, A.M. (2008). Overlapping functions of the starch synthases *SSII* and *SSIII* in amylopectin biosynthesis in *Arabidopsis*. BMC Plant Biology **8,** 96.

Zhang, Y., Guo, Q., Feng, N., Wang, J.-r., Wang, S.-j., and He, Z.-h. (2016). Characterization of A- and B-type starch granules in Chinese wheat cultivars. Journal of Integrative Agriculture **15,** 2203-2214.

Zhao, X., Lv, L., Li, J., Ma, F., Bai, S., Zhou, Y., Zhang, D., Li, S., and Song, C.-p. (2021). Genome-wide association study of grain shapes in *Aegilops tauschii*. Euphytica **217,** 144.

Zhong, Y., Blennow, A., Kofoed-Enevoldsen, O., Jiang, D., and Hebelstrup, K.H. (2018). *PROTEIN TARGETING TO STARCH 1* is essential for starchy endosperm development in barley. Journal of Experimental Botany**70,** 485-496.

Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y., Wang, Y.-g., Liu, S., Jiao, C., Lu, H., Wang, J., Yin, C., Jiao, Y., and Lu, F. (2020). *Triticum* population sequencing provides insights into wheat adaptation. Nature Genetics **52,** 1412-1422.

Zhou, Y., Bai, S., Li, H., Sun, G., Zhang, D., Ma, F., Zhao, X., Nie, F., Li, J., Chen, L., Lv, L., Zhu, L., Fan, R., Ge, Y., Shaheen, A., Guo, G., Zhang, Z., Ma, J., Liang, H., Qiu, X., Hu, J., Sun, T., Hou, J., Xu, H., Xue, S., Jiang, W., Huang, J., Li, S., Zou, C., and Song, C.P. (2021). Introgressing the *Aegilops tauschii* genome into wheat as a basis for cereal improvement. Nature Plants **7,** 774-786.

Zhu, C., Gore, M., Buckler, E.S., and Yu, J. (2008). Status and prospects of association mapping in plants. The Plant Genome **1**.

Ziegler, G.R., Creek, J.A., and Runt, J. (2005). Spherulitic crystallization in starch as a model for starch granule initiation. Biomacromolecules **6,** 1547-1554.

Appendix One: Supplementary Tables

Table S1 Standard components. Components are listed alphabetically.

Component	Composition
JIC cereal mix	65% peat, 25% loam, 10% grit, 3 kg/m^3 dolomitic limestone, 1.3 kg/m^3 pg mix, and 3 kg/m^3 osmocote exact
LB agar	10 g/L peptone from casein, 5 g/L yeast extract, 10 g/L NaCl, 11 g/L agar
LB liquid media	10 g/L peptone from casein, 5 g/L yeast extract, 10 g/L NaCl
SOC	20 g/L tryptone, 5 g/L yeast extract, 580 mg/L NaCl, 186 mg/L KCl, 2.03g/L $MgCl_2$(6H20), 2.56 g/L $MgSO_4$ (7H2), 3.6 g/L glucose

Table S2 Precise location of the peaks identified in my GWAS. The locations of the peaks identified in the GWAS in the TOWWC193/AL8/78 v5 genome are listed with their corresponding size. For B-type granule number peaks, these are the peak locations when a 10 μm threshold was used. Priority peaks are indicated with *.

Peak	Phenotype	Chromosome	Start of peak	End of peak	Peak size (Mb)
A	B-type granule diameter	3	133,390,000	135,380,000	1.99
B*	B-type granule diameter	3	543,060,000	556,500,000	13.44
C*	B-type granule diameter	5	60,120,000	60,700,000	0.58
D*	B-type granule diameter	7	84,010,000	88,820,000	4.81
E	B-type granule diameter	7	382,890,000	382,900,000	0.01
V*	B-type granule number	3	46,120,000	49,070,000	2.95
W*	B-type granule number	3	115,080,000	116,770,000	1.69
X	B-type granule number	3	178,650,000	284,490,000	105.84
Y	B-type granule number	3	490,340,000	513,680,000	23.34
Z	B-type granule number	3	583,040,000	586,790,000	3.75

Table S3 *Ae. tauschii* starch genes. A list of starch-related genes was obtained from Chen *et al.* (2023a) and the protein sequences were used in BLASTp searches the *Ae. tauschii* v4 genome to obtain the *Ae. tauschii* gene IDs in this version of the genome. Genes are listed in alphabetical order.

Gene ID	Description
AET1Gv20088000	Phosphoglucose isomerase (cytosolic)
AET1Gv20245200	Starch synthase 3a
AET1Gv20286900	Brittle1 transporter 2
AET1Gv20295700	MAR-binding filament-like protein 1.1
AET1Gv20357200	Starch synthase 2c
AET1Gv20542100	Beta amylase 3
AET1Gv20601800	Isoamylase 2
AET1Gv20826000	Glucan, water dikinase 1.1
AET1Gv20853900	Starch synthase 4
AET1Gv20995100	ADPG pyrophosphorylase (cytosolic large subunit)
AET1Gv21020100	Phosphoglucan Phosphatase Like SEX4 1
AET2Gv20249300	Disproportionating enzyme 2
AET2Gv20331100	Disproportionating enzyme 1
AET2Gv20457800	Beta amylase 5.1
AET2Gv20458000	Beta amylase 5.2
AET2Gv20469200	Glucose 6-phosphate/P translocator
AET2Gv20657800	Alpha amylase 2
AET2Gv20664800	Starch branching enzyme 2a
AET2Gv20703800	Starch branching enzyme 2b
AET2Gv20836000	Granule-bound starch synthase 2
AET2Gv21035000	Starch synthase 3b
AET3Gv20084800	Glucose transporter
AET3Gv20259600	MAR-binding filament-like protein 1.2
AET3Gv20365100	Beta amylase 10
AET3Gv20571700	Nucleotide transporter 1
AET3Gv20604400	Alpha amylase 3
AET3Gv20821600	Starch phosphorylase 2.2 (cytosolic)
AET3Gv20821900	Starch phosphorylase 2.1 (cytosolic)
AET4Gv20049600	B-granule content 1
AET4Gv20094300	Phosphoglucomutase

Table S3 *Ae. tauschii* starch genes. (continued)

Gene ID	Description
AET4Gv20694900	Alpha amylase 1
AET4Gv20776200	Early Starvation 1
AET4Gv20804800	Beta amylase 1.1
AET4Gv20830800	Phosphoglucan phosphatase SEX4
AET5Gv20381900	Phosphoglucan Phosphatase Like SEX4 2
AET5Gv20443400	ADPG pyrophosphorylase (small subunit 2)
AET5Gv20568500	Alpha amylase 1
AET5Gv20582500	Phosphoglucose isomerase (plastidial)
AET5Gv20586000	Isoamylase 3
AET5Gv20915800	Starch phosphorylase 1 (plastidial)
AET5Gv21077600	Alpha amylase 1
AET5Gv21093900	ADPG pyrophosphorylase (plastidial large subunit)
AET6Gv20268000	Protein Targeting to Starch 1
AET6Gv20442000	Myosin-resembling chloroplast protein
AET6Gv20448800	Nucleotide transporter 2
AET6Gv20453300	Brittle1 transporter 1
AET6Gv20761400	Starch synthase 2b
AET6Gv20786800	Alpha amylase 1
AET6Gv20821300	Alpha amylase 1
AET6Gv20821400	Alpha amylase 1
AET7Gv20100800	Phosphoglucan, water dikinase
AET7Gv20167400	Granule-bound starch synthase 1
AET7Gv20309100	Starch synthase 1
AET7Gv20336800	Limit dextrinase
AET7Gv20483200	Starch synthase 2a
AET7Gv20582500	Glucan, water dikinase 1.2
AET7Gv20615300	Isoamylase 1
AET7Gv20706000	ADPG pyrophosphorylase (small subunit 1)
AET7Gv20718600	Glucose 6-phosphate/P translocator
AET7Gv20938000	Alpha amylase 1
AET7Gv20939100	Alpha amylase 1
AET7Gv20939200	Alpha amylase 1
AET7Gv21207300	Brittle1 transporter 3
AET7Gv21320400	Starch branching enzyme 1.1

Table S4 Genes located under the B-type granule diameter GWAS peaks. Genes underlying the peaks identified in the GWAS for B-type granule diameter (Figure 4.1). Gene names are shown according to the v4 genome nomenclature, except where the gene was not annotated in the v4 genome assembly, in which case the v5 gene name is shown. Genes are listed in the order they appear on the forward strand and the peaks are listed based on chromosome number.

Peak from GWAS	Gene	Predicted protein
B	*AET3Gv20949600*	NB-ARC domain containing protein
	AET3Gv20949700	FBD domain containing protein
	AET3Gv20949800	ATP-dependent DNA helicase
	AET3Gv20949900	FBD domain containing protein
	AET3Gv20950200	Histone deacetylase domain containing protein
	AET3Gv20950500	DUF295 domain containing protein
	AET3Gv20950900	DEE Tnp4 domaining containing protein
	AET3Gv20951000	FBD domain containing protein
	AET3Gv20951100	CBS domain-containing protein CBSX6
	AET3Gv20951900	Leucine rich repeat domain containing protein
	AET3Gv20952200	Uncharacterised - no annotated domains
	AET3Gv20952500	Nucleotide-diphospho-sugar transferase domain-containing protein
	AET3Gv20952600	Nucleotide-diphospho-sugar transferase domain-containing protein
	AET3Gv20952700	Uncharacterised - no annotated domains
	AET3Gv20953200	Uncharacterised - no annotated domains
	AET3Gv20953300	Transcription and mRNA export factor ENY2
	AET3Gv20953700	Methyltransferase
	AET3Gv20954100	Amidohydrolase-related protein
	AET3Gv20954200	Dynamin-related protein
	AET3Gv20954300	Nucleotide-diphospho-sugar transferase domain-containing protein
	AET3Gv20954500	Nucleotide-diphospho-sugar transferase domain-containing protein
	AET3Gv20954600	Nucleotide-diphospho-sugar transferase domain-containing protein

Table S4 Genes located under the B-type granule diameter GWAS peaks. (continued)

Peak from GWAS	Gene	Predicted protein
B (continued)	*PGSB gene 29601*	Nucleotide-diphospho-sugar transferase domain-containing protein
	AET3Gv20954700	Nucleotide-diphospho-sugar transferase domain-containing protein
	AET3Gv20954800	RNA-binding protein
	AET3Gv20954900	UPF0481 protein
	AET3Gv20955300	DDE Tnp4 domain-containing protein
	AET3Gv20955400	Mannosyl-oligosaccharide glucosidase with GH63 domain
	AET3Gv20955500	Exostosin GT47 domain containing protein
	AET3Gv20956200	TFIIB-type domain-containing protein
	AET3Gv20956300	Sulfotransferase-domain containing protein
	AET3Gv20956400	Sulfotransferase-domain containing protein
	AET3Gv20956800	Exocyst complex component Exo70
	AET3Gv20956900	CBS domain-containing protein
	AET3Gv20957000	Mitochondrial import receptor subunit TOM20
	PGSB gene 29622	Nucleotide-diphospho-sugar transferase domain-containing protein
	AET3Gv20958300	F-box protein
	AET3Gv20958400	FAR1-related sequence protein
	AET3Gv20958700	CBS domain-containing protein
	AET3Gv20958800	NB-ARC domain-containing protein
	AET3Gv20959200	Uncharacterised disordered protein
	AET3Gv20960100	Squamosa promoter-binding-like protein
	AET3Gv20960300	Uncharacterised - no annotated domains
	AET3Gv20960400	Uncharacterised - no annotated domains
	AET3Gv20960500	Uncharacterised - no annotated domains
	AET3Gv20960600	Uncharacterised - no annotated domains
	AET3Gv20961500	Rhodanese-like domain containing protein
	AET3Gv20962100	Ribosomal protein S5 domain
	AET3Gv20962900	Histidine kinase

Table S4 Genes located under the B-type granule diameter GWAS peaks. (continued)

Peak from GWAS	Gene	Predicted protein
B (continued)	*AET3Gv20963100*	CG domain containing protein
	AET3Gv20963200	CBS domain containing protein
	AET3Gv20963300	GNK2 domain containing protein
	AET3Gv20963400	DUF3339 domain containing protein
	AET3Gv20963700	Exportin-5 C-terminal domain containing protein
	AET3Gv20963900	Uncharacterised - no annotated domains
	AET3Gv20964200	Exportin-5 C-terminal domain containing protein
	AET3Gv20964500	Exportin-5 C-terminal domain containing protein
	AET3Gv20964800	Transmembrane protein
	AET3Gv20965000	MADS-box domain containing protein
	AET3Gv20965600	DNA/RNA polymerase superfamily
	AET3Gv20966600	DNAJ and DUF3444 domain containing proteins
	AET3Gv20966700	Stress enhanced protein 1
	AET3Gv20966900	F-box only protein 6
	AET3Gv20967200	50S ribosomal protein L27
	AET3Gv20968100	DUF4408 domain containing protein
	AET3Gv20968800	WD40 repeat containing protein
	AET3Gv20969100	TCP domain containing protein
	AET3Gv20969600	PRP8 domain containing protein
	AET3Gv20969900	GYF domain containing protein
	AET3Gv20970300	Uncharacterised - no annotated domains
	AET3Gv20970400	Uncharacterised - no annotated domains
	AET3Gv20970900	DEK C-terminal domain containing protein
	AET3Gv20971200	Uncharacterised - no annotated domains
	AET3Gv20971300	Auxin-inducible protein
	AET3Gv20971400	CID domain containing protein
	AET3Gv20972000	Glutathione S-transferase
	AET3Gv20972100	Glutathione S-transferase
	AET3Gv20972200	Glutathione transferase
	AET3Gv20972300	Peptidase C14 caspase domain-containing protein
	AET3Gv20972500	NB-ARC domain-containing protein

