A roadmap for gene functional characterisation in wheat

Nikolai M. Adamski*, Philippa Borrill*, Jemima Brinton*, Sophie Harrington*, Clemence Marchal* and Cristobal Uauy

John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

* Authors contributed equally to this work

Nikolai M. Adamski: Nikolai.Adamski@jic.ac.uk
Philippa Borrill: Philippa.Borrill@jic.ac.uk
Jemima Brinton: Jemima.Brinton@jic.ac.uk
Sophie Harrington: Sophie.Harrington@jic.ac.uk
Clemence Marchal: Clemence.Marchal@jic.ac.uk
Cristobal Uauy: Cristobal.Uauy@jic.ac.uk (corresponding author)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26877v1 | CC BY 4.0 Open Access | rec: 20 Apr 2018, publ: 20 Apr 2018

Introduction

Research in *Arabidopsis* and other model species has uncovered mechanisms regulating important biological processes in plants. With the advent of high quality functional genomic resources in wheat it is now possible to use this knowledge for crop improvement directly in wheat.

Domesticated wheat can be divided into tetraploid pasta wheat (*Triticum durum*) and hexaploid bread wheat (*Triticum aestivum*). Polyploid wheat is the result of two hybridisation events (reviewed in (Matsuoka, 2011)) and thus each gene can be expected to have two (tetraploid pasta wheat) or three (hexaploid bread wheat) copies. These closely related copies, known as homoeologous genes, are on average >95% similar across their coding region (Figure 1) and have a highly conserved gene structure. Tetraploid and hexaploid wheat have large genomes, 10 and 15 Gb respectively, which consist mostly (~80%) of repetitive elements.

The combination of these factors has for a long time hampered development of genomics tools in wheat. In recent years though, this has changed dramatically and there are now a set of tools and resources available. An additional limitation had been the generation time of wheat, which ranges from four to six months depending on the requirement of cold periods (vernalisation) to induce flowering. Again, recent advances in growth conditions have radically changed these timeframes (Watson *et al.*, 2018). Wheat has now become a tractable system for translational, comparative and functional genomics and thus attractive for both wheat and non-wheat researchers.

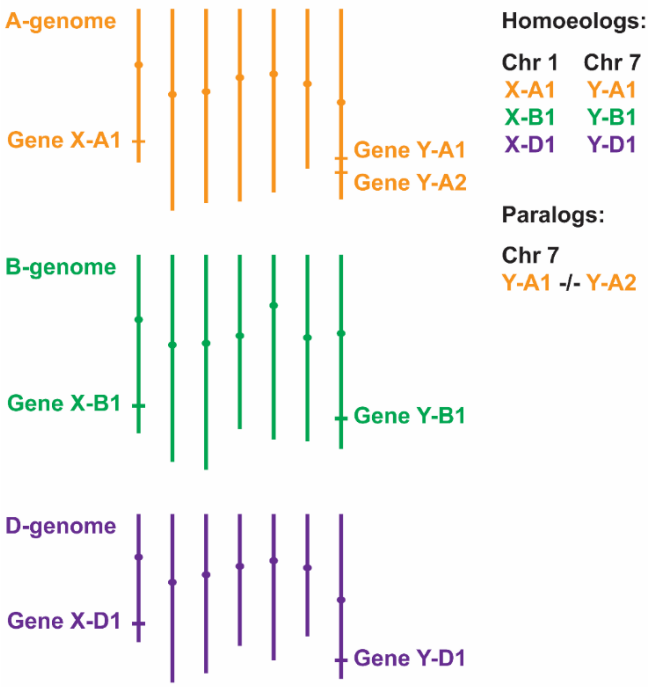

Figure 1: Gene homology within polyploid wheat. Due to two separate hybridisation events, genes in polyploid wheat will be present in multiple copies, called homoeologs, which usually have similar chromosome locations (e.g. Gene X on chromosomes 1A, 1B and 1D). Similar to other species, duplicate genes that are unique to one genome, called paralogs (e.g. two copies of Gene Y on chromosome 7A), have evolved, either within wheat or in one of its ancestral species.

In the current review, we describe some of the recent developments in wheat genomics focussing on published and publicly available resources and tools. We lay out a roadmap on how to make use of them (Figure 2) and include a case study to exemplify them. We hope this review will be a helpful guide for plant scientists who already work on wheat or who are considering expanding their research into wheat.

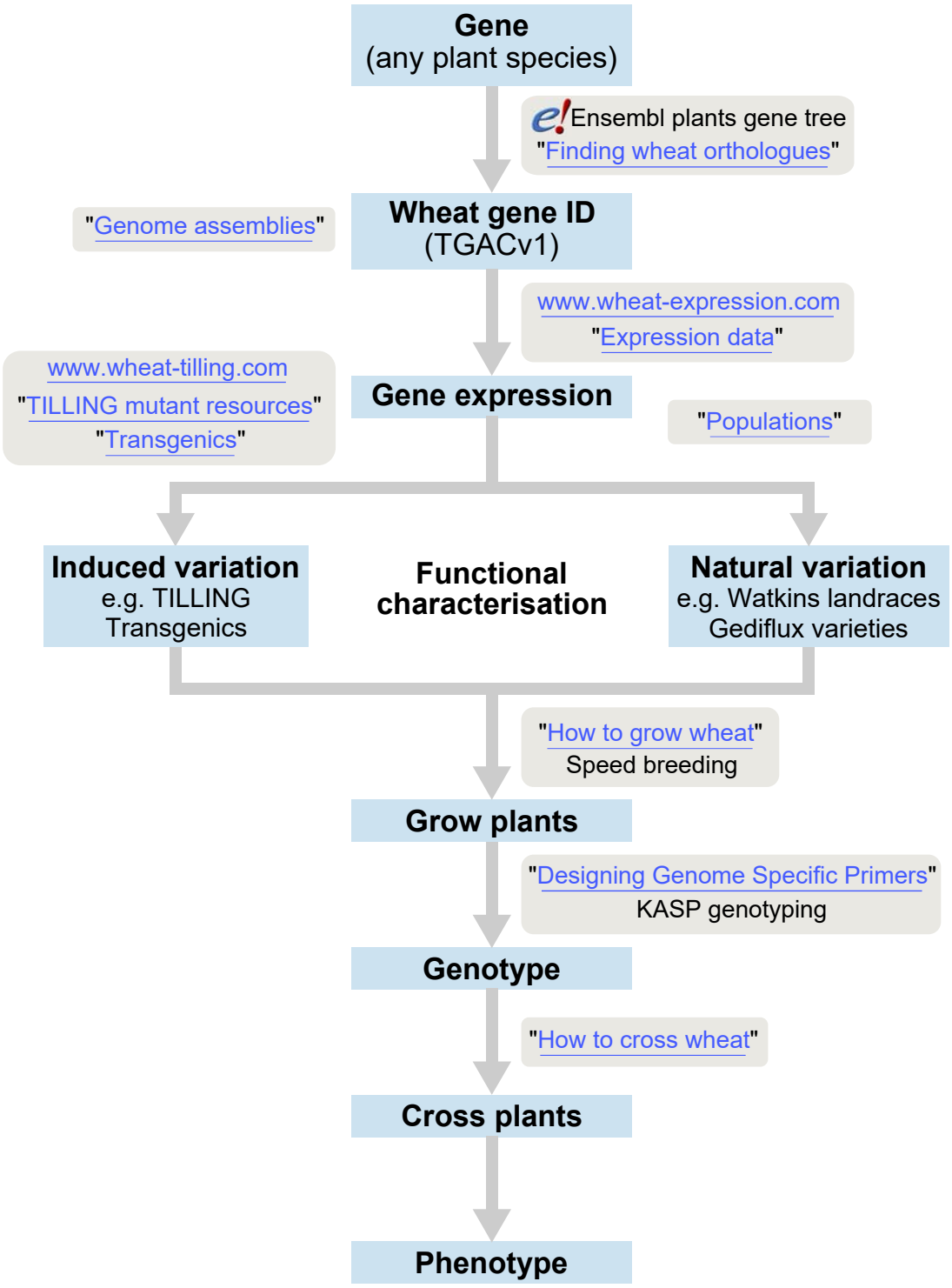

Figure 2: The roadmap for gene characterisation in wheat provides an overview of how to take a gene from any plant species, identify the correct wheat ortholog(s) using EnsemblPlants and determine gene expression using www.wheat-expression.com. Suggestions for functional characterisation are provided including induced variation such as mutants (www.wheat-tilling.com) or transgenics. In addition, publicly available populations incorporating natural variation are available (LINKS). Finally, steps for growing, genotyping and crossing plants are outlined. Links to detailed tutorials and further information for all steps are provided and can be found on www.wheat-training.com.

