A roadmap for gene functional characterisation in wheat

Nikolai M Adamski[1]*, Philippa Borrill[2]*, Jemima Brinton[1]*, Sophie A Harrington[1]*, Clemence Marchal[1]*, Alison R Bentley[3], William D Bovill[4], Luigi Cattivelli[5], James Cockram[3], Bruno Contreras-Moreira[6], Brett Ford[4], Sreya Ghosh[1], Wendy Harwood[1], Keywan Hassani-Pak[7], Sadiye Hayta[1], Lee T Hickey[8], Kostya Kanyuka[7], Julie King[9], Marco Maccaferri[10], Guy Naamati[6], Curtis J Pozniak[11], Ricardo H Ramirez-Gonzalez[1], Carolina Sansaloni[12], Ben Trevaskis[4], Luzie U. Wingen[1], Brande BH Wulff[1] and Cristobal Uauy[1]

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

[2] School of Biosciences, University of Birmingham, Birmingham B15 2TT, United Kingdom

[3] John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge CB3 0LE, United Kingdom

[4] Commonwealth Scientific and Industrial Research Organisation Agriculture and Food (CSIRO), GPO Box 1700, Canberra, ACT 2601, Australia

[5] Council for Agricultural Research and Economics, Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy

[6] The European Bioinformatics Institute, Hinxton CB10 1SD, United Kingdom

[7] Rothamsted Research, Harpenden AL5 2JQ, United Kingdom

[8] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia

[9] Division of Plant and Cop Sciences, The University of Nottingham, Sutton Bonington Campus, Loughborough, United Kingdom

[10] Department of Agricultural and Food Sciences, University of Bologna, 40127 Bologna, Italy

[11] Crop Development Centre, University of Saskatchewan, Saskatoon, SK S7N 5A8, Canada

[12] International Maize and Wheat Improvement Center (CIMMYT), El Batán, Mexico, 56237

* Authors contributed equally to this work

Nikolai M. Adamski: Nikolai.Adamski@jic.ac.uk

Philippa Borrill: p.borrill@bham.ac.uk

Jemima Brinton: Jemima.Brinton@jic.ac.uk

Sophie Harrington: Sophie.Harrington@jic.ac.uk

Clemence Marchal: Clemence.Marchal@jic.ac.uk

Alison R Bentley: Alison.Bentley@niab.com

William Bovill: Bill.Bovill@csiro.au

Luigi Cattivelli: luigi.cattivelli@crea.gov.it

James Cockram: James.Cockram@niab.com

Bruno Contreras-Moreira: bcontreras@ebi.ac.uk
Brett Ford: bxrett76@hotmail.com
Sreya Ghosh: Sreya.Ghosh@jic.ac.uk
Wendy Harwood: Wendy.Harwood@jic.ac.uk
Keywan Hassani-Pak: Keywan.Hassani-Pak@rothamsted.ac.uk
Sadiye Hayta: Sadiye.Hayta@jic.ac.uk
Lee Hickey: l.hickey@uq.edu.au
Kostya Kanyuka: Kostya.Kanyuka@rothamsted.ac.uk
Julie King: Julie.king@nottingham.ac.uk
Marco Maccaferri: marco.maccaferri@unibo.it
Guy Naamati: gnaamati@ebi.ac.uk
Curtis Pozniak: curtis.pozniak@usask.ca
Ricardo Ramirez-Gonzalez Ricardo.Ramirez-Gonzalez@jic.ac.uk
Carolina Sansaloni: C.Sansaloni@cgiar.org
Ben Trevaskis: Ben.Trevaskis@csiro.au
Luzie U. Wingen: Luzie.Wingen@jic.ac.uk
Brande Wulff: brande.wulff@jic.ac.uk
Cristobal Uauy: Cristobal.Uauy@jic.ac.uk (corresponding author)

Abstract

To adapt to the challenges of climate change and the growing world population, it is vital to increase global crop production. Understanding the function of genes within staple crops will accelerate crop improvement by allowing targeted breeding approaches. Despite the importance of wheat, which provides 20 % of the calories consumed by humankind, a lack of genomic information and resources has hindered the functional characterisation of genes in this species. The recent release of a high-quality reference sequence for wheat underpins a suite of genetic and genomic resources that support basic research and breeding. These include accurate gene model annotations, gene expression atlases and gene networks that provide background information about putative gene function. In parallel, sequenced mutation populations, improved transformation protocols and structured natural populations provide rapid methods to study gene function directly. We highlight a case study exemplifying how to integrate these resources to study gene function in wheat and thereby accelerate improvement in this important crop. We hope that this review provides a helpful guide for plant scientists, especially those expanding into wheat research for the first time, to capitalise on the discoveries made in *Arabidopsis* and other plants. This will accelerate the improvement of wheat, a complex polyploid crop, of vital importance for food and nutrition security.

Introduction

Research in *Arabidopsis* and other model species has uncovered mechanisms regulating important biological processes in plants. However, as research in these model species does not always translate directly into crop species such as wheat, understanding gene function in crop species themselves is critical for crop improvement. With the advent of functional genomics resources in wheat, discoveries from model species can rapidly be tested and functional genetic studies can now be performed for agronomically-important traits directly in wheat itself (Borrill, 2019).

The most common forms of domesticated wheat are tetraploid durum wheat (*Triticum turgidum* spp. *durum* L.) and hexaploid bread wheat (*Triticum aestivum L.*). Polyploid wheat is derived from hybridisation events between different ancestral progenitor species (reviewed in Matsuoka (2011)), and thus each gene typically exists as two (tetraploid durum wheat) or three (hexaploid bread wheat) copies. These closely related copies, known as homoeologous genes, are on average >95% similar across their coding regions (Figure 1) and usually have a highly conserved gene structure. Tetraploid and hexaploid wheat have large genomes, 12 and 16 Gb respectively, which consist mostly (>85%) of repetitive elements. The combination of these factors has, for a long time, hampered the development of genomics tools in wheat. Recent advances in sequencing technologies and bioinformatics tools has helped overcome these difficulties, and there are now a wide range of resources available for genomic analysis in wheat. The speed of wheat research has also been limited by its relatively long generation time, which ranges from four to six months depending on the requirement of cold periods (vernalisation) to induce flowering. Again, recent advances in the use of controlled growth conditions have radically changed these timeframes (Watson *et al.*, 2018). Wheat has now become a tractable system for translational, comparative and functional genomics.

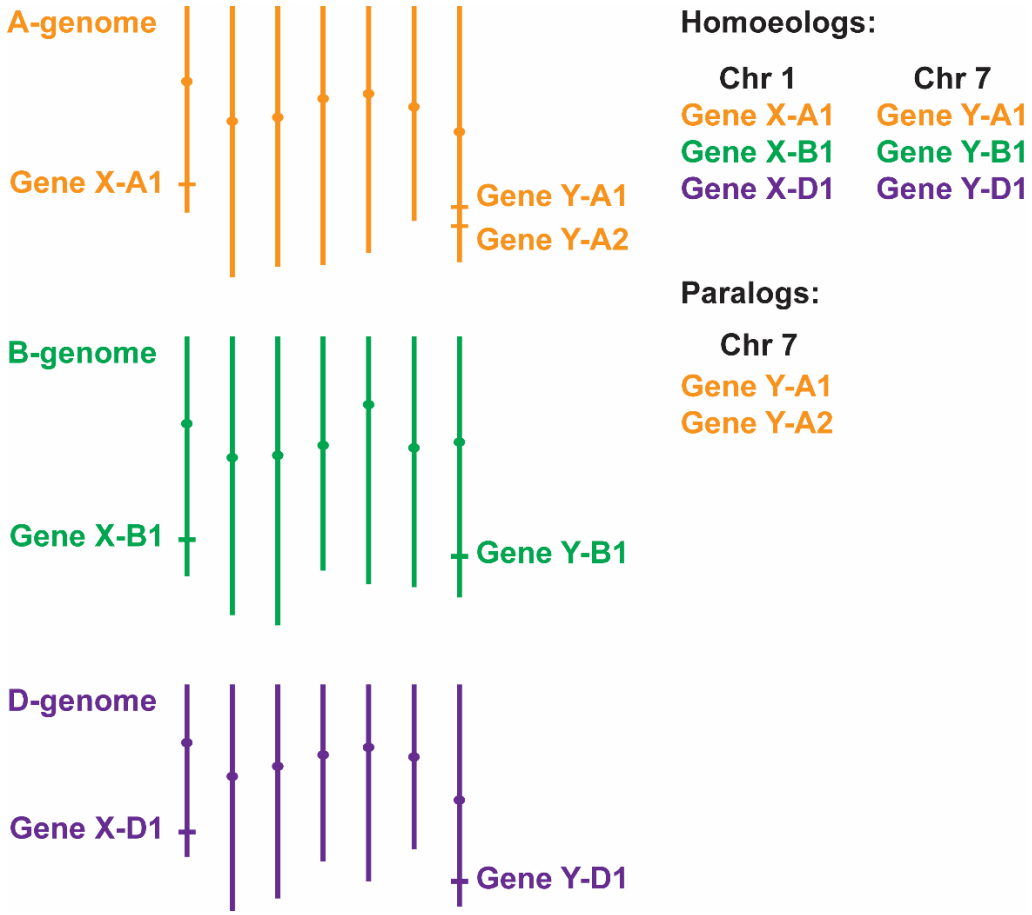

Figure 1: Gene homology within polyploid wheat. Due to two separate hybridisation events, genes in polyploid wheat will be present in multiple copies called homoeologs, which usually have similar chromosome locations. In the example of hexaploid bread wheat illustrated here, Gene X has homoeologs on chromosomes 1A, 1B and 1D. Duplicated genes, called paralogs (e.g. two copies of Gene Y on chromosome 7A), have evolved either within wheat or in one of its ancestral species. Most paralogs arise from intra-chromosomal duplications, although inter-chromosomal duplications can also occur.

Here we describe some of the recent developments in wheat genomics, focussing on published and publicly available resources and tools, and lay out a roadmap for their use (Figure 2). We present available wheat genome assemblies and annotations and discuss a series of approaches to functionally characterise genes. 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 highlights potential pitfalls. We expect this review will be a helpful guide for plant scientists who already work on wheat or who are considering expanding their research into wheat.