Table S4 Genes located under the B-type granule diameter GWAS peaks. (continued)

Peak from GWAS	Gene	Predicted protein
B (continued)	*AET3Gv20972600*	Glutathione transferase
	AET3Gv20972700	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase
	AET3Gv20973000	Enoyl-CoA hydratase
	AET3Gv20973100	Enoyl-CoA hydratase
	AET3Gv20973600	Uncharacterised - no annotated domains
	AET3Gv20973900	NB-ARC domain-containing protein
	PGSB gene 29757	Glutamine cyclotransferase
	AET3Gv20974000	DUF1719 domain containing protein
	AET3Gv20974100	DUF3245 domain containing protein
	AET3Gv20974300	Coiled-coil domain containing protein
	AET3Gv20974400	GDSL esterase/lipase
	AET3Gv20974500	Acyl-CoA synthase domain containing protein
	AET3Gv20974600	NAC domain containing protein
	AET3Gv20975000	2-hydroxy-palmitic acid dioxygenase Mpo1-like
	AET3Gv20975200	Protein kinase domain containing protein
	AET3Gv20975300	DNA/RNA binding protein with an Alba like domain
	AET3Gv20975600	E2 ubiquitin-conjugating enzyme
	AET3Gv20975700	Myb transcription factor
	AET3Gv20976100	Uncharacterised - no annotated domains
	AET3Gv20976200	RING-domain containing protein
	AET3Gv20976400	Transaldolase
	AET3Gv20976500	Exostosin GT47 domain containing protein
	AET3Gv20976600	RING-type E3 ubiquitin transferase
	AET3Gv20976700	Exostosin GT47 domain containing protein
	AET3Gv20976800	Micro-fibrillar-associated-protein 1C-terminal domain-containing protein
	AET3Gv20977700	UDP-Glycosyltransferase
	AET3Gv20977900	CRAL-TRIO domain-containing protein
	AET3Gv20978000	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain-containing protein

Table S4 Genes located under the B-type granule diameter GWAS peaks. (continued)

Peak from GWAS	Gene	Predicted protein
B (continued)	*AET3Gv20978300*	Histone-lysine N-methyltransferase
	AET3Gv20978400	Myb transcription factor
	AET3Gv20978500	Acyl transferase
	AET3Gv20978700	HMA-domain containing protein
	AET3Gv20978900	DNAJ domain containing protein
	AET3Gv20979000	Inactive leucine-rich repeat receptor-like protein kinase CORYNE
	AET3Gv20979200	Auxin response factor
	AET3Gv20979400	Aspartokinase
	AET3Gv20979600	Transcription factor ICE1
	AET3Gv20979800	Coiled-coil domain containing protein SCD2
	AET3Gv20980300	Knottin, scorpion toxin-like protein
	AET3Gv20980400	Knottin, scorpion toxin-like protein
	AET3Gv20980600	BHLH domain-containing protein
	AET3Gv20981100	Knottin, scorpion toxin-like protein
	AET3Gv20981200	Glutamine cyclotransferase
	AET3Gv20981400	DUF3741 domain-containing protein
C	AET5Gv20149000	E3 ubiquitin-protein ligase RMA
	AET5Gv20149200	Transcription factor PIF-1 related
	AET5Gv20149100	Trypsin-like serine protease
	AET5Gv20150100	Uncharacterised - no annotated domains
D	*AET7Gv20332600*	Protein detoxification
	AET7Gv20332900	Zinc finger domain containing protein
	AET7Gv20333100	Adenylyltransferase and sulfurtransferase MOCS3-1
	AET7Gv20333300	NB-ARC domain containing protein
	AET7Gv20333400	Uncharacterised disordered protein
	AET7Gv20333600	Glutathione transferase
	AET7Gv20333700	Zinc finger and CCT domain containing protein
	AET7Gv20333900	NB-ARC domain containing protein
	AET7Gv20334200	DUF569 domain containing protein

Table S4 Genes located under the B-type granule diameter GWAS peaks. (continued)

Peak from GWAS	Gene	Predicted protein
D (continued)	*AET7Gv20334300*	Tetratricopeptide repeat containing protein
	AET7Gv20334400	Cdk-activating kinase assembly factor MAT1 centre domain-containing protein
	AET7Gv20334900	WD repeat containing protein
	AET7Gv20334700	Fanconi-associated nuclease
	AET7Gv20335100	DUF761 domain containing protein
	AET7Gv20335500	DUF761 domain containing protein
	AET7Gv20335600	Suppressor of the G2 allele of *skp1* homolog
	AET7Gv20335700	Pentatricopeptide repeat-containing protein
	AET7Gv20336000	EF-hand domain containing protein
	AET7Gv20336100	Ribosomal protein S30
	AET7Gv20336200	HNH endonuclease
	AET7Gv20336800	Pullulanase/Limit dextrinase
	AET7Gv20336700	Uncharacterised - no annotated domains
	AET7Gv20337500	DMP3-like
	AET7Gv20337600	L-gluconolactone oxidase
	AET7Gv20338500	Calcium binding protein
	AET7Gv20338600	Galactinol-sucrose galactosyltransferase
	AET7Gv20338700	Telomere associated protein Rif1 N terminal domain containing protein
	AET7Gv20338800	AT-hook motif containing protein
	AET7Gv20339100	Exocyst subunit Exo70 family protein
	AET7Gv20339300	Glycoside hydrolase family 31 N-terminal domain-containing protein
	AET7Gv20339500	HAT C-terminal dimerization domain-containing protein
	AET7Gv20339700	DDE Tnp4 domain-containing protein
	AET7Gv20339600	Uncharacterised - no annotated domains
	AET7Gv20340000	Coiled-coil domain containing protein
	AET7Gv20340100	AT- hook motif containing protein
	AET7Gv20340400	DUF569 domain containing protein

Table S4 Genes located under the B-type granule diameter GWAS peaks. (continued)

Peak from GWAS	Gene	Predicted protein
D (continued)	*AET7Gv20340300*	Uncharacterised - no annotated domains
	AET7Gv20340900	Transposase MuDR plant domain containing protein
	AET7Gv20341600	Pentatricopeptide repeat-containing protein
	AET7Gv20341500	Leucine rich repeat protein kinase
	AET7Gv20341700	Uncharacterised - no annotated domains
	AET7Gv20341800	FAR1 DNA binding domain protein
	AET7Gv20342000	Uncharacterised - no annotated domains
	AET7Gv20343000	Myb transcription factor
	AET7Gv20343800	Sulfiredoxin
	AET7Gv20344000	Uncharacterised - no annotated domains
	AET7Gv20344200	Myb transcription factor
	AET7Gv20344800	Sulfiredoxin
	AET7Gv20345000	Developmental and cell death domain containing protein
	AET7Gv20345400	Glycine-rich cell wall structural protein
	AET7Gv20346000	v-SNARE coiled-coil homology domain containing protein
	AET7Gv20345600	Developmental and cell death domain containing protein
	AET7Gv20346100	4-hydroxyphenylacetaldehyde oxime monooxygenase
	AET7Gv20346200	COBRA-like protein 10
	AET7Gv20347000	Wall associated receptor kinase
	AET7Gv20347100	Transmembrane protein

Table S5 Genes located under the B-type granule number GWAS peaks. Genes underlying the peaks identified in the GWAS for B-type granule number (Figure 4.2). Gene names are shown according to the v4 genome nomenclature, except where the gene was not annotated in the v4 genome assembly, in which case the v5 gene name is shown, Genes are listed in the order they appear on the forward strand and the peaks are listed based on which occurs first on the third chromosome.

Peak from GWAS	Gene	Predicted protein
V	*AET3Gv20188700*	RAP domain-containing protein
	AET3Gv20188900	FBD domain containing protein
	AET3Gv20189300	GDSL esterase
	AET3Gv20189500	Transcription factor GTE8
	AET3Gv20190600	Proteinase inhibitor I13
	AET3Gv20190800	Proteinase inhibitor I13
	AET3Gv20190900	Proteinase inhibitor I13
	AET3Gv20191000	Proteinase inhibitor I13
	PGSB gene 23181	Coiled-coil domain protein
	AET3Gv20191100	Histone H2B
	AET3Gv20191300	Uncharacterised - no annotated domains
	AET3Gv20191400	Triacylglycerol lipase OBL1
	AET3Gv20191600	Aminoacyl-transfer RNA synthetases class-II family profile domain-containing protein
	AET3Gv20191700	Signal peptidase complex catalytic subunit SEC11A-like
	AET3Gv20191900	DUF674 domain containing protein
	AET3Gv20192200	Protein trichome birefringence-like, N-terminal domain
	AET3Gv20192600	DUF674 domain containing protein
	AET3Gv20192700	DUF674 domain containing protein
	AET3Gv20193700	DUF674 domain containing protein
	AET3Gv20194600	NB-ARC domain containing protein
	AET3Gv20194700	NB-ARC domain containing protein
	PGSB gene 23212	DUF1677-containing protein
	AET3Gv20195000	THIF-type NAD/FAD binding fold domain-containing protein
	AET3Gv20196600	NADH-ubiquinone oxidoreductase chain 6-like

Table S5 Genes located under the B-type granule number GWAS peaks. (continued)

Peak from GWAS	Gene	Predicted protein
V (continued)	*AET3Gv20197700*	Transmembrane domain containing protein
W	*AET3Gv20318700*	Cytochrome P450 710A1
	AET3Gv20318800	Phospho-N-acetylmuramoyl-pentapeptide-transferase homolog
	AET3Gv20319100	Phytocyanin domain containing protein
	AET3Gv20319200	Potassium channel
	AET3Gv20319400	DUF538 domain containing protein
	AET3Gv20319500	B3 DNA binding domain protein
	AET3Gv20319800	SOUL-haem binding domain containing protein
	AET3Gv20319900	DUF616 domain containing protein
	AET3Gv20320100	Uncharacterised - no annotated domains
	AET3Gv20320700	Cationic amino acid transporter 5
	AET3Gv20320800	Late embryogenesis abundant protein
	AET3Gv20321000	Probable tRNA N6-adenosine threonylcarbamoyltransferase

Table S6 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule diameter. WoLF PSORT was used to predict protein localisation of the protein encoded by the transcript which was labelled in Ensembl Plants as the canonical transcript or in cases where the canonical transcript was poorly predicted, the transcript most similar to the canonical wheat ortholog. For some genes it was not possible to predict the canonical transcript and predicted protein localisation could not be determined, these are shown below as 'Could not predict'.

Peak from GWAS	Gene	Predicted localisation
B	*AET3Gv20949600*	Could not predict
	AET3Gv20949700	Could not predict
	AET3Gv20949800	Could not predict
	AET3Gv20949900	Could not predict
	AET3Gv20950200	Could not predict
	AET3Gv20950500	Could not predict
	AET3Gv20950900	Could not predict
	AET3Gv20951000	Could not predict
	AET3Gv20951100	Plastid/Extracellular
	AET3Gv20951900	Plastid
	AET3Gv20952200	Nucleus
	AET3Gv20952500	Could not predict
	AET3Gv20952600	Could not predict
	AET3Gv20952700	Cytoplasm
	AET3Gv20953200	Could not predict
	AET3Gv20953300	Plastid
	AET3Gv20953700	Plastid
	AET3Gv20954100	Plastid
	AET3Gv20954200	Nucleus
	AET3Gv20954300	Cytoplasm
	AET3Gv20954500	Could not predict
	AET3Gv20954600	Could not predict
	PGSB gene 29601	Could not predict
	AET3Gv20954700	Extracellular

Table S6 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule diameter. (continued)

Peak from GWAS	Gene	Predicted localisation
B (continued)	*AET3Gv20954800*	Could not predict
	AET3Gv20954900	Could not predict
	AET3Gv20955300	Plastid
	AET3Gv20955400	Plastid
	AET3Gv20955500	Plastid
	AET3Gv20956200	Mitochondria
	AET3Gv20956300	Cytoplasm
	AET3Gv20956400	Cytoplasm
	AET3Gv20956800	Cytoplasm
	AET3Gv20956900	Peroxisome
	AET3Gv20957000	Could not predict
	PGSB gene 29622	Could not predict
	AET3Gv20958300	Could not predict
	AET3Gv20958400	Could not predict
	AET3Gv20958700	Cytoplasm
	AET3Gv20958800	Could not predict
	AET3Gv20959200	Nucleus
	AET3Gv20960100	Nucleus
	AET3Gv20960300	Could not predict
	AET3Gv20960400	Plastid
	AET3Gv20960500	Plastid
	AET3Gv20960600	Could not predict
	AET3Gv20961500	Plastid
	AET3Gv20962100	Could not predict
	AET3Gv20962900	Could not predict
	AET3Gv20963100	Cytoplasm
	AET3Gv20963200	Plastid
	AET3Gv20963300	Endoplasmic reticulum
	AET3Gv20963400	Vacuole

Table S6 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule diameter. (continued)

Peak from GWAS	Gene	Predicted localisation
B (continued)	*AET3Gv20963700*	Nucleus
	AET3Gv20963900	Cytoplasm
	AET3Gv20964200	Cytoplasm
	AET3Gv20964500	Cytoplasm
	AET3Gv20964800	Could not predict
	AET3Gv20965000	Nucleus
	AET3Gv20965600	Could not predict
	AET3Gv20966600	Nucleus
	AET3Gv20966700	Plastid
	AET3Gv20966900	Plasma membrane
	AET3Gv20967200	Plastid
	AET3Gv20968100	Plasma membrane
	AET3Gv20968800	Plastid
	AET3Gv20969100	Nucleus
	AET3Gv20969600	Cytoplasm
	AET3Gv20969900	Nucleus
	AET3Gv20970300	Nucleus
	AET3Gv20970400	Could not predict
	AET3Gv20970900	Could not predict
	AET3Gv20971200	Could not predict
	AET3Gv20971300	Plastid
	AET3Gv20971400	Nucleus
	AET3Gv20972000	Plastid
	AET3Gv20972100	Cytoplasm
	AET3Gv20972200	Cytoplasm
	AET3Gv20972300	Plastid
	AET3Gv20972500	Endoplasmic reticulum
	AET3Gv20972600	Cytoplasm
	AET3Gv20972700	Peroxisome
	AET3Gv20973000	Could not predict

