9 Genome-Informed Discovery of Genes and Framework of Functional Genes in Wheat

Awais Rasheed, Humaira Qayyum and Rudi Appels

Abstract

The complete reference genome of wheat was released in 2018 (IWGSC in Science 361:eaar7191, 2018), and since then many wheats genomic resources have been developed in a short period of time. These resources include resequencing of several hundred wheat varieties, exome capture from thousands of wheat germplasm lines, large-scale RNAseq studies, and complete genome sequences with de novo assemblies of 17 important cultivars. These genomic resources provide impetus for accelerated gene discovery and manipulation of genes for genetic improvement in wheat. The groundwork for this prospect includes the discovery of more than 200 genes using classical gene mapping techniques and comparative genomics approaches to explain moderate to major phenotypic variations in wheat. Similarly, QTL repositories are available in wheat which are frequently used by wheat genetics researchers and breeding communities for reference. The current wheat genome annotation is currently lagging in pinpointing the already discovered genes and QTL, and annotation of such information on the wheat genome sequence can significantly improve its value as a reference document to be used in wheat breeding. We aligned the currently discovered genes to the reference genome, provide their position and *TraesID*s, and present a framework to annotate such genes in future.

Keywords

Wheat genomics · Single nucleotide polymorphisms (SNPs) · KASP markers · Gene discovery · Functional markers · Gene networks

9.1 Introduction

Wheat holds a central position among major food crops by providing 20% of the total caloric requirements for the humans around the world. Common wheat (*Triticum aestivum* L.) is an allohexaploid ($2n = 6x = 42$; AABBDD) crop successfully cultivated all over the world covering an area of approximately 220 million ha. Genetic improvement in wheat productivity,

A. Rasheed (✉) · H. Qayyum
Quaid-i-Azam University, Islamabad 45320, Pakistan
e-mail: arasheed@qau.edu.pk

A. Rasheed
Chinese Academy of Agricultural Sciences (CAAS), and CIMMYT-China Office, Beijing, China

R. Appels
University of Melbourne, Food and Nutrition, Parkville, and AgriBio (Latrobe University), Bundoora, Melbourne, Australia

R. Appels et al. (eds.), *The Wheat Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-031-38294-9_9

resilience to climate extremes, and quality are challenges to be met in continuing to feed the global population, mitigate the effects of climate change, and fulfill the end user quality preferences. Since the expansion of wheat production area will not be possible due to the continuous shrinking of arable land, the increase in the grain yield by improved agronomic practices and breeding are feasible approaches. It has been recognized that conventional crop breeding approaches are not able to deliver the target of 70% increase in crop productivity by the end of 2050 (Tester and Langridge 2010). The innovation required in all breeding components includes selection accuracy, selection intensity, deploying new genetic variations, and shortening of the breeding cycles in developing cultivars (Li et al. 2018).

Conventional plant breeding heavily has relied on the selection of key phenotypes related to yield-related traits such as harvest index in wheat (Lopes et al. 2012), and it seems impossible to further improve harvest index using conventional breeding. Secondly, the phenotypic-based selections are labor intensive and time consuming, and off-spring can only be selected at the certain homozygous generation at the later growth stages. The concept of genomics-assisted breeding (GAB) was proposed as an alternate to overcome the selection challenges associated with conventional breeding (Varshney et al. 2005). The marker-assisted selection component dominated in the breeding programs where the diagnostic markers for the genes with major phenotypic effects were developed and successfully used for selection (Liu et al. 2012). However, many complex traits such as yield and adaptability to stressed environments are controlled by many genes with minor effects or quantitative trait loci (QTL), further interacting with environment (Gao et al. 2015). Their individual effects are too small to be efficiently captured by one or few markers (Bernardo and Yu 2007). Therefore, a transition from marker to genome-based breeding is indispensable to achieve the productivity targets (Rasheed and Xia 2019).

The next-generation sequencing (NGS) has revolutionized plant genomics and resulted in development of techniques and resources amenable to plant breeding (Bevan et al. 2017). The ever-growing plant genomic resources have provided plethora of SNP information distributed throughout the plant genomes, which have made them markers of choice for a variety of research applications, especially in breeding and genetics research. Until now, the reference genome sequences are available for most of the crop species, including wheat, while pan-genome sequences are increasing with the rapid pace (see Chap. 14). Characterization of the pan-genome can rapidly identify variations within the candidate genes, which have a direct application in breeding. In this chapter, we discuss different genome-informed scenarios being pursued to discover genes underpinning important phenotypes (Blake et al. 2016). We also provide a framework of functional genes of wheat in the context of the recent reference genome sequence assembly and discuss database resources necessary to reduce redundancy in research.