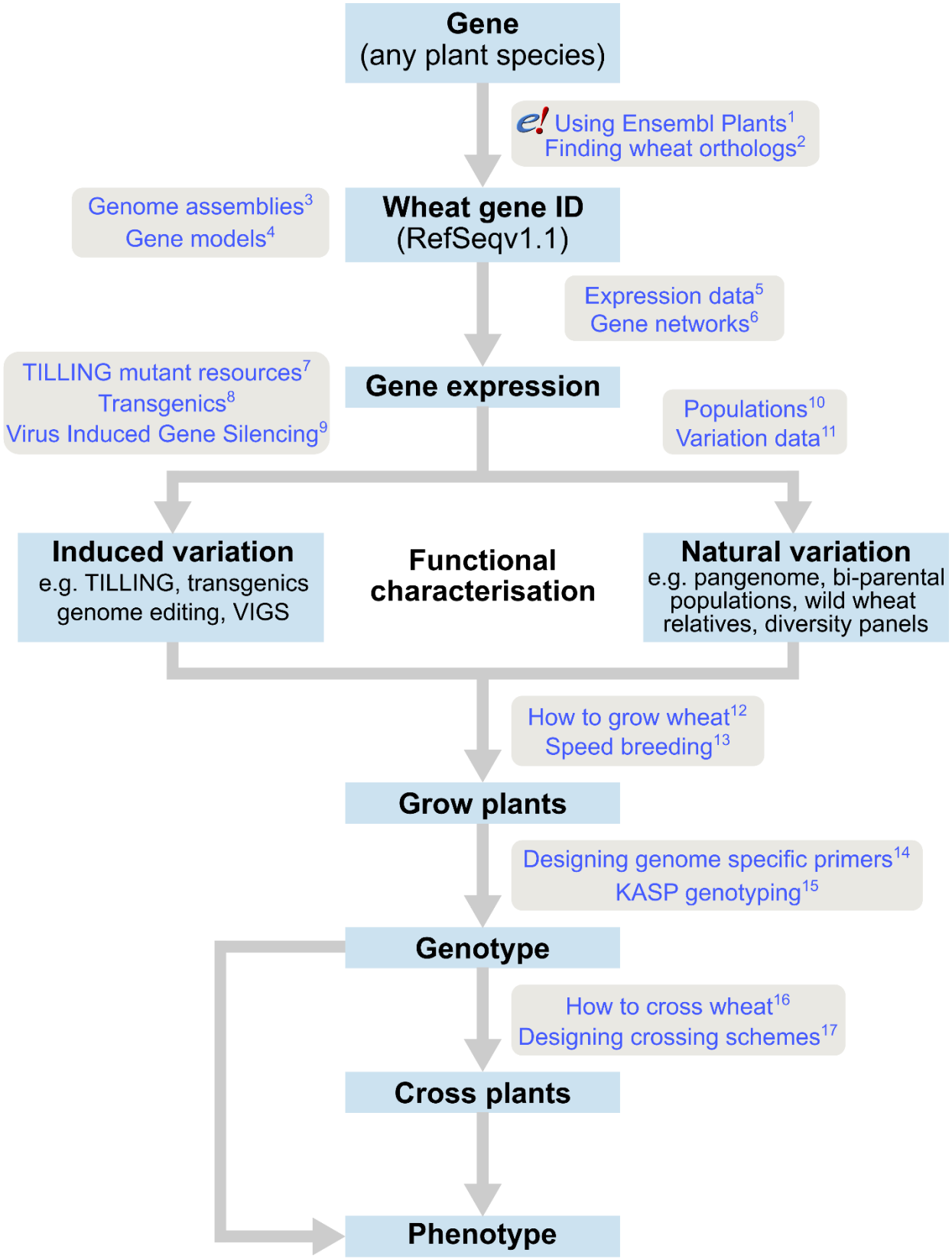
Gene
(any plant species)
Using Ensembl Plants[1]
Finding wheat orthologs[2]
Genome assemblies[3]
Gene models[4]
Wheat gene ID
(RefSeqv1.1)
Expression data[5]
Gene networks[6]
TILLING mutant resources[7]
Transgenics[8]
Virus Induced Gene Silencing[9]
Gene expression
Populations[10]
Variation data[11]
Induced variation
e.g. TILLING, transgenics
genome editing, VIGS
Functional
characterisation
Natural variation
e.g. pangenome, bi-parental
populations, wild wheat
relatives, diversity panels
How to grow wheat[12]
Speed breeding[13]
Grow plants
Designing genome specific primers[14]
KASP genotyping[15]
Genotype
How to cross wheat[16]
Designing crossing schemes[17]
Cross plants
Phenotype

Figure 2: The roadmap for gene characterisation in wheat. Overview of a proposed strategy to take a gene from any plant species, identify the correct wheat ortholog(s) using Ensembl Plants (https://plants.ensembl.org) and determine gene expression using expression browsers and gene networks. Suggestions for functional characterisation are provided including induced variation such as mutants, transgenics or Virus-Induced Gene Silencing (VIGs). In addition, publicly available populations incorporating natural variation are available. Finally steps for growing, genotyping and crossing plants are outlined. Links to detailed tutorials and further information are provided and can be found on www.wheat-training.com.

[1] www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/EnsemblPlants-primer.pdf
[2] www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Finding-wheat-orthologs.pdf
[3] www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Genome_assemblies.pdf
[4] www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Gene-models.pdf
[5] www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Expression-browsers.pdf
[6] www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Gene-networks.pdf
[7] www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Selecting-TILLING-mutants.pdf
[8] www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Transgenics.pdf
[9] www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Virus_Induced_Gene_Silencing.pdf
[10] www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Populations.pdf
[11] www.wheat-training.com/wp-content/uploads/Genomic_resources/Variation-data.pdf
[12] www.wheat-training.com/wp-content/uploads/Wheat_growth/pdfs/Growing_Wheat_final.pdf
[13] www.wheat-training.com/wp-content/uploads/Wheat_growth/pdfs/Speed_breeding.pdf
[14] www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Designing-genome-specific-primers.pdf
[15] https://www.biosearchtech.com/support/education/kasp-genotyping-reagents/running-kasp-genotyping-reactions
[16] http://www.wheat-training.com/wp-content/uploads/Wheat_growth/pdfs/How-to-cross-wheat-pdf.pdf
[17] www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Designing-crossing-schemes.pdf

Wheat genome assemblies

A high-quality genome reference sequence is an essential resource for functional genetics and genomics in any species. Several hexaploid wheat genome assemblies 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). The most comprehensive assembly, called RefSeqv1.0, is a chromosome-level genome assembly annotated with high and low confidence gene models (IWGSC, 2018). An improved RefSeqv2.0 assembly has been generated by incorporating optical mapping data and PacBio SMRT reads (from Zimin *et al.* (2017)), although it has yet to be annotated. Two tetraploid wheat genomes have also been sequenced, assembled, and annotated to the same standard as RefSeqv1.0 — the wild tetraploid progenitor of wheat, wild emmer (Avni *et al.*, 2017), and a modern durum wheat variety (Maccaferri *et al.*, 2019). Diploid ancestral progenitor species have also been assembled to varying levels of completeness (Luo *et al.*, 2017; Zhao *et al.*, 2017; Ling *et al.*, 2018; Miki *et al.*, 2019). We summarize the annotated assemblies for polyploid wheat in Table 1; in this review we will focus mainly on the RefSeqv1.0 assembly.

Table 1. Comparison of annotated genome assemblies in hexaploid and tetraploid wheat. RefSeqv1.0 is the most widely used assembly and annotation of hexaploid wheat (available on Ensembl Plants https://plants.ensembl.org/wheat). The information from previous assemblies and annotations (Chromosome Survey Sequence (CSS) and TGACv1) are also available in the Ensembl Plants archive (https://oct2017-plants.ensembl.org) or as tracks in the Ensembl Plants genome browser interface. Ensembl Plants enables access to additional information such as SNP variation, gene trees, homoeolog assignments, and TILLING mutant information. Through this interface users can also combine knowledge from the bread, durum and wild emmer genomes.

	CSS	TGACv1	RefSeqv1.0	Durum wheat	Wild emmer wheat
Publication	IWGSC (2014)	Clavijo *et al.* (2017)	IWGSC (2018)	Maccaferri *et al.* (2019)	Avni *et al.* (2017)
Contigs/Chromosomes	>1 million	735,943	21 chromosomes + ChrU	14 chromosomes + ChrU	14 chromosomes + ChrU
Mean scaffold size	7.7 kb	88.7 kb	Chromosomes	Chromosomes	Chromosomes
Assembly Size	10.2 Gb	13.4 Gb	14.6 Gb	10.5 Gb	10.5 Gb
Order	Crude order	Large Bins	Physical order	Physical order	Physical order
Coding genes†	133,090 HC 88,998 LC	104,091 HC 103,660 LC	107,891 HC 161,537 LC	66,559 HC 303,404 LC	67,182 HC 271,179 LC
Assembly-related resources	Archive Ensembl Plants	Archive Ensembl Plants	Ensembl Plants GrainGenes, URGI	Ensembl Plants GrainGenes	Ensembl Plants GrainGenes
	TILLING mutants		TILLING mutants		
	expVIP, wheatExp	expVIP	expVIP, eFP		
Cultivar	Chinese Spring	Chinese Spring	Chinese Spring	Svevo	Zavitan

† Number of high confidence (HC) and low confidence (LC) genes which are defined based on multiple criteria outlined in the published papers. Care must be taken when interpreting their nomenclature (see Figure 3).

Note: RefSeqv2.0 was released in August 2019 and is available here: https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies. Briefly, optical mapping and PacBio sequencing data were integrated to further improve contiguity, although this assembly is not yet annotated.

Like most of the previous hexaploid assemblies, RefSeqv1.0 is derived from the wheat landrace 'Chinese Spring'. A combination of multiple Illumina and mate pair libraries were sequenced and assembled into scaffolds. Using a method of chromosome conformation capture called Hi-C, these scaffolds were further connected into pseudomolecules representing the 21 nuclear chromosomes of wheat, plus one additional 'pseudo-chromosome' containing all unassigned sequences (IWGSC, 2018).

The gene models for the RefSeqv1.0 assembly were annotated using two prediction pipelines, which were then consolidated with the previous TGACv1 annotation into a single set of gene models (RefSeqv1.0 models). A subset of these (~2,000 gene models) were later re-annotated manually, resulting in the RefSeqv1.1 gene model set (Figure 3). Over half of high confidence protein coding genes are present as exactly three homoeologous copies (1:1:1 triads), while several other combinations exist (e.g. 2:1:1 whereby there are two paralogs on the A genome, and a single homoeolog each on the B and D genomes as Gene Y in Figure 1).

The RefSeqv1.0 assembly and the RefSeqv1.1 gene models, as well as the durum and wild emmer assemblies and gene models, have been integrated into the publicly available Ensembl Plants genome browser (https://plants.ensembl.org) (Bolser *et al.*, 2015; Howe *et al.*, 2019). Existing variation data, both natural and induced, has been mapped to the RefSeqv1.0 hexaploid assembly and deposited in Ensembl Plants databases for visualisation via the genome browser. Integrating resources into a common reference facilitates their use and in the following sections we will discuss how to best access and utilise these resources.

