

Rothamsted Repository Download

A - Papers appearing in refereed journals

Adamski, N. M., Borril, P., Brinton, J., Harrington, S. A., Marchal, C., Bentley, A. R., Bovill, W. D., Cattivelli, L., Cockram, J., Contreras-Moreira, B., Ford, B., Ghosh, S., Harwood, W., Hassani-Pak, K., Hayta, S., Hickey, L. T., Kanyuka, K., King, J., Maccaferri, M., Naamati, G., Pozniak, C. J., Ramirez-Gonzalez, R. H., Sansaloni, C., Trevaskis, B., Wingen, L. U., Wulff, B. B. H. and Uauy, C. 2020. A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. *eLife*. 9, p. e55646.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.7554/eLife.55646>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/975v0/a-roadmap-for-gene-functional-characterisation-in-crops-with-large-genomes-lessons-from-polyploid-wheat>.

© 24 March 2020, Please contact library@rothamsted.ac.uk for copyright queries.

A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat

Nikolai M Adamski^{1*}, Philippa Borrill^{2*}, Jemima Brinton^{1*}, Sophie A Harrington^{1*}, Clemence Marchal^{1*}, Alison R Bentley³, William D Bovill⁴, Luigi Cattivelli⁵, James Cockram³, Bruno Contreras-Moreira⁶, Brett Ford⁴, Sreya Ghosh¹, Wendy Harwood¹, Keywan Hassani-Pak⁷, Sadiye Hayta¹, Lee T Hickey⁸, Kostya Kanyuka⁷, Julie King⁹, Marco Maccaferri¹⁰, Guy Naamati⁶, Curtis J Pozniak¹¹, Ricardo H Ramirez-Gonzalez¹, Carolina Sansaloni¹², Ben Trevaskis⁴, Luzie Wingen¹, Brande BH Wulff¹ and Cristobal Uauy¹

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 [European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK](#)

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 Crop 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

40 Bruno Contreras-Moreira: bcontreras@ebi.ac.uk
41 Brett Ford: bxrett76@hotmail.com
42 Sreya Ghosh: Sreya.Ghosh@jic.ac.uk
43 Wendy Harwood: Wendy.Harwood@jic.ac.uk
44 Keywan Hassani-Pak: Keywan.Hassani-Pak@rothamsted.ac.uk
45 Sadiye Hayta: Sadiye.Hayta@jic.ac.uk
46 Lee Hickey: l.hickey@uq.edu.au
47 Kostya Kanyuka: Kostya.Kanyuka@rothamsted.ac.uk
48 Julie King: Julie.king@nottingham.ac.uk
49 Marco Maccaferri: marco.maccaferri@unibo.it
50 Guy Naamati: gnaamati@ebi.ac.uk
51 Curtis Pozniak: curtis.pozniak@usask.ca
52 Ricardo Ramirez-Gonzalez Ricardo.Ramirez-Gonzalez@jic.ac.uk
53 Carolina Sansaloni: C.Sansaloni@cgiar.org
54 Ben Trevaskis: Ben.Trevaskis@csiro.au
55 Luzie Wingen: Luzie.Wingen@jic.ac.uk
56 Brande Wulff: brande.wulff@jic.ac.uk
57 Cristobal Uauy: Cristobal.Uauy@jic.ac.uk (corresponding author)
58
59

Abstract

Understanding the function of genes within staple crops will accelerate crop improvement by allowing targeted breeding approaches. Despite their importance, a lack of genomic information and resources has hindered the functional characterisation of major crop genes. The recent release of high-quality reference sequences for these crops underpins a suite of genetic and genomic resources that support basic research and breeding. For wheat, these include gene model annotations, expression atlases and gene networks that provide information about putative function. Sequenced mutant 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. This review provides a helpful guide for plant scientists, especially those expanding into crop research, to capitalise on the discoveries made in *Arabidopsis* and other plants. This will accelerate the improvement of crops 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 and other crops, discoveries from model species can rapidly be tested and functional genetic studies can now be performed for agronomically-important traits directly in the crops themselves (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 Gbp 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 and other crops with large genomes, such as sugarcane (Garsmeur *et al.*, 2018). 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 (Borrill *et al.*, 2019).

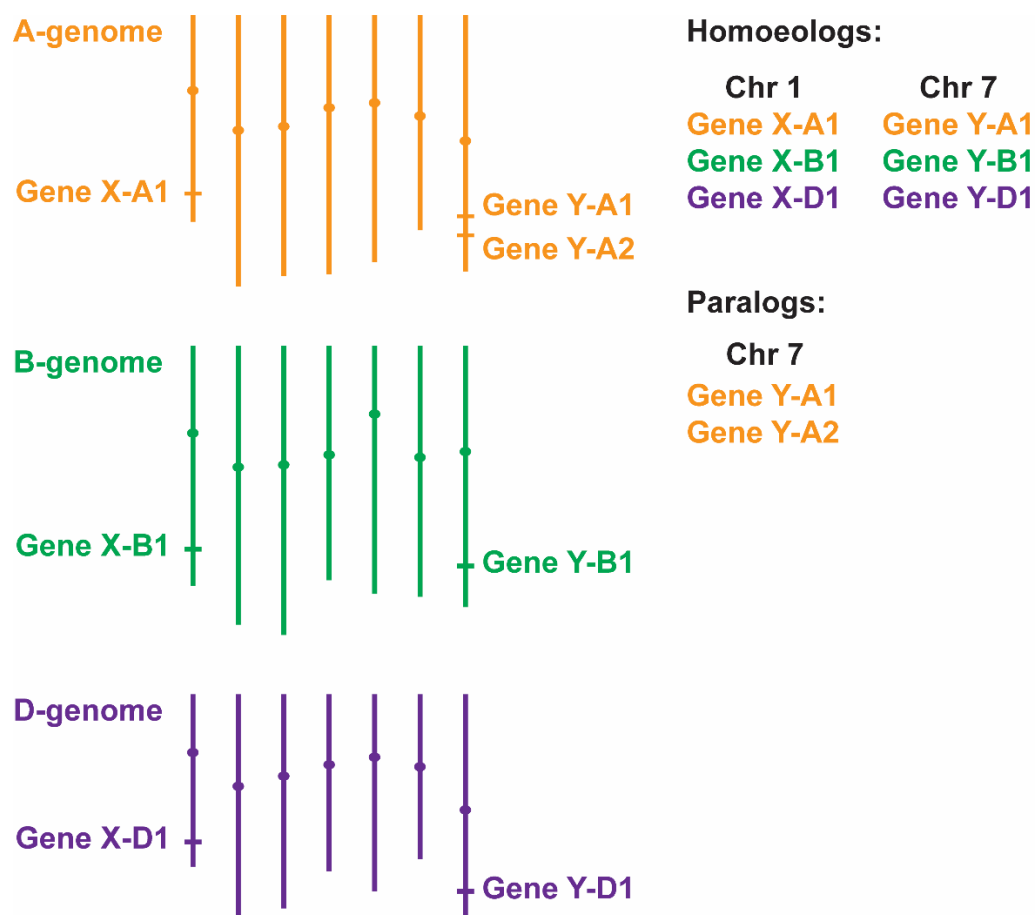
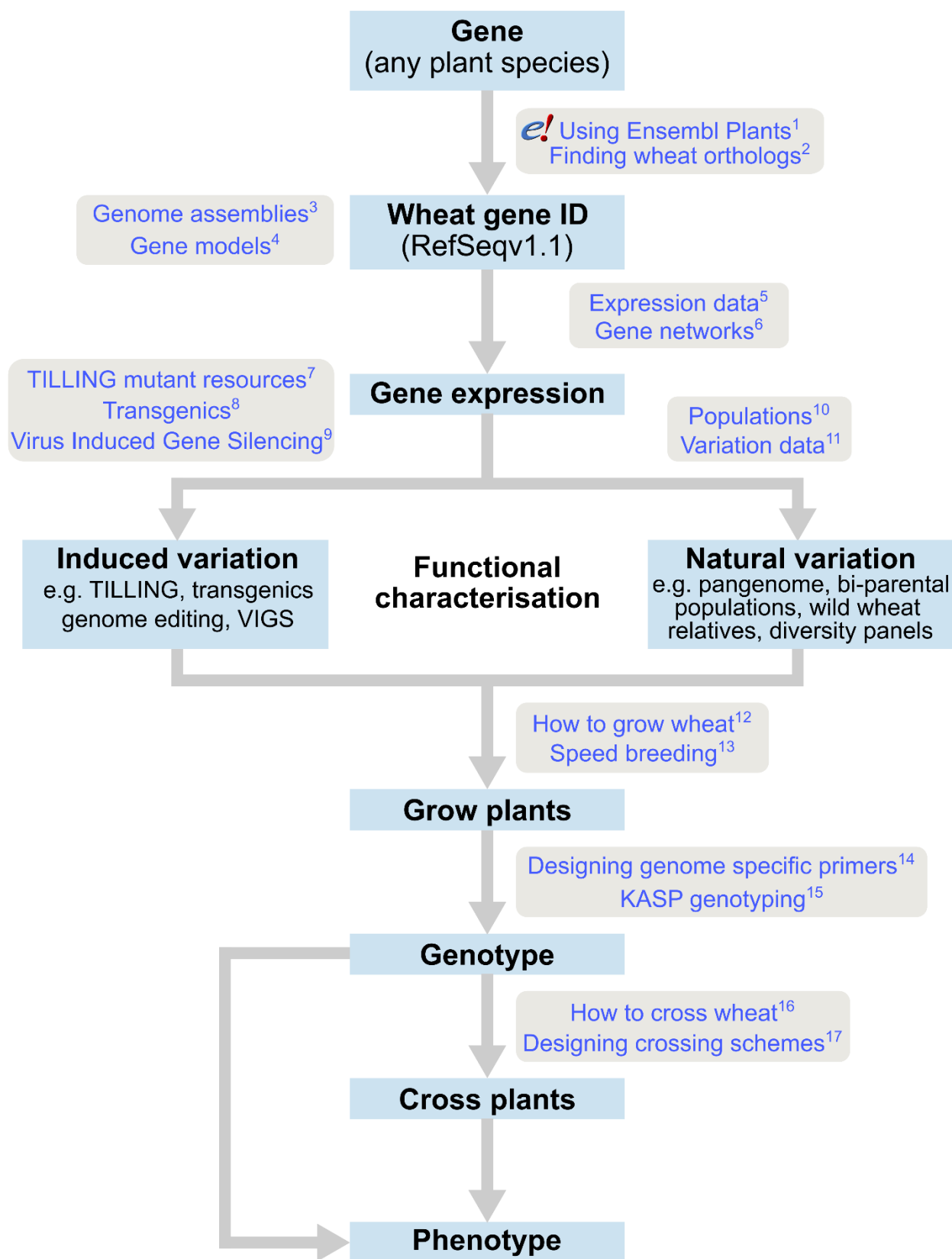


Figure 1: Gene homology within polyploid wheat. Due to two separate hybridisation events, genes in polyploid wheat will be present in multiple copies called homoeologs, which usually have similar chromosome locations. In the example of hexaploid bread wheat illustrated here, Gene X has homoeologs on 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 focus mainly on the Ensembl Plants database, as it integrates many of the publicly available data on wheat. However, other databases such as URGI \(<https://wheat-urgi.versailles.inra.fr/>; \(Alaux *et al.*, 2018\)\), the Wheat Information System \(WheatIS; <http://www.wheatis.org/>\), and GrainGenes \(<https://wheat.pw.usda.gov/GG3/>; \(Blake *et al.*, 2019\)\) also host and integrate similar, but also complementary genetic, genomic and phenomic data for wheat.](#) 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 crops with large genomes such as wheat.



113

114

115

116

117

118

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

119 natural variation are available. Finally steps for growing, genotyping and crossing plants are outlined. Links
 120 to detailed tutorials and further information are provided and can be found on www.wheat-training.com.
 121 ¹ www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/EnsemblPlants-primer.pdf
 122 ² www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Finding-wheat-orthologs.pdf
 123 ³ www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Genome_assemblies.pdf
 124 ⁴ www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Gene-models.pdf
 125 ⁵ www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Expression-browsers.pdf
 126 ⁶ www.wheat-training.com/wp-content/uploads/Genomic_resources/pdfs/Gene-networks.pdf
 127 ⁷ www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Selecting-TILLING-mutants.pdf
 128 ⁸ www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Transgenics.pdf
 129 ⁹ [www.wheat-training.com/wp-](http://www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Virus_Induced_Gene_Silencing.pdf)
 130 [content/uploads/Functional_studies/PDFs/Virus_Induced_Gene_Silencing.pdf](http://www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Virus_Induced_Gene_Silencing.pdf)
 131 ¹⁰ www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Populations.pdf
 132 ¹¹ www.wheat-training.com/wp-content/uploads/Genomic_resources/Variation-data.pdf
 133 ¹² www.wheat-training.com/wp-content/uploads/Wheat_growth/pdfs/Growing_Wheat_final.pdf
 134 ¹³ www.wheat-training.com/wp-content/uploads/Wheat_growth/pdfs/Speed_breeding.pdf
 135 ¹⁴ [www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Designing-genome-specific-](http://www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Designing-genome-specific-primers.pdf)
 136 [primers.pdf](http://www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Designing-genome-specific-primers.pdf)
 137 ¹⁵ [https://www.biosearchtech.com/support/education/kasp-genotyping-reagents/running-kasp-](https://www.biosearchtech.com/support/education/kasp-genotyping-reagents/running-kasp-genotyping-reactions)
 138 [genotyping-reactions](https://www.biosearchtech.com/support/education/kasp-genotyping-reagents/running-kasp-genotyping-reactions)
 139 ¹⁶ http://www.wheat-training.com/wp-content/uploads/Wheat_growth/pdfs/How-to-cross-wheat-pdf.pdf
 140 ¹⁷ www.wheat-training.com/wp-content/uploads/Functional_studies/PDFs/Designing-crossing-schemes.pdf

141 Wheat genome assemblies

142 A high-quality genome reference sequence is an essential resource for functional genetics and genomics in
 143 any species. Several hexaploid wheat genome assemblies have been released over the past six years
 144 (Brenchley *et al.*, 2012; IWGSC, 2014; Chapman *et al.*, 2015; Clavijo *et al.*, 2017; Zimin *et al.*, 2017). The
 145 most comprehensive assembly, called RefSeqv1.0, is a chromosome-level genome assembly annotated with
 146 high and low confidence gene models (IWGSC, 2018). Two tetraploid wheat genomes have also been
 147 sequenced, assembled, and annotated to the same standard as RefSeqv1.0 — the wild tetraploid
 148 progenitor of wheat, wild emmer (Avni *et al.*, 2017), and a modern durum wheat variety (Maccaferri *et al.*,
 149 2019). Diploid ancestral progenitor species have also been assembled to varying levels of completeness
 150 (Luo *et al.*, 2017; Zhao *et al.*, 2017; Ling *et al.*, 2018; Miki *et al.*, 2019). We summarize the annotated
 151 assemblies for polyploid wheat in Table 1; in this review we will focus mainly on the RefSeqv1.0 assembly.