Wheat assemblies

A high-quality genome reference sequence is an essential resource for functional genetics and genomics in any species. Several genome assemblies of wheat have been released over the past six years (Brenchley *et al.*, 2012; IWGSC, 2014; Chapman *et al.*, 2015; Clavijo *et al.*, 2017; Zimin *et al.*, 2017) and are summarized in Table 1. Here, we will focus on the two most widely used publicly available assemblies, namely the Chromosome Survey Sequence (CSS) (IWGSC, 2014) and the TGACv1 (Clavijo *et al.*, 2017). While the assemblies by Chapman *et al.* and Zimin *et al.* are more contiguous than the CSS and TGACv1 respectively, they lack annotation, which limits their use for gene functional characterisation.

Table 1. Comparison of annotated genome assemblies in wheat. Currently available (CSS and TGACv1) and soon to be publicly available (RefSeqv1.0) annotated assemblies are shown. Currently, TGACv1 is the most widely used assembly and is available on *Ensembl*Plants (http://plants.ensembl.org/Triticum_aestivum/Info/Index). *Ensembl*Plants also enables access to SNP variation, gene trees and homoeolog assignments

	CSS	TGACv1	RefSeqv1.0*
Publication/Release date	IWGSC, 2014	Clavijo *et al.*, 2017	IWGSC, 2018
Contigs/ Chromosomes	> 1 million	735,943	21 chromosomes + ChrU
Mean scaffold size	7.7 kb	88.7 kb	Full Chromosomes
Assembly Size	10.2 Gb	13.4 Gb	14.6 Gb
Order	Crude order	Large Bins	"True" physical order
Coding genes†	100,934	104,390	HC and LC genes
Resources‡	Archive *Ensembl*Plants	*Ensembl*Plants	(*Ensembl*Plants)
	TILLING mutants		(TILLING mutants)
	expVIP, wheatExp	expVIP	(expVIP)
Variety	Chinese Spring	Chinese Spring	Chinese Spring

* The RefSeqv1.0 genome is currently available under Toronto agreement from https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies and resources in brackets will be made available upon publication.
† Number of high confidence (HC) genes are shown for CSS and TGACv1. These annotations also include low confidence (LC) genes which are defined based on multiple criteria outlined in the published papers. RefSeqv1.0 also includes HC and LC genes and care must be taken when interpreting their nomenclature (see Figure 2).
‡ TILLING mutants can be accessed through www.wheat-tilling.com and expression data is available at www.wheat-expression.com. More details and tutorials are available at www.wheat-training.com.

To generate the CSS assembly, cytogenetic stocks of the hexaploid wheat landrace Chinese Spring were used. Chromosome arms were separated and purified via flow-sorting and then Illumina-sequenced. The resulting assembly accounted for 10.2 Gb of the wheat genome, consisting of approximately 1.8 million contigs with average lengths (N50 values) ranging from 1.7 to 8.9 kb across chromosome arms. A total of 100,344 high confidence (HC) gene models (v2.2) were predicted (Table 1, Figure 3). The highly fragmented nature of this assembly meant that no physical order of contigs could be obtained, but 44% of contigs and 56% of gene models were anchored genetically using a high-density genetic map (Mascher *et al.*, 2013).

The TGACv1 assembly is also based on the hexaploid wheat landrace Chinese Spring, but it was Illumina-sequenced directly without flow-sorting of chromosomes. Improved assembly algorithms resulted in 735,943 scaffolds with N50 of 88.8 kb and accounting for 13.4 Gb of the wheat genome (Clavijo *et al.*, 2017) and a total of 104,091 HC gene models were predicted (Table 1, Figure 3). Although the TGACv1 also has no physical order, half of the assembly, represented by 63.7% of scaffolds and containing 72.7% of HC gene models, was anchored using an improved high-density genetic map (Chapman *et al.*, 2015).

Whilst no full comparison of the CSS and TGACv1 assemblies exists, the assembly statistics indicate that the CSS is more fragmented than TGACv1. A brief analysis of a few gene families (disease resistance genes, glutenins, and members of the gibberellin biosynthetic and signal transduction pathways) also shows that the TGACv1 gene models are more complete than the CSS models (Clavijo *et al.*, 2017). Furthermore, a

single TGACv1 gene model is often represented by multiple (fragmented) CSS gene models (Brinton *et al.*, 2018). Taken together, the TGACv1 assembly and its gene models are considerably more robust and reliable than the CSS. However, several key resources still make use of the CSS for historical reasons, a prominent example being the wheat Target Induced Local Lesions in Genome (TILLING) database (Krasileva *et al.*, 2017). Hence it is important to understand the merits and limitations of the CSS gene models.

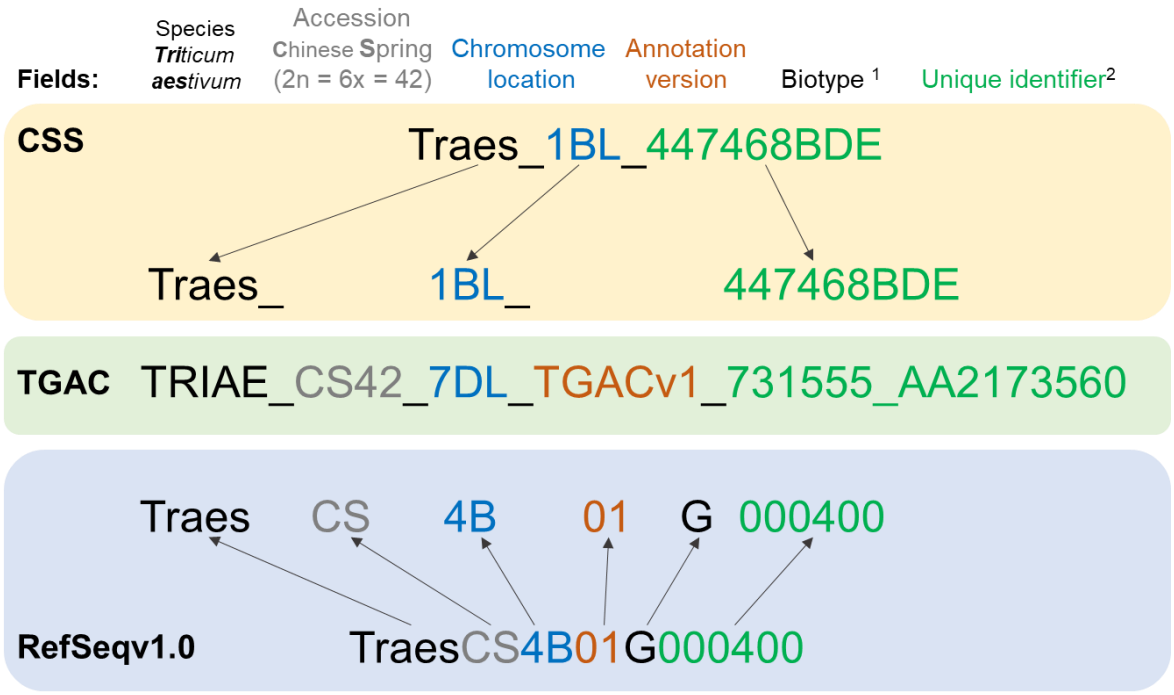

Figure 3. Gene nomenclature description from the three available gene annotations for wheat. Fields represented in the nomenclature are shown at the top with matching colours with the corresponding features in the gene names. Yellow background shows the Chromosome Survey Sequence (CSS) gene names with dark grey arrows pointing towards the corresponding field in the TGAC gene annotation (TGACv1, in green), the reference on *Ensembl*Plants. The gene nomenclature for the upcoming RefSeqv1.0 annotation is depicted in blue. [1] In RefSeqv1.0 biotype is represented as an additional identifier where G= gene. [2] In the RefSeqv1.0 annotation, identifiers are progressive numbers in steps of 100s reflecting the relative position between gene models. For example, gene TraesCS4B01G000400 would be adjacent to gene TraesCS4B01G000500. Note that RefSeqv1.0 comprises High Confidence and Low Confidence gene models. Low Confidence gene models are flagged by the "LC" at the end (not shown). HC and LC genes with the same unique identifier are **not** the same locus and are not in sequential order. Hence TraesCS4B01G000400 and TraesCS4B01G000400**LC** are both on chromosome 4B but are not physically next to each other.