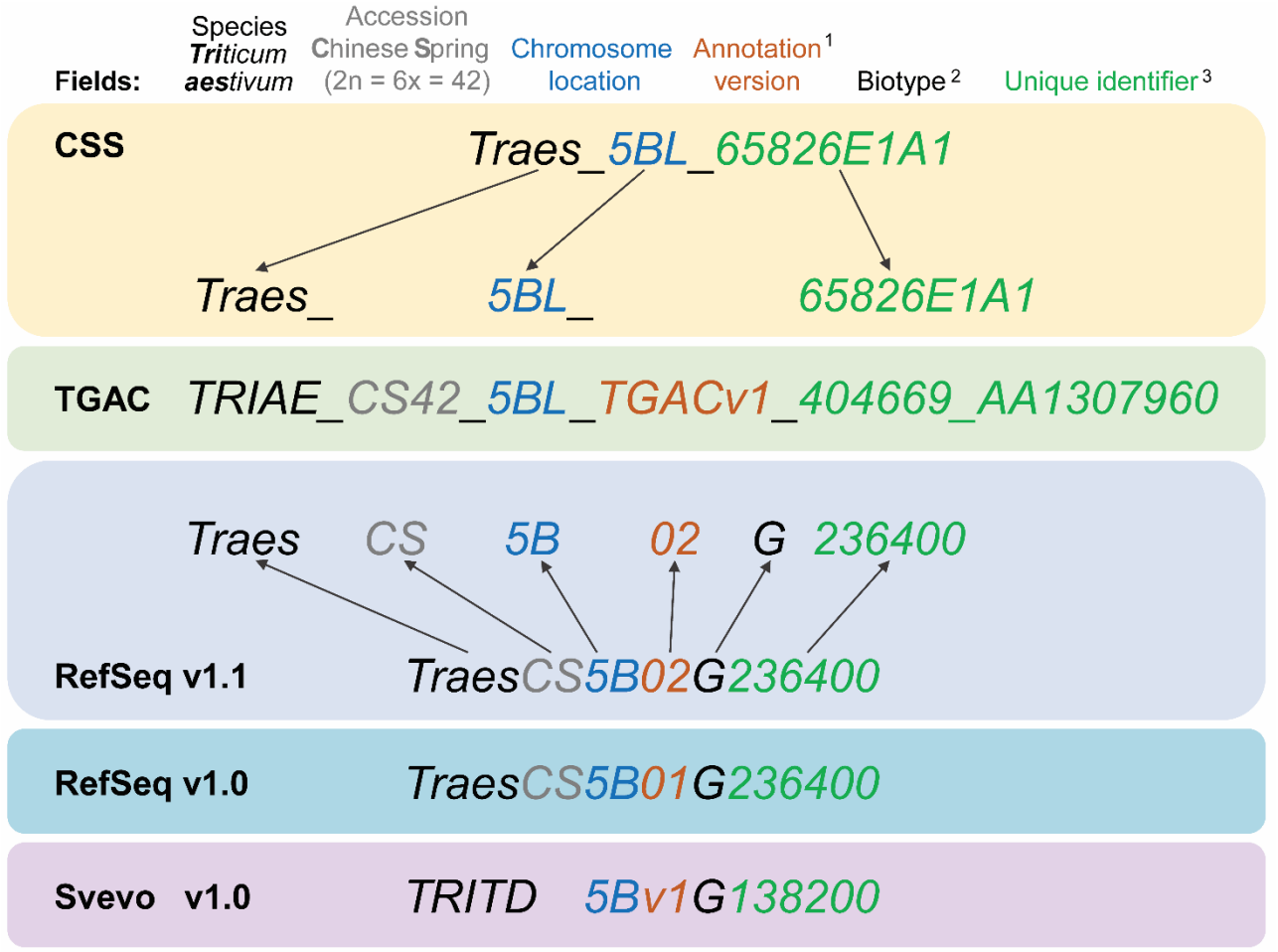

Figure 3. Gene model ID nomenclature description from the five available gene annotations for domesticated polyploid wheat. Here, one gene is used as an example to highlight the differences in gene ID nomenclature. Fields represented in the nomenclature are shown at the top with matching colours for the corresponding features in the gene names. Yellow background shows the CSS gene names with dark grey arrows pointing towards the corresponding field in the TGAC gene annotation (TGACv1, green background). Blue backgrounds show the gene nomenclatures for RefSeqv1.0 and v1.1 annotations (as used in Ensembl Plants), while the lilac background shows the nomenclature for Svevo v1.0 (modern durum wheat).

[1] Two annotation versions are available for the RefSeqv1.0 genome assembly: RefSeqv1.0 (release annotation) and RefSeqv1.1 (improved annotation). These can be differentiated by the annotation version number i; "01" for RefSeqv1.0 and "02" for RefSeqv1.1. Otherwise, the annotations follow the same rules.
[2] In the RefSeq and Svevo annotations, the biotype is represented by an additional identifier, where G = gene.
[3] In the RefSeqv1.0 and v1.1 annotation, identifiers are progressive numbers in steps of 100s reflecting the relative position between gene models. For example, gene *TraesCS5B02G236400* would be adjacent to gene *TraesCS5B02G236500*. In the gene annotation for the tetraploid durum wheat cv. Svevo, the species name is TRITD (*TRITicum Durum*) and gene identifiers increase in steps of 10s, rather than by steps of 100s as in the RefSeq hexaploid wheat annotation.
Note that RefSeqv1.0 and v1.1 comprises High Confidence (HC) and Low Confidence (LC) gene models. Low Confidence gene models are flagged by the "LC" at the end (not shown). HC and LC genes which otherwise display the same unique identifier are **not** the same locus and are not in sequential order. Hence, *TraesCS5B02G236400* and *TraesCS5B02G236400**LC*** are both located on chromosome 5B, but are not the same gene nor are they

physically adjacent. *Similarly, genes from homoeologous chromosomes with the same subsequent numeric identifier are not necessarily homoeologous genes. For example, TraesCS5A02G236400, TraesCS5B02G236400 and TraesCS5D02G236400 are* ***not*** *homoeologous genes*.

Finding wheat orthologs

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

Conveniently, all the data and tools necessary for identifying putative gene orthologs from different plant species are available through the Ensembl Plants website (https://plants.ensembl.org) (Bolser *et al.*, 2015; Howe *et al.*, 2019). The Plant Compara pipeline has been integrated into Ensembl Plants to create "gene trees" that identify and clearly display the likely orthologs of any given gene for all of the species available on its website (Vilella *et al.*, 2009; Herrero *et al.*, 2016). This includes the RefSeqv1.1, *Arabidopsis* TAIR10 and rice (*Oryza sativa*) IGRSP1.0 gene models, amongst others. This represents a quick and reliable way to identify putative wheat orthologs of a given gene (Figure 2). Tutorials for using Ensembl Plants interactively or programmatically can be found on their website or at www.wheat-training.com.

When performing a search for putative wheat orthologs via the Ensembl Plants pipeline, we would expect to find three orthologs in hexaploid wheat for most gene queries. These orthologs would normally be located on homoeologous chromosome groups, e.g. chromosomes 1A, 1B and 1D (Figure 1). 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. three homoeologous genes could be on chromosome arms 4AL, 5BL and 5DL. Furthermore, gene structure of wheat orthologs is often conserved with respect to rice and other closely related monocot species; this comparison can usually be done within Ensembl Plants. If this is not possible, wheat RNA-seq data can be used to determine the gene structure. As an alternative to the Ensembl Plants Gene Trees, one can perform reciprocal protein BLAST searches to identify putative wheat orthologs. We exemplify the above-mentioned approaches along with potential pitfalls in more detail in the 'Case Study' section.

Expression data

Determining if, when, where, and to what level a gene is expressed often constitutes one of the first steps towards its functional characterisation. Gene expression information can also be used to prioritize candidate genes underlying a quantitative trait locus (QTL) or to predict those members of a large gene family most relevant to trait expression. Numerous wheat RNA-Seq datasets have been generated and published. Although the raw data are often publicly available (e.g. via the NCBI sequence read archive, https://www.ncbi.nlm.nih.gov/sra), they are not sufficiently curated for rapid access and their use in direct comparisons is complicated due to the diversity of tissues, treatments, and origins of the samples. Expression browsers aim to centralise these public datasets and analyse them together, ideally allowing retrieval of expression information for a list of genes under

different conditions. Four expression browsers are currently available for wheat: expVIP (http://www.wheat-expression.com; (Borrill *et al.*, 2016)), wheat eFP browser (http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi; (Ramirez-Gonzalez *et al.*, 2018)), EBI Gene Expression Atlas (https://www.ebi.ac.uk/gxa/experiments?species=triticum+aestivum), and WheatExp (https://wheat.pw.usda.gov/WheatExp; (Pearce *et al.*, 2015)). Here we will focus on the first two given that they include a larger and more diverse set of samples and use the RefSeqv1.0 and v1.1 gene models described in Table 1.

Currently, expVIP includes expression data from 36 studies (1,016 RNA-Seq samples) across a diverse range of wheat tissues, developmental stages, cultivars, and environmental conditions including various abiotic and biotic stress treatments. It can display expression data for up to 250 genes at once, which can be particularly useful when working with a gene family, genes within a QTL interval, or genes involved in the same regulatory process. The expression values for each gene homoeolog, based on the same homoeolog assignments as in Ensembl Plants, can also be displayed. The 'homoeolog expression patterns' of triads (genes that are present as exactly three homoeologous copies) can also be displayed through ternary plots and compared across tissues (Ramirez-Gonzalez *et al.*, 2018).

To allow comparisons across studies, the 1,016 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 customize visualization displays and allows users to select data relevant to their experimental comparisons. 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 analyses. Although the default gene model reference is RefSeqv1.1, users can also choose the CSS, TGACv1 and RefSeqv1.0 transcriptome references for legacy reasons. Video and text tutorials describing expVIP are available on https://github.com/Uauy-Lab/expvip-web/wiki/List-of-tutorial-videos and www.wheat-training.com.

An additional resource is the electronic Fluorescent Pictograph (eFP) wheat browser which provides a simple visual assessment of expression data using pictures of wheat coloured according to a gene's relative expression level (http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi). This browser includes 209 RNA-Seq samples (also in expVIP) representing 22 tissue types from grain, root, leaf, and spike samples across multiple time points from a single hexaploid spring wheat cultivar ('Azhurnaya').

Gene networks

The available RNA-Seq data provides the opportunity to identify networks of co-expressed genes. Ramirez-Gonzalez *et al.* (2018) constructed tissue and stress-specific co-expression networks in wheat to determine whether genes from the same triad showed variable spatiotemporal expression. In addition, a GENIE3 network was developed to predict transcription factor targets across the multiple RNA-Seq samples (Huynh-Thu *et al.*, 2010; Ramirez-Gonzalez *et al.*, 2018). Together, these networks provide a powerful set of tools for hypothesis generation using wheat-specific datasets. We have recently validated the GENIE3 network using independent RNA-Seq data from tetraploid wheat (Harrington *et al.*, 2019). Both co-expression and GENIE3 networks are incorporated into KnetMiner (http://knetminer.rothamsted.ac.uk/Triticum_aestivum/).