Table S6 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule diameter. (continued)

Peak from GWAS	Gene	Predicted localisation
B (continued)	*AET3Gv20973100*	Cytoplasm
	AET3Gv20973600	Could not predict
	AET3Gv20973900	Plastid
	PGSB gene 29757	Could not predict
	AET3Gv20974000	Nucleus
	AET3Gv20974100	Nucleus
	AET3Gv20974300	Nucleus
	AET3Gv20974400	Peroxisome
	AET3Gv20974500	Plasma membrane
	AET3Gv20974600	Nucleus
	AET3Gv20975000	Plasma membrane
	AET3Gv20975200	Plastid
	AET3Gv20975300	Cytoplasm
	AET3Gv20975600	Plastid
	AET3Gv20975700	Nucleus
	AET3Gv20976100	Plastid
	AET3Gv20976200	Nucleus
	AET3Gv20976400	Plastid
	AET3Gv20976500	Vacuole
	AET3Gv20976600	Nucleus
	AET3Gv20976700	Vacuole
	AET3Gv20976800	Nucleus
	AET3Gv20977700	Cytoplasm
	AET3Gv20977900	Cytoplasm
	AET3Gv20978000	Extracellular
	AET3Gv20978300	Cytoplasm
	AET3Gv20978400	Nucleus
	AET3Gv20978500	Plastid
	AET3Gv20978700	Nucleus
	AET3Gv20978900	Could not predict
	AET3Gv20979000	Could not predict

Table S6 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule diameter. (continued)

Peak from GWAS	Gene	Predicted localisation
B (continued)	*AET3Gv20979200*	Nucleus
	AET3Gv20979400	Plastid
	AET3Gv20979600	Nucleus
	AET3Gv20979800	Could not predict
	AET3Gv20980300	Extracellular
	AET3Gv20980400	Extracellular
	AET3Gv20980600	Nucleus
	AET3Gv20981100	Extracellular
	AET3Gv20981200	Could not predict
	AET3Gv20981400	Nucleus
C	AET5Gv20149000	Nucleus
	AET5Gv20149200	Nucleus
	AET5Gv20149100	Cytoplasm
	AET5Gv20150100	Cytoplasm
D	*AET7Gv20332600*	Plasma membrane
	AET7Gv20332900	Could not predict
	AET7Gv20333100	Cytoplasm
	AET7Gv20333300	Plastid
	AET7Gv20333400	Nucleus
	AET7Gv20333600	Cytoplasm
	AET7Gv20333700	Could not predict
	AET7Gv20333900	Nucleus
	AET7Gv20334200	Could not predict
	AET7Gv20334300	Could not predict
	AET7Gv20334400	Could not predict
	AET7Gv20334900	Could not predict
	AET7Gv20334700	Plastid
	AET7Gv20335100	Plastid
	AET7Gv20335500	Plastid
	AET7Gv20335600	Could not predict
	AET7Gv20335700	Could not predict

Table S6 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule diameter. (continued)

Peak from GWAS	Gene	Predicted localisation
D (continued)	*AET7Gv20336000*	Plastid
	AET7Gv20336100	Nucleus
	AET7Gv20336200	Plastid
	AET7Gv20336800	Plastid
	AET7Gv20336700	Extracellular
	AET7Gv20337500	Could not predict
	AET7Gv20337600	Could not predict
	AET7Gv20338500	Nucleus
	AET7Gv20338600	Plastid
	AET7Gv20338700	Nucleus
	AET7Gv20338800	Cytoplasm
	AET7Gv20339100	Could not predict
	AET7Gv20339300	Nucleus
	AET7Gv20339500	Plastid
	AET7Gv20339700	Could not predict
	AET7Gv20339600	Plastid
	AET7Gv20340000	Nucleus
	AET7Gv20340100	Nucleus
	AET7Gv20340400	Mitochondria
	AET7Gv20340300	Could not predict
	AET7Gv20340900	Cytoplasm
	AET7Gv20341600	Mitochondria
	AET7Gv20341500	Plasma membrane
	AET7Gv20341700	Could not predict
	AET7Gv20341800	Plastid
	AET7Gv20342000	Could not predict
	AET7Gv20343000	Nucleus
	AET7Gv20343800	Cytoplasm
	AET7Gv20344000	Cytoplasm
	AET7Gv20344200	Nucleus

Table S6 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule diameter. (continued)

Peak from GWAS	Gene	Predicted localisation
D (continued)	*AET7Gv20344800*	Plastid
	AET7Gv20345000	Could not predict
	AET7Gv20345400	Could not predict
	AET7Gv20346000	Plastid
	AET7Gv20345600	Could not predict
	AET7Gv20346100	Cytoplasm
	AET7Gv20346200	Could not predict
	AET7Gv20347000	Plasma membrane
	AET7Gv20347100	Mitochondria

Table S7 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule number. WoLF PSORT was used to predict protein localisation of the protein encoded by the transcript which was labelled in Ensembl Plants as the canonical transcript or in cases where the canonical transcript was poorly predicted, the transcript most similar to the canonical wheat ortholog. For some genes it was not possible to predict the canonical transcript and predicted protein localisation could not be determined, these are shown below as 'Could not predict'.

Peak from GWAS	Gene	Predicted localisation
V	*AET3Gv20188700*	Plastid
	AET3Gv20188900	Could not predict
	AET3Gv20189300	Vacuole
	AET3Gv20189500	Nucleus
	AET3Gv20190600	Plastid
	AET3Gv20190800	Plastid
	AET3Gv20190900	Plastid
	AET3Gv20191000	Could not predict
	PGSB gene 23181	Could not predict
	AET3Gv20191100	Could not predict
	AET3Gv20191300	Could not predict
	AET3Gv20191400	Plasma membrane
	AET3Gv20191600	Could not predict
	AET3Gv20191700	Extracellular
	AET3Gv20191900	Peroxisome
	AET3Gv20192200	Extracellular
	AET3Gv20192600	Could not predict
	AET3Gv20192700	Plastid
	AET3Gv20193700	Could not predict
	AET3Gv20194600	Endoplasmic reticulum
	AET3Gv20194700	Plastid
	PGSB gene 23212	Could not predict
	AET3Gv20195000	Could not predict
	AET3Gv20196600	Cytoplasm
	AET3Gv20197700	Nucleus

Table S7 Localisation of the protein encoded by the genes under the GWAS peaks associated with B-type granule number. (continued)

Peak from GWAS	Gene	Predicted localisation
W	*AET3Gv20318700*	Endoplasmic reticulum
	AET3Gv20318800	Plasma membrane
	AET3Gv20319100	Plastid
	AET3Gv20319200	Plasma membrane
	AET3Gv20319400	Cytoplasm
	AET3Gv20319500	Cytoplasm
	AET3Gv20319800	Extracellular
	AET3Gv20319900	Mitochondria
	AET3Gv20320100	Plastid
	AET3Gv20320700	Plasma membrane
	AET3Gv20320800	Nucleus
	AET3Gv20321000	Plastid

Table S8 Putative wheat orthologs of genes under peak B associated with B-type granule diameter. Blast searches were used to identify the putative *T. aestivum* and *T. turgidum* orthologs of the genes under peak B associated with B-type granule diameter (Figure 4.1 and Table S4). n/a = genes with no putative ortholog, some genes had more than one putative ortholog in wheat and in these cases all putative orthologs are listed. Genes with endosperm expression, for *T. aestivum* where the mean TPM ≥ 1 and for *T. turgidum* where the mean TPM ≥ 1 for at least one timepoint in the dataset from Chen *et al.* (2023a), are in bold.

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20949600	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20949700	*TraesCS3A02G427000*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20949800	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20949900	*n/a*	*n/a*	*TraesCS3D02G422100*	*n/a*	*n/a*
AET3Gv20950200	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20950500	*TraesCS3A02G426900*	*TraesCS3B02G463600*	*TraesCS3D02G422400*	*n/a*	***TRITD3Bv1G235790***
AET3Gv20950900	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20951000	*n/a*	*TraesCS3B02G463700*	*TraesCS3D02G422600*	*n/a*	***TRITD3Bv1G235900***
AET3Gv20951100	*TraesCS3A02G427100*	*TraesCS3B02G463800* *TraesCS3B02G463400*	*TraesCS3D02G422700*	***TRITD3Av1G244480***	***TRITD3Bv1G235920*** ***TRITD3Bv1G235530***
AET3Gv20951900	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20952200	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20952500	*TraesCS3A02G427300*	*TraesCS3B02G464200*	*TraesCS3D02G422900*	*TRITD3Av1G244660*	*TRITD3Bv1G236040*
AET3Gv20952600	*TraesCS3A02G427400*	*TraesCS3B02G464400*	*TraesCS3D02G423000*	*TRITD3Av1G244680*	*TRITD3Bv1G236160*
AET3Gv20952700	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20953200	*TraesCS3A02G427500*	*TraesCS3B02G464600*	*TraesCS3D02G423100*	***TRITD3Av1G244740***	***TRITD3Bv1G236270***
AET3Gv20953300	***TraesCS3A02G427600***	***TraesCS3B02G464700***	***TraesCS3D02G423200***	***TRITD3Av1G244760***	***TRITD3Bv1G236280***
AET3Gv20953700	*n/a*	*n/a*	*TraesCS3D02G423300*	*TRITD3Av1G244830*	*TRITD3Bv1G236290*

Table S8 Putative wheat orthologs of genes under peak B associated with B-type granule diameter. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20954100	***TraesCS3A02G427900***	*TraesCS3B02G465000* *TraesCS3B02G464900*	*TraesCS3D02G423400*	***TRITD3Av1G244950***	*TRITD3Bv1G236500*
AET3Gv20954200	***TraesCS3A02G428000***	***TraesCS3B02G465100***	***TraesCS3D02G423500***	***TRITD3Av1G244960***	***TRITD3Bv1G236520***
AET3Gv20954300	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20954500	*n/a*	*TraesCS3B02G465300*	*TraesCS3D02G423700*	***TRITD3Av1G245020***	*TRITD3Bv1G236580*
AET3Gv20954600	*TraesCS3A02G428100*	*TraesCS3B02G465200*	*TraesCS3D02G423800*	*TRITD3Av1G244980*	*TRITD3Bv1G236560*
PGSB gene 29601	***TraesCS3A02G438700***	***TraesCS3B02G472600***	***TraesCS3D02G431100***	***TRITD3Av1G248980***	***TRITD3Bv1G240580***
AET3Gv20954700	*TraesCS3A02G428300*	*TraesCS3B02G465700*	*TraesCS3D02G424000*	*TRITD3Av1G245040*	*TRITD3Bv1G236660*
AET3Gv20954800	***TraesCS3A02G428400***	***TraesCS3B02G465800***	***TraesCS3D02G424100***	***TRITD3Av1G245050***	***TRITD3Bv1G236670***
AET3Gv20954900	*TraesCS3A02G428500*	*TraesCS3B02G466000*	*TraesCS3D02G424200*	*TRITD3Av1G245060*	*n/a*
AET3Gv20955300	*TraesCS3A02G428700*	*TraesCS3B02G466300*	*TraesCS3D02G424300*	*TRITD3Av1G245100*	*TRITD3Bv1G236860*
AET3Gv20955400	***TraesCS3A02G428800***	***TraesCS3B02G466400***	***TraesCS3D02G424400***	***TRITD3Av1G245260***	***TRITD3Bv1G236900***
AET3Gv20955500	***TraesCS3A02G428900***	***TraesCS3B02G466500***	***TraesCS3D02G424500***	***TRITD3Av1G245270***	***TRITD3Bv1G236910***
AET3Gv20956200	***TraesCS3A02G429100***	*TraesCS3B02G466800*	***TraesCS3D02G424600***	***TRITD3Av1G245290***	*n/a*
AET3Gv20956300	*TraesCS3A02G429200*	*TraesCS3B02G466900*	*TraesCS3D02G424700*	*TRITD3Av1G245330*	*n/a*
AET3Gv20956400	*TraesCS3A02G429300*	*TraesCS3B02G467000*	*TraesCS3D02G424800*	*TRITD3Av1G245340*	*TRITD3Bv1G236970*
AET3Gv20956800	***TraesCS3A02G429600***	***TraesCS3B02G467500***	***TraesCS3D02G424900***	***TRITD3Av1G245590***	***TRITD3Bv1G237330***
AET3Gv20956900	*TraesCS3A02G429700*	*TraesCS3B02G467600*	*TraesCS3D02G425000*	*TRITD3Av1G245600*	*TRITD3Bv1G237340*
AET3Gv20957000	***TraesCS3A02G429800***	***TraesCS3B02G467700***	***TraesCS3D02G425100***	***TRITD3Av1G245610***	***TRITD3Bv1G237360***
PGSB gene 29622	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20958300	***TraesCS3A02G429900***	***TraesCS3B02G467900***	***TraesCS3D02G425300***	***TRITD3Av1G245750***	***TRITD3Bv1G237570***
AET3Gv20958400	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20958700	*TraesCS3A02G432100*	*TraesCS3B02G468100*	*TraesCS3D02G425400*	*TRITD3Av1G245990*	*TRITD3Bv1G237930*