It is important to mention that an improved assembly and annotation, called Refseqv1.0, is currently available, but under the Toronto agreement (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies). It consists of 21 chromosome pseudomolecules with annotated HC and LC gene models and will likely be publicly available within the next six months. We describe the impact that this new assembly will have on wheat genomics in the 'Future Directions' section at the end of this review.

Given that RefSeqv1.0 is not publicly available and has not been integrated into the different public databases and resources yet, we will focus on TGACv1 which currently represents the best assembly to use for functional genomics in wheat. This review will therefore focus on how to use the TGACv1 assembly and annotation to:

- Identify the correct wheat ortholog(s) for any given gene from another plant species
- Obtain expression data of wheat genes
- Perform functional genomics like transgenics, TILLING, CRISPR and VIGS
- Make use of natural variation and available wheat populations

We also outline strategies for growing, crossing and genotyping wheat using the latest available tools and techniques. Finally, we present a case study that encapsulates the above steps and that highlights potential pitfalls.

Finding wheat orthologs

Although sequence homology does not equate to functional homology, it represents a good starting point for translational and/or comparative genomics. Correctly identifying the right ortholog in another plant species can be a difficult task however, especially between distantly related species like *Arabidopsis* and polyploid wheat. These two species are separated by ~200 million years of evolution and as a result both nucleotide and protein similarities are low.

Conveniently, all the data and tools necessary for obtaining gene orthologs from different plant species are available through the *Ensembl*Plants website (http://plants.ensembl.org/index.html), which is a publicly available resource (Bolser *et al.*, 2016). In addition, *Ensembl*Plants have implemented the Plant Compara pipeline to create gene trees between all the species available on its website (Vilella *et al.*, 2009; *Ensembl*Plants, 2018). This includes the TGACv1 gene models, *Arabidopsis* TAIR10 and rice IGRSP1.0, amongst others. This represents a quick and reliable way to identify the wheat orthologs of a given gene (Figure 2). Tutorials for using *Ensembl*Plants can be found on their website or at www.wheat-training.com.

When performing a search for wheat orthologs via the *Ensembl*Plants gene trees we would expect to find three orthologs in hexaploid wheat for every gene query from another species. These orthologs would normally be located on homoeologous chromosome group, e.g. chromosomes 1A, 1B and 1D. A well-documented exception to this rule is the long arm of chromosome 4A (4AL), which has undergone translocation events with chromosome arms 5AL and 7BS (Devos *et al.*, 1995; Ma *et al.*, 2013). Therefore, orthologs within these translocated regions will be physically located on different chromosome groups, e.g. the three homoeologous genes will be on chromosome arms 4AL, 5BL and 5DL. Furthermore, we would expect that the gene structure of the wheat orthologs is conserved with respect to rice and other closely related species. If this is not the case, we would suggest exploring the transcriptomic evidence of the wheat gene models or comparing to additional well-annotated grass species such as *Brachypodium* and maize, which can be done using *Ensembl*Plants.

As an alternative to the *Ensembl*Plants Plant Compara pipeline one can also perform reciprocal protein BLASTs to identify wheat orthologs. We exemplify the above-mentioned approaches along with potential pitfalls in more detail in the 'Case Study' section below.

Expression data

Determining if, when and where a given candidate gene is expressed often constitutes one of the first steps towards its functional characterisation. Expression information is also a criterion to help in narrowing down a list of candidate genes underlying a QTL or to predict which members of a large gene family are relevant to the trait of interest. Due to the reduction in sequencing costs over the last decade, numerous RNA-seq datasets for wheat have been generated and published. Although the raw data are publicly available, it is no easy task to access and use them in direct comparisons due to the diversity in the samples' origins. Expression browsers aim to centralise these available datasets and analyse them together, ideally allowing retrieval of expression information for a given list of genes under different conditions. Two expression browsers, WheatExp (https://wheat.pw.usda.gov/WheatExp) and expVIP (http://www.wheat-expression.com), are currently available for wheat (Pearce *et al.*, 2015; Borrill *et al.*, 2016). Here we will focus on the latter

(Figure 2) given that it includes a larger and more diverse set of samples and uses the TGACv1 gene models described in Table 1.

Currently, expVIP includes expression data from 16 studies (~400 RNA-seq samples) covering different tissues across different developmental time points as well as across various abiotic and biotic stress conditions. It displays the expression data for a given gene (bar graphs) or a list of up to 50 genes across the different samples (heatmaps). This list option is particularly useful for investigating expression of a gene family or genes involved in the same regulatory process. There is also the possibility of displaying the expression values for the homoeologs of a given gene, which were defined based on the *Ensembl*Plants Plant Compara pipeline (Vilella *et al.*, 2009; *Ensembl*Plants, 2018). To allow comparisons across studies, the ~400 RNA-Seq samples in expVIP were classified according to four high-level categories based on variety, tissue, developmental stage and stress. These high-level categories are themselves divided into more detailed subcategories. These categories can be used to filter what is displayed on the main interface and select data relevant to the user. Data can be displayed both as transcripts per million (TPM) or as raw counts and can be directly downloaded to carry out differential gene expression analysis using programmes such as DESeq2 (using the csv file for raw counts) (Love *et al.*, 2014). Images of the graphs and heatmaps can also be retrieved. Although the default reference is TGACv1, users can also choose the CSS transcriptome reference for legacy reasons. Video and text tutorials describing expVIP are available on the www.wheat-training.com website.

Functional studies

After identifying a gene of interest in wheat there are now many opportunities and resources available for functional characterisation and validation (Figure 2). These include resources based both on natural and induced variation and can involve both transgenic and non-transgenic approaches. It is important to remember that due to the polyploid nature of wheat, there is often functional redundancy between homoeologs (Borrill *et al.*, 2015). This means that it may be necessary to manipulate all homoeologs simultaneously to obtain a phenotype (see the 'Strategies for use' section below for more information).

Induced variation

TILLING

Polyploid species, such as wheat, are well suited to mutational approaches as the functional redundancy in the genome allows for the tolerance of a higher mutational load compared with diploid species (Tsai *et al.*, 2013; Uauy *et al.*, 2017). Bespoke mutant populations can be developed and screened for desired mutations in a gene of interest, however this screening process is arduous and time-consuming. To overcome this barrier, an *in-silico* wheat TILLING resource has been developed through a collaboration between UK and US institutes (Krasileva *et al.*, 2017). This resource consists of two ethyl methanesulphonate (EMS) mutagenized populations: 1,535 lines of tetraploid durum wheat variety 'Kronos' and 1,200 lines of hexaploid bread wheat variety 'Cadenza'. The exome sequences of all 2,735 mutant lines have been captured and re-sequenced using Illumina next-generation sequencing (NGS); the raw data was aligned to the CSS reference, mutations were identified and their effects predicted based on the CSS gene models (IWGSC, 2014). Deleterious alleles have been identified in these populations for ~90 % of the captured wheat genes (Krasileva *et al.*, 2017), thus making this a powerful resource for rapidly identifying mutations in a gene of interest (Figure 2).

It is important to stress that currently mutation effects are predicted based on the CSS gene models, which, as discussed in the 'Wheat Assemblies' section, can be unreliable. It is therefore important to check the predicted effect of mutations in the context of a complete and correct gene model when selecting mutant lines. Another important consideration is that crossing is necessary to combine mutations in order to generate a complete null individual. In addition, mutant lines will contain a high level of background mutations. Depending on the phenotype of interest several rounds of backcrossing may be required before

the phenotype can be assessed (see 'Strategies for use'). All data for these populations are publicly available on www.wheat-tilling.com and practical information about selecting mutant lines and downstream analyses can be found at www.wheat-training.com/tilling-mutant-resources.