KnetMiner is a web-application for searching and visualising genome-scale knowledge networks (Hassani-Pak *et al.*, 2016). It aims to provide research leads for scientists who are investigating the molecular basis of complex traits. KnetMiner accepts keywords in combination with a gene list and/or genomic regions as input. KnetMiner

searches the underlying knowledge network to identify links between user-provided genes and keywords. A network-based visualisation, named Network View, allows users to examine complex relationships between gene networks and traits. The networks contain nodes that represent different entities such as genes, single nucleotide polymorphisms (SNPs), publications, and traits (e.g. heat or drought tolerance) that are linked via different relation types (e.g. co-expression, GENIE3-targets, protein-protein interaction, published-in). Together, KnetMiner and the integrated gene networks provide a powerful resource for gene discovery and hypothesis generation in wheat (see Case Study below).

Functional studies

After identifying a gene of interest in wheat there are now several options 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 and paralogs simultaneously to measure a strong phenotypic effect (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 their genomes 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 (Krasileva *et al.*, 2017). This resource consists of two ethyl methanesulphonate (EMS) mutagenized populations: 1,535 lines of the tetraploid durum wheat variety 'Kronos' and 1,200 lines of the hexaploid bread wheat variety 'Cadenza'. Exome capture and Illumina sequencing of these 2,735 mutant lines was then carried out. The raw data was originally aligned to the CSS reference, mutations were identified, and their effects predicted based on the CSS gene models (Krasileva *et al.*, 2017). Alleles predicted *in silico* to be deleterious (e.g. premature stop codons, splice site mutations, non-synonymous amino acid substitutions with SIFT score < 0.05), were identified 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). The raw data has now been aligned to the RefSeqv1.0 genome, allowing mutation identification and effect prediction based on the RefSeqv1.1 gene models. These updated data are publicly available on Ensembl Plants (see Case Study for details). For legacy purposes, the mutations called against the CSS reference remain available via www.wheat-tilling.com. However, caution should be exercised as the mutation effects here are predicted based on the CSS gene models, which are known to be less reliable than the RefSeq gene models (Brinton *et al.*, 2018).

There are several important considerations when selecting a mutant line for characterisation. First, it is essential to check the predicted effect of mutations in the context of a complete and experimentally validated gene model. Second, in most cases, crossing is necessary to combine mutations in homoeologous genes in order to generate a complete null individual. Third, mutant lines will contain a high level of background mutations: a typical mutant line has between 50 (tetraploid) and 110 (hexaploid) mutations predicted to result in a truncated protein. Depending on the phenotype of interest (i.e. qualitative vs. quantitative) several rounds of backcrossing may be required before the phenotype can be assessed (see 'Strategies for Use'). Lastly, if the gene of interest is missing or is already a null allele in Kronos or Cadenza, which can be determined using the full genome sequences of the two cultivars, mutant populations of other genotypes are available (e.g. Dong *et al.* (2009); Chen *et al.* (2012);

Bovina *et al.* (2014); Sestili *et al.* (2015); Colasuonno *et al.* (2016)), although these would need to be screened using conventional PCR-based approaches. Additional practical information about selecting mutant lines and downstream analyses can be found at www.wheat-training.com/tilling-mutant-resources and in Uauy *et al.* (2017).

Transgenic approaches

Stable transformation of wheat is possible and 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). Generating stable transgenic lines in wheat most commonly involves transforming immature wheat embryos and subsequent callus regeneration (Harwood, 2012). Reports in the literature of *Agrobacterium*-mediated wheat transformation generally describe low transformation efficiencies with average efficiencies of around 5%. An efficient, but patented transformation system is available through licence from Japan Tobacco (www.jti.co.jp). Transformation by overexpression of transcription factors such as maize *Baby Boom* and *Wuschel2* has also yielded improved transformation efficiencies in monocots (Lowe *et al.*, 2016), although there are no formal reports yet in wheat. Recently, an open-access wheat transformation system with transformation efficiencies of up to 25% was published (Hayta *et al.*, 2019), albeit for a single cultivar.

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) for downstream experiments such as ChIP-seq (Deng *et al*., 2015) or localisation studies (Harwood *et al*., 2005). However, these are still not commonly employed in wheat research. As transformation methods have only been optimised for a limited number of wheat varieties (e.g. Richardson *et al.* (2014)), it is important to understand whether the gene is expressed/functional in the chosen variety when defining transgenic strategies (see 'Strategies for Use').

Recent developments in genome editing technologies provide new opportunities for manipulating genes in wheat. TALEN and CRISPR/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.*, 2014b; Luo *et al.*, 2019), 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 for complex applications such as allele swapping or gene insertion, as reviewed by Puchta (2017). Similar to RNAi, constructs for Cas9-mediated gene editing can be designed to target all homoeologs simultaneously (Zhang *et al.*, 2016; Howells *et al.*, 2018). 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 (0.9%; (Zhang *et al.*, 2016) and subsequent crosses to combine multiple edited targets are likely to be required.

A major limitation of using transgenic approaches to manipulate agronomically relevant traits is the associated legal and regulatory constraints. To overcome these, the nuclease transgene can be segregated away from the edited gene(s) in subsequent generations. However, in Europe, and in contrast to many other countries in the world, the resulting plants would be regulated as transgenics due to the 2018 ruling on genome editing by the European Court of Justice (ECJ). Some studies have documented CRISPR/Cas9-editing in wheat without transgene integration, for example, by delivering the CRISPR/Cas9 components as ribonucleoproteins (RNPs). As no foreign

DNA is used in CRISPR/Cas9 RNP-mediated genome editing, the wheat mutants obtained are completely transgene free (Liang *et al.*, 2017), although still not exempt from the ECJ regulation.

Virus Induced Gene Silencing

Virus-Induced Gene Silencing (VIGS) involves transient knock-down of expression of target genes followed by assessment of the resulting phenotype (Lee *et al.*, 2012). The most widely used vectors for VIGS in wheat are those derived from barley stripe mosaic virus (BSMV), a plant virus with a tripartite RNA genome that readily spreads throughout tissues following mechanical rub-inoculation onto the leaves. All three BSMV genomic RNAs, RNAα, RNAβ and RNAγ, are required to cause infection. RNAγ has been modified to allow insertion of short (up to 350 bp) plant mRNA derived sequences. Infection of plants with the resulting recombinant virus induces a natural post-transcriptional gene silencing defence mechanism that targets the viral RNA, but also the endogenous plant mRNA having high level (>70%) nucleotide identity with the plant sequence inserted into RNAγ, for degradation. A detailed protocol for VIGS is available at www.wheat-training.com (Figure 2).

VIGS in wheat has been used primarily to investigate disease resistance in a range of varieties, and has been restricted to a few tissue types such as leaf (Lee *et al.*, 2015), young seedlings (Zhang *et al.*, 2017a) and spikes (Ma *et al.*, 2012). However, in principle, BSMV-mediated VIGS can be applied to any wheat genotype and to almost any gene of interest. This functional genomics tool is particularly useful when analysing multiple candidate genes, for example in map-based cloning projects (i.e. when physical intervals contain several candidate genes) or from RNA-Seq differentially expressed datasets. VIGS is also useful in wheat genotypes that are difficult to transform and in those for which mutant/TILLING populations are unavailable. VIGS can be used for simultaneous silencing of all homoeologs or, in principle, entire small gene families without the need for further genetic crosses.

Natural Variation

Although using induced variation presents a clear route to understand the function of specific genes in wheat, the wealth of natural variation in wheat lines, and populations based on this variation, present an alternative route to discover genes and correlate them with function. For example, populations differing for alleles of the gene of interest could be used to rapidly infer the role of the gene. In order to capture the diversity within wheat and create populations to test gene function, natural variation has been extensively documented. Most studies have focused on SNPs between varieties that can be quickly assayed through SNP arrays designed from gene coding sequences and untranslated regions (UTRs) (Wang *et al.*, 2014a; Winfield *et al.*, 2016; Allen *et al.*, 2017), 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 *in silico* allelic series are readily available for many genes of interest.

Beyond SNP variation, two recent studies (He *et al.*, 2019; Pont *et al.*, 2019) applied exome capture to diverse wheat lines to characterise the natural variation throughout the coding region of wheat. These studies identified millions of SNPs within coding sequences in over 1,000 wheat lines, including hexaploid cultivars and landraces, and tetraploid and diploid relatives. The data (available at http://wheatgenomics.plantpath.ksu.edu/1000EC and https://urgi.versailles.inra.fr) will allow rapid characterisation of the extent of variation within genes of interest. These changes in coding sequences may have direct phenotypic consequences, however the impact of most of these variants remains unknown.

Therefore, despite this wealth of data, the challenge remains 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 and include links to access genotypes, sequences and seeds in Table 2. Further details are available at www.wheat-training.com.

Wild wheat relatives and progenitor species:

Relatively low genetic variation in elite bread wheat varieties, especially on the D genome, typically reflects (i) adaptation and selection from landraces over a long time period, combined with (ii) the genetic bottleneck effects associated with the rare natural hybridisation events between the diploid and tetraploid ancestral wheat species that lead to the evolution of hexaploid wheat. Wheat is related to several other grass species, many of which are wild and uncultivated. These wild relatives provide a vast and largely untapped reservoir of genetic variation for many agronomically important traits. A wealth of cytogenetic stocks for these wild relatives have been created over the last 100 years by researchers globally (reviewed by Mujeeb-Kazi *et al.* (2013)). The recent genotyping and sequencing of some of these resources makes them especially suitable for gene functional characterisation (Table 2).

Synthetic hexaploid wheat:

Another approach to capture variation in wheat progenitors is via 're-synthesis', the process used to create synthetic hexaploid wheat (SHW). SHWs are typically created by crossing tetraploid durum wheat with the diploid D-genome progenitor *Aegilops tauschii.* Approximately 400 SHWs were developed at CIMMYT in Mexico during the 1990s (Mujeeb-Kazi *et al.*, 1996) and these have been extensively utilised in CIMMYT and international wheat breeding programmes (e.g. Gororo *et al.* (2002); Ogbonnaya *et al.* (2007)). More recently, NIAB (UK) have developed a new SHW resource encompassing 50 SHWs along with pre-breeding derivatives. This germplasm, alongside marker data, is publicly available (Table 2).

Wheat diversity panels:

Numerous collections of wheat landraces, varieties and breeders' lines are available from research centres around the world. These panels represent valuable sources of potential genetic variation for targeted exploitation within wheat research and pre-breeding pipelines, especially when associated with existing genotypic and phenotypic datasets (Table 2). Further details are available at www.wheat-training.com.