9.2 Wheat Reference Genome Sequence and Other Genomic Resources

9.2.1 The Reference Genome Sequence of cv. CHINESE SPRING

Wheat has a history in being used a model plant for understanding cytogenetics, physical mapping of genes, and to facilitate pre-breeding to introduce inter-specific and intergenic diversity. For example, the array of wheat aneuploid stocks, unequaled in any other crop, was developed by Sears (1954). All these genetic stocks were developed using wheat cv. CHINESE SPRING (Sears and Sears 1978). Such aneuploids include all the possible chromosome addition or deletion lines in the form of nullisomics, trisomics, monosomics, and tetrasomics. These cytogenetic stocks greatly facilitated

the genetic studies which were not possible in many of the higher organisms at that time. These stocks were used to identify major genes controlling important traits and physically map their positions along chromosomes, including the genes related to waxiness, maturity, endosperm proteins, and vernalization (Driscoll and Jensen 1964; Shepherd 1968; Halloran and Boydell 1967; Law 1966). Later, these efforts provided the basis for starting a 'Catalogue of Gene Symbols for Wheat' to catalogue wheat genes (McIntosh 1973). Since a wide array of genetic stocks were available in the CHINESE SPRING wheat background, this cultivar was selected to develop the first reference genome sequence in wheat. The International Wheat Genome Sequencing Consortium (IWGSC) was established in 2005, and after 13 years of its establishment, the high-quality reference sequence was released in 2018 (IWGSC 2018).

9.2.2 Other Genomic Resources in Wheat

All genome sequence resources available in wheat to date are provided in Table 9.1 and include population-level whole-genome resequencing, exome sequencing, and to lesser extent some SNP genotyping resources. The analysis of the CHINESE SPRING reference genome is now complemented by de novo sequences of ten important wheat cultivars from global breeding programs and has allowed the documentation of breeding histories, wild introgressions in the cultivated wheat, and chromosomal structural rearrangements that facilitated wheat breeding (Walkowiak et al. 2020); Jayakodi et al. 2021). Apart from the sequencing efforts in cultivated wheat, the genome sequences of diploid and tetraploid progenitors of bread wheat including *Ae. Tauschii* (Zhao et al. 2017), *T. monococcum* (Ling et al. 2018), *Ae. Speltoides* (Avni et al. 2022), and *T. dicoccoides* (Avni et al. 2017) are available. Recently, a population-level genome sequence resource of global *Ae. Tauschii* accessions was provided for use in trait discovery and functional genetic validation of D-genome introgressions in bread wheat (Gaurav et al. 2022). The shared utility of all such resources is underpinning the assignment of functional attributes to genes through association genetics or by selective sweeps. For example, 120 Chinese wheat cultivars and landraces were resequenced, and it was identified that the D-subgenome of modern cultivars is mostly derived from landraces, while A- and B-subgenomes were mainly derived from European landraces (Chen et al. 2019). Strong signals of selective sweeps were restricted to 48 high-confidence (HC) genes selected during modern wheat breeding. The strongest signals were for genes *TaNPF6.1-6B*, *TaNAC24*, and *TaRVE3*, which are associated with nitrogen use efficiency, drought and heat stress tolerance, and flowering time, respectively (Chen et al. 2019).

The exome capture of more than 500 global wheat accessions was conducted to identify the genes underpinning selection of adaptation of modern-day bread wheat during last 10,000 years (Pont et al. 2019). The authors concluded that dispersion of wheat and human migration patterns were consistent with an origin out of the Fertile Crescent and Egypt to Maghreb (Northern Africa) with a coastal route. The major driving forces in wheat adaptation were the vernalization requirement, historical groupings, and geographic origins (Europe, Asia, Africa, and America) and thus resulted in the partitioning of the genetic diversity in wheat. Furthermore, a total of 168 Mb of genome regions on different chromosomes contained selective sweeps which were identical between the Asian and European germplasm, even though European wheats had more frequent introgressions compared to wheats from Eastern Asia (He et al. 2019; Zhou et al. 2020), based on the resequencing of 890 bread and durum wheat accessions and the identification of introgressions from wild species favoring global wheat adaptation. Another globally important genomic resource is the DArTseq database of 44,624 wheat accessions from the International Maize