Transgenic approaches

Transformation of wheat can be performed using a variety of methods including both particle bombardment (Vasil *et al.*, 1992; Sparks and Jones, 2009) and *Agrobacterium*-mediated transformation (Cheng *et al.*, 1997; Sparks *et al.*, 2014). The generation of stable transgenic lines in wheat most commonly involves the transformation of immature wheat embryos and subsequent callus regeneration (Harwood, 2012). Using transgenic approaches, gene expression can be altered in a variety of ways such as overexpressing or ectopically expressing the gene of interest using either constitutive, tissue-specific or inducible promoters (Hensel *et al.*, 2011). Similarly, RNA-interference (RNAi) has been used successfully in wheat to reduce gene expression with the added benefit that constructs can be designed to target all homoeologous genes simultaneously, thereby overcoming the potential drawback of functional redundancy among homoeologs (Fu *et al.*, 2007). In addition to altering expression patterns, modified proteins can also be introduced (e.g. including tags) that can be used for downstream experiments, such as ChIP-seq (Deng *et al.*, 2015) or localisation studies (Harwood *et al.*, 2005), although these are still not commonly employed in wheat. As transformation methods have only been optimised for a limited number of wheat varieties it is important to understand whether the gene is expressed/functional in the chosen variety when defining transgenic strategies (see 'Strategies for use'). Transient gene silencing through Virus Induced Gene Silencing (VIGS) has also be performed in wheat primarily to investigate disease resistance. VIGS has been carried out in a range of varieties, but is generally restricted to a few tissue types such as leaf tissues (Lee *et al.*, 2015), young seedlings (Zhang *et al.*, 2017a) and spikes (Ma *et al.*, 2012).

In addition to these more traditional transgenic approaches, the recent developments in genome editing technologies provide new opportunities for manipulating genes in wheat. Cas9-mediated genome editing has been successfully demonstrated in wheat both in transient expression systems (Shan *et al.*, 2014) and stably transformed plants (Wang *et al.*, 2014), using a range of methods (reviewed in (Uauy *et al.*, 2017)). Currently, most studies have introduced specific point mutations or small deletions leading to subsequent protein disruption, although the technology holds the potential to have more complex applications such as allele swapping or gene insertion in the future (Puchta, 2017). Similar to RNAi, constructs for Cas9-mediated gene editing can be designed to target all homoeologs simultaneously (Zhang *et al.*, 2016). Due to the current efficiency of genome editing however, the likelihood of obtaining mutations in all homoeologs in a single T_0 plant remains low and subsequent crosses to combine edits are likely required. One of the major limitations of using transgenic approaches to manipulate agronomically relevant traits are the associated regulatory constraints. To overcome this, the nuclease transgene can be segregated away from the edited gene(s) in subsequent generations and studies have also documented methods of Cas9-editing in wheat that avoid transgene integration altogether (Liang *et al.*, 2017).

Natural Variation

Natural variation has been extensively documented in wheat. Most studies have focused on single nucleotide polymorphisms (SNPs) between varieties that can be quickly assayed through SNP arrays designed from gene coding sequences (described in (Borrill *et al.*, 2015) and www.wheat-training.com). Thousands of varieties and landraces have been processed using these arrays and datasets are available through websites such as TCAP (https://triticeaetoolbox.org/wheat) (Blake *et al.*, 2016) and CerealsDB (http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB) (Wilkinson *et al.*, 2016). Given that all SNPs from the latter have been incorporated into *Ensembl*Plants, this means that large allelic series are readily available *in silico* for many genes of interest.

The challenge remains, however, to define the functional significance of this variation. Traditionally, mapping populations or association panels would need to be developed or assembled, and then

genotyped, to assess how particular SNPs or haplotypes affect the trait of interest. In wheat, many of these resources are now publicly available (Figure 2), thus facilitating the functional characterisation of genes of interest. We describe some of these resources below:

Watkins landraces: The Watkins core collection constitutes wheat landraces from around the world that were grown as local farmer-saved seed before the 1930's. A core set of 107 landraces represents the majority of the variation found in this collection (Wingen *et al.*, 2014). These 107 Watkins landraces have been genotyped with SNP arrays (data in CerealsDB and *Ensembl*Plants) and used to generate $F_{5:6}$ mapping populations against the common parent Paragon (Wingen *et al.*, 2017). Paragon is a spring hexaploid wheat that has recently been fully sequenced and is available at Grassroot Genomics (http://www.earlham.ac.uk/grassroots-genomics). This means that for any gene of interest, users can identify natural variation using the CerealsDB/*Ensembl*Plants SNP data and determine which of the 107 Watkins landraces has potentially significant natural variation with respect to the sequenced Paragon line. Seeds for the $F_{5:6}$ mapping population can be directly ordered from the John Innes Centre alongside a basic genetic map of ~150-200 SNP markers, for the corresponding Watkins x Paragon cross. Full details are available at http://wisplandracepillar.jic.ac.uk/results_resources.htm.

Gediflux varieties: As for the Watkins landraces, populations for over 110 wheat varieties from Europe and CIMMYT (released between 1920 and 2000's) were created using the same common parent Paragon. Again, all SNP data is available through CerealsDB and *Ensembl*Plants and populations at different stages are available (http://wisplandracepillar.jic.ac.uk/results_resources.htm). Although genetic maps have not been developed, they can be quickly generated using available SNP platforms.

For both Watkins and Gediflux populations, however, it might not be necessary to generate full genetic maps if the intention is to assess lines with contrasting alleles/haplotypes at a candidate gene. Lines from any mapping population can be genotyped for the DNA variant(s) using SNP markers and then assessed for the phenotypic effects of the contrasting haplotypes in the $F_{5:6}$ populations. Half of the lines each will carry the alternative haplotypes at the target loci, while segregating for background alleles elsewhere in the genome, providing functional validation of the effects of the natural variation on the phenotype of interest. These populations are all free and publicly available as part of the UK Designing Future Wheat Programme (http://wisplandracepillar.jic.ac.uk/results_resources.htm).

Multiparent Advanced Generation Inter-Cross (MAGIC) population: An eight-parent MAGIC population has also been developed from UK/North European winter wheat genotypes. This population consists of over 700 genotyped F_7 lines (Mackay *et al.*, 2014) and a high-density genetic map (Gardner *et al.*, 2016), with seeds also being publicly available from NIAB (http://www.niab.com/pages/id/402/NIAB_MAGIC_population_resources). The multiple generations of inter-crossing to make MAGIC populations means they have highly recombined chromosomes which enables the use of GWAS to define small genetic intervals for traits of interest. Likewise, the use of eight parents allows more allelic variation to be examined compared to a bi-parental population. Two of the eight parents have already been sequenced by the Earlham Institute (http://www.earlham.ac.uk/grassroots-genomics), while the other six are currently under way.

Moving towards a wheat pangenome: The increase in sequencing output has meant that varieties different from Chinese Spring have been fully sequenced to a relatively high standard of contiguity. These include several hexaploid (Cadenza, Claire, Robigus, Paragon, Julius, Landmark, Jagger, Arina*LrFor*) and tetraploid (Kronos, Svevo, Zavitan) varieties/accessions (Table 2). Annotation of these varieties is ongoing through the 10+ Genome Project (http://www.10wheatgenomes.com). The re-sequencing of these varieties provides information on variation across non-coding regions including promoter sequences for this defined set of germplasm. Importantly, several of these genotypes are part of the structured populations outlined above, i.e. the MAGIC population.

Table 2: Currently available wheat genome assemblies for varieties different to the reference Chinese Spring landrace.

Variety	Habit	Origin	Availability *
Hexaploid bread wheat			
CDC Landmark	spring	Canada	10+ Genome Project
Arina*LrFor*	winter	Switzerland	10+ Genome Project
Julius	winter	Germany	10+ Genome Project
Jagger	winter	US	10+ Genome Project
Paragon	spring	UK	10+ Genome Project
Cadenza	spring	UK	10+ Genome Project
Synthetic W7984	spring	Mexico	(Chapman *et al.*, 2015)
Robigus	winter	UK	10+ Genome Project
Claire	winter	UK	10+ Genome Project
Tetraploid pasta wheat			
Zavitan†	-	Israel	(Avni *et al.*, 2017)
Svevo	spring	Italy	Interomics
Kronos	spring	US	10+ Genome Project

† Zavitan is a tetraploid wild emmer (*T. dicoccoides*) accession.
* Varieties included within the 10+ Genome Project can be accessed through the Earlham Grassroot Genomics portal (https://wheatis.tgac.ac.uk/grassroots-portal/blast) and the 10+ Genome project portal (http://webblast.ipk-gatersleben.de/wheat_ten_genomes) (subset of varieties in each). The Svevo genome can be accessed through https://www.interomics.eu/durum-wheat-genome subject to Toronto agreement. Synthetic W7984 and Zavitan can be accessed through the Grassroot and 10+ Genome portal, respectively.