152 **Table 1. Comparison of annotated genome assemblies in hexaploid and tetraploid wheat.** RefSeqv1.0 is
 153 the most widely used assembly and annotation of hexaploid wheat (available on Ensembl Plants
 154 <https://plants.ensembl.org/wheat>). The information from previous assemblies and annotations
 155 (Chromosome Survey Sequence (CSS) and TGACv1) are also available in the Ensembl Plants archive
 156 (<https://oct2017-plants.ensembl.org>) or as tracks in the Ensembl Plants genome browser interface.
 157 Ensembl Plants enables access to additional information such as SNP variation, gene trees, homoeolog
 158 assignments, and TILLING mutant information. Through this interface users can also combine knowledge
 159 from the bread, durum and wild emmer genomes.
 160

	CSS	TGACv1	RefSeqv1.0	Durum wheat	Wild emmer wheat
Publication	IWGSC (2014)	Clavijo <i>et al.</i> (2017)	IWGSC (2018)	Maccaferri <i>et al.</i> (2019)	Avni <i>et al.</i> (2017)
Contigs/Chromosomes	>1 million	735,943	21 chromosomes + ChrU	14 chromosomes + ChrU	14 chromosomes + ChrU
Mean scaffold size	7.7 kbp	88.7 kbp	Chromosomes	Chromosomes	Chromosomes
Assembly Size	10.2 Gbp	13.4 Gbp	14.6 Gbp	10.5 Gbp	10.5 Gbp
Order	Synteny/genetic order*	Large Bins	Physical order	Physical order	Physical order
Coding genes†	133,090 HC	104,091 HC	107,891 HC	66,559 HC	67,182 HC
	88,998 LC	103,660 LC	161,537 LC	303,404 LC	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

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

163 * Chromosome arm assignment was derived from chromosome flow-sorting, while approximate intra-chromosomal ordering was
 164 established using synteny derived from grasses (GenomeZipper) and genetic mapping (POPSEQ) (Mascher *et al.*, 2013; IWGSC,
 165 2014).
 166

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

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

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

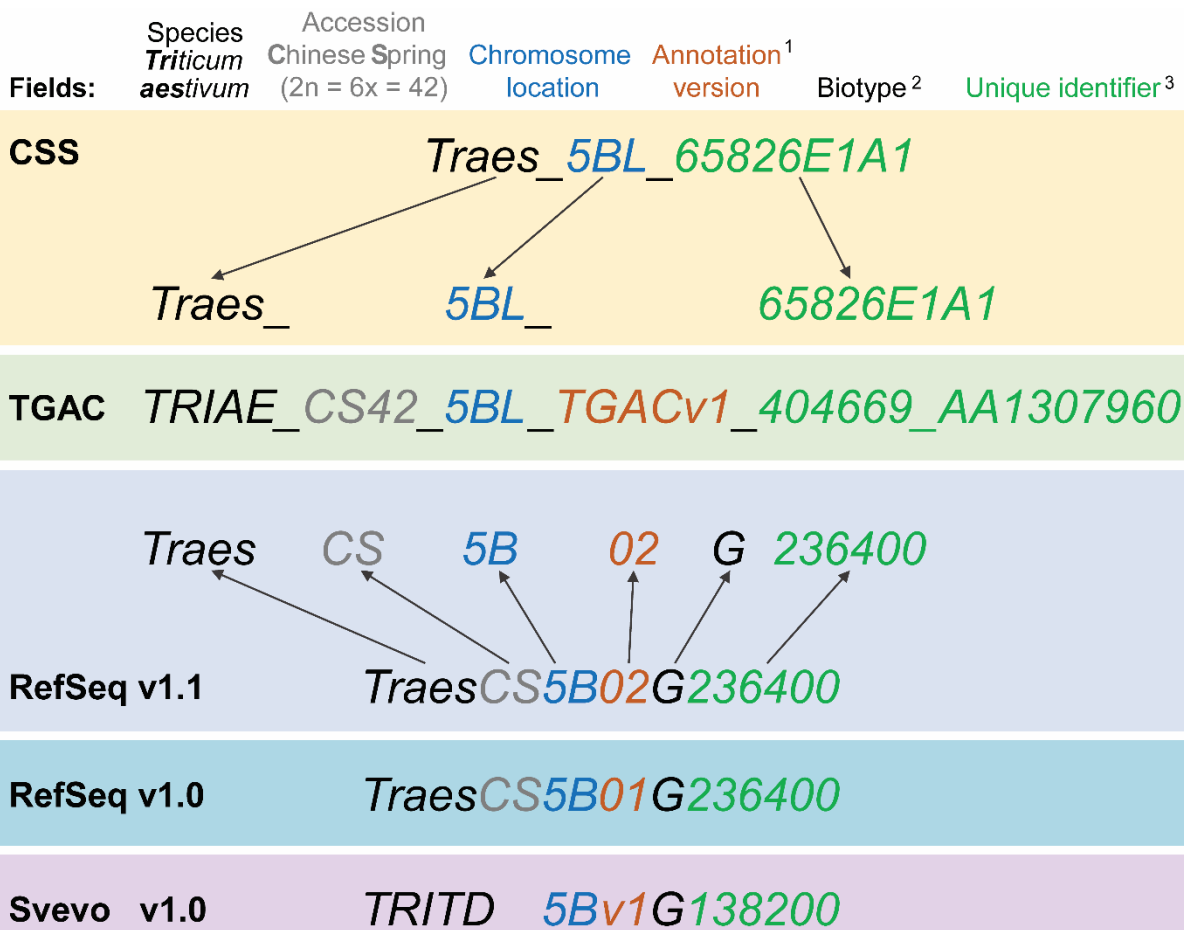


Figure 3. Gene model ID nomenclature description from the five available gene annotations for domesticated polyploid wheat. Here, one gene is used as an example to highlight the differences in gene ID nomenclature. Fields represented in the nomenclature are shown at the top with matching colours for 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; “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*. However, it is important to note that the relative positions of genes may change in future genome releases as the assembly is improved, for example, if scaffolds are rearranged. In these cases, the gene order would no longer be retained. 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 *TraesCS5B02G236400LC* are both located on chromosome 5B, but are not the same gene nor are they physically adjacent. Similarly, genes from homoeologous chromosomes with the

212 same subsequent numeric identifier are not necessarily homoeologous genes. For example,
213 *TraesCS5A02G236400*, *TraesCS5B02G236400* and *TraesCS5D02G236400* are **not** homoeologous genes.
214

215 Finding wheat orthologs

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

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

230 When performing a search for putative wheat orthologs via the Ensembl Plants pipeline, we would expect
231 to find three orthologs in hexaploid wheat for most gene queries. These orthologs would normally be
232 located on homoeologous chromosome groups, e.g. chromosomes 1A, 1B and 1D (Figure 1). A well-
233 documented exception to this rule is the long arm of chromosome 4A (4AL), which has undergone
234 translocation events with chromosome arms 5AL and 7BS (Devos *et al.*, 1995; Ma *et al.*, 2013). Therefore,
235 orthologs within these translocated regions will be physically located on different chromosome groups, e.g.
236 three homoeologous genes could be on chromosome arms 4AL, 5BL and 5DL. Furthermore, gene structure
237 of wheat orthologs is often conserved with respect to rice and other closely related monocot species; this
238 comparison can usually be done within Ensembl Plants. If this is not possible, wheat RNA-seq data can be
239 used to determine the gene structure. As an alternative to the Ensembl Plants Gene Trees, one can perform
240 reciprocal protein BLAST searches to identify putative wheat orthologs. We exemplify the above-mentioned
241 approaches along with potential pitfalls in more detail in the ‘Case Study’ section.

242 Expression data

243 Determining if, when, where, and to what level a gene is expressed often constitutes one of the first steps
244 towards its functional characterisation. Gene expression information can also be used to prioritize
245 candidate genes underlying a quantitative trait locus (QTL) or to predict those members of a large gene
246 family most relevant to trait expression. Numerous RNA-Seq datasets for wheat and many other crops have

247 been generated and published. Although the raw data are often publicly available (e.g. via the NCBI
 248 sequence read archive, <https://www.ncbi.nlm.nih.gov/sra>), they are not sufficiently curated for rapid
 249 access and their use in direct comparisons is complicated due to the diversity of tissues, treatments, and
 250 origins of the samples. Expression browsers aim to centralise these public datasets and analyse them
 251 together, ideally allowing retrieval of expression information for a list of genes under different conditions.
 252 For wheat, four expression browsers are currently available: expVIP (<http://www.wheat-expression.com>;
 253 (Borrill *et al.*, 2016)), wheat eFP browser (http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi; (Ramirez-
 254 Gonzalez *et al.*, 2018)), EBI Gene Expression Atlas
 255 (<https://www.ebi.ac.uk/gxa/experiments?species=triticum+aestivum>), and WheatExp
 256 (<https://wheat.pw.usda.gov/WheatExp>; (Pearce *et al.*, 2015)). Here we will focus on the first two given that
 257 they include a larger and more diverse set of samples and use the RefSeqv1.0 and v1.1 gene models
 258 described in Table 1.

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

267 To allow comparisons across studies, the 1,016 RNA-Seq samples in expVIP were classified according to four
 268 high-level categories based on variety, tissue, developmental stage and stress. These high-level categories
 269 are themselves divided into more detailed subcategories. These categories can be used to customize
 270 visualization displays and allows users to select data relevant to their experimental comparisons. Data can
 271 be displayed both as transcripts per million (TPM) or as raw counts and can be directly downloaded to carry
 272 out differential gene expression analyses. Although the default gene model reference is RefSeqv1.1, users
 273 can also choose the CSS, TGACv1 and RefSeqv1.0 transcriptome references for legacy reasons. Tutorials
 274 describing expVIP are available on <https://github.com/Uauy-Lab/expvip-web/wiki> and [www.wheat-](http://www.wheat-training.com)
 275 [training.com](http://www.wheat-training.com). Recently, expVIP was implemented for berry fruit species (Thole *et al.*, 2019).

276 An additional resource is the electronic Fluorescent Pictograph (eFP) browser, which provides a simple
 277 visual assessment of expression data using pictures coloured according to a gene’s relative expression level.
 278 The eFP expression browser is available for several crops (e.g. potato, soybean, barley) and most recently
 279 wheat (http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi). The wheat interface includes 209 RNA-Seq
 280 samples (also in expVIP) representing 22 tissue types from grain, root, leaf, and spike samples across
 281 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 (https://knetminer.org/Triticum_aestivum/).

KnetMiner is a web-application for searching and visualising genome-scale knowledge networks of e.g. *Arabidopsis*, wheat, and human diseases (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 and searches the underlying knowledge network to identify links between these 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).

Epigenomics

With the availability of the wheat genome, increasing interest has turned towards the wheat epigenome, i.e. heritable modifications to the genome that do not affect the DNA sequence itself, such as histone and DNA methylation. The global DNA methylome of polyploid wheat has been explored in multiple studies (Gardiner *et al.*, 2015; Gardiner *et al.*, 2018; Li *et al.*, 2019). The methylome of the reference cultivar Chinese Spring was initially captured at the seedling stage (Gardiner *et al.*, 2015), with more recent work focussing on the variation present in the seedling methylome of the 104 landraces from the Watkins core collection (Table 2) (Gardiner *et al.*, 2018). Researchers have also examined the changes in DNA methylation status as a result of biotic stress in wheat seedlings (Geng *et al.*, 2019). The raw bisulfite sequencing data from these experiments is available through public archives, however, it is not immediately accessible on genome browsers. More recently, new epigenomic data from Chinese Spring seedlings was released, which includes a wide variety of epigenetic marks such as DNA methylation, seven histone modifications, and chromatin accessibility (Li *et al.*, 2019). This data has been made publicly

316 [available through a bespoke genome browser \(http://bioinfo.sibs.ac.cn/cs_epigenome\)](http://bioinfo.sibs.ac.cn/cs_epigenome) and can be readily
317 [accessed by researchers to gain insight into the epigenomic landscape surrounding their genes of interest.](#)

318 Functional studies

319 After identifying a gene of interest there are now several options and resources available for functional
320 characterisation and validation in wheat (Figure 2). These include resources based both on natural and
321 induced variation and can involve both transgenic and non-transgenic approaches. It is important to
322 remember that due to the polyploid nature of wheat, there is often functional redundancy between
323 homoeologs (Borrill *et al.*, 2015). This means that it may be necessary to manipulate all homoeologs and
324 paralogs simultaneously to measure a strong phenotypic effect (see the ‘Strategies for Use’ section below
325 for more information).