Table S8 Putative wheat orthologs of genes under peak B associated with B-type granule diameter. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20958800	*TraesCS3A02G432200*	*TraesCS3B02G468200*	*TraesCS3D02G425500*	*TRITD3Av1G246000*	*TRITD3Bv1G237940*
AET3Gv20959200	*TraesCS3A02G432400*	*TraesCS3B02G468300*	*TraesCS3D02G425600*	***TRITD3Av1G246020***	***TRITD3Bv1G238310***
AET3Gv20960100	*TraesCS3A02G432500*	*n/a*	*TraesCS3D02G425800*	*TRITD3Av1G246060*	*n/a*
AET3Gv20960300	*TraesCS3A02G432600*	*TraesCS3B02G468500*	*TraesCS3D02G425900*	*TRITD3Av1G246080*	*TRITD3Bv1G238110*
AET3Gv20960400	*n/a*	*TraesCS3B02G468600*	*TraesCS3D02G426000*	*n/a*	*TRITD3Bv1G238090*
AET3Gv20960500	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20960600	*TraesCS3A02G432700* *TraesCS3A02G432800*	*TraesCS3B02G468700* *TraesCS3B02G468800*	*TraesCS3D02G426200*	*n/a*	*TRITD3Bv1G238340* *TRITD3Bv1G238360*
AET3Gv20961500	*TraesCS3A02G433000*	*TraesCS3B02G468900*	*TraesCS3D02G426300*	***TRITD3Av1G246220***	*TRITD3Bv1G238520*
AET3Gv20962100	*n/a*	*n/a*	*n/a*	*n/a*	***TRITD2Bv1G212980***
AET3Gv20962900	***TraesCS3A02G433200***	***TraesCS3B02G469000***	***TraesCS3D02G426600***	***TRITD3Av1G246370***	*TRITD3Bv1G238860*
AET3Gv20963100	***TraesCS3A02G433300***	***TraesCS3B02G469100***	***TraesCS3D02G426700***	***TRITD3Av1G246380***	***TRITD3Bv1G238870***
AET3Gv20963200	***TraesCS3A02G433400***	***TraesCS3B02G469200***	*TraesCS3D02G426800*	***TRITD3Av1G246420***	***TRITD3Bv1G238940***
AET3Gv20963300	*TraesCS3A02G433500*	*TraesCS3B02G469400*	*TraesCS3D02G426900*	*TRITD3Av1G246460*	*TRITD3Bv1G239040*
AET3Gv20963400	***TraesCS3A02G433700***	*TraesCS3B02G469500*	*TraesCS3D02G427000*	***TRITD3Av1G246490***	***TRITD3Bv1G239050***
AET3Gv20963700	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20963900	*n/a*	*n/a*	*TraesCS3D02G427200*	*n/a*	*n/a*
AET3Gv20964200	***TraesCS3A02G433900***	*n/a*	***TraesCS3D02G427300***	***TRITD3Av1G246680***	*n/a*
AET3Gv20964500	*n/a*	*n/a*	*n/a*	*TRITD3Av1G246720*	*n/a*
AET3Gv20964800	*TraesCS3A02G434300*	***TraesCS3B02G469600***	*TraesCS3D02G427500*	***TRITD3Av1G246730***	***TRITD3Bv1G239110***
AET3Gv20965000	*TraesCS3A02G435000*	*TraesCS3B02G470000*	*TraesCS3D02G428000*	*TRITD3Av1G247000*	*TRITD3Bv1G239410*
AET3Gv20965600	*n/a*	*n/a*	*n/a*	***TRITD1Av1G045900***	*n/a*
AET3Gv20966600	*TraesCS3A02G435100*	*TraesCS3B02G470100*	*n/a*	*TRITD3Av1G247220*	*TRITD3Bv1G239490*

Table S8 Putative wheat orthologs of genes under peak B associated with B-type granule diameter. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20966700	*TraesCS3A02G435300*	*TraesCS3B02G470200*	*TraesCS3D02G428100*	***TRITD3Av1G247480***	*n/a*
AET3Gv20966900	***TraesCS3A02G435400***	***TraesCS3B02G470300***	***TraesCS3D02G428200***	***TRITD3Av1G247490***	***TRITD3Bv1G239650***
AET3Gv20967200	***TraesCS3A02G435700***	***TraesCS3B02G470400***	***TraesCS3D02G428300***	***TRITD3Av1G247910***	***TRITD3Bv1G239710***
AET3Gv20968100	*TraesCS3A02G435800*	***TraesCS3B02G470500***	*TraesCS3D02G428400*	*TRITD3Av1G247930*	***TRITD3Bv1G239720***
AET3Gv20968800	***TraesCS3A02G436000***	***TraesCS3B02G470600***	***TraesCS3D02G428600***	*TRITD3Av1G247970*	***TRITD3Bv1G239740***
AET3Gv20969100	***TraesCS3A02G436100***	***TraesCS3B02G470700***	***TraesCS3D02G428700***	***TRITD3Av1G248050***	***TRITD3Bv1G239750***
AET3Gv20969600	*n/a*	*TraesCS1B02G415700* *TraesCS1B02G416400* *TraesCS1B02G368800*	***TraesCS3D02G428800***	*n/a*	***TRITD1Bv1G209260*** ***TRITD1Bv1G193770*** *TRITD1Bv1G209430*
AET3Gv20969900	***TraesCS3A02G436500***	***TraesCS3B02G470800***	***TraesCS3D02G428900***	***TRITD3Av1G248320***	***TRITD3Bv1G239880***
AET3Gv20970300	*TraesCS3A02G436600*	***TraesCS3B02G470900***	***TraesCS3D02G429100***	***TRITD3Av1G248460***	***TRITD3Bv1G239920***
AET3Gv20970400	*n/a*	*TraesCS3B02G471000*	*n/a*	*n/a*	*TRITD3Bv1G239930*
AET3Gv20970900	***TraesCS3A02G436900***	***TraesCS3B02G471100***	***TraesCS3D02G429200***	***TRITD3Av1G248490***	*n/a*
AET3Gv20971200	*n/a*	*TraesCS3B02G471300*	*TraesCS3D02G429300*	*n/a*	*n/a*
AET3Gv20971300	*TraesCS3A02G437100*	*n/a*	*TraesCS3D02G429400*	*TRITD3Av1G248570*	*TRITD3Bv1G240120*
AET3Gv20971400	***TraesCS3A02G437300***	***TraesCS3B02G471400***	***TraesCS3D02G429500***	***TRITD3Av1G248640***	***TRITD3Bv1G240130***
AET3Gv20972000	*n/a*	*n/a*	*TraesCS3D02G429600*	*n/a*	*n/a*
AET3Gv20972100	*TraesCS3A02G437400*	*TraesCS3B02G471500*	*TraesCS3D02G429700*	*TRITD3Av1G248730*	*TRITD3Bv1G240200*
AET3Gv20972200	*TraesCS3A02G437500*	*TraesCS3B02G471600*	*TraesCS3D02G429800*	*TRITD3Av1G248740*	*TRITD3Bv1G240210*
AET3Gv20972300	*TraesCS3A02G437700*	*TraesCS3B02G471700*	*TraesCS3D02G430000*	*TRITD3Av1G248760*	*TRITD3Bv1G240240*
AET3Gv20972500	*TraesCS3A02G437800*	*TraesCS3B02G471800*	*TraesCS3D02G430100*	*TRITD3Av1G248790*	*TRITD3Bv1G240250*
AET3Gv20972600	*TraesCS3A02G437900*	*TraesCS3B02G471900*	*TraesCS3D02G430200*	*TRITD3Av1G248840*	*TRITD3Bv1G240270*
AET3Gv20972700	***TraesCS3A02G438000***	***TraesCS3B02G472000***	***TraesCS3D02G430300***	***TRITD3Av1G248850***	***TRITD3Bv1G240280***

Table S8 Putative wheat orthologs of genes under peak B associated with B-type granule diameter. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20973000	***TraesCS3A02G438100***	*n/a*	***TraesCS3D02G430400***	*n/a*	*n/a*
AET3Gv20973100	*TraesCS3A02G438300*	*TraesCS3B02G472200*	*TraesCS3D02G430500*	*TRITD3Av1G248880*	*TRITD3Bv1G240470*
AET3Gv20973600	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20973900	*TraesCS3A02G438600*	*n/a*	*n/a*	*TRITD3Av1G248960*	*n/a*
PGSB gene 29757	*n/a*	*TraesCS3B02G105400*	*TraesCS3D02G090100*	*n/a*	*TRITD3Bv1G029960*
AET3Gv20974000	*TraesCS3A02G438800*	*TraesCS3B02G472700*	*TraesCS3D02G431200*	*n/a*	*TRITD3Bv1G240600*
AET3Gv20974100	*TraesCS4A02G399800*	*n/a*	*TraesCS3D02G431300*	***TRITD4Av1G231640***	*n/a*
AET3Gv20974300	*TraesCS4A02G399700*	*n/a*	*TraesCS3D02G431400*	*TRITD4Av1G231630*	*n/a*
AET3Gv20974400	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20974500	*n/a*	*n/a*	*TraesCS3D02G431600*	*n/a*	*n/a*
AET3Gv20974600	***TraesCS3A02G438900***	***TraesCS3B02G472800***	***TraesCS3D02G431700***	***TRITD3Av1G249030***	***TRITD3Bv1G240700***
AET3Gv20975000	***TraesCS3A02G439000***	***TraesCS3B02G473000***	***TraesCS3D02G431800***	***TRITD3Av1G249250***	***TRITD3Bv1G240920***
AET3Gv20975200	*TraesCS3A02G439200*	*TraesCS3B02G473100*	*TraesCS3D02G432000*	*TRITD3Av1G249280*	*TRITD3Bv1G240940*
AET3Gv20975300	*TraesCS3A02G439300*	*TraesCS3B02G473200*	*TraesCS3D02G432100*	*TRITD3Av1G249290*	*TRITD3Bv1G240950*
AET3Gv20975600	***TraesCS3A02G439400***	***TraesCS3B02G473300***	***TraesCS3D02G432200***	***TRITD3Av1G249320***	***TRITD3Bv1G240990***
AET3Gv20975700	*TraesCS3A02G439500*	*TraesCS3B02G473600*	*TraesCS3D02G432300*	*TRITD3Av1G249330*	*TRITD3Bv1G241070*
AET3Gv20976100	*n/a*	*n/a*	*TraesCS7D02G148600*	*n/a*	*n/a*
AET3Gv20976200	***TraesCS3A02G439700***	***TraesCS3B02G473700***	***TraesCS3D02G432500***	***TRITD3Av1G249570***	***TRITD3Bv1G241170***
AET3Gv20976400	***TraesCS3A02G439800***	***TraesCS3B02G473800***	***TraesCS3D02G432600***	***TRITD3Av1G249580***	***TRITD3Bv1G241180***
AET3Gv20976500	*TraesCS3A02G439900*	*TraesCS3B02G474000*	*TraesCS3D02G432700*	*TRITD3Av1G249610*	*TRITD0Uv1G025560*
AET3Gv20976600	***TraesCS3A02G440000***	***TraesCS3B02G474100***	***TraesCS3D02G432800***	***TRITD3Av1G249700***	***TRITD3Bv1G241230***
AET3Gv20976700	***TraesCS3A02G440100***	***TraesCS3B02G474200***	***TraesCS3D02G432900***	***TRITD3Av1G249750***	*n/a*
AET3Gv20976800	***TraesCS3A02G440200***	***TraesCS3B02G474300***	***TraesCS3D02G433000***	***TRITD3Av1G249780***	***TRITD3Bv1G241350***

Table S8 Putative wheat orthologs of genes under peak B associated with B-type granule diameter. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20977700	*n/a*	*n/a*	*n/a*	*n/a*	*TRITD3Bv1G241510*
AET3Gv20977900	*TraesCS3A02G441100*	*TraesCS3B02G475100*	*TraesCS3D02G433800*	***TRITD3Av1G249950***	*TRITD3Bv1G241570*
AET3Gv20978000	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20978300	*TraesCS3A02G441200*	*TraesCS3B02G475200*	***TraesCS3D02G433900***	***TRITD3Av1G249990***	***TRITD3Bv1G241610***
AET3Gv20978400	*TraesCS3A02G441300*	*TraesCS3B02G475300*	*TraesCS3D02G434100*	*TRITD3Av1G250000*	***TRITD3Bv1G241620***
AET3Gv20978500	*TraesCS3A02G441400*	*n/a*	*TraesCS3D02G434200*	*TRITD3Av1G250010*	*n/a*
AET3Gv20978700	*TraesCS3A02G441500*	*TraesCS3B02G475400*	*TraesCS3D02G434300*	*TRITD3Av1G250030*	***TRITD3Bv1G241690***
AET3Gv20978900	*TraesCS3A02G441600*	*TraesCS3B02G475500*	*TraesCS3D02G434400*	***TRITD3Av1G250050***	***TRITD3Bv1G241700***
AET3Gv20979000	*TraesCS3A02G441700*	*TraesCS3B02G475600*	*TraesCS3D02G434500*	***TRITD3Av1G250060***	***TRITD3Bv1G241720***
AET3Gv20979200	***TraesCS3A02G442000***	***TraesCS3B02G475800***	***TraesCS3D02G434700***	***TRITD3Av1G250150***	***TRITD3Bv1G241970***
AET3Gv20979400	***TraesCS3A02G442100***	*TraesCS3B02G475900*	*TraesCS3D02G434800*	***TRITD3Av1G250160***	***TRITD3Bv1G241980***
AET3Gv20979600	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20979800	***TraesCS3A02G442300***	***TraesCS3B02G476100***	***TraesCS3D02G435000***	***TRITD3Av1G250250***	***TRITD3Bv1G242080***
AET3Gv20980300	***TraesCS3A02G442400***	***TraesCS3B02G476200***	***TraesCS3D02G435100***	***TRITD3Av1G250280***	***TRITD3Bv1G242100***
AET3Gv20980400	***TraesCS3A02G442500***	***TraesCS3B02G476300***	*n/a*	***TRITD3Av1G250300***	*n/a*
AET3Gv20980600	*TraesCS3A02G442600*	*n/a*	*TraesCS3D02G435200*	*n/a*	*n/a*
AET3Gv20981100	***TraesCS3A02G442700***	*n/a*	***TraesCS3D02G435300***	***TRITD3Av1G250390***	*n/a*
AET3Gv20981200	***TraesCS3A02G442900***	*n/a*	*TraesCS3D02G435400*	***TRITD3Av1G250420***	*n/a*
AET3Gv20981400	*TraesCS3A02G443000* *TraesCS3A02G443100*	*TraesCS3B02G476600* *TraesCS3B02G476700*	*TraesCS3D02G435600*	*TRITD3Av1G250440* *TRITD3Av1G250470*	*TRITD3Bv1G242230* *TRITD3Bv1G242250*

Table S9 Putative wheat orthologs of genes under peak C associated with B-type granule diameter. Blast searches were used to identify the putative *T. aestivum* and *T. turgidum* orthologs of the genes under the peak C associated with B-type granule diameter (Figure 4.1 and Table S4). n/a = genes with no putative ortholog, some genes had more than one putative ortholog in wheat and in these cases all putative orthologs are listed. Genes with endosperm expression, for *T. aestivum* where the mean TPM ≥ 1 and for *T. turgidum* where the mean TPM ≥ 1 for at least one timepoint in the dataset from Chen *et al.* (2023a), are in bold.