Strategies for use

Variety selection and growth conditions

Whilst resources are now available for the functional validation of target genes in wheat, practical knowledge is also required to maximise the value of these resources. Firstly, wheat varieties are adapted to different growing conditions making it important to consider the conditions under which functional validation will be carried out. If functional validation will be conducted in greenhouse or controlled environment room conditions then most varieties will be suitable, although varieties without vernalisation requirements are faster to grow (details on wheat growth conditions at www.wheat-training.com). The TILLING populations in Kronos (tetraploid) and Cadenza (hexaploid) do not require vernalisation and come from different regions of the world, facilitating field trial validation in addition to greenhouse experiments. Kronos is a Californian variety adapted to warm dry weather whereas Cadenza is a UK variety adapted to cooler conditions.

For CRISPR and other transgenic approaches several varieties may be used, although only a few wheat varieties have high enough transformation efficiencies to be practical. This means that most transgenic studies in wheat are limited to a few varieties, such as Fielder, Cadenza, Bobwhite, Kenong 199 and Kronos (Li *et al.*, 2012; Richardson *et al.*, 2014; Liang *et al.*, 2017). This is now changing thanks to work by groups around the world including Emma Wallington and colleagues who have expanded this portfolio to 39 varieties (http://www.niab.com/pages/id/90/crop_transformation), alongside Xingguo Ye whose group has transformed 15 Chinese varieties (Wang *et al.*, 2017). Transformation efficiencies however still differ between varieties. Correct varietal selection for transformation is important given that some varieties might not be suitable to study a particular phenotype (e.g. if the variety is resistant to a disease and hence cannot be used for complementation studies). Similarly, it is important to assess whether the gene of interest is present/functional in the chosen variety, for example through PCR amplification of the gene. For Kronos and Cadenza, this can be done quickly by examining their genome sequence available through the Earlham Institute (https://wheatis.tgac.ac.uk/grassroots-portal/blast).

Combining mutations for complete knock-outs in polyploid wheat

The polyploid nature of wheat means that it normally has three homoeologous copies of every gene. These copies are highly similar and may have redundant functions (Borrill *et al.*, 2015). Therefore, to characterise the function of a gene in wheat it is often necessary to knock out all three homoeologs. This may be achieved by simultaneously targeting all three copies using either RNAi (Uauy *et al.*, 2006) or CRISPR (Zhang *et al.*, 2017b). A large number of transformants need to be screened to identify a null in all three genomes from a CRISPR construct (Zhang *et al.*, 2017b). If the targets are more divergent it may not even be possible to use one guide RNA to target all three homoeologs, in which case several guide RNAs may be used through multiplexing. Alternatively, separate knock-outs for each homoeolog can be generated using CRISPR or identified in TILLING mutant populations. The mutations in each homoeolog can be combined by crossing (for details see www.wheat-training.com), with two crosses necessary to combine knock-out mutations in hexaploid wheat (Figure 4). Tetraploid wheat, with only two homoeologs, can be used to accelerate functional characterisation as it requires just one cross to create complete knock-out mutants (Figure 4). After self-pollination of this F_1, phenotyping of the trait of interest can be carried out in the F_2 generation. It is important to note that TILLING lines contain many background mutations and backcrossing may be required depending on the phenotype. More details on these strategies are published in (Uauy *et al.*, 2017).

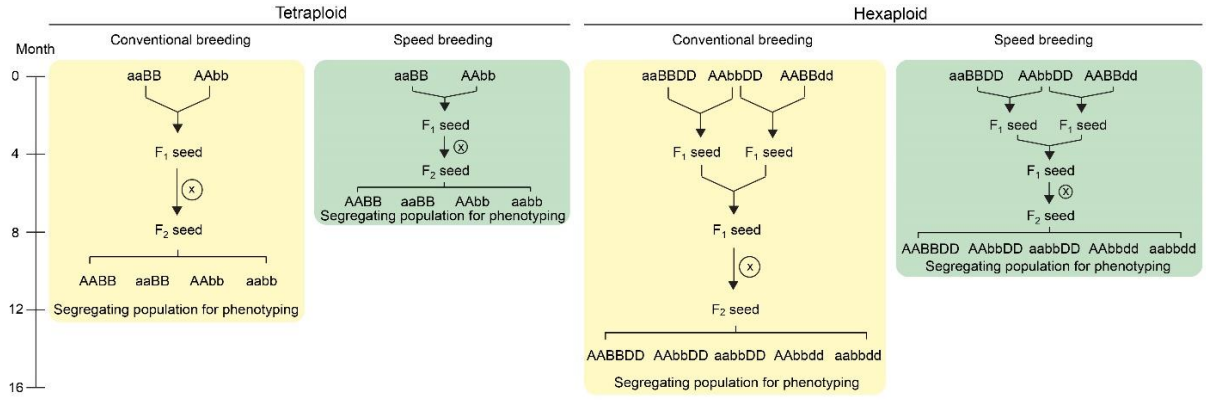

Figure 4. Crossing scheme to combine TILLING or CRISPR single mutants. In tetraploids, mutations in the A and B genome homoeologs of a gene of interest can be combined through a single cross. The F_1 plants are self-pollinated to produce a segregating F_2 population which contains homozygous double and single mutants, as well as wild type plants (screening by PCR required; only four genotypes shown). These F_2 progeny can be characterised for the phenotype of interest. The use of speed breeding reduces the time taken to reach this phenotyping stage from 12 (yellow) to 7.5 months (green). In the hexaploid population a second round of crossing is required to combine the three mutant alleles. The F_2 progeny segregating for the three mutant alleles can be genotyped using PCR to select the required combination of mutant alleles (only 5 genotypes shown; all factorial combinations are possible). Speed breeding reduces the time taken to generate triple homozygous mutants for phenotyping to 10 months (green), compared to 16 months in conventional conditions (yellow). Self-pollination is represented by an X inside a circle.

Speed breeding to accelerate crossing and phenotyping

The need to combine multiple mutations/alleles and carry out backcrossing to remove background mutations takes a considerable amount of time, with at least four months required per generation. Recently, the “speed breeding” technique has been implemented in wheat, which uses extended day lengths of 22 hours and improved light quality to accelerate the generation time in wheat (Watson *et al.*, 2018). Reduction of generation times to 8-10 weeks is achieved through an accelerated growth rate and harvesting of immature seeds 2-3 weeks post anthesis. The immature seeds are dried and then imbibed in the cold, resulting in nearly 100% germination. Incorporating speed breeding within crossing programmes

can reduce the time required to produce and phenotype double mutants in tetraploid wheat to less than 7.5 months and triple mutants in hexaploid wheat to less than 10 months (Figure 4). In addition to reducing generation times, it has been shown that several traits of interest such as disease resistance, height and flowering time can be properly characterised under speed breeding conditions (Watson *et al.*, 2018).

Homoeolog specific PCR markers

To carry out a successful crossing programme it is essential to select for the mutations of interest. In polyploid wheat it is necessary to track mutations in each homoeolog separately, which can be achieved using homoeolog-specific PCR. Primers can be made homoeolog-specific by including a SNP at the 3' end of the primer (www.wheat-tilling.com and www.wheat-training.com). The primer will amplify the targeted homoeolog much more efficiently than the non-targeted homoeolog resulting in genome-specific amplification. Rapid design of homoeolog specific primers can be achieved using the PolyMarker automated pipeline (Ramirez-Gonzalez *et al.*, 2015) and webserver (http://polymarker.tgac.ac.uk). Routinely, genotyping of SNPs is carried out using Kompetitive Allele Specific PCR (KASP) markers which are high throughput, inexpensive and can be used in individual lab settings equipped with PCR machines and widely available fluorescence plate readers (Allen *et al.*, 2011). The SNP to be genotyped (e.g. between mutant and wild type) will be located at the 3' end of the forward primer, with two alternative forward primers used in the reaction (one for the mutant and one for the wild type allele), whilst the homoeolog-specific SNP is located at the 3' end of the universal reverse primer (or *vice versa*). Amplification will thus be homoeolog-specific and allele-specific. As the primers amplify the DNA, an allele-specific tail on the allele-specific primer will also be amplified. This allele-specific tail binds one of two different FRET cassettes, which is no longer quenched and emits fluorescence. Which of the FRET cassettes becomes unquenched depends upon the amplified tail and reveals which SNP allele(s) is present. This fluorescent signal can be read in a fluorescence plate reader. Guidance on the design of genome-specific primers and KASP markers is available (www.wheat-tilling.com and www.wheat-training.com).