Multiparent Advanced Generation Inter-Cross (MAGIC) populations:

MAGIC populations have been developed for many crop species (Huang *et al.*, 2015; Cockram and Mackay, 2018). The multiple generations of inter-crossing required to create MAGIC populations results in highly recombined chromosomes which enables the use of approaches such as genome wide association scans (GWAS) and whole-genome average interval mapping (WGAIM; (Verbyla *et al.*, 2007)) to define small genetic intervals for traits of interest (reviewed by Verbyla *et al.* (2014)). Likewise, the use of multiple parents in MAGIC allows more allelic variation to be examined compared to typical bi-parental populations (Cockram and Mackay, 2018). In wheat, six MAGIC populations are currently publicly available constructed from 4, 8 or 16 founders. Parent information and further details can be found in Table 2.

Combining induced and natural variation for a holistic picture of gene function

To date natural variation has largely been used for forward genetics, to map genetic regions underlying a phenotypic trait of interest. However, there is now an opportunity to extend the use of natural variation in wheat into reverse genetics, to complement transgenic, gene editing and induced variation approaches. Using natural populations that differ in a target gene would allow characterisation of the effect sizes of natural alleles and could be compared to the effects of induced variation such as TILLING mutants. There are also synergies between forward and genetic approaches in wheat, for example the development of TILLING mutants in a gene of interest may then coincide with a region identified by a QTL mapping approach and so help to validate the QTL. Researchers now have at their disposal a powerful toolkit to combine induced and natural variation to study gene function in wheat.

Moving towards a wheat pangenome

Increases in DNA sequencing outputs have allowed the assembly of multiple wheat cultivars to a similar standard as the reference Chinese Spring genome. These include eight spring and eight winter hexaploid and three tetraploid varieties/accessions (Table 3). Annotation of some of these varieties is ongoing through the 10+ Wheat Genomes Project (http://www.10wheatgenomes.com) and will provide information on the core (genes shared by all assembled varieties) and dispensable genes (genes shared among a few varieties). In addition, presence absence variation, copy number variation, structural rearrangements (inversions/translocations), and variation across non-coding regions are being quantified. Importantly, several of these genotypes are part of the resources outlined above, e.g. sequenced TILLING population (Kronos and Cadenza). These assemblies will be integrated into Ensembl Plants and are available for download under Toronto Agreement (https://wheat.ipk-gatersleben.de/).

Table 3: Tetraploid and hexaploid wheat genome assemblies that are currently available, in addition to the Chinese Spring reference hexaploid genome.

Variety	Habit	Origin	Availability *
Hexaploid wheat			
CDC Landmark	spring	Canada	10+ Genome Project
CDC Stanley	spring	Canada	10+ Genome Project
Paragon	spring	UK	10+ Genome Project
Cadenza	spring	UK	10+ Genome Project
Lancer	spring	Australia	10+ Genome Project
Mace	spring	Australia	10+ Genome Project
Synthetic W7984	spring	Mexico	Chapman *et al.* (2015)
Weebil	spring	Mexico	10+ Genome Project
Arina*LrFor*	winter	Switzerland	10+ Genome Project
Julius	winter	Germany	10+ Genome Project
Jagger	winter	US	10+ Genome Project
Robigus	winter	UK	10+ Genome Project
Claire	winter	UK	10+ Genome Project
Norin61	winter	Japan	10+ Genome Project
SY Mattis	winter	France	10+ Genome Project
Spelt (PI190962)	winter	Europe	10+ Genome Project
Tetraploid wheat			
Zavitan†	-	Israel	Avni *et al.* (2017)
Svevo	spring	Italy	Maccaferri *et al.* (2019)
Kronos	spring	US	10+ Genome Project

† 'Zavitan' is a tetraploid wild emmer (*T. dicoccoides*) accession.

* Varieties included within the 10+ Wheat Genomes Project can be accessed through the Earlham Grassroot Genomics portal (https://wheatis.tgac.ac.uk/grassroots-portal/blast) and the 10+ Wheat Genomes 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 and Ensembl Plants. 'Synthetic W7984' and 'Zavitan' can be accessed through the Grassroot Genomics, and Ensembl Plants, 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 (e.g. daylength and vernalisation requirements) making it important to consider the conditions under which functional validation will be conducted. If phenotyping will be undertaken in greenhouse or controlled environment 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). If field trials are required for phenotypic characterisation (e.g. yield-related traits), local adaptation is often necessary for correct interpretation of results given genotype x environment interactions. For example, the sequenced TILLING populations (Kronos and Cadenza) do not require vernalisation, facilitating greenhouse experiments, and originate from different regions of the world, allowing field trials under different environments (Kronos is a Californian variety adapted to warm dry weather whereas Cadenza is a UK variety adapted to cooler conditions).

For CRISPR/Cas9 and other non-transient transgenic approaches several varieties may be used, although relatively few wheat varieties have been shown to display high enough transformation efficiencies to be practical. This means that traditionally most transgenic studies in wheat have been 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; Hayta *et al.*, 2019). This is now changing thanks to work by groups at NIAB (UK), CAAS (China) and CSIRO (Australia) who have successfully transformed 39 (Wallington, 2015), 15 (Wang *et al.*, 2017) and six (Richardson *et al.*, 2014) varieties, respectively. However, the *Agrobacterium*-mediated transformation efficiencies in all these studies still differ between varieties. Correct varietal selection for transformation is critical for functional studies, 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 to test a candidate resistance gene). Similarly, it is important to assess whether the gene of interest is present/functional in the chosen variety, for example through PCR amplification and sequencing of the gene. For several varieties this can now be done quickly by direct examination of their genome sequence (Table 3).

Combining mutations for complete knock-outs in polyploid wheat

As we noted earlier, the polyploid nature of wheat means that it normally has multiple homoeologous copies of every gene. These copies typically have highly similar coding DNA sequence 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 e.g. (Uauy *et al.*, 2006) or CRISPR/Cas9 e.g. (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; Howells *et al.*, 2018). If the targets are more divergent it may not even be possible to use a single guide RNA to target all three homoeologs, in which case several guides may be used through multiplexing. Alternatively, separate knock-outs for each homoeolog can be generated by CRISPR/Cas9 or identified in TILLING 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 each of the three homoeologs 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 initiated in the F_2 generation by comparing homozygous double knock-out mutants to the sibling wild type plants. It is important to note that TILLING lines contain many background mutations and

backcrossing may be required to overcome the confounding effects of background mutations on target phenotype. More details on these strategies are published in (Uauy *et al.*, 2017).

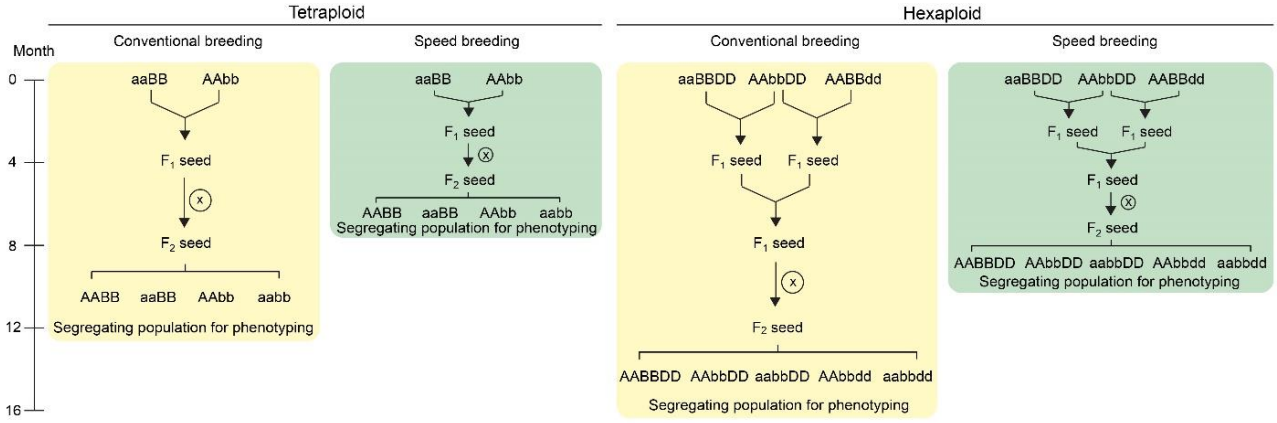

Figure 4. Crossing scheme to combine TILLING or CRISPR/Cas9 single mutants in wheat. In tetraploid wheat, mutations in the A and B genome homoeologs can be combined through a single cross. The F_1 plants are self-pollinated to produce a segregating F_2 population which contains homozygous double and single mutants, as well as wild type plants (screening using molecular markers required; only four genotypes shown). These F_2 progeny can be characterised for the phenotype of interest. The use of 'speed breeding' (Watson *et al.*, 2018), reduces the time taken to reach this phenotyping stage from 12 (yellow) to 7.5 months (green). In hexaploid wheat, a second round of crossing is required to combine the mutant alleles from all three homoeologs. The F_2 progeny segregating for the three mutant alleles can be genotyped using molecular markers to select the required combination of mutant alleles (only five 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. Combinations of wild type alleles from the A (AA), B (BB) and D (DD) genomes, as well as the mutant alleles from each genome (aa, bb and dd, respectively) are indicated.

Accelerating crossing, generation time, 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 in a spring wheat genetic background. 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 (Ghosh *et al.*, 2018; 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 the crossing schemes described above, it is essential to be able 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 genetic markers. Primers can be designed to include a homoeolog-specific SNP at the 3' end of the primer. The primer will amplify the targeted homoeolog more efficiently than the non-targeted homoeolog(s) resulting in genome-specific amplification. Rapid design of homoeolog-specific primers can be achieved using the PolyMarker pipeline (Ramirez-Gonzalez *et al.*, 2015) and webserver (http://www.polymarker.info/). Routinely, genotyping of SNPs is carried out using Kompetitive Allele Specific PCR (KASP) markers which are relatively 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 two alternative allele-specific primers used in the KASP 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 common primer. Amplification should thus be both homoeolog-specific and allele-specific. Further guidance on the design of genome-specific primers and KASP markers is available at 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 heat shock factor-like transcription factor *TBF1*, also known as *HsfB1*, is a critical regulator of the plant growth-to-defence transition (Pajerowska-Mukhtar *et al.*, 2012), and the response to heat stress (Guo *et al.*, 2016). We therefore hypothesize that its wheat orthologs may have a similar role in regulating defence and/or abiotic stress responses (Ikeda *et al.*, 2011; Arora *et al.*, 2019). The first step to test this hypothesis is to identify wheat *TBF1* orthologs, which can be done using the Ensembl Plants Gene Tree (Bolser *et al.*, 2015), which displays predicted orthologs for all species included in Ensembl Plants. *TBF1* is one of five *HSFB* orthologs, named *HSFB1, 2A, 2B, 4,* and *5*, respectively. Examination of the Ensembl Plants Gene Tree shows a single wheat triad that falls within the *HSFB1* clade, located on the group 5 chromosomes (Figure 5A).