326 Induced variation

327 TILLING

328 Polyploid species, such as wheat, are well suited to mutational approaches as the functional redundancy in
329 their genomes allows for the tolerance of a higher mutational load compared with diploid species (Tsai *et al.*,
330 2013; Uauy *et al.*, 2017). Bespoke mutant populations can be developed and screened for desired
331 mutations in a gene of interest, though the screening process is arduous and time-consuming. To overcome
332 this barrier, an *in-silico* wheat TILLING resource has been developed (Krasileva *et al.*, 2017). This resource
333 consists of two ethyl methanesulphonate (EMS) mutagenized populations: 1,535 lines of the tetraploid
334 durum wheat variety ‘Kronos’ and 1,200 lines of the hexaploid bread wheat variety ‘Cadenza’. Exome
335 capture and Illumina sequencing of these 2,735 mutant lines was then carried out. The raw data was
336 originally aligned to the CSS reference, mutations were identified, and their effects predicted based on the
337 CSS gene models (Krasileva *et al.*, 2017). Alleles predicted *in silico* to be deleterious (e.g. premature stop
338 codons, splice site mutations, non-synonymous amino acid substitutions with SIFT score < 0.05), were
339 identified for ~90% of the captured wheat genes (Krasileva *et al.*, 2017), thus making this a powerful
340 resource for rapidly identifying mutations in a gene of interest (Figure 2). The raw data has now been
341 aligned to the RefSeqv1.0 genome, allowing mutation identification and effect prediction based on the
342 RefSeqv1.1 gene models. These updated data are publicly available on Ensembl Plants (see Case Study for
343 details). For legacy purposes, the mutations called against the CSS reference remain available via
344 www.wheat-tilling.com. However, caution should be exercised as the mutation effects here are predicted
345 based on the CSS gene models, which are known to be less reliable than the RefSeq gene models (Brinton
346 *et al.*, 2018).

347 There are several important considerations when selecting a mutant line for characterisation. First, it is
348 essential to check the predicted effect of mutations in the context of a complete and experimentally
349 validated gene model. Second, in most cases, crossing is necessary to combine mutations in homoeologous
350 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;

386 Luo *et al.*, 2019), using a range of methods (reviewed in Uauy *et al.* (2017)). Currently, most studies have
387 introduced specific point mutations or small deletions leading to subsequent protein disruption, although
388 the technology holds the potential for complex applications such as allele swapping or gene insertion, as
389 reviewed by Puchta (2017). Similar to RNAi, constructs for Cas9-mediated gene editing can be designed to
390 target all homoeologs simultaneously (Zhang *et al.*, 2016; Howells *et al.*, 2018). Due to the current
391 efficiency of genome editing however, the likelihood of obtaining mutations in all homoeologs in a single T₀
392 plant remains low (0.9%; (Zhang *et al.*, 2016) and subsequent crosses to combine multiple edited targets
393 are likely to be required.

394 A major limitation of using transgenic approaches to manipulate agronomically relevant traits is the
395 associated legal and regulatory constraints. To overcome these, the nuclease transgene can be segregated
396 away from the edited gene(s) in subsequent generations. However, in Europe, and in contrast to many
397 other countries in the world, the resulting plants would be regulated as transgenics due to the 2018 ruling
398 on genome editing by the European Court of Justice (ECJ). Some studies have documented CRISPR/Cas9-
399 editing in wheat without transgene integration, for example, by delivering the CRISPR/Cas9 components as
400 ribonucleoproteins (RNPs). As no foreign DNA is used in CRISPR/Cas9 RNP-mediated genome editing, the
401 wheat mutants obtained are completely transgene free (Liang *et al.*, 2017), although still not exempt from
402 the ECJ regulation.

403 Virus Induced Gene Silencing

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

414 VIGS in wheat has been used primarily to investigate disease resistance in a range of varieties, and has been
415 restricted to a few tissue types such as leaf (Lee *et al.*, 2015), young seedlings (Zhang *et al.*, 2017a) and
416 spikes (Ma *et al.*, 2012). However, in principle, BSMV-mediated VIGS can be applied to any wheat genotype
417 and to almost any gene of interest. This functional genomics tool is particularly useful when analysing
418 multiple candidate genes, for example in map-based cloning projects (i.e. when physical intervals contain
419 several candidate genes) or from RNA-Seq differentially expressed datasets. VIGS is also useful in wheat
420 genotypes that are difficult to transform and in those for which mutant/TILLING populations are

421 unavailable. VIGS can be used for simultaneous silencing of all homoeologs or, in principle, entire small
422 gene families without the need for further genetic crosses.

423 Natural Variation

424 Although using induced variation presents a clear route to understand the function of specific genes in
425 wheat, the wealth of natural variation in wheat lines, and populations based on this variation, present an
426 alternative route to discover genes and correlate them with function. For example, populations differing for
427 alleles of the gene of interest could be used to rapidly infer the role of the gene. In order to capture the
428 diversity within wheat and create populations to test gene function, natural variation has been extensively
429 documented. Most studies have focused on SNPs between varieties that can be quickly assayed through
430 SNP arrays designed from gene coding sequences and untranslated regions (UTRs) (Wang *et al.*, 2014a;
431 Winfield *et al.*, 2016; Allen *et al.*, 2017), described in Borrill *et al.* (2015) and www.wheat-training.com.
432 Thousands of varieties and landraces have been processed using these arrays and datasets are available
433 through websites such as TCAP (<https://triticeaetoolbox.org/wheat>) (Blake *et al.*, 2016) and CerealsDB
434 (<http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB>) (Wilkinson *et al.*, 2016). Given that all SNPs from
435 the latter have been incorporated into Ensembl Plants, this means that large *in silico* allelic series are
436 readily available for many genes of interest.

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

444 Therefore, despite this wealth of data, the challenge remains to define the functional significance of this
445 variation. Traditionally, mapping populations or association panels would need to be developed or
446 assembled, and then genotyped, to assess how particular SNPs or haplotypes affect the trait of interest. In
447 wheat, many of these resources are now publicly available (Figure 2), thus facilitating the functional
448 characterisation of genes of interest. We describe some of these resources below and include links to
449 access genotypes, sequences and seeds in Table 2. Further details are available at [www.wheat-](http://www.wheat-training.com)
450 [training.com](http://www.wheat-training.com).

451 **Table 2: Natural variation resources available in wheat.**

Collection	Short description	Number of accessions	Genotyping	Data/seed availability	More information/Reference
<u>Wild wheat relatives and progenitor species</u>					
Seeds of Discovery	Wheat and wild relative accessions held by ICARDA and CIMMYT	80,000 accessions: 56,342 domesticated hexaploid (8 taxa); 18,946 domesticated tetraploid (8 taxa); 3,903 crop wild relatives included all known 27 wild species from <i>Aegilops-Triticum</i> species complex and 11 genomic constitutions.	DArT-seq	CIMMYT Dataverse http://hdl.handle.net/11529/10548030 Germinate data warehouse http://germinate.cimmyt.org/wheat . Records for all germplasm accessions can also be accessed at https://ssl.fao.org/glis/	https://seedsofdiscovery.org/
Open Wild Wheat	Accessions of <i>Aegilops tauschii</i> (D genome progenitor)	265 accessions	Whole genome shotgun sequenced (10-30x)	Sequencing: https://opendata.earlham.ac.uk/wheat/under_license/toronto/ ; Seed: https://www.seedstor.ac.uk/search-browseaccessions.php?idCollection=38	www.openwildwheat.org ; Arora <i>et al.</i> , 2019
Wild wheat introgression lines	Introgression lines from <i>Aegilops caudata</i> , <i>Aegilops speltoides</i> , <i>Amblyopyrum muticum</i> , <i>Thinopyrum bessarabicum</i> , <i>Thinopyrum elongatum</i> , <i>Thinopyrum intermedium</i> , <i>Thinopyrum ponticum</i> , <i>Triticum timopheevii</i> , <i>Triticum urartu</i> , rye and wheat cultivars (Chinese Spring, Highbury, Paragon, Pavon 76)	153 stable homozygous introgression lines available	35K Axiom Wheat Relative Genotyping array + 710 KASP markers (Grewal <i>et al.</i> , 2019)	Genotype: https://www.nottingham.ac.uk/wrc/germplasm-resources/genotyping.aspx ; Seed: https://www.seedstor.ac.uk/ (accessions WR0001-WR0155)	www.nottingham.ac.uk/WISP ; Grewal <i>et al.</i> , 2018a; Grewal <i>et al.</i> , 2018b, King <i>et al.</i> , 2017a, King <i>et al.</i> , 2017b
<u>Synthetic hexaploid wheat</u>					

Synthetic hexaploid wheat	Synthetic hexaploid wheats generated using <i>Aegilops tauschii</i> (DD) + European tetraploid (AABB) wheat	50 synthetic hexaploid wheats + pre-breeding accessions; backcross populations with Robigus and Paragon also available	35K Axiom breeders array	Genotype: https://www.cerealsdb.uk.net/cerealsdb/genomics/CerealsDB/axiom_download.php Seed: https://www.seedstor.ac.uk/ (store codes WS0001-WS0232)	https://www.niab.com/research/research-projects/designing-future-wheat
<i>Wheat diversity panels</i>					
Watkins historic collection of landrace wheats	World collection of wheat landraces grown as farmer saved seed before the 1930s. Genetically stable collection developed by two generations of single seed descent	829 accessions (core set of 119 represent majority of assayed genotypic variation). F _{4,5} mapping populations against Paragon, mainly for the core set.	35K Axiom breeders array (Allen <i>et al.</i> , 2017); subset exome sequenced (Gardiner <i>et al.</i> , 2018)	Genotype: https://www.cerealsdb.uk.net/cerealsdb/genomics/CerealsDB/axiom_download.php Seed: https://www.seedstor.ac.uk/ (store codes WATDE0001-WATDE1063)	http://wisplandracepillar.jic.ac.uk/results_resources.htm ; Wingen <i>et al.</i> , 2014; Wingen <i>et al.</i> , 2017
GEDIFLUX (Genetic Diversity Flux) collection	Western European winter wheat varieties that individually occupied over 5% of national acreage from 1945 to 2000. Biparental populations with Paragon (ongoing)	479 accessions	35K Axiom breeders array	Genotype: https://www.cerealsdb.uk.net/cerealsdb/genomics/CerealsDB/axiom_download.php ; Seed: https://www.seedstor.ac.uk/ (store codes WGED0001-WGED0729)	http://wisplandracepillar.jic.ac.uk/results_resources.htm ; Wingen <i>et al.</i> , 2014
NIAB wheat association mapping panel	Bread wheat varieties released between 1916-2007. Predominantly UK varieties (68%), also other North Western European countries e.g. France (10%) and Germany (8%)	480 accessions	90k SNP array	Seed, Genotype and Pedigree: https://www.niab.com/research/research-projects/resources	Fradgley <i>et al.</i> , 2019
OzWheat diversity panel	Genetic diversity in Australian wheat breeding (colonial landraces 1860s, first Australian-bred cultivars 1890s, CIMMYT-derived semi dwarfs 1960s, post 2000 wheat)	285 accessions	90k SNP array + additional 26K SNPs from transcriptome data	Seed and Genotype: contact Shannon Dillon from CSIRO (Shannon.Dillon@csiro.au)	

Vavilov wheat collection	Hexaploid wheat accessions including landraces, historic breeding lines and cultivars. Pure lines generated by single seed descent	295 accessions	DART-seq (34,311 polymorphic markers)	Genotype: Dr Lee Hickey at The University of Queensland (l.hickey@uq.edu.au); Seed: Australian Grains Genebank (sally.norton@ecodev.vic.gov.au)	Riaz <i>et al.</i> , 2017
WHEALBI wheat panel	Worldwide wheat accessions including diploid and tetraploid wild relatives, old hexaploid landraces and modern elite cultivars	487 accessions	Exome capture (~600,000 genetic variants in ~40,000 genes; 12,000 genes identified as putative presence/absence variation compared to RefSeqv1.0)	Genotype: https://urgi.versailles.inra.fr/download/iwgc/IWGC_RefSeq_Annotations/v1.0/iwgc_refseqv1.0_Whealbi_GWAS.zip ; Seed: https://www.gbif.org/dataset/a52ca10a-136a-4072-a6de-3ec6e7852365	Pont <i>et al.</i> , 2019
Global Durum Wheat (GDP) panel	Diversity used in durum wheat breeding programs globally, including landraces and modern varieties	1,056 accessions	90k SNP array	Genotype: ms in preparation; Seed: ICARDA genebank http://indms.icarda.org Filippo Bassi, F.Bassi@cgiar.org	
Tetraploid wheat Global Collection (TGC)	Wild emmer wheat, domesticated emmer, durum wheat landraces and other tetraploid wheat sub-species (<i>Triticum aethiopicum</i> , <i>Triticum carthlicum</i> , <i>Triticum polonicum</i> , <i>Triticum turanicum</i> , <i>Triticum turgidum</i> , <i>Triticum karamyshevii</i> and <i>Triticum petropavlovskyi</i>)	1,856	90k SNP array	Genotype: GrainGenes; Seed: on request for non-commercial use from University of Bologna (marco.maccaferri@unibo.it and roberto.tuberosa@unibo.it)	Maccaferri <i>et al.</i> , 2019
<i>MAGIC populations</i>					
CSIRO, Aus	4-way (parents Baxter, Chara, Westonia, Yitpi); 8-way (parents Baxter, Westonia, Yitpi, AC Barrie (Canada), Xiaoya54 (China), Volcani (Israel), Pastor	1,500 (4-way) and 3,000 (8-way) RILs	90k SNP array, microsatellite and DART markers > 20,000 SNPs mapped in each	Seed and Genotype: on request from CSIRO (Bill.Bovill@csiro.au)	Huang <i>et al.</i> , 2012; Shah <i>et al.</i> , 2019