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET5Gv20149000	***TraesCS5A02G049400***	***TraesCS5B02G054600***	***TraesCS5D02G060200***	***TRITD5Av1G021950***	***TRITD5Bv1G022310***
AET5Gv20149200	*TraesCS5A02G049600*	*TraesCS5B02G054800*	*TraesCS5D02G060300*	*TRITD5Av1G021980*	*TRITD5Bv1G022380*
AET5Gv20149100	*TraesCS5A02G049500*	*n/a*	*n/a*	*n/a*	*n/a*
AET5Gv20150100	*TraesCS5A02G049700*	*TraesCS5B02G055500*	*TraesCS5D02G060500*	*TRITD5Av1G022080*	*TRITD5Bv1G022700*

Table S10 Putative wheat orthologs of genes under peak D associated with B-type granule diameter. Blast searches were used to identify the putative *T. aestivum* and *T. turgidum* orthologs of the genes under peak D associated with B-type granule diameter (Figure 4.1 and Table S4). n/a = genes with no putative ortholog, some genes had more than one putative ortholog in wheat and in these cases all putative orthologs are listed. Genes with endosperm expression, for *T. aestivum* where the mean TPM ≥ 1 and for *T. turgidum* where the mean TPM ≥ 1 for at least one timepoint in the dataset from Chen *et al.* (2023a), are in bold.

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET7Gv20332600	*TraesCS7A02G131500*	*TraesCS7B02G032100*	*TraesCS7D02G130700*	*TRITD7Av1G039210*	*TRITD7Bv1G012480*
AET7Gv20332900	***TraesCS7A02G131700***	***TraesCS7B02G032400***	***TraesCS7D02G131000***	***TRITD7Av1G039310***	***TRITD7Bv1G012530***
AET7Gv20333100	***TraesCS7A02G131800***	***TraesCS7B02G032500***	***TraesCS7D02G131100***	***TRITD7Av1G039320***	***TRITD7Bv1G012550***
AET7Gv20333300	*n/a*	*TraesCS5B02G038300*	*TraesCS7D02G131200* *TraesCS7D02G131300*	*TRITD4Av1G259220* *TRITD4Av1G259230* *TRITD4Av1G259210*	*TRITD5Bv1G015620*
AET7Gv20333400	*TraesCS7A02G131900*	*TraesCS7B02G032600*	*TraesCS7D02G131400*	*n/a*	*n/a*
AET7Gv20333600	*TraesCS7A02G132000*	*TraesCS7B02G032700*	*TraesCS7D02G131600*	*TRITD7Av1G039330*	*TRITD7Bv1G012620*
AET7Gv20333700	*TraesCS7A02G132100*	*TraesCS7B02G032800*	*TraesCS7D02G131700*	*n/a*	*TRITD7Bv1G012670*
AET7Gv20333900	*TraesCS7A02G132200*	*n/a*	*TraesCS7D02G131800*	*TRITD7Av1G039360*	*TRITD7Bv1G012680*
AET7Gv20334200	*TraesCS7A02G132300*	*TraesCS7B02G032900*	*n/a*	*TRITD7Av1G039370*	*TRITD7Bv1G012690*
AET7Gv20334300	***TraesCS7A02G132400***	***TraesCS7B02G033000***	***TraesCS7D02G131900***	***TRITD7Av1G039380***	***TRITD7Bv1G012780***
AET7Gv20334400	***TraesCS7A02G132500***	***TraesCS7B02G033100***	***TraesCS7D02G132000***	*n/a*	***TRITD7Bv1G012790***
AET7Gv20334900	***TraesCS7A02G132700***	***TraesCS7B02G033300***	***TraesCS7D02G132200***	***TRITD7Av1G039430***	***TRITD7Bv1G012830***
AET7Gv20334700	*TraesCS7A02G132600*	*TraesCS7B02G033200*	***TraesCS7D02G132100***	*TRITD7Av1G039400*	*TRITD7Bv1G012800*
AET7Gv20335100	*TraesCS7A02G132800*	*TraesCS7B02G033700*	*TraesCS7D02G132300*	*TRITD7Av1G039440*	***TRITD7Bv1G012840***
AET7Gv20335500	*TraesCS7A02G132900*	*n/a*	*TraesCS7D02G132400*	*n/a*	*n/a*

Table S10 Putative wheat orthologs of genes under peak D associated with B-type granule diameter. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET7Gv20335600	***TraesCS7A02G133000***	***TraesCS7B02G033800***	***TraesCS7D02G132600***	***TRITD7Av1G039470***	***TRITD7Bv1G012870***
AET7Gv20335700	*TraesCS7A02G133100*	*TraesCS7B02G033900*	*TraesCS7D02G132643*	*TRITD7Av1G039480*	*TRITD7Bv1G012880*
AET7Gv20336000	*TraesCS7A02G133200*	*TraesCS7B02G034000*	*TraesCS7D02G132700*	*TRITD7Av1G039490*	*TRITD7Bv1G012890*
AET7Gv20336100	*n/a*	***TraesCS7B02G034200***	*n/a*	*n/a*	***TRITD7Bv1G012910***
AET7Gv20336200	*TraesCS7A02G133300*	*TraesCS7B02G034400*	*TraesCS7D02G132900*	***TRITD7Av1G039500***	*TRITD7Bv1G013020*
AET7Gv20336800	***TraesCS7A02G133500***	***TraesCS7B02G034600***	***TraesCS7D02G133100***	***TRITD7Av1G039520***	***TRITD7Bv1G013060***
AET7Gv20336700	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET7Gv20337500	*TraesCS7A02G133600*	*TraesCS7B02G034700*	*TraesCS7D02G133200*	*TRITD7Av1G039530*	*TRITD7Bv1G013080*
AET7Gv20337600	*TraesCS7A02G133700*	*TraesCS7B02G034800*	*TraesCS7D02G133300*	*n/a*	*TRITD7Bv1G013320*
AET7Gv20338500	***TraesCS7A02G134000***	***TraesCS7B02G035100***	***TraesCS7D02G133400***	*n/a*	***TRITD7Bv1G013400***
AET7Gv20338600	***TraesCS7A02G134100***	***TraesCS7B02G035200***	***TraesCS7D02G133500***	***TRITD7Av1G039640***	***TRITD7Bv1G013420***
AET7Gv20338700	*TraesCS7A02G134200*	*TraesCS7B02G035300*	*TraesCS7D02G133600*	*TRITD7Av1G039650*	*n/a*
AET7Gv20338800	***TraesCS7A02G134300***	*TraesCS7B02G035400*	***TraesCS7D02G133700***	***TRITD7Av1G039660***	***TRITD7Bv1G013510***
AET7Gv20339100	*TraesCS7A02G134400*	*TraesCS7B02G035500*	*TraesCS7D02G133800*	*TRITD7Av1G039680*	***TRITD7Bv1G013540***
AET7Gv20339300	*TraesCS7A02G134500*	*TraesCS7B02G035700*	*TraesCS7D02G133900*	*TRITD7Av1G039710*	*n/a*
AET7Gv20339500	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET7Gv20339700	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET7Gv20339600	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET7Gv20340000	*TraesCS7A02G134600*	*TraesCS7B02G035800*	*TraesCS7D02G134000*	*TRITD7Av1G039790*	*TRITD7Bv1G013580*
AET7Gv20340100	*n/a*	*TraesCS7B02G035900*	*TraesCS7D02G134100*	*n/a*	*n/a*
AET7Gv20340400	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET7Gv20340300	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET7Gv20340900	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*

Table S10 Putative wheat orthologs of genes under peak D associated with B-type granule diameter. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET7Gv20341600	*TraesCS7A02G134800*	*TraesCS7B02G036200*	*TraesCS7D02G134400*	*TRITD7Av1G040160*	*TRITD7Bv1G013680*
AET7Gv20341500	*TraesCS7A02G134700*	*TraesCS7B02G036100*	*TraesCS7D02G134300*	*TRITD7Av1G040150*	*TRITD7Bv1G013660*
AET7Gv20341700	*TraesCS7A02G134900*	*TraesCS7B02G036300*	*TraesCS7D02G134500*	*n/a*	*TRITD7Bv1G013690*
AET7Gv20341800	*n/a*	*n/a*	*n/a*	***TRITD7Av1G046280***	*n/a*
AET7Gv20342000	*TraesCS1A02G216400*	*n/a*	*n/a*	*TRITD7Av1G271600* *TRITD1Av1G139350*	*n/a*
AET7Gv20343000	***TraesCS7A02G135000***	***TraesCS7B02G036500***	***TraesCS7D02G135000***	***TRITD7Av1G040190***	***TRITD7Bv1G013800***
AET7Gv20343800	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET7Gv20344000	***TraesCS7A02G135600***	*n/a*	***TraesCS7D02G135600***	***TRITD7Av1G040280***	*n/a*
AET7Gv20344200	*TraesCS7A02G135400*	***TraesCS7B02G036700***	***TraesCS7D02G135700***	***TRITD7Av1G040260***	***TRITD7Bv1G013860***
AET7Gv20344800	***TraesCS7A02G135700***	***TraesCS7B02G036800***	***TraesCS7D02G135800***	***TRITD7Av1G040300***	*TRITD7Bv1G013900*
AET7Gv20345000	*TraesCS7A02G135800*	*TraesCS7B02G036900*	*TraesCS7D02G135900*	*TRITD7Av1G040390*	***TRITD7Bv1G013910***
AET7Gv20345400	*n/a*	*n/a*	*n/a*	*TRITD7Av1G040440*	*n/a*
AET7Gv20346000	***TraesCS7A02G136300***	***TraesCS7B02G037300***	***TraesCS7D02G136300***	***TRITD7Av1G040510***	***TRITD7Bv1G014050***
AET7Gv20345600	*n/a*	*TraesCS7B02G037200*	*TraesCS7D02G136200*	*TRITD7Av1G040460*	*TRITD7Bv1G014030*
AET7Gv20346100	*TraesCS7A02G136500*	*TraesCS7B02G037400* *TraesCS7B02G037600*	*TraesCS7D02G136400*	*TRITD7Av1G040530*	*TRITD7Bv1G014070*
AET7Gv20346200	*TraesCS7A02G136600*	*TraesCS7B02G037700*	*TraesCS7D02G136500*	*TRITD7Av1G040550*	*TRITD7Bv1G014150*
AET7Gv20347000	*TraesCS1A02G035700*	*n/a*	*TraesCS7D02G136700*	*TRITD1Av1G007560*	*TRITD1Bv1G009800*
AET7Gv20347100	***TraesCS1A02G035800***	***TraesCS1B02G045700***	***TraesCS7D02G136800***	*n/a*	*n/a*

Table S11 Putative wheat orthologs of genes under the peak V associated with B-type granule number. Blast searches were used to identify the putative *T. aestivum* and *T. turgidum* orthologs of the genes under peak V associated with B-type granule number (Figure 4.2 and Table S5). n/a = genes with no putative ortholog, some genes had more than one putative ortholog in wheat and in these cases all putative orthologs are listed. Genes with endosperm expression, for *T. aestivum* where the mean TPM ≥ 1 and for *T. turgidum* where the mean TPM ≥ 1 for at least one timepoint in the dataset from Chen *et al.* (2023a), are in bold.