To support the predicted *Arabidopsis*-wheat orthologs obtained from Ensembl Plants, we recommend carrying out comparisons between wheat and rice to establish orthology between these cereal species. Both the wheat homoeologs and the rice gene model *Os09g0456800* have the same gene structure, consisting of two exons with a conserved intron/exon boundary position. To further support the relationship of the rice gene to the wheat homoeologs, the predicted rice protein can be used as a query for BLASTp analysis of the wheat proteome in Ensembl Plants; the expected wheat orthologs are the top three hits for the A, B, and D genomes (Figure 5B).

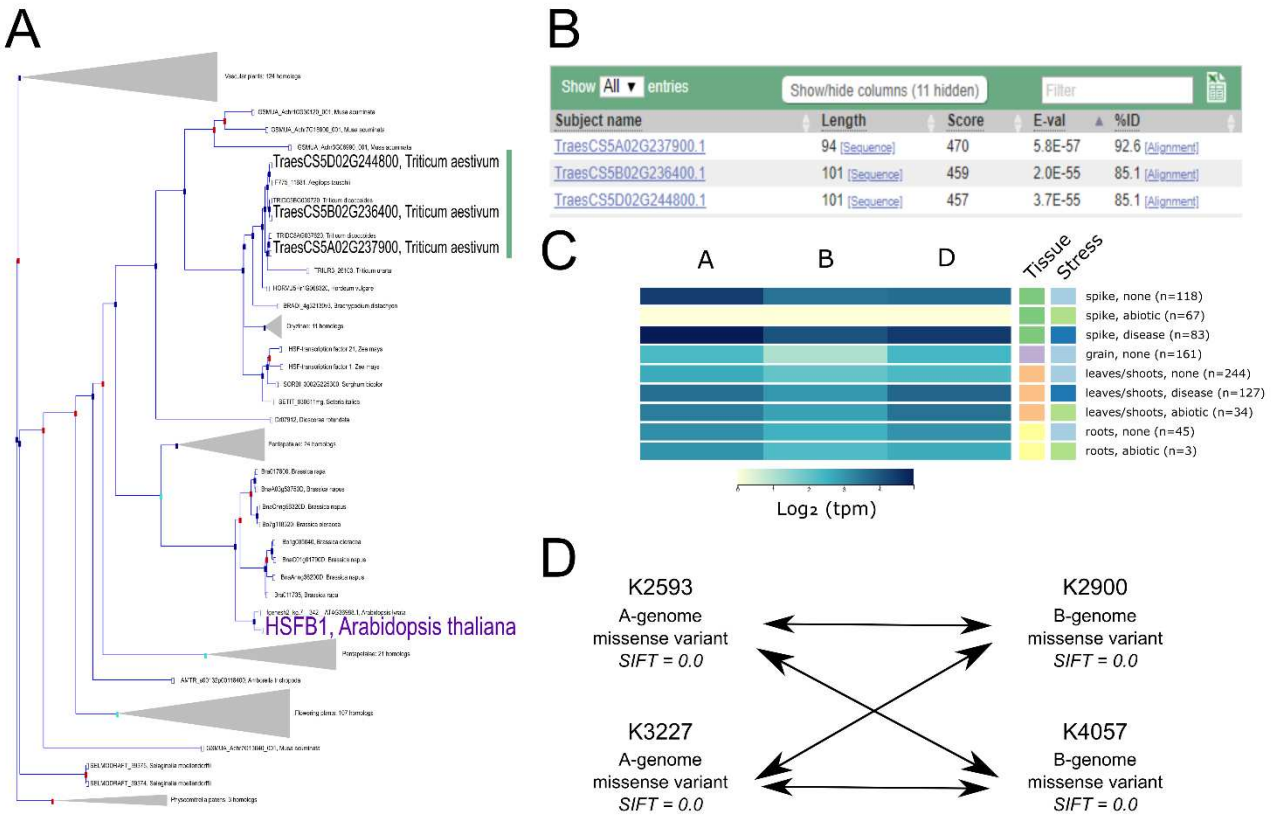

Figure 5: Case study exemplifying use of available gene functional characterisation in wheat. (A) The Ensembl Plants Gene Tree illustrates the identification of the wheat triad (green bar) most closely related to *AtHSFB1* (shown in purple). (**B**) Using *Os09g0456800* (the rice ortholog of *AtHSFB1*) as a BLASTp query against wheat predicted proteins independently identifies the same wheat triad. (**C**) Examination of RNA expression data from www.wheat-expression.com shows that the wheat triad is most highly expressed in the spike, with differential expression in abiotic and disease stress conditions. The samples are identified by tissue of origin (spike, green; grain, purple; leaves/shoots, orange; roots, yellow) and stress (none, light blue; abiotic, green; disease, dark blue) as they are on the website. (**D**) After identification of suitable wheat TILLING mutants, A and B genome homoeologs are combined via this example crossing scheme, demonstrating the four crosses required between the two selected mutations in each homoeolog. Note that the functional validation proposed in (**D**) is carried out using the tetraploid mutant population.

Having identified the wheat orthologs of *Arabidopsis TBF1*, we can examine and compare expression profiles using the expVIP browser (www.wheat-expression.com) (Borrill *et al.*, 2016; Ramirez-Gonzalez *et al.*, 2018) (Figure 5C). All three wheat homoeologs have similar expression profiles, with expression changes in the spike under disease and abiotic stress. This is consistent with the eFP browser data which shows high expression in the spikelet and awns of the non-stressed plants, as well as in more mature leaf tissues (Winter *et al.*, 2007; Ramirez-Gonzalez *et al.*, 2018). The expression data suggests that the wheat *TBF1* homoeologs are most strongly expressed in the spike and may have differential expression in response to biotic and abiotic stress.

Further investigation of these homoeologs can be performed using the KnetMiner knowledge network. For wheat *TBF1* orthologs, this includes homology, co-expression data, and associated TILLING mutants, combined with other wheat-specific information such as GENIE3 networks, wheat related publications, gene-phenotype relations extracted from the literature, GWAS data and *Arabidopsis* protein-protein interactions. Here the wheat genes, referred to as *HSFB1,* are orthologous to the *Arabidopsis* gene *TBF1* as demonstrated earlier, and the three wheat homoeologs fall into a module associated with responses to abiotic stresses (Figure 6). In addition, the *HSFB1* B and D homoeologs are predicted in the GENIE3 network to target the *LRK10* and *PPD* genes, which have known links to drought tolerance and sensitivity (Figure 6). The Knetminer database also recapitulates the relationship between the wheat *HSFB1* homoeologs and their rice and *Arabidopsis* orthologs which regulate heat stress responses (Figure 6). Considered as a whole, these data support the hypothesis that the *HSFB1* wheat genes are involved in the response to abiotic stress, perhaps specifically in drought response.

After evaluating *in silico* expression levels, we can then characterise the phenotype of wheat *TBF1* mutants using the exome-sequenced wheat TILLING mutant populations (Figure 2). We suggest to initially use the Kronos population, as it is based on a tetraploid line and thus contains only two copies of the gene (A and B homoeologs). This means that only two mutants need to be crossed to generate a full knockout. The hexaploid Cadenza TILLING population could also be used, but this would require an additional generation to combine mutant alleles across all three homoeologs.

All TILLING mutations can be accessed directly from Ensembl Plants in the “Genetic Variation” section. Although the TILLING mutants were originally called against the CSS assembly (Krasileva *et al.*, 2017), those available on Ensembl Plants have been re-called against the more recent RefSeqv1.0 genome. Available mutations in the gene of interest can be visualised as a table or positioned along the gene using the “Variant Image” or “Variant Table” option. We can thus rapidly identify mutations that are predicted to lead to a premature termination codon (PTC). However, if no appropriate PTC mutations are available, splice-site mutations predicted to lead to downstream frameshifts, or missense mutations in highly conserved amino acid residues with low SIFT (Sorting Intolerant from Tolerant; (Ng and Henikoff, 2003)) scores are good alternatives. 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.

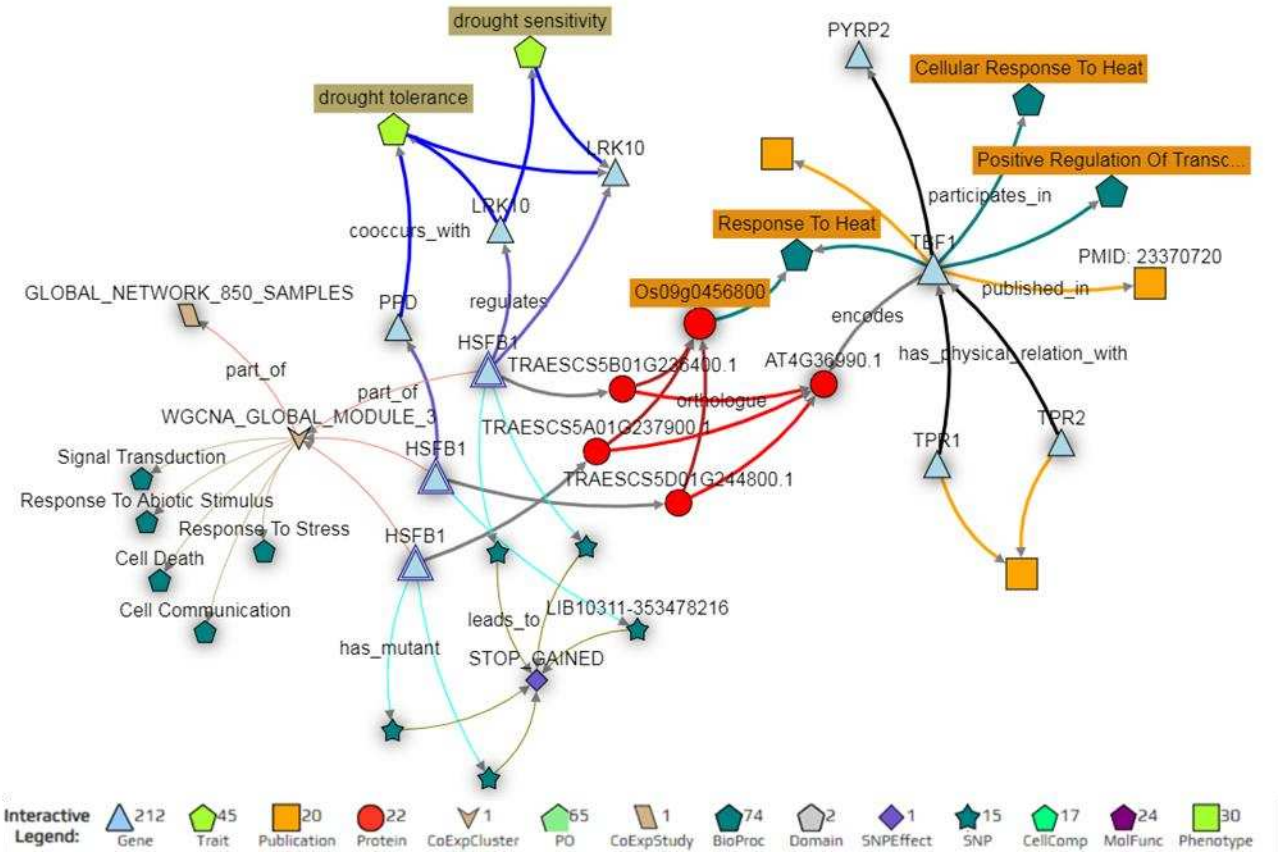

Figure 6: The KnetMiner network illustrates the putative role of the wheat *TBF1* orthologs in responding to abiotic stress. The wheat orthologs of the *Arabidopsis* gene *TBF1*, here depicted as three copies of the gene *HSFB1* (light blue triangles) fall in expression module three (brown arrow; WGCNA module 3). The genes in this module are enriched for GO terms such as "Response to Stress" and "Response to Abiotic Stimulus" (dark green pentagons). The *HFSB1* homoeologs are predicted to regulate other genes (blue triangles) in the GENIE3 network (purple connecting arrows) which are associated with the drought tolerance trait ontology terms (light green pentagon). PTC mutations are available for all three *HFSB1* homoeologs (dark green stars connecting with STOP GAINED SNP effect).