	(Mexico), Alsen (USA))		population		
NIAB, UK	8-way (parents Alchemy, Brompton, Claire, Hereward, Rialto, Robigus, Xi19, Soissions); 16-way (Banco, Bersee, Brigadier, Copain, Cordiale, Flamingo, Gladiator, Holdfast, Klokka, Maris Fundin, Robigus, Slejpner, Soissons, Spark, Steadfast, Stetson)	NIAB 8-way MAGIC: >1,000 RILs; NIAB 16-way MAGIC: ~600 RILs	35K breeders array. Genome sequence (Claire, Robigus, others underway). Exome capture sequence of 16-way parents. Skim-seq of all RILs underway.	Claire and Robigus genomes: https://opendata.earlham.ac.uk/opendata/data/Triticum_aestivum/EI/v1.1/ ; Genotyping and Seed: https://www.niab.com/research/research-projects/resources	Mackay <i>et al.</i> , 2014; Gardner <i>et al.</i> , 2016
Germany	8-way (Event, Format, BAYP4535, Potenzial, Ambition, Bussard, Firl3565, Julius)	394 F _{6:8} RILs	5,435 SNPs from SNP array	Genotype and pedigree: http://doi.org/10.14459/2018mp1435172 (click the “open attachment browser” link); Seed: Bavarian State Research Centre for Agriculture (Freising, Germany)	Stadlmeier <i>et al.</i> , 2018
Germany	WM-800, 8-way (Patras, Meister, Linus, JB Asano, Tobak, Bernstein, Safari, Julius)	910 F _{4:6} RILs	15k Infinium iSelect SNP array	Genotype and pedigree: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6069784 ; Seed: on request from Martin Luther University, Germany (klaus.pillen@landw.uni-halle.de)	Sannemann <i>et al.</i> , 2018
Durum	4-way (Claudio (Italy), Colosseo (Italy), Neodur (France), Rascon/2*Tarro (advanced CIMMYT line))	334 F _{7:8} RILs	90k SNP array	Genotype and pedigree: https://onlinelibrary.wiley.com/doi/full/10.1111/pbi.12424 ; Seed: on request for non-commercial use from University of Bologna (marco.maccaferri@unibo.it and roberto.tuberosa@unibo.it)	Milner <i>et al.</i> , 2016

452 Wild wheat relatives and progenitor species:

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

462 Synthetic hexaploid wheat:

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

470 Wheat diversity panels:

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

475 Multiparent Advanced Generation Inter-Cross (MAGIC) populations:

476 MAGIC populations have been developed for many crop species (Huang *et al.*, 2015; Cockram and Mackay,
477 2018). The multiple generations of inter-crossing required to create MAGIC populations results in highly
478 recombined chromosomes which enables the use of approaches such as genome wide association scans
479 (GWAS) and whole-genome average interval mapping (WGAIM; (Verbyla *et al.*, 2007)) to define small
480 genetic intervals for traits of interest (reviewed by Verbyla *et al.* (2014)). Likewise, the use of multiple
481 parents in MAGIC allows more allelic variation to be examined compared to typical bi-parental populations
482 (Cockram and Mackay, 2018). In wheat, [seven](#) MAGIC populations are currently publicly available
483 constructed from 4, 8 or 16 founders. Parent information and further details can be found in Table 2.

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

485 [To date natural variation has largely been used for forward genetics approaches such as mapping genetic](#)
486 [regions underlying a phenotypic trait of interest. However, there is now an opportunity to apply natural](#)

variation in wheat for reverse genetics studies to complement transgenic, gene editing and induced variation approaches. For example, the pre-harvest sprouting locus *Phs-A1* was reported by two independent studies to be underpinned by different genes: in one case by a pair of tandem duplicated *Plasma Membrane 19* (*PM19-A1* and *PM19-A2*) genes (Barrero *et al.*, 2015), and in the other by a *mitogen-activated protein kinase kinase 3* (*TaMKK3-A*) gene (Torada *et al.*, 2016). Transgenic approaches seemed to validate the role of both *PM19* and *TaMKK3-A* to influence pre-harvest sprouting. However, by using 11 biparental populations and a MAGIC population segregating for the *Phs-A1* locus, it was possible to break the linkage with the polymorphism in *PM19* and confirm that the causal gene in all populations was *TaMKK3-A* (Shorinola *et al.*, 2017). This example illustrates the power of natural variation to validate the causal variants underpinning phenotypes in wheat.

Populations exploiting natural variation can also be used to validate gene function. For example, *TEOSINTE BRANCHED1* (*TB1*) was identified to regulate wheat spike architecture using a 4-parent Australian MAGIC population, and this function was confirmed using induced variation (TILLING and transgenic overexpression) and natural variation in the 8-parent UK MAGIC population (Dixon *et al.*, 2018). Interestingly, whilst *TB1* was important in both MAGIC populations, different homoeologs underpinned the variation: *TB1-D1* in the Australian population and *TB1-B1* in the UK population. This study suggests that by using natural variation, we can start to understand the nuanced regulation of phenotypes in wheat elicited by individual homoeologs. Together, these examples show that 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 and related technologies have allowed the assembly of chromosome scale assemblies for multiple cultivars in major crops such as maize (<https://nam-genomes.org/>), rice (Zhou *et al.*, 2019) or oilseed rape (Song *et al.*, 2020). For wheat, eight spring, eight winter hexaploid, and three tetraploid varieties/accessions have been assembled, several to a similar standard as the reference Chinese Spring genome (Table 3). Annotation of most 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/>).

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

Variety	Habit	Origin	Availability *
<i>Hexaploid wheat</i>			
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 <i>et al.</i> (2015)
Weebil	spring	Mexico	10+ Genome Project
ArinaLrFor	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
<i>Tetraploid wheat</i>			
Zavitan†	-	Israel	Avni <i>et al.</i> (2017)
Svevo	spring	Italy	Maccaferri <i>et al.</i> (2019)
Kronos	spring	US	10+ Genome Project

522 † ‘Zavitan’ is a tetraploid wild emmer (*T. dicoccoides*) accession.

523 * Varieties included within the 10+ Wheat Genomes Project can be accessed through the Earlham Grassroot Genomics portal
524 (<https://wheatis.tgac.ac.uk/grassroots-portal/blast>) and the 10+ Wheat Genomes project portal ([http://webblast.ipk-](http://webblast.ipk-gatersleben.de/wheat-ten-genomes)
525 [gatersleben.de/wheat-ten-genomes](http://webblast.ipk-gatersleben.de/wheat-ten-genomes)) (subset of varieties in each). The ‘Svevo’ genome can be accessed through
526 <https://www.interomics.eu/durum-wheat-genome> and Ensembl Plants. ‘Synthetic W7984’ and ‘Zavitan’ can be accessed through
527 the Grassroot Genomics, and Ensembl Plants, respectively.
528

529 Strategies for use

530 Variety selection and growth conditions

531 Whilst resources are now available for the functional validation of target genes in wheat, practical
532 knowledge is also required to maximise the value of these resources. Firstly, wheat varieties are adapted to
533 different growing conditions (e.g. daylength and vernalisation requirements) making it important to
534 consider the conditions under which functional validation will be conducted. If phenotyping will be
535 undertaken in greenhouse or controlled environment conditions then most varieties will be suitable,
536 although varieties without vernalisation requirements are faster to grow (details on wheat growth
537 conditions at www.wheat-training.com). If field trials are required for phenotypic characterisation (e.g.
538 yield-related traits), local adaptation is often necessary for correct interpretation of results given genotype
539 x environment interactions. For example, the sequenced TILLING populations (Kronos and Cadenza) do not
540 require vernalisation, facilitating greenhouse experiments, and originate from different regions of the
541 world, allowing field trials under different environments (Kronos is a Californian variety adapted to warm
542 dry weather whereas Cadenza is a UK variety adapted to cooler conditions).

543 For CRISPR/Cas9 and other non-transient transgenic approaches several varieties may be used, although
544 relatively few wheat varieties have been shown to display high enough transformation efficiencies to be
545 practical. This means that traditionally most transgenic studies in wheat have been limited to a few
546 varieties, such as 'Fielder', Cadenza, 'Bobwhite', 'Kenong 199' and Kronos (Li *et al.*, 2012; Richardson *et al.*,
547 2014; Liang *et al.*, 2017; Hayta *et al.*, 2019). This is now changing thanks to work by groups at NIAB (UK),
548 CAAS (China) and CSIRO (Australia) who have successfully transformed 39 (Wallington, 2015), 15 (Wang *et al.*,
549 2017) and six (Richardson *et al.*, 2014) varieties, respectively. However, the *Agrobacterium*-mediated
550 transformation efficiencies in all these studies still differ between varieties. Correct varietal selection for
551 transformation is critical for functional studies, given that some varieties might not be suitable to study a
552 particular phenotype (e.g. if the variety is resistant to a disease and hence cannot be used to test a
553 candidate resistance gene). Similarly, it is important to assess whether the gene of interest is
554 present/functional in the chosen variety, for example through PCR amplification and sequencing of the
555 gene. For several varieties this can now be done quickly by direct examination of their genome sequence
556 (Table 3).

557 Combining mutations for complete knock-outs in polyploid wheat

558 As we noted earlier, the polyploid nature of wheat means that it normally has multiple homoeologous
559 copies of every gene. These copies typically have highly similar coding DNA sequence and may have
560 redundant functions (Borrill *et al.*, 2015). Therefore, to characterise the function of a gene in wheat it is
561 often necessary to knock out all three homoeologs. This may be achieved by simultaneously targeting all
562 three copies using either RNAi (e.g. (Uauy *et al.*, 2006)) or CRISPR/Cas9 (e.g. (Zhang *et al.*, 2017b)). A large
563 number of transformants need to be screened to identify a null in all three genomes from a CRISPR
564 construct (Zhang *et al.*, 2017b; Howells *et al.*, 2018). If the targets are more divergent it may not even be
565 possible to use a single guide RNA to target all three homoeologs, in which case several guides may be used
566 through multiplexing. Alternatively, separate knock-outs for each homoeolog can be generated by
567 CRISPR/Cas9 or identified in TILLING populations. The mutations in each homoeolog can be combined by
568 crossing (for details see www.wheat-training.com), with two crosses necessary to combine knock-out
569 mutations in each of the three homoeologs in hexaploid wheat (Figure 4). Tetraploid wheat, with only two
570 homoeologs, can be used to accelerate functional characterisation as it requires just one cross to create
571 complete knock-out mutants (Figure 4). After self-pollination of this F₁, phenotyping of the trait of interest
572 can be initiated in the F₂ generation by comparing homozygous double knock-out mutants to the sibling
573 wild type plants. It is important to note that TILLING lines contain many background mutations and
574 backcrossing may be required to overcome the confounding effects of background mutations on target
575 phenotype. More details on these strategies are published in (Uauy *et al.*, 2017).

576

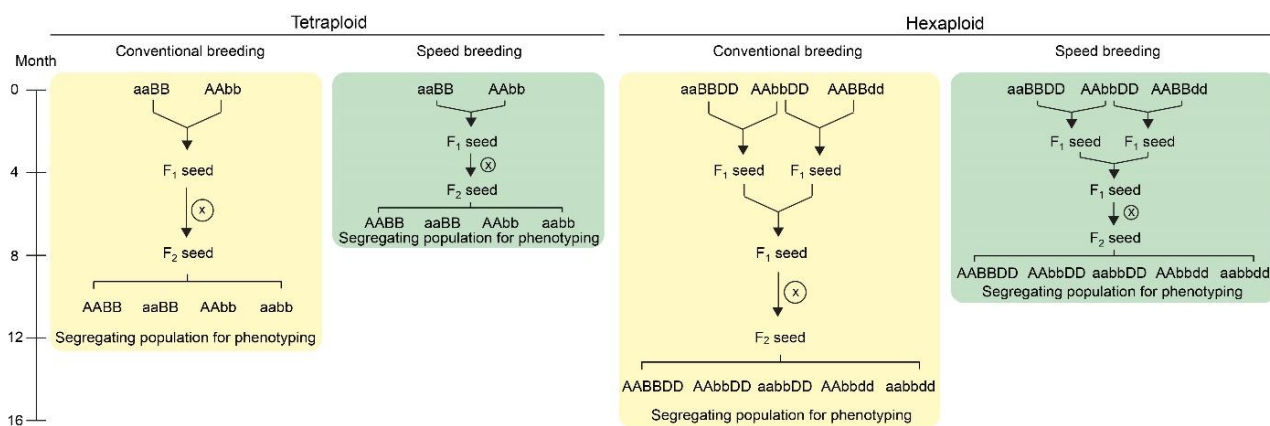


Figure 4. Crossing scheme to combine TILLING or CRISPR/Cas9 single mutants in wheat. In tetraploid wheat, mutations in the A and B genome homoeologs can be combined through a single cross. The F₁ plants are self-pollinated to produce a segregating F₂ population which contains homozygous double and single mutants, as well as wild type plants (screening using molecular markers required; only four genotypes shown). These F₂ 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 F₂ 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 (and other crops such as barley, canola and chickpea), 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). 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). It is important to note that most gene models where annotated in an automated manner and hence gene structures are likely to contain some errors, pending manual curation. We would thus recommend that researchers manually inspect the annotation of their genes of interest before proceeding further with their analyses.