Case study

To put the previous resources into context, we present a case study for obtaining wheat mutants and expression data using a gene of interest from *Arabidopsis thaliana*. The MADS-box transcription factor *AGL21* influences the development of lateral roots in *Arabidopsis* and we hypothesize that the wheat orthologs have the same function (Yu *et al.*, 2014). The first step in testing this hypothesis is to identify wheat orthologs of *AGL21*, which can be done using *Ensembl*Plants (Bolser *et al.*, 2016). The *Ensembl*Plants gene tree displays predicted orthologs for all species included on *Ensembl*Plants. *AGL21* is one of four closely-related *Arabidopsis* MADS-box transcription factors in what will be called the "*AGL21* clade" for the purpose of this case study (Figure 5A). The *AGL21* tree contains two orthologous gene triads (sets of three homoeologous genes) in wheat, one on chromosome 2 and one on chromosome 6 (Figure 5A). As this duplication is also seen in *O. sativa*, *Z. mays* and *B. distachyon*, these two triads likely represent a duplication in the cereal ancestor.

To confirm the predicted orthologs obtained from the *Ensembl*Plants gene tree, we also recommend carrying out comparisons between the wheat and rice orthologs. In our case, the rice orthologs of our two triads have slightly different gene structures than the wheat orthologs. In each case, the wheat orthologs have one additional exon annotated compared to the rice orthologs. To check whether these rice genes are the correct orthologs, we can use BLASTp on *Ensembl*Plants to search the wheat proteome for each rice protein sequence. After doing this, we see that the expected wheat orthologs are the top three hits for the A-, B-, and D-genome (Figure 5B). We can also see that the other wheat triad constitutes the next best hit, again supporting the close evolutionary relationship between the two triads.

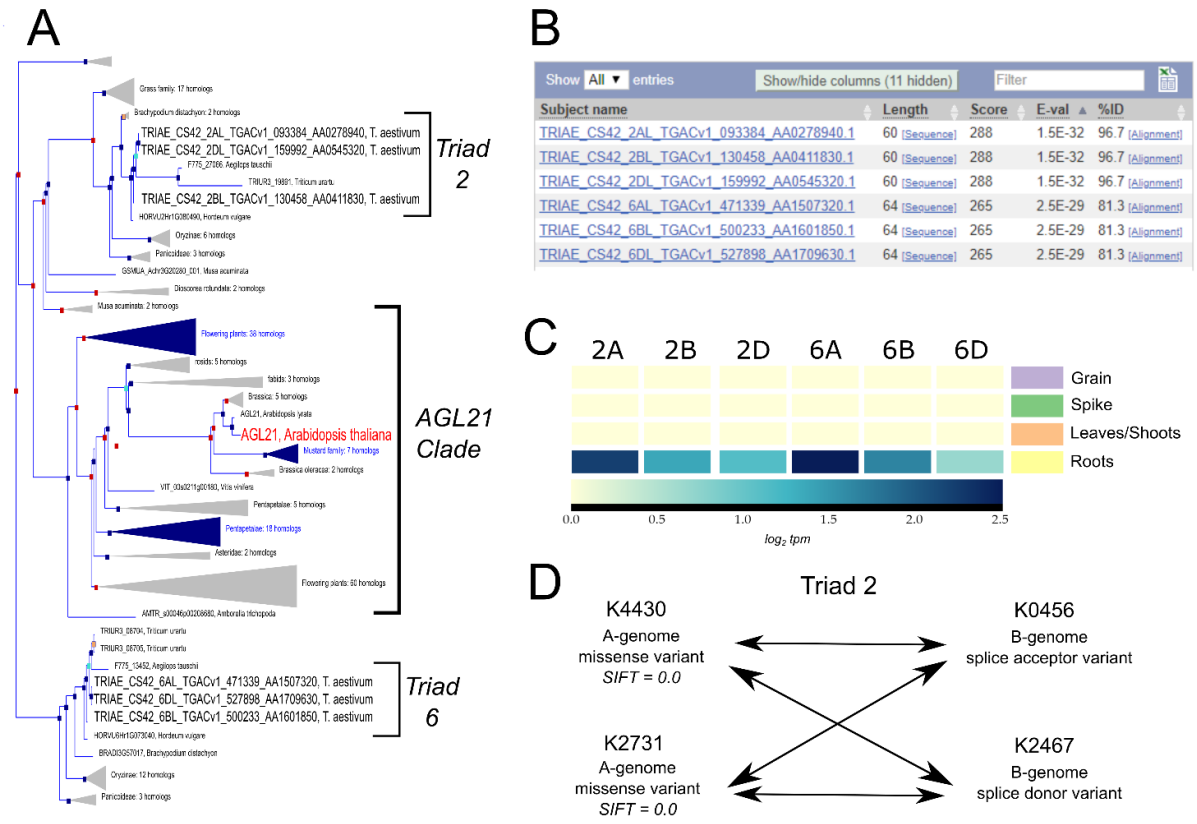

Figure 5: Case study exemplifying use of available gene functional characterisation in wheat. The *Ensembl*Plants Gene Tree (A) illustrates the identification of the two wheat triads (Triad 2 and Triad 6) that are related to *AGL21* (shown in red). Running BLASTp on the Triad 6 rice ortholog Os02g0731200 hits Triad 6 followed by the closely related Triad 2 (B). Examination of expression data from www.wheat-expression.com (C) shows that the two triads have very similar expression patterns across tissues; both triads are predominantly expressed in the roots. An example crossing scheme for Triad 2 is illustrated in (D), demonstrating the four crosses required between the two selected mutations in each homoeolog.

We can now take our rice orthologs and BLAST back against *Arabidopsis* to confirm that we are using the correct orthologs. Here, using BLASTp, we find that the top four hits for the triad 2 ortholog in rice, Os04g0461300, are within the *AGL21* clade, including *AGL21*. Similarly, the triad 6 ortholog, Os02g0731200, also has *AGL21* as one of its top hits. Notice, however, that the top hit for Os02g0731200 is a different protein, AGL13, which does not fall within the *AGL21* clade. This might indicate that the putative rice ortholog of *AGL21* is incorrect. We can investigate this by looking at the gene tree of *AGL13*, which places it in the context of an entirely separate set of *Agamous-like* (*AGL*) genes. Going back to the BLASTp output, we can also see that while AGL13 has the highest result, the % ID of the hit is lower than that of AGL21, directly below it, and the remaining BLAST hits are principally to other genes in the *AGL21* clade. Finally, we can BLAST the AGL13 protein back against rice, where we see that Os02g0731200 is not a top hit. Based on the combination of BLAST and *Ensembl*Plants Plant Compara information, we can conclude that Os02g0731200 is most likely an ortholog of AGL21 rather than AGL13. Further explanation about how to correctly identify wheat orthologs is available on www.wheat-training.com.

Having identified two wheat triads that are likely orthologs of *AGL21*, we can use the expVIP browser (www.wheat-expression.com) to compare their expression profiles (Borrill *et al.*, 2016) (Figure 2). Both wheat triads are highly expressed in the root and show very similar expression patterns (Figure 5C). Based on this expression profile, it seems that the two triads are equally likely to be functional orthologs of

AGL21. The best option is therefore to take both triads forward for mutant generation and phenotyping as discussed below.

After evaluating expression levels, the next step is to begin the process of phenotypic characterisation of the wheat orthologs (Figure 2). One option is to use the exome-sequenced TILLING mutant populations (Krasileva *et al.*, 2017). As the aim of this study is to functionally characterise the genes of interest, we would suggest using the Kronos population, as it is a tetraploid line and hence each 'triad' is only composed of two genes (A and B homoeologs). This means that only two mutants need to be crossed to generate a full knock-out of each "triad". The Cadenza population could also be used, but this would require an additional generation to combine all three mutant alleles.