For both the A and the B genome *TBF1* homoeologs in Kronos, no PTC mutations are available. However, we identified missense mutations in highly conserved residues with SIFT scores of 0 suggesting that these mutations are likely to have a deleterious effect on protein function (Figure 5D). In addition to SIFT, 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 missense mutations on conserved protein domains.

TILLING lines from both population can be ordered via the GRU (https://www.seedstor.ac.uk/shopping-cart-tilling.php) in the UK or from the Dubcovsky lab (https://dubcovskylab.ucdavis.edu/wheat-tilling) in the USA. To maximise the chance of having selected functionally important mutants, we recommend choosing two independent mutant lines for each homoeolog and carrying out crosses between each mutant in the A and B genomes (four crosses shown 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 on www.wheat-training.com. For most TILLING mutations genome-specific primers have been predesigned and are available in Ensembl Plants. If there are no suitable predesigned primers, online tools such as PolyMarker can be used (Ramirez-Gonzalez *et al.*, 2015), or if needed, can be designed manually. After carrying out the initial cross, we grow the F_1 individuals under speed breeding conditions, and self-pollinate to obtain the F_2 seed. We then grow F_2 individuals and select via genetic markers individuals homozygous for one or both mutant alleles, as well as homozygous wild type control individuals (Figure 4). We can then carry out our first phenotypic evaluation on the F_2 plants using the homozygous wild type lines as controls without the need for backcrossing to Kronos. We can do this because the background mutations in the chosen lines will be segregating within both the mutant and the wild type lines, leading to an equivalent background mutation load between the sibling genotypes (Uauy *et al.*, 2017). Backcrossing to 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 yield components (Simmonds *et al.*, 2016), and when plants are intended for field phenotyping.

Concluding remarks

In recent years there has been a dramatic expansion in the number and accessibility of functional genomics resources in wheat. A step-change has been achieved from a highly fragmented assembly with incomplete gene models to a full pseudomolecule reference sequence with detailed annotation. This facilitates discovery and functional characterisation of genes using a series of well-established methodologies. Most resources described in this review are integrated with the bread wheat reference genome sequence including the expVIP expression browser, TILLING mutants, natural variation, co-expression networks and Ensembl Plants analyses and display tools. As a result, it is now easier than ever to use these resources as they are unified by a common reference genome and gene models. Furthermore, a pangenome of wheat is now available providing high quality genome sequences for multiple varieties of wheat. These genomes will facilitate functional studies in a range of different genetic backgrounds and enhance the value of the populations containing natural variation captured from diverse wheat varieties. Whilst wheat functional genomic resources have been in a state of flux for the past five years, the groundwork to accelerate gene discovery and characterisation in polyploid wheat has now been laid. This foundation provides exciting opportunities to accelerate wheat improvement and to help secure food production for the future.

Acknowledgments

This work was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) through the Designing Future Wheat (BB/P016855/1) and GEN 710 (BB/P013511/1) ISPs, grants BB/M008908/1, BB/M011666/1 and BB/P010741/1, an Anniversary Future Leader Fellowship BB/M014045/1, and funding for the Ensembl4breeders workshop (to BCM). This work was also supported by the Rank Prize Funds New Lecturer Award (to PB) and a Royal Society Research Grant (RGS\R1\191163). Support was also received from the John Innes Foundation (to SAH).

Author contributions

NMA, PB, JB, SAH, CM and CU conceived, designed and coordinated the manuscript. NMA, PB, JB, SAH, CM, KHP and CU designed the figures. All authors wrote and edited the 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 Biotechnol J* **9**, 1086-1099.

Allen AM, Winfield MO, Burridge AJ, Downie RC, Benbow HR, Barker GLA, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, Griffiths S, Bentley AR, Alda M, Jack P, Phillips AL, Edwards KJ. 2017. Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnol J* **15**, 390-401.

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 LJ, Poland J, Bariana H, Jones JDG, Bentley AR, Ayliffe M, Olson E, Xu SS, Steffenson BJ, Lagudah E, Wulff BBH. 2019. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat Biotechnol* **37**, 139-143.

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 DM, Kerhornou A, Walts B, Kersey P. 2015. Triticeae resources in Ensembl Plants. *Plant Cell Physiol* **56**, e3-e3.

Borrill P. 2019. Blurring the boundaries between cereal crops and model plants. *New Phytol*.

Borrill P, Adamski N, Uauy C. 2015. Genomics as the key to unlocking the polyploid potential of wheat. *New Phytol* **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.

Bovina R, Brunazzi A, Gasparini G, Sestili F, Palombieri S, Botticella E, Lafiandra D, Mantovani P, Massi A. 2014. Development of a TILLING resource in durum wheat for reverse- and forward-genetic analyses. *Crop Pasture Sci* **65**, 112-124.

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.

Chen L, Huang L, Min D, Phillips A, Wang S, Madgwick PJ, Parry MAJ, Hu Y-G. 2012. Development and characterization of a new TILLING population of common bread wheat (*Triticum aestivum* L.). *PLoS One* **7**, e41570.

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 Physiol* **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.

Cockram J, Mackay I. 2018. Genetic mapping populations for conducting high-resolution trait mapping in plants. In: Varshney RK, Pandey MK, Chitikineni A, eds. *Plant Genetics and Molecular Biology*. Cham: Springer International Publishing, 109-138.

Colasuonno P, Incerti O, Lozito ML, Simeone R, Gadaleta A, Blanco A. 2016. DHPLC technology for high-throughput detection of mutations in a durum wheat TILLING population. *BMC genetics* **17**, 43-43.

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. *Nat Commun* **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.

Dong C, Dalton-Morgan J, Vincent K, Sharp P. 2009. A modified TILLING method for wheat breeding. *Plant Genome* **2**, 39-47.

Fradgley N, Gardner KA, Cockram J, Elderfield J, Hickey JM, Howell P, Jackson R, Mackay IJ. 2019. A large-scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders. *PLoS biology* **17**, e3000071-e3000071.

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

Gardiner L-J, Joynson R, Omony J, Rusholme-Pilcher R, Olohan L, Lang D, Bai C, Hawkesford M, Salt D, Spannagl M, Mayer KFX, Kenny J, Bevan M, Hall N, Hall A. 2018. Hidden variation in polyploid wheat drives local adaptation. *Genome Res* **28**, 1319-1332.

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 Biotechnol J* **14**, 1406-1417.

Ghosh S, Watson A, Gonzalez-Navarro OE, Ramirez-Gonzalez RH, Yanes L, Mendoza-Suárez M, Simmonds J, Wells R, Rayner T, Green P, Hafeez A, Hayta S, Melton RE, Steed A, Sarkar A, Carter J, Perkins L, Lord J, Tester M, Osbourn A, Moscou MJ, Nicholson P, Harwood W, Martin C, Domoney C, Uauy C, Hazard B, Wulff BBH, Hickey LT. 2018. Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *bioRxiv*.

Gororo NN, Eagles HA, Eastwood RF, Nicolas ME, Flood RG. 2002. Use of *Triticum tauschii* to improve yield of wheat in low-yielding environments. *Euphytica* **123**, 241-254.

Grewal S, Hubbart-Edwards S, Yang C, Devi U, Baker L, Heath J, Ashling S, Scholefield D, Howells C, Yarde J, Isaac P, King IP, King J. 2019. Rapid identification of homozygosity and site of wild relative introgressions in wheat through chromosome-specific KASP genotyping assays. *Plant Biotechnol J*.

Grewal S, Hubbart-Edwards S, Yang C, Scholefield D, Ashling S, Burridge A, Wilkinson PA, King IP, King J. 2018a. Detection of *T. urartu* introgressions in wheat and development of a panel of interspecific introgression lines. *Front Plant Sci* **9**.

Grewal S, Yang C, Edwards SH, Scholefield D, Ashling S, Burridge AJ, King IP, King J. 2018b. Characterisation of *Thinopyrum bessarabicum* chromosomes through genome-wide introgressions into wheat. *Theor Appl Genet* **131**, 389-406.

Guo M, Liu J-H, Ma X, Luo D-X, Gong Z-H, Lu M-H. 2016. The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Front Plant Sci* **7**.