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).

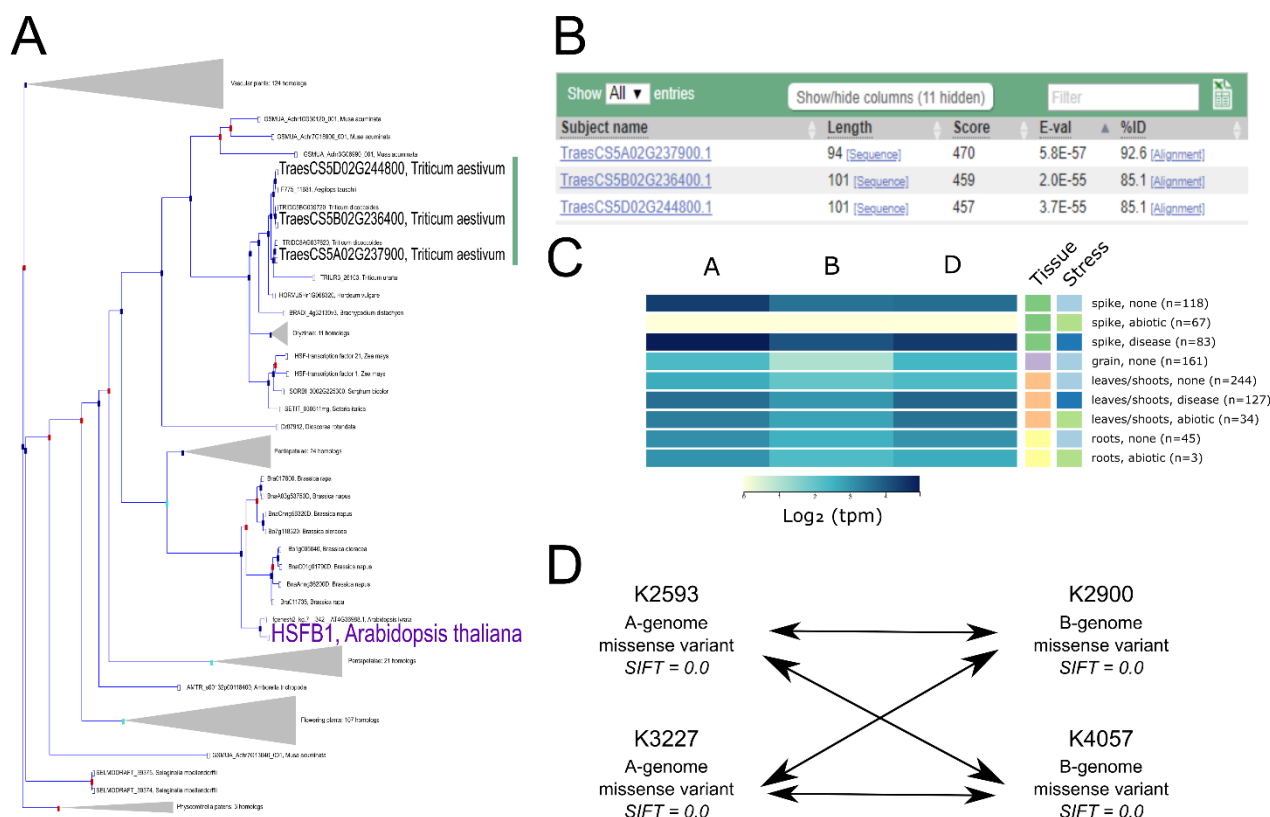


Figure 5: Case study exemplifying use of available gene functional characterisation in wheat. (A) The Ensembl Plants Gene Tree illustrates the identification of the wheat triad (green bar) most closely related to *AtHSFB1* (shown in purple). **(B)** Using *Os09g0456800* (the rice ortholog of *AtHSFB1*) 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. We can also explore the epigenetic environment of the three homoeologs using the bread wheat epigenomic map (http://bioinfo.sibs.ac.cn/cs_epigenome; (Li *et al.*, 2019)). A large peak for the H3K9ac histone modification at the 5' end of the homoeologs is indicative of active transcription from the promoter, corresponding with the observed gene expression. In contrast, the A-homoeolog

667 [TraesCS5A02G237900 is flanked by two genes which have low expression at the seedling stage, and](#)
668 [correspondingly low levels of H3K9ac modifications in their promoters. It is worth noting that the](#)
669 [epigenomic browser uses RefSeqv1.0 gene models, rather than the RefSeqv1.1 gene models used on](#)
670 [Ensembl Plants.](#)

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

683 After evaluating *in silico* expression levels, we can then characterise the phenotype of wheat *TBF1* mutants
684 using the exome-sequenced wheat TILLING mutant populations (Figure 2). We suggest to initially use the
685 Kronos population, as it is based on a tetraploid line and thus contains only two copies of the gene (A and B
686 homoeologs). This means that only two mutants need to be crossed to generate a full knockout. The
687 hexaploid Cadenza TILLING population could also be used, but this would require an additional generation
688 to combine mutant alleles across all three homoeologs (Figure 4).

689 All TILLING mutations re-called against the more recent RefSeqv1.0 genome can be accessed directly from
690 Ensembl Plants in the “Genetic Variation” section. Available mutations in the gene of interest can be
691 visualised as a table or positioned along the gene using the “Variant Image” or “Variant Table” option. We
692 can thus rapidly identify mutations that are predicted to lead to a premature termination codon (PTC).
693 However, if no appropriate PTC mutations are available, splice-site mutations predicted to lead to
694 downstream frameshifts, or missense mutations in highly conserved amino acid residues with low SIFT
695 (Sorting Intolerant from Tolerant; (Ng and Henikoff, 2003)) scores are good alternatives. SIFT scores predict
696 the effect of a mutation on protein function and are based on the physical properties of the alternative
697 amino acid as well as sequence homology.

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 F₁ individuals under speed breeding conditions, and self-pollinate to obtain the F₂ seed. We then grow F₂ 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 F₂ 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 F₂ 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 the last few years there has been a dramatic expansion in the resources available to carry out functional genomics in wheat, largely based upon improvements in the available reference sequences. Within a few years a step-change has been achieved from a highly fragmented assembly with incomplete gene models to a full pseudomolecule reference sequence alongside a detailed gene model annotation. This reference sequence allows the physical anchoring of genes in complete chromosomal order and provides [improved](#) gene models facilitating transgenic constructs and primers design. Most resources described in this review are integrated with the recent bread wheat reference genome sequence including the expVIP and eFP expression browsers, TILLING mutants and Ensembl Plants sequence analyses and display tools. As a result, it is now easier to use these resources as they are unified by a common reference genome and gene models. Furthermore, a pan-genome of wheat is being produced which will provide 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.

Future directions

Whilst many major advances have been made in the last five years to lay the groundwork for gene discovery and functional characterisation in polyploid wheat, looking to the future several key challenges remain.

- i) Polyploidy is a common challenge amongst crop species. In wheat we frequently assume that due to functional redundancy it will be necessary to knock-out all three homoeologs of a gene to assess its phenotypic impact. Yet the extent of homoeolog functional redundancy is still unclear (Borrill *et al.*, 2019). Transcriptomics and proteomics approaches will help generate hypotheses as to the extent of homoeolog redundancy in wheat and allow researchers to specifically target the most phenotypically relevant homoeolog for genetic manipulation.
- ii) Defining accessible (open) chromatin regions allows the identification of *cis*-regulatory sequences of potential functional significance. In animals and plants, genetic variants associated with quantitative traits are significantly enriched in these open chromatin sequences (Maurano *et al.*, 2012; Rodgers-Melnick *et al.*, 2016). In wheat, where over 98% of the genome is non-coding, it will be critical to identify open chromatin regions to more precisely define non-coding variation that may be of functional relevance. Work in tomato has elegantly shown how a wide range of phenotypic variation for quantitative traits can be engineered by genome editing of *cis*-regulatory regions of transcription factors (Rodríguez-Leal *et al.*, 2017).
- iii) To more readily test these hypotheses, increased transformation efficiency and reduced costs will also reshape the future of wheat research, perhaps one day becoming as accessible for wheat researchers as floral dip transformation is for *Arabidopsis*. It is becoming clear from research in wheat and other species that genetic background can have a strong influence on gene function. Therefore, it is essential to develop new protocols to transform multiple wheat varieties to account for these effects and to ensure that the potential of gene editing approaches is fulfilled.
- iv) Genomic databases have been powerful in integrating data from multiple studies and international efforts are now bringing together phenotypic data alongside genotypic data (e.g. Blake *et al.* (2016) and Howe *et al.* (2019)). Challenges remain to standardise phenotype collection protocols and ontologies, which will realise the full power of this information. Expanding these databases to include environmental conditions will allow assessments of interactions between genotypes, phenotypes and the environment.

High quality genome sequences facilitate moving beyond gene-based analysis, revealing the effects of non-genic regions on phenotype. Whilst working in crops with complex genomes will remain challenging, the advance of genomic techniques has enabled the wheat community to leverage lessons learnt in model species. The approaches taken in wheat provide a framework to understand biologically important traits in other species with large genomes.

786 Acknowledgments

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

793 Competing interests

794 The authors declare no competing interests.

795 Author contributions

796 NMA, PB, JB, SAH, CM and CU conceived, designed and coordinated the manuscript. NMA, PB, JB, SAH, CM,
797 KHP and CU designed the figures. All authors wrote and edited the manuscript.

798 References

- 799 **Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, Mohellibi N, Durand S, Kimmel E, Michotey**
800 **C, Guerche C, Loaec M, Lainé M, Steinbach D, Choulet F, Rimbart H, Leroy P, Guilhot N, Salse J, Feuillet C,**
801 **Paux E, Eversole K, Adam-Blondon A-F, Quesneville H, International Wheat Genome Sequencing C.** 2018.
802 Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence
803 to wheat genetic and phenomic data. *Genome Biol* **19**, 111.
804 **Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R,**
805 **McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ.** 2011. Transcript-specific, single-nucleotide
806 polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant*
807 *Biotechnol J* **9**, 1086-1099.
808 **Allen AM, Winfield MO, Burridge AJ, Downie RC, Benbow HR, Barker GLA, Wilkinson PA, Coghill J,**
809 **Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, Griffiths S, Bentley AR, Alda M,**
810 **Jack P, Phillips AL, Edwards KJ.** 2017. Characterization of a wheat breeders' array suitable for high-
811 throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant*
812 *Biotechnol J* **15**, 390-401.
813 **Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, Johnson R, Enk J, Periyannan S,**
814 **Singh N, Asyraf Md Hatta M, Athiyannan N, Cheema J, Yu G, Kangara N, Ghosh S, Szabo LJ, Poland J,**
815 **Bariana H, Jones JDG, Bentley AR, Ayliffe M, Olson E, Xu SS, Steffenson BJ, Lagudah E, Wulff BBH.** 2019.
816 Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat*
817 *Biotechnol* **37**, 139-143.
818 **Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K,**
819 **Jordan KW, Golan G, Deek J, Ben-Zvi B, Ben-Zvi G, Himmelbach A, MacLachlan RP, Sharpe AG, Fritz A,**
820 **Ben-David R, Budak H, Fahima T, Korol A, Faris JD, Hernandez A, Mikel MA, Levy AA, Steffenson B,**
821 **Maccaferri M, Tuberosa R, Cattivelli L, Faccioli P, Ceriotti A, Kashkush K, Pourkheirandish M, Komatsuda**
822 **T, Eilam T, Sela H, Sharon A, Ohad N, Chamovitz DA, Mayer KFX, Stein N, Ronen G, Peleg Z, Pozniak CJ,**
823 **Akhunov ED, Distelfeld A.** 2017. Wild emmer genome architecture and diversity elucidate wheat evolution
824 and domestication. *Science* **357**, 93-97.
825 **Barrero JM, Cavanagh C, Verbyla KL, Tibbits JFG, Verbyla AP, Huang BE, Rosewarne GM, Stephen S, Wang**
826 **P, Whan A, Rigault P, Hayden MJ, Gubler F.** 2015. Transcriptomic analysis of wheat near-isogenic lines
827 identifies *PM19-A1* and *A2* as candidates for a major dormancy QTL. *Genome Biol* **16**, 93.
828 **Blake VC, Birkett C, Matthews DE, Hane DL, Bradbury P, Jannink JL.** 2016. The Triticeae Toolbox:
829 combining phenotype and genotype data to advance small-grains breeding. *Plant Genome* **9**.

830 **Blake VC, Woodhouse MR, Lazo GR, Odell SG, Wight CP, Tinker NA, Wang Y, Gu YQ, Birkett CL, Jannink J-**
831 **L, Matthews DE, Hane DL, Michel SL, Yao E, Sen TZ.** 2019. GrainGenes: centralized small grain resources
832 and digital platform for geneticists and breeders. *Database* **2019**.