All TILLING mutations are available from www.wheat-tilling.com and detailed instructions for the use of the website are available on http://www.wheat-training.com/tilling-mutant-resources. In brief, we can take the gene sequence of each homoeolog retrieved from *Ensembl*Plants and BLASTn against the TILLING mutant database. This returns a list of all SNPs that were called against the corresponding CSS scaffolds (see 'Wheat Assembles'). As mentioned previously, be aware that the predicted consequences of the mutations are based on the CSS gene models, and should be checked against the current TGAC gene models to ensure your SNPs of interest are annotated accurately. Mutations are chosen based on their predicted likelihood to cause a loss of function in the protein. Premature termination codons (PTCs) early within the protein sequence are often the best mutations to use. If there are no PTC mutants available, splice-site mutations that lead to downstream frameshifts, followed by missense mutations in highly conserved residues with low SIFT scores are good alternatives. For Triad 2, we find that the B homoeolog has various splice acceptor and donor variants, while the A homoeolog has multiple missense mutations with SIFT scores of 0, indicating that these mutations are likely to have a deleterious functional impact (Table 3). In the case of Triad 6, several low SIFT-score missense mutations are available for the B homoeolog, while a PTC mutant is available for the A homoeolog (Table 3). SIFT scores predict the effect of a mutation on protein function and are based on the physical properties of the alternative amino acid as well as sequence homology (Ng and Henikoff, 2003). In addition to SIFT as a measure of the potential effects of a missense mutation, we also recommend using the PSSM viewer (https://www.ncbi.nlm.nih.gov/Class/Structure/pssm/pssm_viewer.cgi) to help predict the effect of specific mutations on conserved protein domains.

Table 3. Kronos TILLING mutant lines in wheat orthologs of *AtAGL21*. Mutants were selected based on the likelihood that the mutation will cause a loss of function in the respective proteins for the A and B genome homoeologs.

Chr.	Line	Effect	SIFT
2A	K4430	Missense variant	0.00
	K2731	Missense variant	0.00
2B	K0456	Splice acceptor variant	-
	K2467	Splice donor variant	-
6A	K0875	Stop gained	-
	K2277	Splice donor variant	-
6B	K3591	Missense variant	0.00
	K4321	Missense variant	0.00

TILLING lines from both population can be ordered via SeedStor (https://www.seedstor.ac.uk/shopping-cart-tilling.php) or from the Dubcovsky lab (https://dubcovskylab.ucdavis.edu/wheat-tilling). To maximise the chance of having selected functionally important mutants, we recommend choosing two independent mutant lines for each homoeolog and carrying out four crosses between each mutant in the A and B genomes (shown for Triad 2 in Figure 5D). Detailed guides on growing wheat plants, genotyping TILLING mutants, and crossing mutants can be found on www.wheat-training.com.

Seedlings are genotyped to confirm that the correct mutation is present and to select for homozygous individuals for crossing. To do this, we design genome-specific primers to use in a KASP assay as outlined above and detailed on www.wheat-training.com. For most of the TILLING mutations genome-specific primers have already been designed, and can be accessed on the TILLING website. If there are no suitable primers designed already, online tools such as PolyMarker can be used (Ramirez-Gonzalez *et al.*, 2015). After carrying out the initial cross, we grow the F_1 individuals under speed breeding conditions, self-pollinate, and obtain the F_2 populations. Here, we select for homozygous double and single mutants as well as the homozygous wild type (WT) control (Figure 4). We can then carry out our first phenotypic screen on the F_2 plants using the homozygous WT lines as controls without the need for backcrossing to Kronos WT. We can do this because the background mutations in the chosen lines will be segregating within both the mutant and the WT lines, leading to an equivalent background mutation load between the genotypes (Uauy *et al.*, 2017). Backcrossing to WT Kronos can be started either with the single mutants while carrying out the initial cross and/or with the F_2 double mutant at a later stage. Backcrossing to remove background mutations is especially important when studying quantitative traits, such as grain size (Simmonds *et al.*, 2016), and when plants are intended for field phenotyping.

Future directions

In the last few years there has been a dramatic expansion in the resources available to carry out functional genomics in wheat, largely based upon improvements in the available reference sequences. Within the next six months a full pseudomolecule reference sequence alongside a detailed annotation (RefSeqv1.0) will most likely be released publicly (currently available under the Toronto agreement; Table 2), a step change from working with a fragmented assembly. This reference sequence will physically anchor genes in a complete chromosomal order and provide accurately annotated gene models, enabling more accurate design of transgenic constructs and primers. The majority of resources described in this review are in the process of being updated to use this new reference sequence including the expVIP expression browser, TILLING mutants and *Ensembl*Plants. As a result, it will be easier to use these resources as they will be unified by a common reference genome and gene models. Furthermore, in the case of the TILLING mutants, the mutations will be re-identified using highly reliable gene models, eliminating the need for multiple steps of manual curation before deployment in crossing programmes. Whilst wheat functional genomic resources are currently in a state of flux, the groundwork to accelerate gene discovery and characterisation in polyploid wheat has been laid.

Statements

Funding

This works was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) Designing Future Wheat programme BB/P016855/1, International Wheat Yield Partnership (IWYP) grant IWYP76, BBSRC Future Leader Fellowship BB/M014045/1 to PB, and the John Innes Foundation (SH). CM was funded by a PhD studentship from Group Limagrain.

Competing Interest

The authors declare that they have no competing interests.

Author contributions

All authors contributed to the design, analysis, and the preparation of figures. All authors wrote and approved the final manuscript.

References

Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ. 2011. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal* **9**, 1086-1099.

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, Jordan KW, Golan G, Deek J, Ben-Zvi B, Ben-Zvi G, Himmelbach A, MacLachlan RP, Sharpe AG, Fritz A, Ben-David R, Budak H, Fahima T, Korol A, Faris JD, Hernandez A, Mikel MA, Levy AA, Steffenson B, Maccaferri M, Tuberosa R, Cattivelli L, Faccioli P, Ceriotti A, Kashkush K, Pourkheirandish M, Komatsuda T, Eilam T, Sela H, Sharon A, Ohad N, Chamovitz DA, Mayer KFX, Stein N, Ronen G, Peleg Z, Pozniak CJ, Akhunov ED, Distelfeld A. 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**, 93-97.

Blake VC, Birkett C, Matthews DE, Hane DL, Bradbury P, Jannink JL. 2016. The Triticeae Toolbox: Combining Phenotype and Genotype Data to Advance Small-Grains Breeding. *Plant Genome* **9**.

Bolser D, Staines DM, Pritchard E, Kersey P. 2016. *Ensembl* Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. *Methods Mol Biol* **1374**, 115-140.

Borrill P, Adamski N, Uauy C. 2015. Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist* **208**, 1008-1022.

Borrill P, Ramirez-Gonzalez R, Uauy C. 2016. expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiol* **170**, 2172-2186.

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705-710.

Brinton J, Simmonds J, Uauy C. 2018. Ubiquitin-related genes are differentially expressed in isogenic lines contrasting for pericarp cell size and grain weight in hexaploid wheat. *BMC Plant Biol* **18**, 22.

Chapman JA, Mascher M, Buluc A, Barry K, Georganas E, Session A, Strnadova V, Jenkins J, Sehgal S, Oliker L, Schmutz J, Yelick KA, Scholz U, Waugh R, Poland JA, Muehlbauer GJ, Stein N, Rokhsar DS. 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol* **16**, 26.

Cheng M, Fry JE, Pang S, Zhou H, Hironaka CM, Duncan DR, Conner TW, Wan Y. 1997. Genetic Transformation of Wheat Mediated by *Agrobacterium tumefaciens*. *Plant Physiology* **115**, 971-980.

Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, Lipscombe J, Barker T, Lu FH, McKenzie N, Raats D, Ramirez-Gonzalez RH, Coince A, Peel N, Percival-Alwyn L, Duncan O, Trosch J, Yu G, Bolser DM, Namaati G, Kerhornou A, Spannagl M, Gundlach H, Haberer G, Davey RP, Fosker C, Palma FD, Phillips AL, Millar AH, Kersey PJ, Uauy C, Krasileva KV, Swarbreck D, Bevan MW, Clark MD. 2017. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res* **27**, 885-896.

Deng W, Casao MC, Wang P, Sato K, Hayes PM, Finnegan EJ, Trevaskis B. 2015. Direct links between the vernalization response and other key traits of cereal crops. *Nature Communications* **6**, 5882.