Harrington SA, Backhaus AE, Singh A, Hassani-Pak K, Uauy C. 2019. Validation and characterisation of a wheat GENIE3 network using an independent RNA-Seq dataset. *bioRxiv*, 684183.
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.
Hassani-Pak K, Castellote M, Esch M, Hindle M, Lysenko A, Taubert J, Rawlings C. 2016. Developing integrated crop knowledge networks to advance candidate gene discovery. *Appl Transl Genom* **11**, 18-26.
Hayta S, Smedley MA, Demir SU, Blundell R, Hinchliffe A, Atkinson N, Harwood WA. 2019. An efficient and reproducible *Agrobacterium*-mediated transformation method for hexaploid wheat (*Triticum aestivum* L.). *Plant Methods* **15**, 121.
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 PL, Davies JP, Webb SR, Spangenberg G, Hayes B, Daetwyler H, Tibbits J, Hayden M, Akhunov E. 2019. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat Genet* **51**, 896-904.
Hensel G, Himmelbach A, Chen W, Douchkov DK, Kumlehn J. 2011. Transgene expression systems in the Triticeae cereals. *J Plant Physiol* **168**, 30-44.
Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, Vilella AJ, Searle SMJ, Amode R, Brent S, Spooner W, Kulesha E, Yates A, Flicek P. 2016. Ensembl comparative genomics resources. *Database-Oxford* **2016**, bav096.
Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, Carbajo M, Chakiachvili M, Christensen M, Cummins C, Cuzick A, Davis P, Fexova S, Gall A, George N, Gil L, Gupta P, Hammond-Kosack KE, Haskell E, Hunt SE, Jaiswal P, Janacek SH, Kersey PJ, Langridge N, Maheswari U, Maurel T, McDowall MD, Moore B, Muffato M, Naamati G, Naithani S, Olson A, Papatheodorou I, Patricio M, Paulini M, Pedro H, Perry E, Preece J, Rosello M, Russell M, Sitnik V, Staines DM, Stein J, Tello-Ruiz MK, Trevanion SJ, Urban M, Wei S, Ware D, Williams G, Yates AD, Flicek P. 2019. Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res*.
Howells RM, Craze M, Bowden S, Wallington EJ. 2018. Efficient generation of stable, heritable gene edits in wheat using CRISPR/Cas9. *BMC Plant Biol* **18**, 215.
Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J* **10**, 826-839.
Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P, Leung H, Varshney RK, Cavanagh CR. 2015. MAGIC populations in crops: current status and future prospects. *Theor Appl Genet* **128**, 999-1017.
Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. 2010. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, e12776.
Ikeda M, Mitsuda N, Ohme-Takagi M. 2011. *Arabidopsis* HsfB1 and HsfB2b act as repressors of the expression of heat-inducible *Hsfs* but positively regulate the acquired thermotolerance. *Plant Physiol* **157**, 1243.
IWGSC. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788.
IWGSC. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**.
King J, Grewal S, Yang C-y, Hubbart Edwards S, Scholefield D, Ashling S, Harper JA, Allen AM, Edwards KJ, Burridge AJ, King IP. 2017a. Introgression of *Aegilops speltoides* segments in *Triticum aestivum* and the effect of the gametocidal genes. *Ann Bot* **121**, 229-240.
King J, Grewal S, Yang C-Y, Hubbart S, Scholefield D, Ashling S, Edwards KJ, Allen AM, Burridge A, Bloor C, Davassi A, da Silva GJ, Chalmers K, King IP. 2017b. A step change in the transfer of interspecific variation into wheat from *Amblyopyrum muticum*. *Plant Biotechnol J* **15**, 217-226.

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* **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 Genet Biol* **79**, 84-88.
Lee WS, Hammond-Kosack KE, Kanyuka K. 2012. Barley stripe mosaic virus-mediated tools for investigating gene function in cereal plants and their pathogens: virus-induced gene silencing, host-mediated gene silencing, and virus-mediated overexpression of heterologous protein. *Plant Physiol* **160**, 582-590.
Li J, Ye X, An B, Du L, Xu H. 2012. Genetic transformation of wheat: current status and future prospects. *Plant Biotechnol Rep* **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. *Nat Commun* **8**, 14261.
Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, Yu Y, Du H, Qi M, Li Y, Lu H, Yu H, Cui Y, Wang N, Chen C, Wu H, Zhao Y, Zhang J, Li Y, Zhou W, Zhang B, Hu W, van Eijk MJT, Tang J, Witsenboer HMA, Zhao S, Li Z, Zhang A, Wang D, Liang C. 2018. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* **557**, 424-428.
Lowe K, Wu E, Wang N, Hoerster G, Hastings C, Cho M-J, Scelonge C, Lenderts B, Chamberlin M, Cushatt J, Wang L, Ryan L, Khan T, Chow-Yiu J, Hua W, Yu M, Banh J, Bao Z, Brink K, Igo E, Rudrappa B, Shamseer P, Bruce W, Newman L, Shen B, Zheng P, Bidney D, Falco C, Register J, Zhao Z-Y, Xu D, Jones T, Gordon-Kamm W. 2016. Morphogenic regulators *Baby boom* and *Wuschel* improve monocot transformation. *Plant Cell* **28**, 1998-2015.
Luo M-C, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu S, Long H, Ramasamy RK, Rodriguez JC, Van SL, Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S, Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Müller H-G, Kugler K, Rivarola-Duarte L, Spannagl M, Mayer KFX, Lu F-H, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T, Salzberg SL, Devos KM, Dvořák J. 2017. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* **551**, 498.
Luo M, Li H, Chakraborty S, Morbitzer R, Rinaldo A, Upadhyaya N, Bhatt D, Louis S, Richardson T, Lahaye T, Ayliffe M. 2019. Efficient TALEN-mediated gene editing in wheat. *Plant Biotechnol J* **0**.
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 Biol* **12**, 141.
Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG, Himmelbach A, Mascher M, Bagnaresi P, Faccioli P, Cozzi P, Lauria M, Lazzari B, Stella A, Manconi A, Gnocchi M, Moscatelli M, Avni R, Deek J, Biyiklioglu S, Frascaroli E, Corneti S, Salvi S, Sonnante G, Desiderio F, Mare C, Crosatti C, Mica E, Ozkan H, Kilian B, De Vita P, Marone D, Joukhadar R, Mazzucotelli E, Nigro D, Gadaleta A, Chao S, Faris JD, Melo ATO, Pumphrey M, Pecchioni N, Milanesi L, Wiebe K, Ens J, MacLachlan RP, Clarke JM, Sharpe AG, Koh CS, Liang KYH, Taylor GJ, Knox R, Budak H, Mastrangelo AM, Xu SS, Stein N, Hale I, Distelfeld A, Hayden MJ, Tuberosa R, Walkowiak S, Mayer KFX, Ceriotti A, Pozniak CJ, Cattivelli L. 2019. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet* **51**, 885-895.
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 Genom Genet* **4**, 1603-1610.
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.

Miki Y, Yoshida K, Mizuno N, Nasuda S, Sato K, Takumi S. 2019. Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of *Aegilops*. *DNA Res* **26**, 171-182.
Milner SG, Maccaferri M, Huang BE, Mantovani P, Massi A, Frascaroli E, Tuberosa R, Salvi S. 2016. A multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. *durum*). *Plant Biotechnol J* **14**, 735-748.
Mujeeb-Kazi A, Kazi AG, Dundas I, Rasheed A, Ogbonnaya F, Kishii M, Bonnett D, Wang RRC, Xu S, Chen P, Mahmood T, Bux H, Farrakh S. 2013. Chapter Four - Genetic diversity for wheat improvement as a conduit to food security. In: Sparks DL, ed. *Advances in Agronomy*, Vol. 122: Academic Press, 179-257.
Mujeeb-Kazi A, Rosas V, Roldan S. 1996. Conservation of the genetic variation of *Triticum tauschii* (Coss.) Schmalh. (*Aegilops squarrosa* auct. non L.) in synthetic hexaploid wheats (*T. turgidum* L. s.lat. x *T. tauschii*; 2n=6x=42, AABBDD) and its potential utilization for wheat improvement. *Genet Resour Crop Ev* **43**, 129-134.
Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814.
Ogbonnaya FC, Ye G, Trethowan R, Dreccer F, Lush D, Shepperd J, van Ginkel M. 2007. Yield of synthetic backcross-derived lines in rainfed environments of Australia. *Euphytica* **157**, 321-336.
Pajerowska-Mukhtar KM, Wang W, Tada Y, Oka N, Tucker CL, Fonseca JP, Dong XN. 2012. The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Current Biology* **22**, 103-112.
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.
Pont C, Leroy T, Seidel M, Tondelli A, Duchemin W, Armisen D, Lang D, Bustos-Korts D, Goué N, Balfourier F, Molnár-Láng M, Lage J, Kilian B, Özkan H, Waite D, Dyer S, Letellier T, Alaux M, Russell J, Keller B, van Eeuwijk F, Spannagl M, Mayer KFX, Waugh R, Stein N, Cattivelli L, Haberer G, Charmet G, Salse J, Wheat, Barley Legacy for Breeding Improvement consortium. 2019. Tracing the ancestry of modern bread wheats. *Nat Genet* **51**, 905-911.
Puchta H. 2017. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Curr Opin Plant Biol* **36**, 1-8.
Ramirez-Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, Khedikar Y, Robinson SJ, Cory AT, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang D, Ridout CJ, Chalhoub B, Mayer KFX, Benhamed M, Latrasse D, Bendahmane A, International Wheat Genome Sequencing C, Wulff BBH, Appels R, Tiwari V, Datla R, Choulet F, Pozniak CJ, Provart NJ, Sharpe AG, Paux E, Spannagl M, Brautigam A, Uauy C. 2018. The transcriptional landscape of polyploid wheat. *Science* **361**.
Ramirez-Gonzalez RH, Uauy C, Caccamo M. 2015. PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics* **31**, 2038-2039.
Riaz A, Hathorn A, Dinglasan E, Ziems L, Richard C, Singh D, Mitrofanova O, Afanasenko O, Aitken E, Godwin I, Hickey L. 2017. Into the vault of the Vavilov wheats: old diversity for new alleles. *Genet Resour Crop Ev* **64**, 531-544.
Richardson T, Thistleton J, Higgins TJ, Howitt C, Ayliffe M. 2014. Efficient *Agrobacterium* transformation of elite wheat germplasm without selection. *PLANT CELL TISS ORG* **119**, 647-659.
Sestili F, Palombieri S, Botticella E, Mantovani P, Bovina R, Lafiandra D. 2015. TILLING mutants of durum wheat result in a high amylose phenotype and provide information on alternative splicing mechanisms. *Plant Sci* **233**, 127-133.
Shah R, Huang BE, Whan A, Newberry M, Verbyla K, Morell MK, Cavanagh CR. 2019. The complex genetic architecture of recombination and structural variation in wheat uncovered using a large 8-founder MAGIC population. *bioRxiv*, 594317.
Shan Q, Wang Y, Li J, Gao C. 2014. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat Protoc* **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. *Theor Appl Genet* **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*. 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.
Stadlmeier M, Hartl L, Mohler V. 2018. Usefulness of a multiparent advanced generation intercross population with a greatly reduced mating design for genetic studies in winter wheat. *Front Plant Sci* **9**.
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 Physiol* **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. *Nat Biotechnol* **10**, 667.
Verbyla AP, Cullis BR, Thompson R. 2007. The analysis of QTL by simultaneous use of the full linkage map. *Theor Appl Genet* **116**, 95-111.
Verbyla AP, George AW, Cavanagh CR, Verbyla KL. 2014. Whole-genome QTL analysis for MAGIC. *Theor Appl Genet* **127**, 1753-1770.
Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-335.
Wallington E. 2015. NIAB crop transformation.
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 Biotechnol J* **15**, 614-623.
Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Consortium IWGS, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo M-C, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E. 2014a. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J* **12**, 787-796.
Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L. 2014b. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* **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. *Nat 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.
Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley AR, Edwards KJ. 2016. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J* **14**, 1195-1206.

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.
Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**.
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. *Front Plant Sci* **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. *Plant J* **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. *Nat Commun* **7**, 12617.
Zhao G, Zou C, Li K, Wang K, Li T, Gao L, Zhang X, Wang H, Yang Z, Liu X, Jiang W, Mao L, Kong X, Jiao Y, Jia J. 2017. The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat Plants* **3**, 946-955.
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.