833 **Bolser DM, Kerhornou A, Walts B, Kersey P.** 2015. Triticeae resources in Ensembl Plants. *Plant Cell Physiol*
834 **56**, e3-e3.

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

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

838 **Borrill P, Harrington SA, Uauy C.** 2019. Applying the latest advances in genomics and phenomics for trait
839 discovery in polyploid wheat. *Plant J* **97**, 56-72.

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

842 **Bovina R, Brunazzi A, Gasparini G, Sestili F, Palombieri S, Botticella E, Lafiandra D, Mantovani P, Massi A.**
843 **2014.** Development of a TILLING resource in durum wheat for reverse- and forward-genetic analyses. *Crop*
844 *Pasture Sci* **65**, 112-124.

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

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

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

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

858 **Cheng M, Fry JE, Pang S, Zhou H, Hironaka CM, Duncan DR, Conner TW, Wan Y.** 1997. Genetic
859 transformation of wheat mediated by *Agrobacterium tumefaciens*. *Plant Physiol* **115**, 971-980.

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

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

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

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

874 **Devos KM, Dubcovsky J, Dvorak J, Chinoy CN, Gale MD.** 1995. Structural evolution of wheat chromosomes
875 4A, 5A, and 7B and its impact on recombination. *Theor Appl Genet* **91**, 282-288.

876 **Dixon LE, Greenwood JR, Bencivenga S, Zhang P, Cockram J, Mellers G, Ramm K, Cavanagh C, Swain SM,**
877 **Boden SA.** 2018. *TEOSINTE BRANCHED1* regulates inflorescence architecture and development in bread
878 wheat *Triticum aestivum*. *Plant Cell* **30**, 563.

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

881 **Fradgley N, Gardner KA, Cockram J, Elderfield J, Hickey JM, Howell P, Jackson R, Mackay IJ.** 2019. A large-
882 scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders. *PLoS biology*
883 **17**, e3000071-e3000071.

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

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

889 **Gardiner L-J, Quinton-Tulloch M, Olohan L, Price J, Hall N, Hall A.** 2015. A genome-wide survey of DNA
890 methylation in hexaploid wheat. *Genome Biol* **16**, 273.

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

894 **Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G, Charron C, Hervouet**
895 **C, Costet L, Yahiaoui N, Healey A, Sims D, Cherukuri Y, Sreedasyam A, Kilian A, Chan A, Van Sluys M-A,**
896 **Swaminathan K, Town C, Bergès H, Simmons B, Glaszmann JC, van der Vossen E, Henry R, Schmutz J,**
897 **D'Hont A.** 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat*
898 *Commun* **9**, 2638.

899 **Geng S, Kong X, Song G, Jia M, Guan J, Wang F, Qin Z, Wu L, Lan X, Li A, Mao L.** 2019. DNA methylation
900 dynamics during the interaction of wheat progenitor *Aegilops tauschii* with the obligate biotrophic fungus
901 *Blumeria graminis* f. sp. *tritici*. *New Phytol* **221**, 1023-1035.

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

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

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

912 **Grewal S, Hubbard-Edwards S, Yang C, Scholefield D, Ashling S, Burrridge A, Wilkinson PA, King IP, King J.**
913 2018a. Detection of *T. urartu* introgressions in wheat and development of a panel of interspecific
914 introgression lines. *Front Plant Sci* **9**.

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

918 **Guo M, Liu J-H, Ma X, Luo D-X, Gong Z-H, Lu M-H.** 2016. The plant Heat Stress transcription Factors (HSFs):
919 structure, regulation, and function in response to abiotic stresses. *Front Plant Sci* **7**.

920 **Harrington SA, Backhaus AE, Singh A, Hassani-Pak K, Uauy C.** 2019. Validation and characterisation of a
921 wheat GENIE3 network using an independent RNA-Seq dataset. *bioRxiv*, 684183.

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

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

926 **Hassani-Pak K, Castellote M, Esch M, Hindle M, Lysenko A, Taubert J, Rawlings C.** 2016. Developing
927 integrated crop knowledge networks to advance candidate gene discovery. *Appl Transl Genom* **11**, 18-26.

928 **Hayta S, Smedley MA, Demir SU, Blundell R, Hinchliffe A, Atkinson N, Harwood WA.** 2019. An efficient and
929 reproducible *Agrobacterium*-mediated transformation method for hexaploid wheat (*Triticum aestivum* L.).
930 *Plant Methods* **15**, 121.

931 **He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, Forrest K, Fritz A, Hucl P, Wiebe K, Knox R,**
932 **Cuthbert R, Pozniak C, Akhunova A, Morrell PL, Davies JP, Webb SR, Spangenberg G, Hayes B, Daetwyler**

933 **H, Tibbits J, Hayden M, Akhunov E.** 2019. Exome sequencing highlights the role of wild-relative
934 introgression in shaping the adaptive landscape of the wheat genome. *Nat Genet* **51**, 896-904.

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

937 **Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, Vilella AJ, Searle SMJ, Amode R, Brent
938 S, Spooner W, Kulesha E, Yates A, Flicek P.** 2016. Ensembl comparative genomics resources. *Database-
939 Oxford* **2016**, bav096.

940 **Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser
941 DM, Cambell L, Carbajo M, Chakiachvili M, Christensen M, Cummins C, Cuzick A, Davis P, Fexova S, Gall A,
942 George N, Gil L, Gupta P, Hammond-Kosack KE, Haskell E, Hunt SE, Jaiswal P, Janacek SH, Kersey PJ,
943 Langridge N, Maheswari U, Maurel T, McDowall MD, Moore B, Muffato M, Naamati G, Naithani S, Olson
944 A, Papatheodorou I, Patricio M, Paulini M, Pedro H, Perry E, Preece J, Rosello M, Russell M, Sitnik V,
945 Staines DM, Stein J, Tello-Ruiz MK, Trevanion SJ, Urban M, Wei S, Ware D, Williams G, Yates AD, Flicek P.**
946 2019. Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res.*

947 **Howells RM, Craze M, Bowden S, Wallington EJ.** 2018. Efficient generation of stable, heritable gene edits in
948 wheat using CRISPR/Cas9. *BMC Plant Biol* **18**, 215.

949 **Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR.** 2012. A multiparent
950 advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J* **10**, 826-839.

951 **Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P, Leung H, Varshney RK, Cavanagh CR.**
952 2015. MAGIC populations in crops: current status and future prospects. *Theor Appl Genet* **128**, 999-1017.

953 **Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P.** 2010. Inferring regulatory networks from expression data
954 using tree-based methods. *PLoS One* **5**, e12776.

955 **Ikeda M, Mitsuda N, Ohme-Takagi M.** 2011. *Arabidopsis* HsfB1 and HsfB2b act as repressors of the
956 expression of heat-inducible *Hsfs* but positively regulate the acquired thermotolerance. *Plant Physiol* **157**,
957 1243.

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

960 **IWGSC.** 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome.
961 *Science* **361**.

962 **King J, Grewal S, Yang C-y, Hubbart Edwards S, Scholefield D, Ashling S, Harper JA, Allen AM, Edwards KJ,
963 Burrridge AJ, King IP.** 2017a. Introgression of *Aegilops speltoides* segments in *Triticum aestivum* and the
964 effect of the gametocidal genes. *Ann Bot* **121**, 229-240.

965 **King J, Grewal S, Yang C-Y, Hubbart S, Scholefield D, Ashling S, Edwards KJ, Allen AM, Burrridge A, Bloor C,
966 Davassi A, da Silva GJ, Chalmers K, King IP.** 2017b. A step change in the transfer of interspecific variation
967 into wheat from *Amblyopyrum muticum*. *Plant Biotechnol J* **15**, 217-226.

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

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

973 **Lee WS, Hammond-Kosack KE, Kanyuka K.** 2012. Barley stripe mosaic virus-mediated tools for investigating
974 gene function in cereal plants and their pathogens: virus-induced gene silencing, host-mediated gene
975 silencing, and virus-mediated overexpression of heterologous protein. *Plant Physiol* **160**, 582-590.

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

978 **Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, Zhuang Y, Teng W, Ran X, Tong Y, Xue Y, Zhang W, Zhang Y.** 2019.
979 The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of
980 functional genetic elements. *Genome Biol* **20**, 139.

981 **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
982 genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat Commun* **8**, 14261.

983 **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,
984 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,**

985 **Witsenboer HMA, Zhao S, Li Z, Zhang A, Wang D, Liang C.** 2018. Genome sequence of the progenitor of
986 wheat A subgenome *Triticum urartu*. *Nature* **557**, 424-428.

987 **Lowe K, Wu E, Wang N, Hoerster G, Hastings C, Cho M-J, Scelonge C, Lenderts B, Chamberlin M, Cushatt J,**
988 **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,**
989 **Bruce W, Newman L, Shen B, Zheng P, Bidney D, Falco C, Register J, Zhao Z-Y, Xu D, Jones T, Gordon-**
990 **Kamm W.** 2016. Morphogenic regulators *Baby boom* and *Wuschel* improve monocot transformation. *Plant*
991 *Cell* **28**, 1998-2015.

992 **Luo M-C, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu**
993 **S, Long H, Ramasamy RK, Rodriguez JC, Van SL, Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S,**
994 **Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Müller H-G, Kugler K, Rivarola-Duarte**
995 **L, Spannagl M, Mayer KFX, Lu F-H, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T,**
996 **Salzberg SL, Devos KM, Dvořák J.** 2017. Genome sequence of the progenitor of the wheat D genome
997 *Aegilops tauschii*. *Nature* **551**, 498.

998 **Luo M, Li H, Chakraborty S, Morbitzer R, Rinaldo A, Upadhyaya N, Bhatt D, Louis S, Richardson T, Lahaye**
999 **T, Ayliffe M.** 2019. Efficient TALEN-mediated gene editing in wheat. *Plant Biotechnol J* **17**, 2026-2028.

1000 **Ma J, Stiller J, Berkman PJ, Wei Y, Rogers J, Feuillet C, Dolezel J, Mayer KF, Eversole K, Zheng YL, Liu C.**
1001 **2013. Sequence-based analysis of translocations and inversions in bread wheat (*Triticum aestivum* L.). *PLoS***
1002 ***One* **8**, e79329.**

1003 **Ma M, Yan Y, Huang L, Chen M, Zhao H.** 2012. Virus-induced gene-silencing in wheat spikes and grains and
1004 its application in functional analysis of HMW-GS-encoding genes. *BMC Plant Biol* **12**, 141.

1005 **Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T,**
1006 **Prade VM, Milner SG, Himmelbach A, Mascher M, Bagnaresi P, Faccioli P, Cozzi P, Lauria M, Lazzari B,**
1007 **Stella A, Manconi A, Gnocchi M, Moscatelli M, Avni R, Deek J, Biyiklioglu S, Frascaroli E, Corneti S, Salvi S,**
1008 **Sonnante G, Desiderio F, Mare C, Crosatti C, Mica E, Ozkan H, Kilian B, De Vita P, Marone D, Joukhadar R,**
1009 **Mazzucotelli E, Nigro D, Gadaleta A, Chao S, Faris JD, Melo ATO, Pumphrey M, Pecchioni N, Milanese L,**
1010 **Wiebe K, Ens J, MacLachlan RP, Clarke JM, Sharpe AG, Koh CS, Liang KYH, Taylor GJ, Knox R, Budak H,**
1011 **Mastrangelo AM, Xu SS, Stein N, Hale I, Distelfeld A, Hayden MJ, Tuberosa R, Walkowiak S, Mayer KFX,**
1012 **Cerioti A, Pozniak CJ, Cattivelli L.** 2019. Durum wheat genome highlights past domestication signatures
1013 and future improvement targets. *Nat Genet* **51**, 885-895.

1014 **Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, Greenland AJ, Horsnell R,**
1015 **Howells R, O'Sullivan DM, Rose GA, Howell PJ.** 2014. An eight-parent multiparent advanced generation
1016 inter-cross population for winter-sown wheat: creation, properties, and validation. *G3-Genes Genom Genet*
1017 **4**, 1603-1610.

1018 **Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Munoz-Amatriain M, Close TJ,**
1019 **Wise RP, Schulman AH, Himmelbach A, Mayer KF, Scholz U, Poland JA, Stein N, Waugh R.** 2013. Anchoring
1020 and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J* **76**, 718-727.

1021 **Matsuoka Y.** 2011. Evolution of polyploid *Triticum* wheats under cultivation: the role of domestication,
1022 natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiol* **52**, 750-764.

1023 **Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H,**
1024 **Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M,**
1025 **Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N,**
1026 **Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA.** 2012. Systematic localization of common disease-
1027 associated variation in regulatory DNA. *Science* **337**, 1190-1195.

1028 **Miki Y, Yoshida K, Mizuno N, Nasuda S, Sato K, Takumi S.** 2019. Origin of wheat B-genome chromosomes
1029 inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of *Aegilops*. *DNA Res*
1030 **26**, 171-182.

1031 **Milner SG, Maccaferri M, Huang BE, Mantovani P, Massi A, Frascaroli E, Tuberosa R, Salvi S.** 2016. A
1032 multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum*
1033 *ssp. durum*). *Plant Biotechnol J* **14**, 735-748.

1034 **Mujeeb-Kazi A, Kazi AG, Dundas I, Rasheed A, Ogbonnaya F, Kishii M, Bonnett D, Wang RRC, Xu S, Chen P,**
1035 **Mahmood T, Bux H, Farrakh S.** 2013. Chapter Four - Genetic diversity for wheat improvement as a conduit
1036 to food security. In: Sparks DL, ed. *Advances in Agronomy*, Vol. 122: Academic Press, 179-257.