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20188700	*TraesCS3A02G087500*	*TraesCS3B02G102900*	*TraesCS3D02G087600*	*TRITD3Av1G025330*	***TRITD3Bv1G029060***
AET3Gv20188900	*n/a*	*n/a*	*TraesCS3D02G087700*	*n/a*	*n/a*
AET3Gv20189300	*TraesCS3A02G087700*	*TraesCS3B02G103100*	*TraesCS3D02G087800*	*TRITD3Av1G025410*	*TRITD3Bv1G029070*
AET3Gv20189500	***TraesCS3A02G087800***	***TraesCS3B02G103200***	***TraesCS3D02G087900***	***TRITD3Av1G025470***	***TRITD3Bv1G029200***
AET3Gv20190600	*n/a*	*n/a*	*TraesCS3D02G088400*	*n/a*	*n/a*
AET3Gv20190800	*TraesCS3A02G088500*	*n/a*	*TraesCS3D02G088500* *TraesCS3D02G088600*	*TRITD3Av1G025580* *TRITD3Av1G025600*	*n/a*
AET3Gv20190900	*TraesCS3A02G088300*	*TraesCS3B02G103800* *TraesCS3B02G103900* *TraesCS3B02G104000* *TraesCS3B02G103700*	*TraesCS3D02G088700*	*n/a*	*TRITD3Bv1G029400* *TRITD3Bv1G029430* *TRITD3Bv1G029420*
AET3Gv20191000	*TraesCS3A02G088700*	*TraesCS3B02G104100*	*TraesCS3D02G088800*	*TRITD3Av1G025630*	*TRITD3Bv1G029570*
PGSB gene 23181	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20191100	*n/a*	*n/a*	***TraesCS3D02G088900***	*n/a*	*n/a*
AET3Gv20191300	***TraesCS3A02G088900***	***TraesCS3B02G104200***	***TraesCS3D02G089000***	***TRITD3Av1G025670***	***TRITD3Bv1G029590***
AET3Gv20191400	*n/a*	*TraesCS3B02G104300*	*TraesCS3D02G089100*	*n/a*	*TRITD3Bv1G029640*
AET3Gv20191600	***TraesCS3A02G089000***	***TraesCS3B02G104400***	***TraesCS3D02G089200***	***TRITD3Av1G025700***	***TRITD3Bv1G029660***

Table S11 Putative wheat orthologs of genes under the peak V associated with B-type granule number. (continued)

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20191700	*n/a*	*n/a*	*TraesCS3D02G089300*	*n/a*	*n/a*
AET3Gv20191900	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20192200	*TraesCS3A02G089300*	*TraesCS3B02G104600*	*TraesCS3D02G089600*	*TRITD3Av1G025770*	*TRITD3Bv1G029760*
AET3Gv20192600	*n/a*	*n/a*	*TraesCS3D02G089700*	*n/a*	*n/a*
AET3Gv20192700	*n/a*	*TraesCS3B02G104900*	*TraesCS3D02G089800*	*n/a*	*n/a*
AET3Gv20193700	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20194600	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20194700	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
PGSB gene 23212	*n/a*	*TraesCS3B02G105400*	*TraesCS3D02G090100*	*n/a*	*TRITD3Bv1G029960*
AET3Gv20195000	***TraesCS3A02G090000***	***TraesCS3B02G105300***	***TraesCS3D02G090000***	***TRITD3Av1G025880***	*n/a*
AET3Gv20196600	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
AET3Gv20197700	***TraesCS3A02G090100***	***TraesCS3B02G105600***	***TraesCS3D02G090200***	***TRITD3Av1G025890***	***TRITD3Bv1G029980***

Table S12 Putative wheat orthologs of genes under the peak W associated with B-type granule number. Blast searches were used to identify the putative *T. aestivum* and *T. turgidum* orthologs of the genes under peak W associated with B-type granule number (Figure 4.2 and Table S5). n/a = genes with no putative ortholog, some genes had more than one putative ortholog in wheat and in these cases all putative orthologs are listed. Genes with endosperm expression, for *T. aestivum* where the mean TPM ≥ 1 and for *T. turgidum* where the mean TPM ≥ 1 for at least one timepoint in the dataset from Chen *et al.* (2023a), are in bold.

Gene	*T. aestivum*			*T. turgidum*	
	A genome	B genome	D genome	A genome	B genome
AET3Gv20318700	*TraesCS3A02G132000*	*TraesCS3B02G167400*	*TraesCS3D02G148300*	*n/a*	*n/a*
AET3Gv20318800	***TraesCS3A02G132100***	*TraesCS3B02G167300*	*TraesCS3D02G148400*	***TRITD3Av1G053970***	***TRITD3Bv1G063170***
AET3Gv20319100	*TraesCS3A02G132300*	*TraesCS3B02G167200*	*TraesCS3D02G148500*	*TRITD3Av1G053900*	*TRITD3Bv1G063110*
AET3Gv20319200	*TraesCS3A02G132400*	*TraesCS3B02G167100*	*TraesCS3D02G148600*	*TRITD3Av1G053830*	*TRITD3Bv1G063070*
AET3Gv20319400	***TraesCS3A02G132500*** ***TraesCS3A02G132600***	***TraesCS3B02G167000***	*TraesCS3D02G148900*	***TRITD3Av1G053790***	***TRITD3Bv1G062940***
AET3Gv20319500	*n/a*	*TraesCS5B02G541900*	*TraesCS3D02G148800*	*n/a*	*TRITD5Bv1G248560*
AET3Gv20319800	***TraesCS3A02G132700***	***TraesCS3B02G166900***	***TraesCS3D02G149000***	*n/a*	*n/a*
AET3Gv20319900	***TraesCS3A02G132800***	*TraesCS3B02G166800*	*TraesCS3D02G149100*	***TRITD3Av1G053630***	*TRITD3Bv1G062890*
AET3Gv20320100	*TraesCS3A02G132900*	*TraesCS3B02G166700*	*TraesCS3D02G149200*	*TRITD3Av1G053610*	*TRITD3Bv1G062880*
AET3Gv20320700	*TraesCS3A02G133100*	*TraesCS3B02G166500*	*TraesCS3D02G149300*	***TRITD3Av1G053550***	*n/a*
AET3Gv20320800	***TraesCS3A02G133200***	*TraesCS3B02G166400*	***TraesCS3D02G149400***	***TRITD3Av1G053530***	*TRITD3Bv1G062840*
AET3Gv20321000	***TraesCS3A02G133300***	***TraesCS3B02G166200***	***TraesCS3D02G149500***	***TRITD3Av1G053490***	***TRITD3Bv1G062820***

Table S13 *Clust* analysis grouped the putative *T. turgidum* orthologs of the genes under the GWAS peaks, and starch granule initiation genes into five clusters. The putative *T. turgidum* orthologs in each cluster are listed, starch granule initiation genes are in bold with the common name in brackets. Blank cells are due to non-equal cluster sizes, *T. turgidum* putative orthologs not present in this table did not fall into a cluster, *n* = number of genes.

Cluster One (*n* = 34)	Cluster Two (*n* = 4)	Cluster Three (*n* = 9)	Cluster Four (*n* = 30)	Cluster Five (*n* = 7)
TRITD1Bv1G062760 (*MFP1.1-B*)	TRITD3Av1G025880	**TRITD1Av1G054690 (*MFP1.1-A*)**	**TRITD0Uv1G034540 (*BGC1-B*)**	TRITD3Av1G246730
TRITD3Av1G244480	TRITD3Av1G245260	TRITD3Av1G025670	TRITD1Av1G045900	TRITD3Av1G249570
TRITD3Av1G244760	TRITD3Bv1G239920	TRITD3Av1G053550	TRITD2Bv1G212980	TRITD3Bv1G236270
TRITD3Av1G244950	TRITD7Av1G040260	TRITD3Av1G247910	TRITD3Av1G053630	TRITD3Bv1G240700
TRITD3Av1G244960		TRITD3Av1G250300	TRITD3Av1G053970	TRITD7Av1G039430
TRITD3Av1G245020		TRITD3Bv1G241980	TRITD3Av1G244740	TRITD7Av1G040510
TRITD3Av1G245050		TRITD3Bv1G242100	TRITD3Av1G245270	TRITD7Bv1G012830
TRITD3Av1G247480		TRITD7Av1G040190	TRITD3Av1G245750	
TRITD3Av1G248050		TRITD7Bv1G013800	TRITD3Av1G246370	
TRITD3Av1G248460			TRITD3Av1G246380	
TRITD3Av1G248490			TRITD3Av1G246490	
TRITD3Av1G248980			TRITD3Av1G247490	
TRITD3Av1G249950			TRITD3Av1G249750	
TRITD3Av1G249990			TRITD3Av1G250390	
TRITD3Av1G250420			TRITD3Bv1G029060	
TRITD3Bv1G029590			TRITD3Bv1G236910	
TRITD3Bv1G235790			TRITD3Bv1G239650	
TRITD3Bv1G235920			TRITD3Bv1G239720	
TRITD3Bv1G236520			TRITD3Bv1G239880	
TRITD3Bv1G236670			TRITD3Bv1G241170	

Table S13 *Clust* analysis grouped the putative *T. turgidum* orthologs of the genes under the GWAS peaks, and starch granule initiation genes into five clusters. (continued)

Cluster One (*n* = 34)	Cluster Two (*n* = 4)	Cluster Three (*n* = 9)	Cluster Four (*n* = 30)	Cluster Five (*n* = 7)
TRITD3Bv1G236900			TRITD3Bv1G241230	
TRITD3Bv1G237360			TRITD3Bv1G241690	
TRITD3Bv1G240280			**TRITD4Av1G198830 (*BGC1-A*)**	
TRITD3Bv1G240580			TRITD5Av1G021950	
TRITD3Bv1G240920			**TRITD5Av1G205670 (*PHS1-A*)**	
TRITD3Bv1G240990			TRITD5Bv1G022310	
TRITD3Bv1G241610			TRITD7Av1G039500	
TRITD4Av1G231640			TRITD7Av1G039520	
TRITD6Av1G081580 (*MRC-A*)			TRITD7Av1G046280	
TRITD7Av1G039380			TRITD7Bv1G013060	
TRITD7Bv1G012550				
TRITD7Bv1G012840				
TRITD7Bv1G012910				
TRITD7Bv1G013510				

Appendix Two: Synthesized DNA

pET-28a with *LDA* cDNA. Yellow = His-tag, Purple = *LDA*, Pink = T7 promoter, Green = Kanamycin resistance gene, Orange = NdeI site, Turquoise = NotI site

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTA
CGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCT
TCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCT
CCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGG
GTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTT
GGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTA
TCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAA
ATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTC
AGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATAC
ATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAA
ACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAAT
GAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCT
GCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGG
TTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTT
TATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCA
CTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGC
GATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACAC
TGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGC
TGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAA
TGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATC
TGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGG
GCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCAT
TTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTT
TCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTT
ATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCC
GTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTT
GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTC
TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCT
CGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACC
GGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGG
GGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGT
ATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGG

AAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCC
TTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATC
CCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGA
TGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCT
CAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACG
TGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCT
GCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAA
GCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCA
GCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTT
AAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTC
ATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATG
ATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCG
GCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTA
GGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTG
CAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATT
CATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCG
CGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCC
TCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATG
GCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAG
GGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCG
AAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTG
CATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCG
GAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTA
ATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGA
AACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTG
CGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATT
GCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCC
CCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTC
TTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCG
GTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTG
GGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCC
AGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAG
CCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATT
TGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGG
GAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAA
CATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAAT

GATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTC
GACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCG
AGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGG
CAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAA
TGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTG
GCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGC
GACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGC
GCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGAC
GCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
GAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCC
CGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAG
TGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCA
CCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTC
GATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCC
TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATC
ATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCGGCCGGCGAG
ACCGGCGCCTCCGTCTCCGTCTCCGCCGCCGAGGCCGAGGCCGAGTCCACCCAGG
CGTTCATGCCGGACGCCAGGGCGTACTGGGTGACGAGCGACCTCATCGCTTGGAAC
GTCGGCGAGCAGGAAGCGGCGTCCGTCTGCCTGTACGCCAGCAGAGCCGCCGAGAT
GGGCCTCTCGCCGTCGAATGGCGGCATCCAAGGCTACGACTCCAAGGTTGAGCTGCA
ACCGGAGAGCGCCGGGCTCCCGGAAACCGTGACCCAGAAGTTCCCCTTCATCAGCA
GTTACAGAGCGTTCAGTGTCCCGAGCTCTGTCGACGTCGCCAGCCTCGTGAAATGCC
AACTGGTCGTGGCTTCTTTTGGTGCTGACGGGAAACACGTAGATGTTACAGGACTGCA
GTTACCCGGCGTACTGGATGACATGTTCGCGTACACAGGACCGCTTGGTGCGGTTTT
CAGCGAGGAATCCGTGAGCCTCCACCTTTGGGCTCCTACAGCACAGGGTGTGAGTGT
GTGCTTCTTTGATGGTCCAGCGGGCCCTGTGCTGGAGACGGTTCAGCTCGAGGAGTC
AAATGGAGTTTGGAGTGTTACTGGACCAAGAGAGTGGGAAAACCGGTATTATTTGTATG
AAGTCGACGTGTATCATCCAACCAAGGCGCAAGTTCTGAAATGTTTAGCTGGTGACCC
GTATGCTAGAGGCCTTTCCGCAAATGGCGCGCGGACCTGGTTGGTTGACATTAACAAT
GAAACACTGAAGCCAGCTTCCTGGGATGAATTGGCCGATGAGAAGCCAAAACTTGATT
CCTTCTCTGACATAACCATCTACGAATTGCACATTCGTGATTTTAGCGCCCATGATGGC
ACAGTGGACAGTGACTCTCGTGGAGGATTTCGTGCATTTGCATATCAGGCCTCGGCAG
GAATGCAGCACCTACGGAAATTGTCTGATGCTGGTTTGACGCATGTGCATTTGTTGCC
AAGCTTTCATTTTGCTGGCGTTGACGACATTAAGAGCAACTGGAAATTTGTCGATGAGT
GTGAACTGGCAACATTCCCTCCAGGATCAGATAAGCAACAAGCAGCAGTAGTAGCTAT
TCAGGAAGAGGACCCTTATAATTGGGGGTATAACCCTGTACTCTGGGGGGTTCCAAAA
GGAAGCTATGCAAGTGACCCTGATGGCCCGAGTCGCATTATTGAATACCGTCAGATGG
TTCAGGCCCTCAATCGCATAGGCCTTCGTGTTGTCATGGATGTTGTATACAATCATCTA
GACTCAAGTGGCCCCTGCGGTATCAGCTCAGTGCTTGACAAGATCGTTCCTGGGTACT