Devos KM, Dubcovsky J, Dvorak J, Chinoy CN, Gale MD. 1995. Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theor Appl Genet* **91**, 282-288.

***Ensembl*Plants**. 2018. Protein trees and orthologies.

Fu D, Uauy C, Blechl A, Dubcovsky J. 2007. RNA interference for wheat functional gene analysis. *Transgenic Res* **16**, 689-701.

Gardner KA, Wittern LM, Mackay IJ. 2016. A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnology Journal* **14**, 1406-1417.

Harwood WA. 2012. Advances and remaining challenges in the transformation of barley and wheat. *J Exp Bot* **63**, 1791-1798.

Harwood WA, Bilham LJ, Travella S, Salvo-Garrido H, Snape JW. 2005. Fluorescence *in situ* hybridization to localize transgenes in plant chromosomes. *Methods Mol Biol* **286**, 327-340.

Hensel G, Himmelbach A, Chen W, Douchkov DK, Kumlehn J. 2011. Transgene expression systems in the Triticeae cereals. *Journal of Plant Physiology* **168**, 30-44.

IWGSC. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788.

Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang X, Borrill P, Fosker C, Ayling S, Phillips AL, Uauy C, Dubcovsky J. 2017. Uncovering hidden variation in polyploid wheat. *Proc Natl Acad Sci U S A* **114**, E913-E921.

Lee W-S, Rudd JJ, Kanyuka K. 2015. Virus induced gene silencing (VIGS) for functional analysis of wheat genes involved in *Zymoseptoria tritici* susceptibility and resistance. *Fungal Genetics and Biology* **79**, 84-88.

Li J, Ye X, An B, Du L, Xu H. 2012. Genetic transformation of wheat: current status and future prospects. *Plant Biotechnology Reports* **6**, 183-193.

Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, Liu J, Zhang H, Liu C, Ran Y, Gao C. 2017. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications* **8**, 14261.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.

Ma J, Stiller J, Berkman PJ, Wei Y, Rogers J, Feuillet C, Dolezel J, Mayer KF, Eversole K, Zheng YL, Liu C. 2013. Sequence-based analysis of translocations and inversions in bread wheat (*Triticum aestivum* L.). *PLoS One* **8**, e79329.

Ma M, Yan Y, Huang L, Chen M, Zhao H. 2012. Virus-induced gene-silencing in wheat spikes and grains and its application in functional analysis of HMW-GS-encoding genes. *BMC Plant Biology* **12**, 141.

Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, Greenland AJ, Horsnell R, Howells R, O'Sullivan DM, Rose GA, Howell PJ. 2014. An Eight-Parent Multiparent Advanced Generation Inter-Cross Population for Winter-Sown Wheat: Creation, Properties, and Validation. *G3: Genes|Genomes|Genetics* **4**, 1603-1610.

Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Munoz-Amatriain M, Close TJ, Wise RP, Schulman AH, Himmelbach A, Mayer KF, Scholz U, Poland JA, Stein N, Waugh R. 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J* **76**, 718-727.

Matsuoka Y. 2011. Evolution of polyploid triticum wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiol* **52**, 750-764.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814.

Pearce S, Vazquez-Gross H, Herin SY, Hane D, Wang Y, Gu YQ, Dubcovsky J. 2015. WheatExp: an RNA-seq expression database for polyploid wheat. *BMC Plant Biol* **15**, 299.

Puchta H. 2017. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Current Opinion in Plant Biology* **36**, 1-8.

Ramirez-Gonzalez RH, Uauy C, Caccamo M. 2015. PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics* **31**, 2038-2039.

Richardson T, Thistleton J, Higgins TJ, Howitt C, Ayliffe M. 2014. Efficient *Agrobacterium* transformation of elite wheat germplasm without selection. *Plant Cell, Tissue and Organ Culture (PCTOC)* **119**, 647-659.

Shan Q, Wang Y, Li J, Gao C. 2014. Genome editing in rice and wheat using the CRISPR/Cas system. *Nature Protocols* **9**, 2395.

Simmonds J, Scott P, Brinton J, Mestre TC, Bush M, del Blanco A, Dubcovsky J, Uauy C. 2016. A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theoretical and Applied Genetics* **129**, 1099-1112.

Sparks CA, Doherty A, Jones HD. 2014. Genetic Transformation of Wheat via *Agrobacterium*-Mediated DNA Delivery. In: Henry RJ, Furtado A, eds. *Cereal Genomics: Methods and Protocols*. Totowa, NJ: Humana Press, 235-250.

Sparks CA, Jones HD. 2009. Biolistics Transformation of Wheat. In: Jones HD, Shewry PR, eds. *Transgenic Wheat, Barley and Oats: Production and Characterization Protocols*. Totowa, NJ: Humana Press, 71-92.

Tsai H, Missirian V, Ngo KJ, Tran RK, Chan SR, Sundaresan V, Comai L. 2013. Production of a High-Efficiency TILLING Population through Polyploidization. *Plant Physiology* **161**, 1604-1614.

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. 2006. A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. *Science* **314**, 1298-1301.

Uauy C, Wulff BBH, Dubcovsky J. 2017. Combining Traditional Mutagenesis with New High-Throughput Sequencing and Genome Editing to Reveal Hidden Variation in Polyploid Wheat. *Annu Rev Genet* **51**, 435-454.

Vasil V, Castillo AM, Fromm ME, Vasil IK. 1992. Herbicide Resistant Fertile Transgenic Wheat Plants Obtained by Microprojectile Bombardment of Regenerable Embryogenic Callus. *Bio/Technology* **10**, 667.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. *Ensembl*Compara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-335.

Wang K, Liu H, Du L, Ye X. 2017. Generation of marker-free transgenic hexaploid wheat via an Agrobacterium-mediated co-transformation strategy in commercial Chinese wheat varieties. *Plant Biotechnology Journal* **15**, 614-623.

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology* **32**, 947.

Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey M-D, Asyraf Md Hatta M, Hinchliffe A, Steed A, Reynolds D, Adamski NM, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan M, Edwards D, Batley J, Raman H, Carter J, Rogers C, Domoney C, Moore G, Harwood W, Nicholson P, Dieters MJ, DeLacy IH, Zhou J, Uauy C, Boden SA, Park RF, Wulff BBH, Hickey LT. 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* **4**, 23-29.

Wilkinson PA, Winfield MO, Barker GL, Tyrrell S, Bian X, Allen AM, Burridge A, Coghill JA, Waterfall C, Caccamo M, Davey RP, Edwards KJ. 2016. CerealsDB 3.0: expansion of resources and data integration. *BMC Bioinformatics* **17**, 256.

Wingen LU, Orford S, Goram R, Leverington-Waite M, Bilham L, Patsiou TS, Ambrose M, Dicks J, Griffiths S. 2014. Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat. *Theor Appl Genet* **127**, 1831-1842.

Wingen LU, West C, Leverington-Waite M, Collier S, Orford S, Goram R, Yang CY, King J, Allen AM, Burridge A, Edwards KJ, Griffiths S. 2017. Wheat Landrace Genome Diversity. *Genetics* **205**, 1657-1676.

Yu L-H, Miao Z-Q, Qi G-F, Wu J, Cai X-T, Mao J-L, Xiang C-B. 2014. MADS-Box Transcription Factor AGL21 Regulates Lateral Root Development and Responds to Multiple External and Physiological Signals. *Molecular Plant* **7**, 1653-1669.

Zhang J, Yu D, Zhang Y, Liu K, Xu K, Zhang F, Wang J, Tan G, Nie X, Ji Q, Zhao L, Li C. 2017a. Vacuum and Co-cultivation Agroinfiltration of (Germinated) Seeds Results in Tobacco Rattle Virus (TRV) Mediated Whole-Plant Virus-Induced Gene Silencing (VIGS) in Wheat and Maize. *Frontiers in Plant Science* **8**.

Zhang Y, Bai Y, Wu G, Zou S, Chen Y, Gao C, Tang D. 2017b. Simultaneous modification of three homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *The Plant Journal* **91**, 714-724.

Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, Chen K, Qiu J-L, Gao C. 2016. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature Communications* **7**, 12617.

Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL. 2017. The first near-complete assembly of the hexaploid bread wheat genome, Triticum aestivum. *Gigascience* **6**, 1-7.