1037 **Mujeeb-Kazi A, Rosas V, Roldan S.** 1996. Conservation of the genetic variation of *Triticum tauschii* (Coss.)
 1038 Schmalh. (*Aegilops squarrosa* auct. non L.) in synthetic hexaploid wheats (*T. turgidum* L. s.lat. x *T. tauschii*;
 1039 2n=6x=42, AABBDD) and its potential utilization for wheat improvement. *Genet Resour Crop Ev* **43**, 129-
 1040 134.

1041 **Ng PC, Henikoff S.** 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*
 1042 **31**, 3812-3814.

1043 **Ogbonnaya FC, Ye G, Trethowan R, Dreccer F, Lush D, Shepperd J, van Ginkel M.** 2007. Yield of synthetic
 1044 backcross-derived lines in rainfed environments of Australia. *Euphytica* **157**, 321-336.

1045 **Pajerowska-Mukhtar KM, Wang W, Tada Y, Oka N, Tucker CL, Fonseca JP, Dong XN.** 2012. The HSF-like
 1046 transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Current*
 1047 *Biology* **22**, 103-112.

1048 **Pearce S, Vazquez-Gross H, Herin SY, Hane D, Wang Y, Gu YQ, Dubcovsky J.** 2015. WheatExp: an RNA-seq
 1049 expression database for polyploid wheat. *BMC Plant Biol* **15**, 299.

1050 **Pont C, Leroy T, Seidel M, Tondelli A, Duchemin W, Armisen D, Lang D, Bustos-Korts D, Goué N, Balfourier**
 1051 **F, Molnár-Láng M, Lage J, Kilian B, Özkan H, Waite D, Dyer S, Letellier T, Alaux M, Russell J, Keller B, van**
 1052 **Eeuwijk F, Spannagl M, Mayer KFX, Waugh R, Stein N, Cattivelli L, Haberer G, Charmet G, Salse J, Wheat,**
 1053 **Barley Legacy for Breeding Improvement consortium.** 2019. Tracing the ancestry of modern bread wheats.
 1054 *Nat Genet* **51**, 905-911.

1055 **Puchta H.** 2017. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Curr Opin*
 1056 *Plant Biol* **36**, 1-8.

1057 **Ramirez-Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F,**
 1058 **Pasha A, Khedikar Y, Robinson SJ, Cory AT, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang**
 1059 **D, Ridout CJ, Chalhoub B, Mayer KFX, Benhamed M, Latrasse D, Bendahmane A, International Wheat**
 1060 **Genome Sequencing C, Wulff BBH, Appels R, Tiwari V, Datla R, Choulet F, Pozniak CJ, Provart NJ, Sharpe**
 1061 **AG, Paux E, Spannagl M, Brautigam A, Uauy C.** 2018. The transcriptional landscape of polyploid wheat.
 1062 *Science* **361**.

1063 **Ramirez-Gonzalez RH, Uauy C, Caccamo M.** 2015. PolyMarker: a fast polyploid primer design pipeline.
 1064 *Bioinformatics* **31**, 2038-2039.

1065 **Riaz A, Hathorn A, Dinglasan E, Ziems L, Richard C, Singh D, Mitrofanova O, Afanasenko O, Aitken E,**
 1066 **Godwin I, Hickey L.** 2017. Into the vault of the Vavilov wheats: old diversity for new alleles. *Genet Resour*
 1067 *Crop Ev* **64**, 531-544.

1068 **Richardson T, Thistleton J, Higgins TJ, Howitt C, Ayliffe M.** 2014. Efficient *Agrobacterium* transformation of
 1069 elite wheat germplasm without selection. *PLANT CELL TISS ORG* **119**, 647-659.

1070 **Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES.** 2016. Open chromatin reveals the functional maize
 1071 genome. *Proc Natl Acad Sci* **113**, E3177-E3184.

1072 **Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB.** 2017. Engineering quantitative trait
 1073 variation for crop improvement by genome editing. *Cell* **171**, 470-480.e478.

1074 **Sannemann W, Lisker A, Maurer A, Léon J, Kazman E, Cöster H, Holzapfel J, Kempf H, Korzun V, Ebmeyer**
 1075 **E, Pillen K.** 2018. Adaptive selection of founder segments and epistatic control of plant height in the MAGIC
 1076 winter wheat population WM-800. *BMC Genomics* **19**, 559-559.

1077 **Sestili F, Palombieri S, Botticella E, Mantovani P, Bovina R, Lafiandra D.** 2015. TILLING mutants of durum
 1078 wheat result in a high amylose phenotype and provide information on alternative splicing mechanisms.
 1079 *Plant Sci* **233**, 127-133.

1080 **Shah R, Huang BE, Whan A, Newberry M, Verbyla K, Morell MK, Cavanagh CR.** 2019. The complex genetic
 1081 architecture of recombination and structural variation in wheat uncovered using a large 8-founder MAGIC
 1082 population. *bioRxiv*, 594317.

1083 **Shan Q, Wang Y, Li J, Gao C.** 2014. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat*
 1084 *Protoc* **9**, 2395.

1085 **Shorinola O, Balcárková B, Hyles J, Tibbits JFG, Hayden MJ, Holuřova K, Valárik M, Distelfeld A, Torada A,**
 1086 **Barrero JM, Uauy C.** 2017. Haplotype analysis of the pre-harvest sprouting resistance locus *Phs-A1* reveals a
 1087 causal role of *TaMKK3-A* in global germplasm. *Front Plant Sci* **8**.

1088 **Simmonds J, Scott P, Brinton J, Mestre TC, Bush M, del Blanco A, Dubcovsky J, Uauy C.** 2016. A splice
 1089 acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat
 1090 through wider and longer grains. *Theor Appl Genet* **129**, 1099-1112.

1091 **Song J-M, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R, Xie W-Z, Cheng Y, Zhang Y, Liu**
 1092 **K, Yang Q-Y, Chen L-L, Guo L.** 2020. Eight high-quality genomes reveal pan-genome architecture and
 1093 ecotype differentiation of *Brassica napus*. *Nat Plants* **6**, 34-45.

1094 **Sparks CA, Doherty A, Jones HD.** 2014. Genetic transformation of wheat via *Agrobacterium*-mediated DNA
 1095 delivery. In: Henry RJ, Furtado A, eds. *Cereal Genomics*. Totowa, NJ: Humana Press, 235-250.

1096 **Sparks CA, Jones HD.** 2009. Biolistics transformation of wheat. In: Jones HD, Shewry PR, eds. *Transgenic*
 1097 *wheat, barley and oats: production and characterization protocols*. Totowa, NJ: Humana Press, 71-92.

1098 **Stadlmeier M, Hartl L, Mohler V.** 2018. Usefulness of a multiparent advanced generation intercross
 1099 population with a greatly reduced mating design for genetic studies in winter wheat. *Front Plant Sci* **9**.

1100 **Thole V, Bassard J-E, Ramírez-González R, Trick M, Ghasemi Afshar B, Breitel D, Hill L, Foito A, Shepherd L,**
 1101 **Freitag S, Nunes dos Santos C, Menezes R, Bañados P, Naesby M, Wang L, Sorokin A, Tikhonova O,**
 1102 **Shelenga T, Stewart D, Vain P, Martin C.** 2019. RNA-seq, *de novo* transcriptome assembly and flavonoid
 1103 gene analysis in 13 wild and cultivated berry fruit species with high content of phenolics. *BMC Genomics* **20**,
 1104 995.

1105 **Torada A, Koike M, Ogawa T, Takenouchi Y, Tadamura K, Wu J, Matsumoto T, Kawaura K, Ogiwara Y.**
 1106 2016. A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase kinase. *Current*
 1107 *Biology* **26**, 782-787.

1108 **Tsai H, Missirlian V, Ngo KJ, Tran RK, Chan SR, Sundaresan V, Comai L.** 2013. Production of a high-efficiency
 1109 TILLING population through polyploidization. *Plant Physiol* **161**, 1604-1614.

1110 **Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J.** 2006. A NAC gene regulating senescence improves
 1111 grain protein, zinc, and iron content in wheat. *Science* **314**, 1298-1301.

1112 **Uauy C, Wulff BBH, Dubcovsky J.** 2017. Combining traditional mutagenesis with new high-throughput
 1113 sequencing and genome editing to reveal hidden variation in polyploid wheat. *Annu Rev Genet* **51**, 435-454.

1114 **Vasil V, Castillo AM, Fromm ME, Vasil IK.** 1992. Herbicide resistant fertile transgenic wheat plants obtained
 1115 by microprojectile bombardment of regenerable embryogenic callus. *Nat Biotechnol* **10**, 667.

1116 **Verbyla AP, Cullis BR, Thompson R.** 2007. The analysis of QTL by simultaneous use of the full linkage map.
 1117 *Theor Appl Genet* **116**, 95-111.

1118 **Verbyla AP, George AW, Cavanagh CR, Verbyla KL.** 2014. Whole-genome QTL analysis for MAGIC. *Theor*
 1119 *Appl Genet* **127**, 1753-1770.

1120 **Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E.** 2009. EnsemblCompara GeneTrees:
 1121 complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-335.

1122 **Wallington E.** 2015. NIAB crop transformation.

1123 **Wang K, Liu H, Du L, Ye X.** 2017. Generation of marker-free transgenic hexaploid wheat via an
 1124 *Agrobacterium*-mediated co-transformation strategy in commercial Chinese wheat varieties. *Plant*
 1125 *Biotechnol J* **15**, 614-623.

1126 **Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L,**
 1127 **Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Consortium IWGS, Lillemo M,**
 1128 **Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M,**
 1129 **Pozniak C, Luo M-C, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I,**
 1130 **Cavanagh C, Edwards KJ, Hayden M, Akhunov E.** 2014a. Characterization of polyploid wheat genomic
 1131 diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J* **12**, 787-796.

1132 **Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L.** 2014b. Simultaneous editing of three
 1133 homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol*
 1134 **32**, 947.

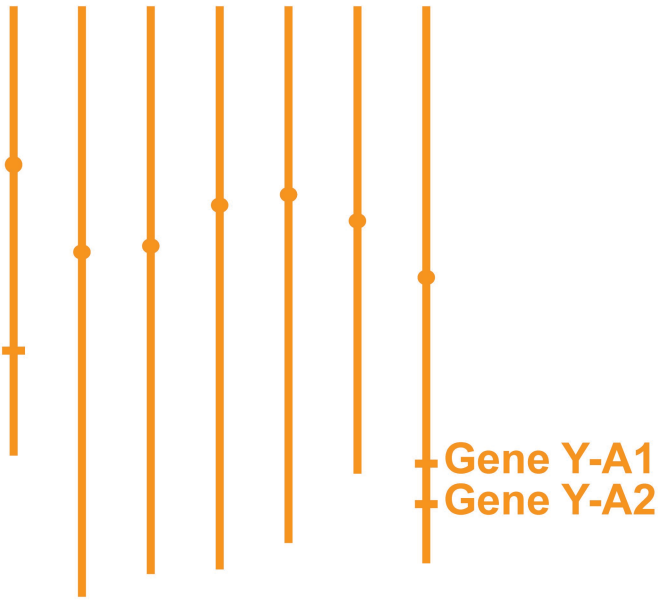
1135 **Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey M-D, Asyraf Md Hatta M, Hinchliffe A,**
 1136 **Steed A, Reynolds D, Adamski NM, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan**
 1137 **M, Edwards D, Batley J, Raman H, Carter J, Rogers C, Domoney C, Moore G, Harwood W, Nicholson P,**
 1138 **Dieters MJ, DeLacy IH, Zhou J, Uauy C, Boden SA, Park RF, Wulff BBH, Hickey LT.** 2018. Speed breeding is a
 1139 powerful tool to accelerate crop research and breeding. *Nat Plants* **4**, 23-29.

1140 **Wilkinson PA, Winfield MO, Barker GL, Tyrrell S, Bian X, Allen AM, Burrridge A, Coghill JA, Waterfall C,**
 1141 **Caccamo M, Davey RP, Edwards KJ.** 2016. CerealsDB 3.0: expansion of resources and data integration. *BMC*
 1142 *Bioinformatics* **17**, 256.
 1143 **Winfield MO, Allen AM, Burrridge AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J, Waterfall C,**
 1144 **Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley AR,**
 1145 **Edwards KJ.** 2016. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary
 1146 gene pool. *Plant Biotechnol J* **14**, 1195-1206.
 1147 **Wingen LU, Orford S, Goram R, Leverington-Waite M, Bilham L, Patsiou TS, Ambrose M, Dicks J, Griffiths**
 1148 **S.** 2014. Establishing the A. E. Watkins landrace cultivar collection as a resource for systematic gene
 1149 discovery in bread wheat. *Theor Appl Genet* **127**, 1831-1842.
 1150 **Wingen LU, West C, Leverington-Waite M, Collier S, Orford S, Goram R, Yang CY, King J, Allen AM,**
 1151 **Burrridge A, Edwards KJ, Griffiths S.** 2017. Wheat landrace genome diversity. *Genetics* **205**, 1657-1676.
 1152 **Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ.** 2007. An "electronic fluorescent
 1153 pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**.
 1154 **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-
 1155 cultivation agroinfiltration of (germinated) seeds results in tobacco rattle virus (TRV) mediated whole-plant
 1156 virus-induced gene silencing (VIGS) in wheat and maize. *Front Plant Sci* **8**.
 1157 **Zhang Y, Bai Y, Wu G, Zou S, Chen Y, Gao C, Tang D.** 2017b. Simultaneous modification of three
 1158 homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *Plant J* **91**, 714-
 1159 724.
 1160 **Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, Chen K, Qiu J-L, Gao C.** 2016. Efficient and transgene-free genome
 1161 editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat Commun* **7**, 12617.
 1162 **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**
 1163 **J.** 2017. The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat Plants* **3**, 946-955.
 1164 **Zhou Y, Chebotarov D, Kudrna D, Llaca V, Lee S, Rajasekar S, Mohammed N, Al-Bader N, Sobel-Sorenson**
 1165 **C, Parakkal P, Arbelaez LJ, Franco N, Alexandrov N, Hamilton NRS, Leung H, Mauleon R, Lorieux M,**
 1166 **Zuccolo A, McNally K, Zhang J, Wing RA.** 2019. Twelve Platinum-Standard Reference Genomes Sequences
 1167 (PSRefSeq) that complete the full range of genetic diversity of Asian rice. *bioRxiv*, 2019.2012.2029.888347.
 1168 **Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL.** 2017. The first near-complete assembly of the
 1169 hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience* **6**, 1-7.