ATGTTAGAAGAGATACTAATGGCCAGATTGAGAACAGTGCAGCTATGAACAATACAGCA
AGTGAGCATTTCATGGTTGATAGGTTAATCGTGGATGACCTTTTGAACTGGGCAGTAAA
CTACAAAATTGATGGGTTCAGATTTGATCTTATGGGCCATATCATGAAACACACCATGAT
GAGAGCAAAATCTGCTCTCCAAAGCCTTACAAGGGATGCACATGGAGTTGATGGTTCA
AAAATATACTTGTATGGTGAAGGATGGGACTTCGCTGAAGTTGCACGCAATCAACGTG
GAATAAATGGATCCCAGCTTAATATGAGTGGAACGGGGATTGGTAGCTTCAATGATAGA
ATCCGGGATGCTGTTAATGGGGGTAATCCATTTGGTAACCCACTACAGCAAGGCTTTAA
TACTGGTCTGTTCTTAGAGCCGAATGGGTTTTATCAGGGCAATGAAGCAGATACCAGG
CGCTCGCTCGCTACTTACGCTGACCAAATACAGATTGGACTAGCTGGTAATCTGAGGG
ATTATGTACTAATAACTCATACTGGAGAAGCTAAGGAGGGATCAGAAATTCACACTTTCG
ATGGATTACCAGTCGGCTATACTTCGTCCCCGATAGAAATAATAAACTATGTTTCTGCTC
ATGACAATGAGACTCTGTTTGATGTTATCAGTGTAAAGACCCCAATGAACCTTTCAGTT
GATGAGAGATGCAGGATAAATCATTTGGCCTCCAGCATGATGGCATTATCCCAGGGAAT
ACCCTTTTTCCATGCTGGTGACGAGATACTAAGATCTAAGTCCATTGATCGAGATTCATA
CAACTCTGGTGATTGGTTTAACAAGCTTGATTTTACCTATGAAACAAACAATTGGGGTG
TTGGGCTTCCTCCAAGTGAAAAGAACGAAGATAATTGGCCCCTGATGAAACCAAGATT
GGAAAATCCATCTTTCAAACCTGCAAAAGGACACATTCTTGCTTCCCTAGACAGTTTTG
TTGATATCTTGAAGATCAGATACTCATCTCCGCTTTTTCGTCTTAGTACAGCAAGTGACA
TTAAGCAAAGGGTTCGCTTTCACAACACAGGGCCCTCCTCAGTCCCAGGTGTTATTGT
CATGGGCATTGAAGATGCACGAGATGAGAAGCCCGAGGTGGCTCAGTTAGATGCGAA
CTTCTCTTATGTTGTAACTGTCTTCAATGTGTGTCCGCACGAAGTATCCATGGATATTCC
AGCCCTCGCTTCAATGCGGCTTGAATTGCATCCTGTGCAGGTGAATTCATCAGATGCT
TTGGTGGGGAAATCTGTGTACGAGGCCGCGACGGGCAGGTTCACTGTGCCCAGAAG
AACCGTGTCAGTCTTTGTTGAACCTCGGTGTTGAGCGGCCGCACTCGAGCACCACCA
CCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC
TGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGG
GGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

Appendix Three: Supplementary Figures

Figure S1 Alignment for amino acids 79-266 of LDA for accessions with low and high B-type granule diameters. Accessions are arranged from lowest (top) to highest (bottom). The A140T and E248D mutations are marked with *, mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments.

Figure S2 Alignment for amino acids 267-454 of LDA for accessions with low and high B-type granule diameters. Accessions are arranged from lowest (top) to highest (bottom). Mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments.

Figure S3 Alignment for amino acids 455-642 of LDA for accessions with low and high B-type granule diameters. Accessions are arranged from lowest (top) to highest (bottom). Mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments.

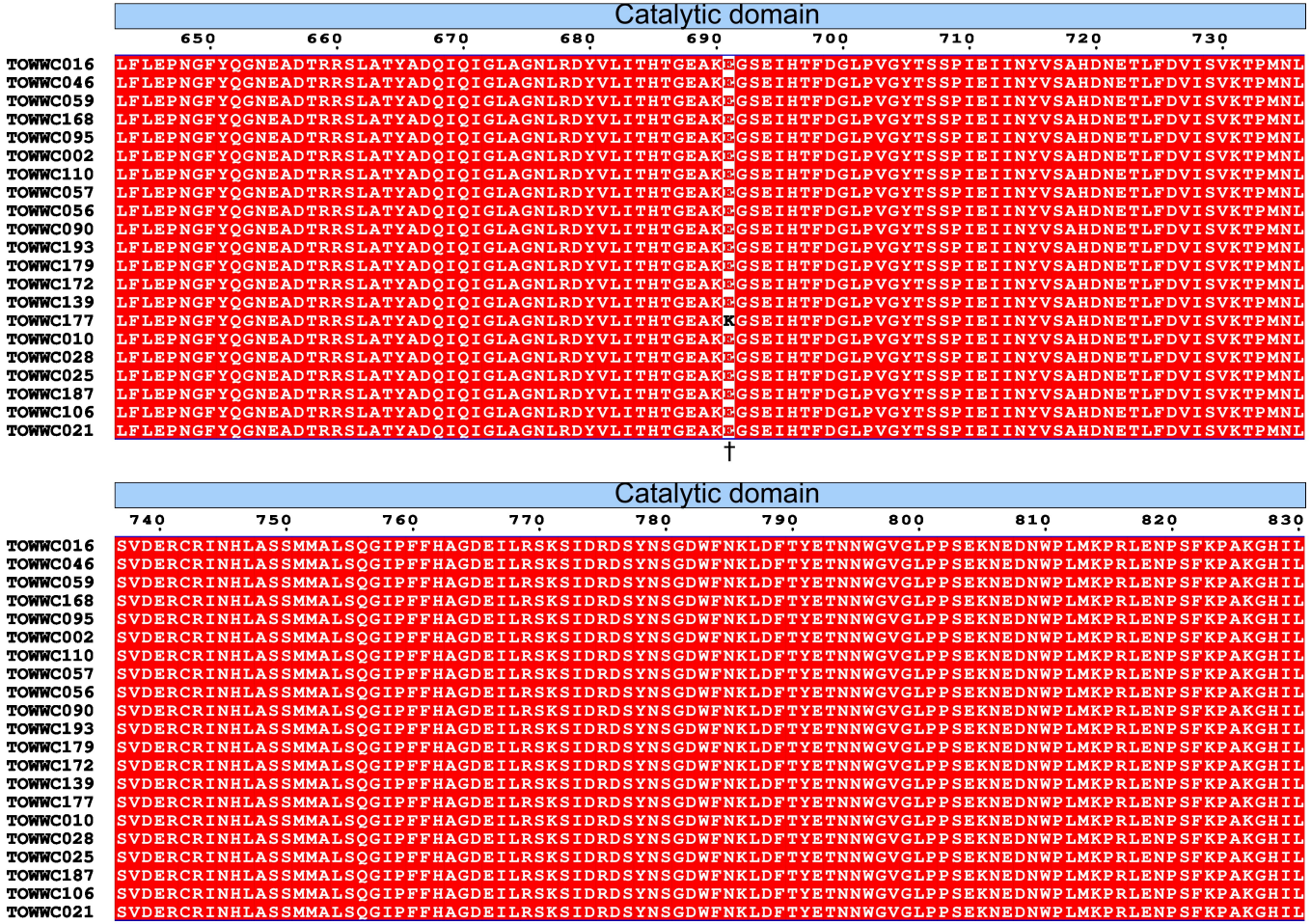


Figure S4 Alignment for amino acids 643-830 of LDA for accessions with low and high B-type granule diameters. Accessions are arranged from lowest (top) to highest (bottom). Mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments.

Figure S5 Alignment for amino acids 831-963 of LDA for accessions with low and high B-type granule diameters. Accessions are arranged from lowest (top) to highest (bottom). Mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments. Numbering is relative to the *Ae. tauschii* sequence.

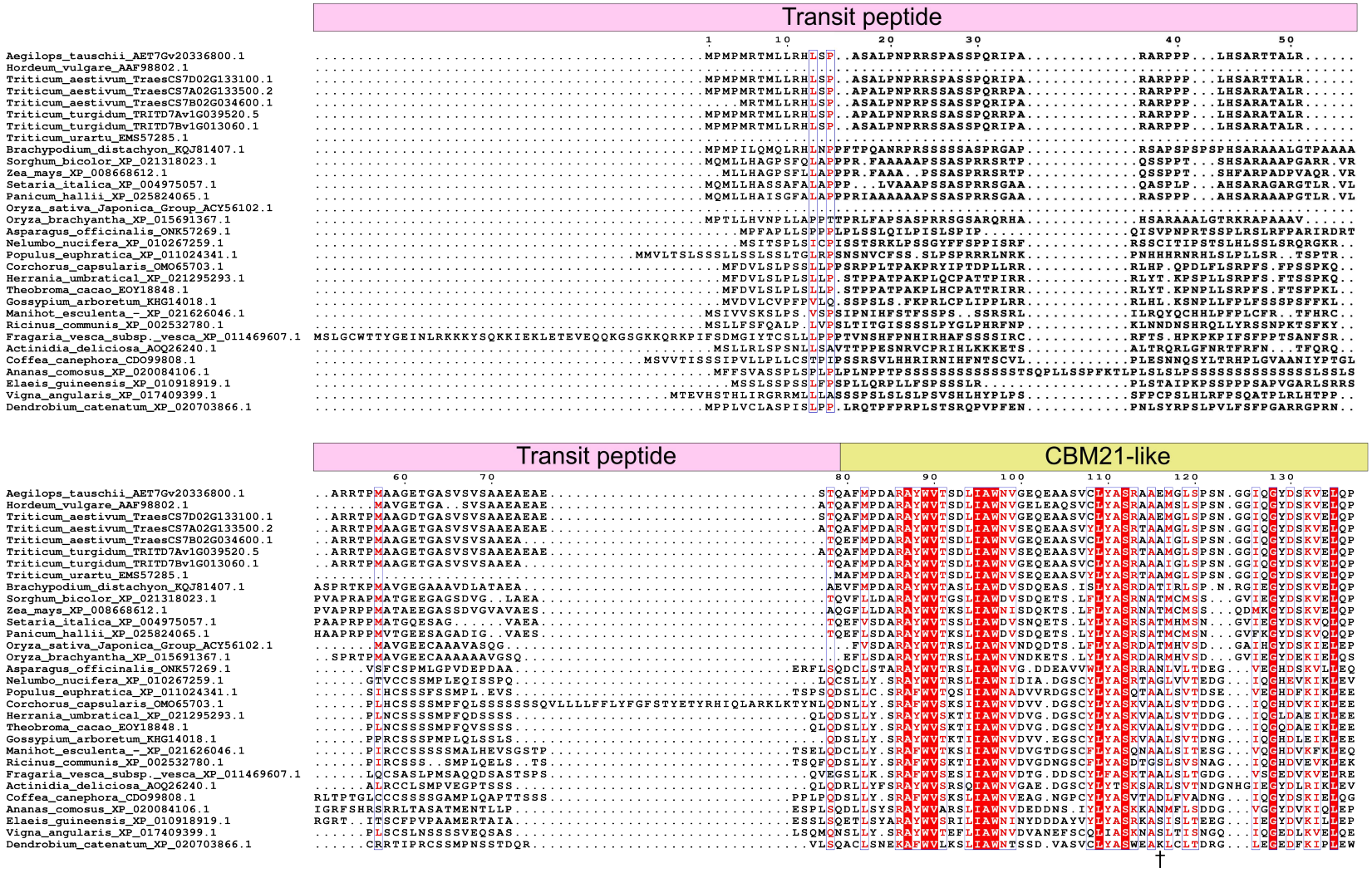


Figure S6 Alignment of amino acids 1-137 in LDA from different species. Accessions are arranged in the same order as in Andersen *et al.* (2020). Locations of the mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments. Numbering is relative to the *Ae. tauschii* sequence.

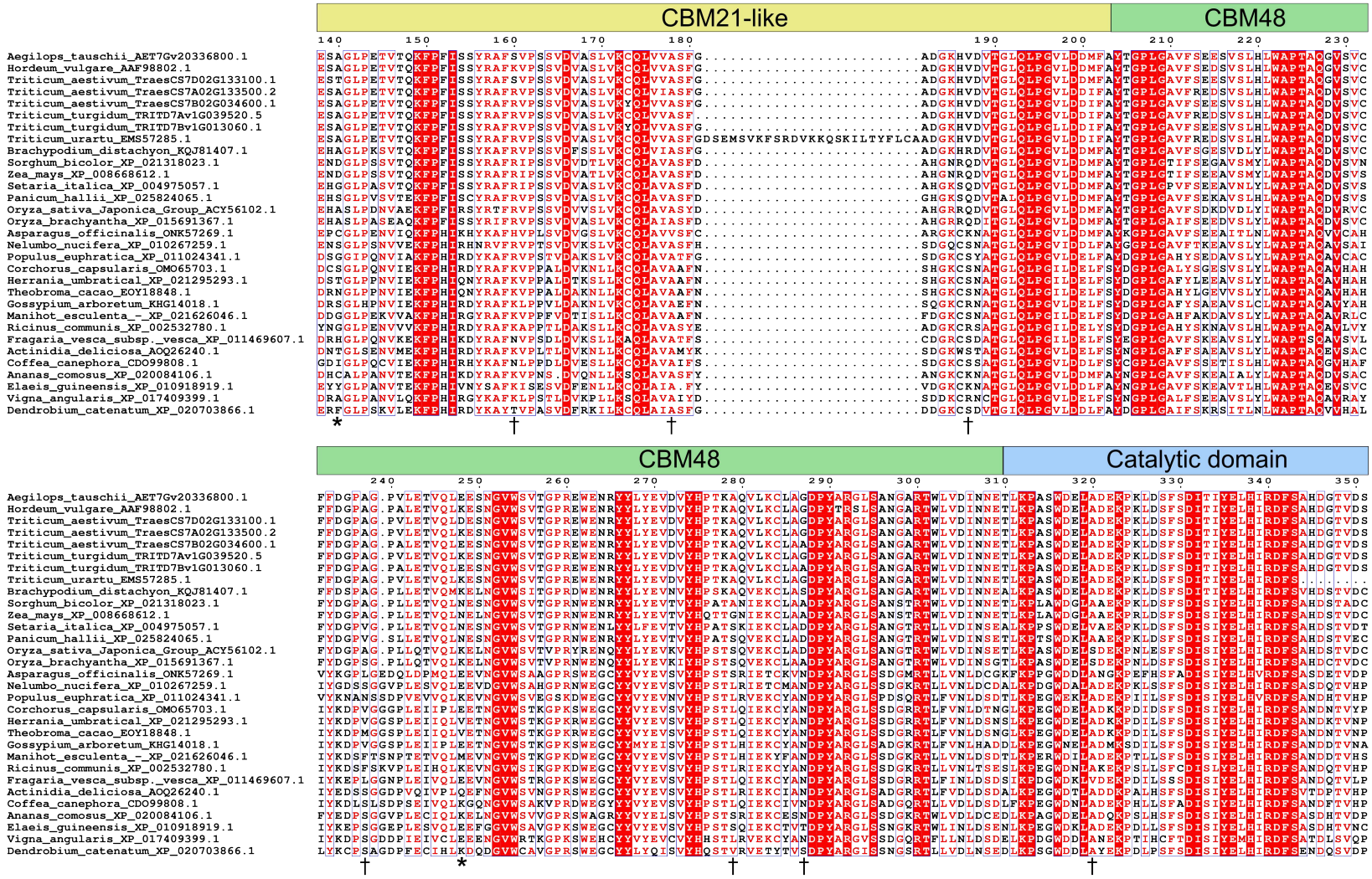


Figure S7 Alignment of amino acids 138-351 in LDA from different species. Accessions are arranged in the same order as in Andersen *et al.* (2020). Locations 140 and 248 are marked with *, locations of the mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments. Numbering is relative to the *Ae. tauschii* sequence.