1170

A-genome

Gene X-A1



Homoeologs:

Chr 1

Gene X-A1

Gene X-B1

Gene X-D1

Chr 7

Gene Y-A1

Gene Y-B1

Gene Y-D1

Paralogs:

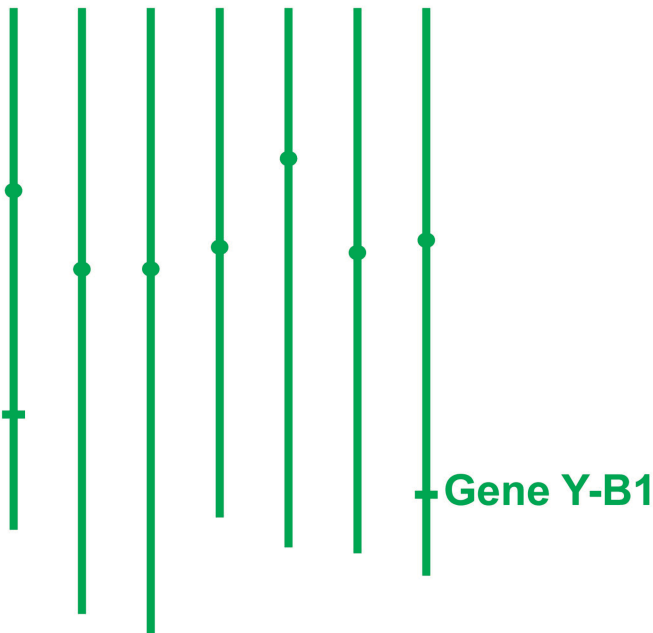
Chr 7

Gene Y-A1

Gene Y-A2

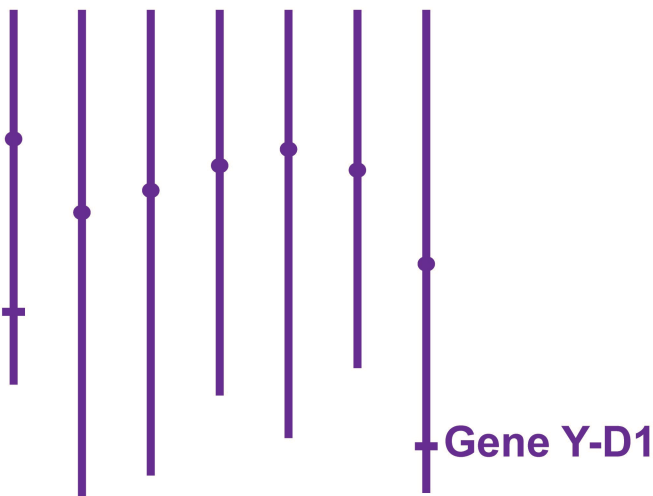
B-genome

Gene X-B1

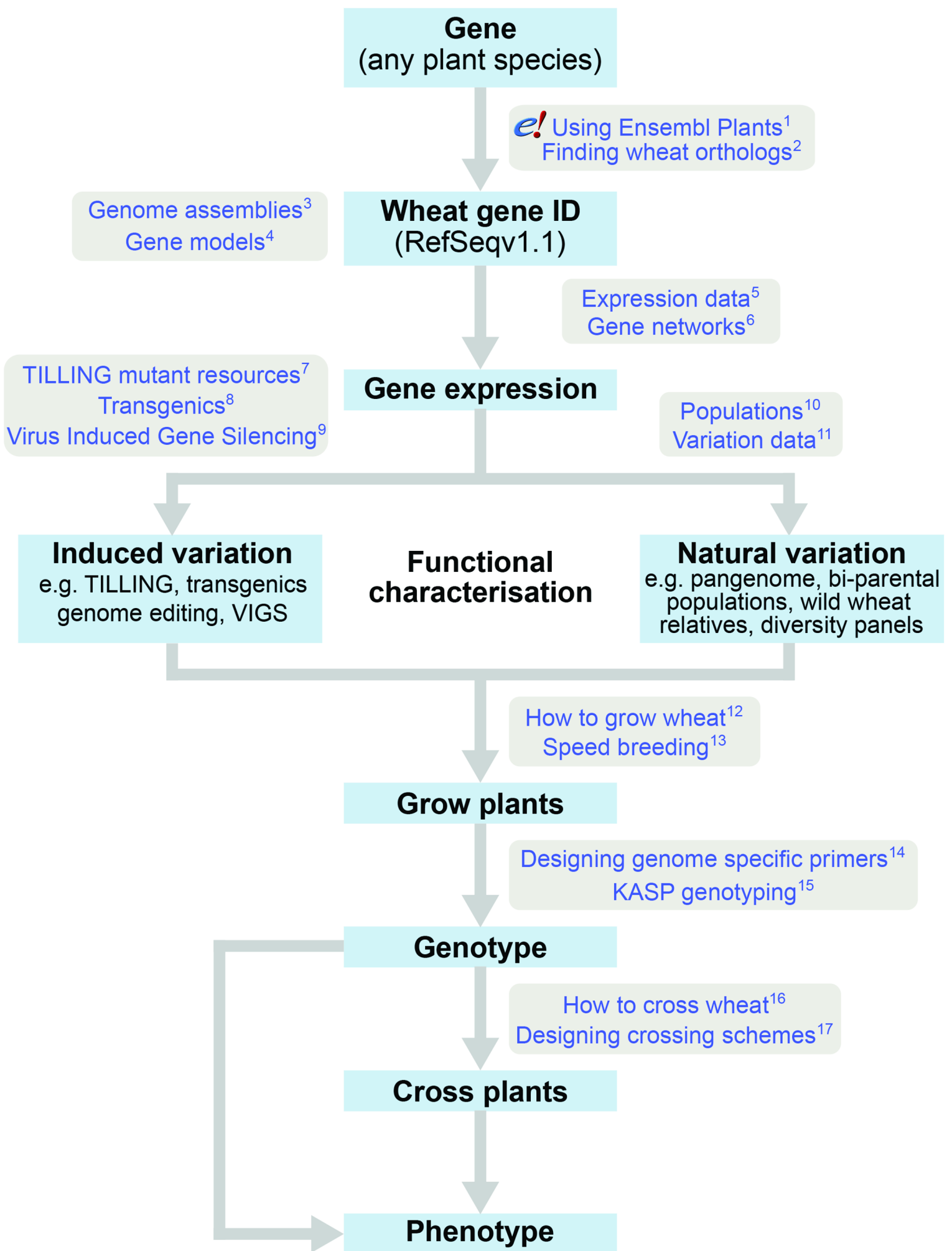


D-genome

Gene X-D1



Gene Y-D1



Fields:	Species <i>Triticum aestivum</i>	Accession Chinese Spring (2n = 6x = 42)	Chromosome location	Annotation ¹ version	Biotype ²	Unique identifier ³
---------	-------------------------------------	---	------------------------	------------------------------------	----------------------	--------------------------------

CSS

Traes_5BL_65826E1A1

Traes_ 5BL_ 65826E1A1

TGAC

TRIAE_CS42_5BL_TGACv1_404669_AA1307960

RefSeq v1.1

Traes CS 5B 02 G 236400

TraesCS5B02G236400

RefSeq v1.0

TraesCS5B01G236400

Svevo v1.0

TRITD 5Bv1G138200

Tetraploid

Hexaploid

Month

0

4

8

12

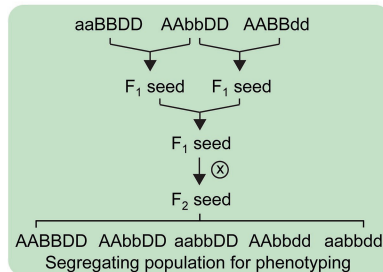
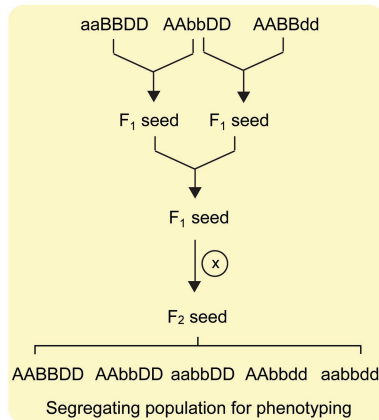
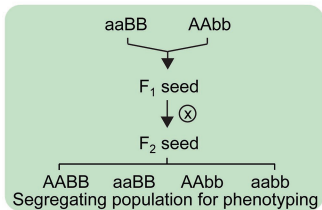
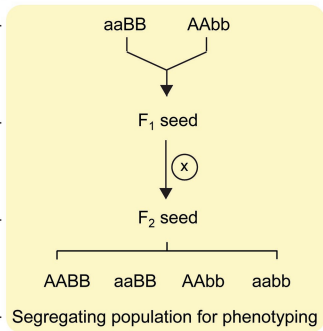
16

Conventional breeding

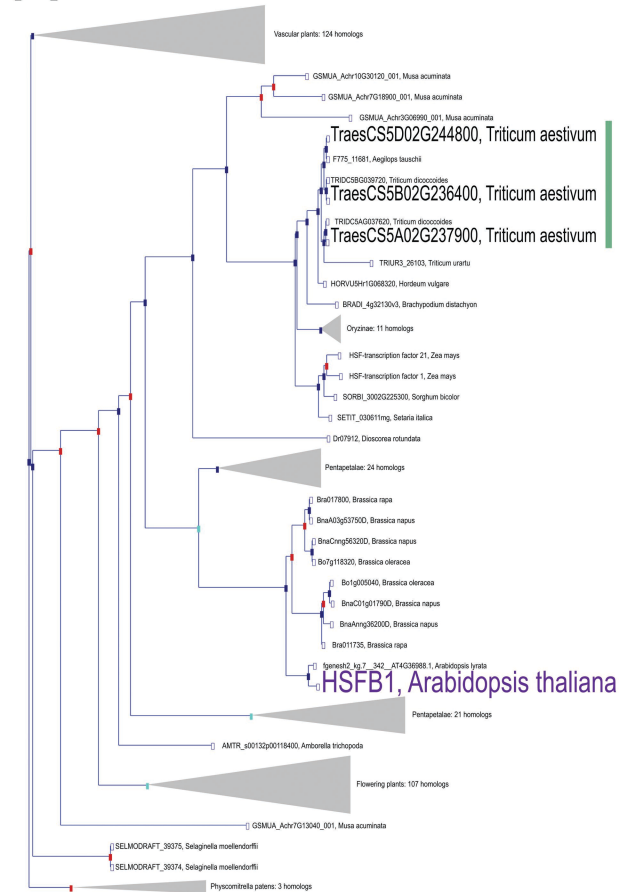
Speed breeding

Conventional breeding

Speed breeding



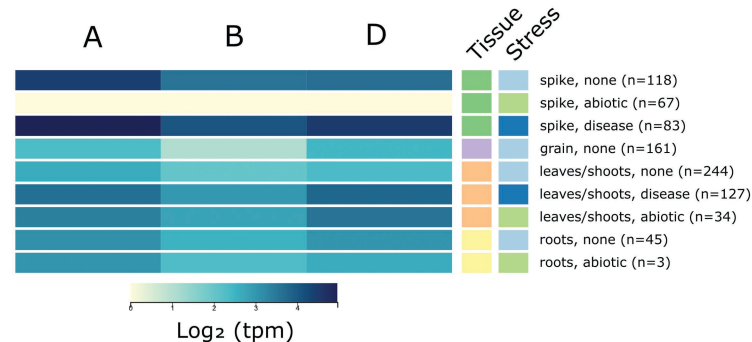
A



B

Show	All ▾ entries	Show/hide columns (11 hidden)	Filter	
Subject name	Length	Score	E-val	%ID
TraesCS5A02G237900.1	94 [Sequence]	470	5.8E-57	92.6 [Alignment]
TraesCS5B02G236400.1	101 [Sequence]	459	2.0E-55	85.1 [Alignment]
TraesCS5D02G244800.1	101 [Sequence]	457	3.7E-55	85.1 [Alignment]

C



D

K2593
A-genome
missense variant
SIFT = 0.0

K2900
B-genome
missense variant
SIFT = 0.0

K3227
A-genome
missense variant
SIFT = 0.0

K4057
B-genome
missense variant
SIFT = 0.0