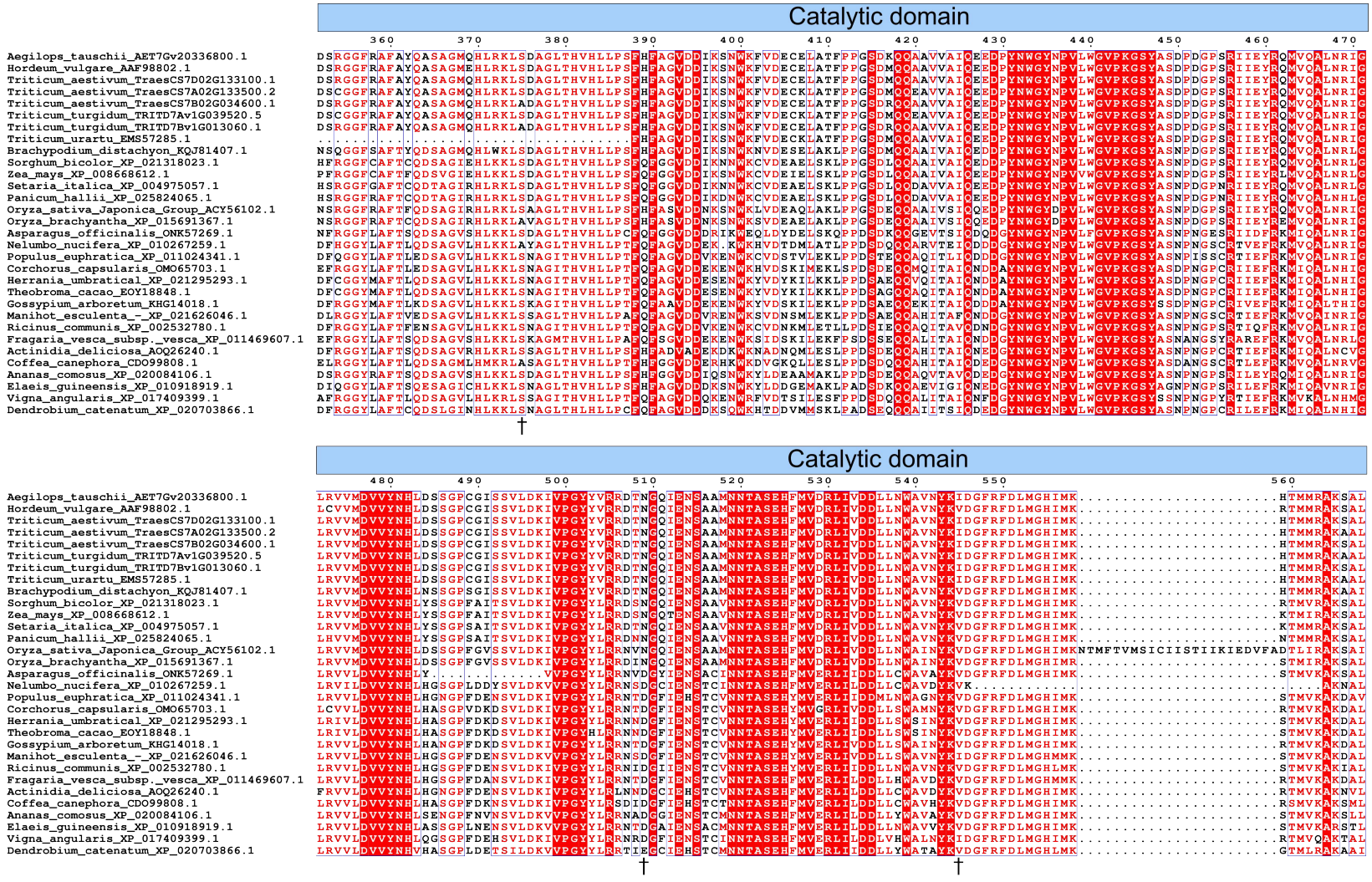


Figure S8 Alignment of amino acids 352-568 in LDA from different species. Accessions are arranged in the same order as in Andersen *et al.* (2020). Locations of the mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments. Numbering is relative to the *Ae. tauschii* sequence.

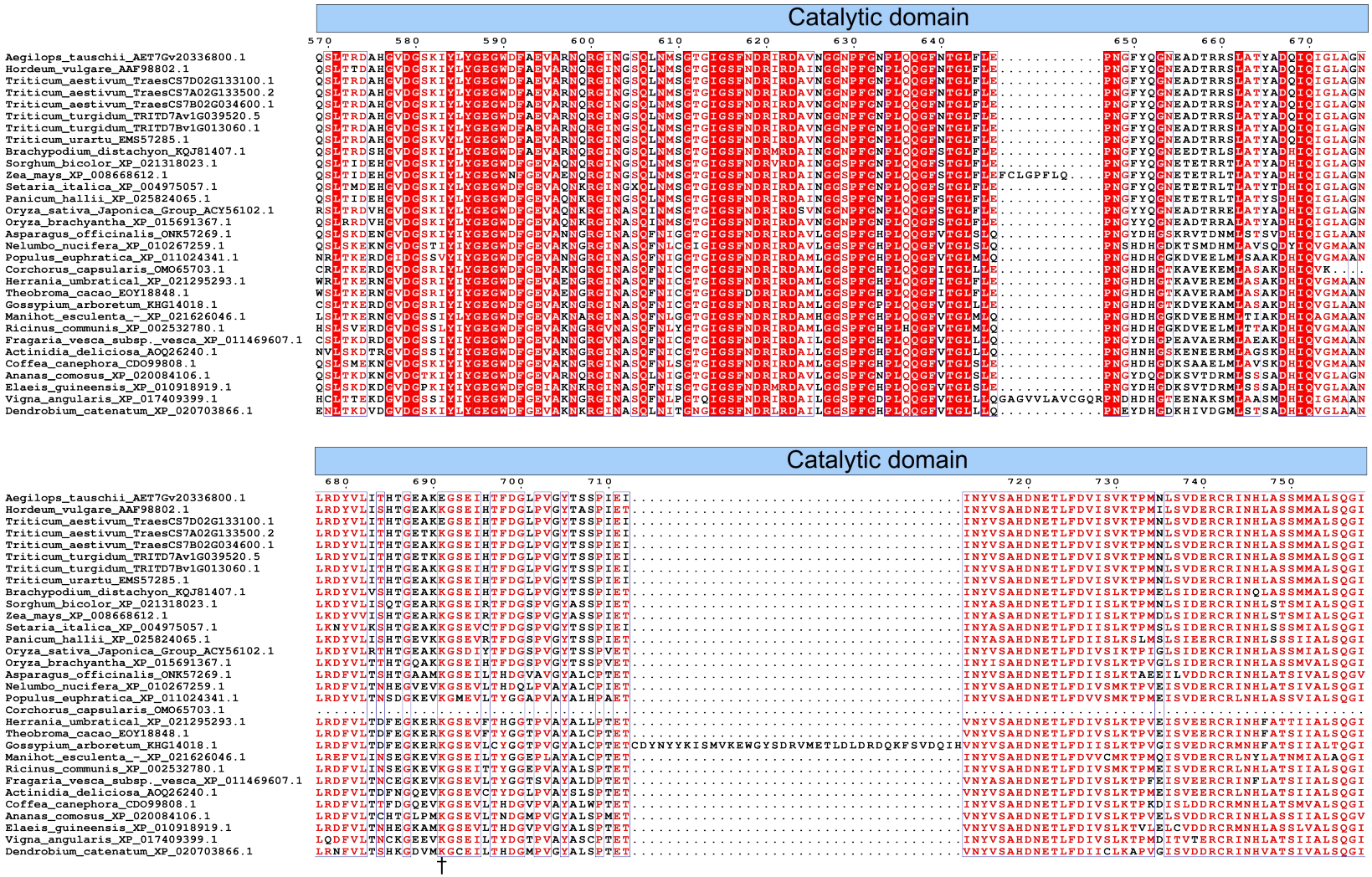


Figure S9 Alignment of amino acids 569-758 in LDA from different species. Accessions are arranged in the same order as in Andersen *et al.* (2020). Locations of the mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments. Numbering is relative to the *Ae. tauschii* sequence.

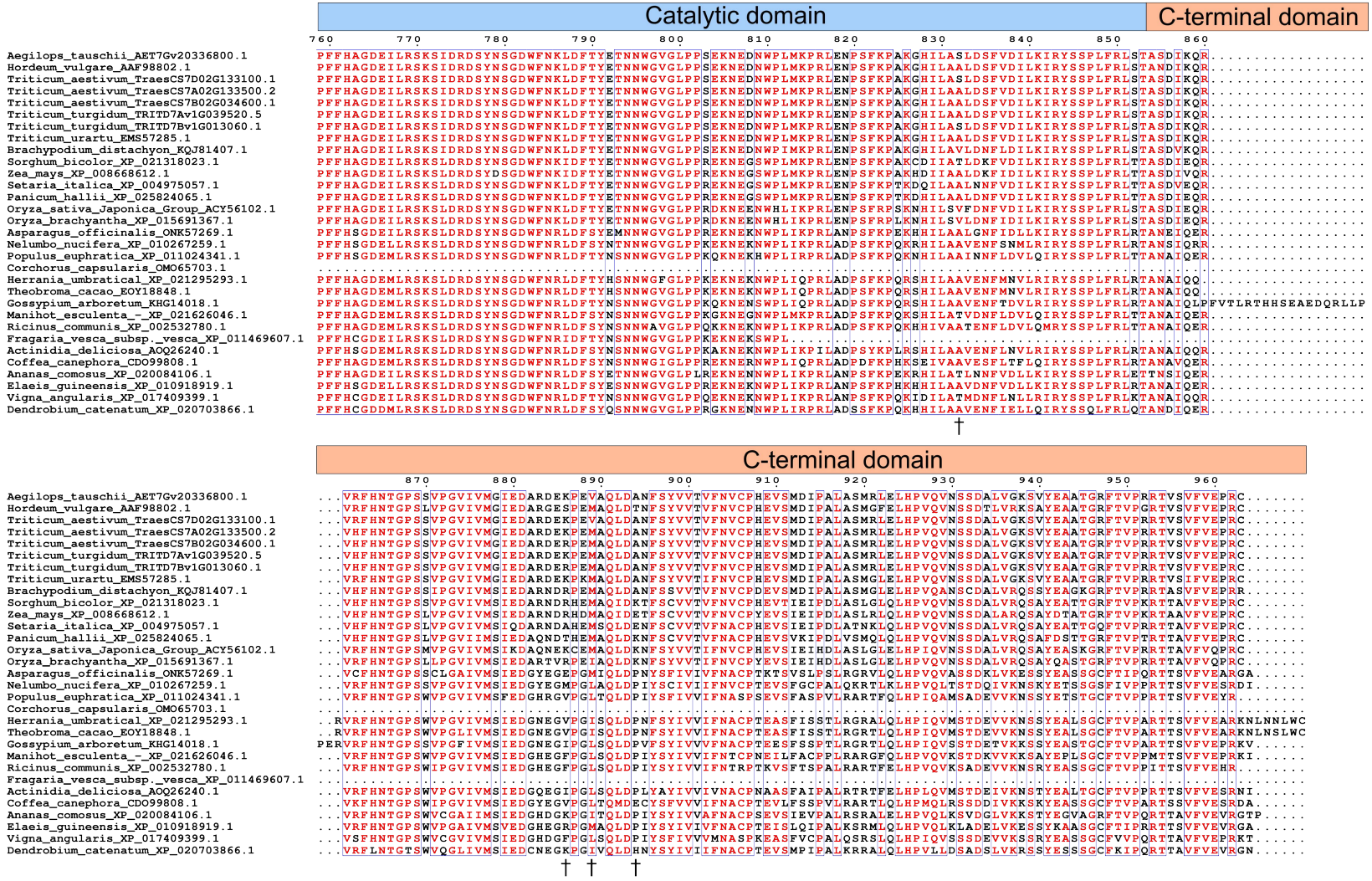


Figure S10 Alignment of amino acids 759-963 in LDA from different species. Accessions are arranged in the same order as in Andersen *et al.* (2020). Locations of the mutations exclusively found in TOWWC177 are marked with †. Residue colouring is the same as described in Figure 5.9a, and the corresponding protein domains are above all alignments. Numbering is relative to the *Ae. tauschii* sequence.