Table 9.1 Wheat genomic resources post-reference genome sequence

Resource	Number of accessions	Sequencing strategy	Objective	Reference
Pan-genome	10	WGS	Build a pan-genome of wheat	Walkowiak et al. (2020)
Chinese accessions	120	WGR	Identify the selection regions during wheat breeding	Chen et al. (2019)
Global landraces and cultivated wheats	4506	280 K SNP array	Wheat phylogeography and genetic diversity	Balfourier et al. (2019)
Global wheat accessions	500	Exome sequencing	Years of hybridization, selection, adaptation, and plant breeding has shaped the genetic makeup of modern bread wheats	Pont et al. (2019)
Hexaploid/tetraploid accessions	890	Exome sequencing	Identify the wild-relative introgressions favoring global wheat adaptation	He et al. (2019)
Chinese wheat accessions	770	DArTseq/660 K	Dispersion history, adaptive evolution, and selection of wheat in China	Zhou et al. (2018)
CIMMYT germplasm	44,624	DArTseq	Genomic predictabilities of 35 key traits and demonstrate the potential of genomic selection for wheat end-use quality	Juliana et al. (2019)
Ae. tauschii global collection	242	WGR	D-genome diversity for gene discovery	Gaurav et al. 2022
Chinese minicore collection	287	Exome sequencing	Identify genetic regions associated yield and adaptability	Li et al. (2022a)
Elite cultivars of China	145	WGR	Seventy years of breeder-driven selection	Hao et al. (2020)
25 wild wheat populations	414	WGR	Introgression from wild populations	Zhou et al. (2020)
Aegilops tauschii	278	WGR	Novel haplotypes with potential applications in wheat improvement	Zhou et al. (2021)

WGS: Whole-genome sequencing; WGR: Whole-genome resequencing

and Wheat Improvement Center (CIMMYT) GenBank (Juliana et al. 2019). The DArTseq data was used to conduct genome-wide association studies (GWAS) for 50 different traits of breeding interest and identified important loci for end-use quality, biotic, and abiotic stress resistances. These studies provide a deep insight into genetic diversity and genetic regions in wheat under artificial and natural selection and will keep proving important resources for use of such information in breeding.

9.3 Wheat Functional Genes Discovery: Strategies and Inventory

Quantitative trait loci (QTL) mapping and GWAS have dominated wheat genomics research to date. These studies identify the favorable alleles and their diagnostic markers which can be then used in wheat breeding to introgress important QTL or genes (Rasheed and Xia 2019). In Table 9.2, we provide a near-to-complete framework of the functional genes discovered so far by such approaches. However, such genetic dissection especially in case of GWAS can be ambiguous due to the confounding effects of population structure or low-accuracy genotype calls at some loci (Browning and Yu 2009), or due to the small population size (Finno et al. 2014). It is, therefore, necessary to further validate the phenotypic effects of such loci in biparental mapping populations or other genetic backgrounds, as well as by other biological means such as genetic transformation, gene silencing or gene knockout, and gene editing. The population-level whole-genome resequencing or exome capture data facilitated the discovery of several genes for economically important traits. From the resequencing data of 145 Chinese wheat accessions, Hao et al. (2020) identified that *TaFRK2-7A* gene contained three non-synonymous mutations compared to CS allele and was strongly associated with starch and amylose contents in mature seeds. The exome sequence of 287 wheat accessions identified the causal variations in *TaARF12* encoding an auxin response factor and *TaDEP1* encoding the G-protein γ-subunit, pleiotropically regulating both plant height and grain weight in wheat (Li et al. 2022a, b).

In recent years, several loci were identified simultaneously by GWAS and biparental mapping strategies. Liu et al. (2017) identified marker-trait association for black point resistance. Loci underpinning flour color (Zhai et al. 2016), kernel number per spike (Shi et al. 2017), and thousand grain weight (Sehgal et al. 2020; Wang et al. 2021) were also identified following a similar strategy. A functional gene, *TaRPP13L1* associated with flour color, was identified by GWAS in wheat cultivars from China and two KRONOS wheat mutants carrying premature stop codons of the *TaRPP13L1* gene and was thus validated as a gene influencing flour color (Chen et al. 2019).

Another gene discovery approach which is now widely used is bulk segregation analysis (BSA), where DNA from individuals of a population showing contrasting, extreme, and phenotypes is pooled and then RNAseq, exome sequencing, or whole-genome resequencing is applied (Zou et al. 2016). This is a rapid method to identify consistent polymorphic regions between contrasting pools of wheat lines. In addition to the discovery of SNPs between contrasting pools, differentially expressed genes can also be identified in the case of RNAseq analysis of tissues. Using this approach, a QTL interval with four candidate genes has been discovered on chr4A underpinning resistance against orange wheat blossom midge (OWBM) affecting wheat production in many countries (Hao et al. 2019). Likewise, resistance to yellow rust in wheat cultivar ZHOUMAI 22 was delimited to a physical interval of 4 Mb using BSA and RNAseq approach (Wang et al. 2017a). Other studies where this approach has been effective in discovering candidate genes include

Table 9.2 Framework of functional genes characterized in wheat with positions in wheat genome and associated traits

Gene	Chr	Phenotype	Crop ontology	Position	Traes ID
TaNAAT1-A	1A	GFe/GZn	CO_321:0000224	chr1A:487330367.0.487333932	TraesCS1A02G291100
TaNAAT2-A	1A	GFe/GZn	CO_321:0000224	chr1A:487463045.0.487466385	TraesCS1A02G291200
Glu-A1	1A	Gluten/end-use quality	CO_321:0000152	chr1A:508723999.0.508726319	TraesCS1A02G317311
Glu-A3	1A	Gluten/end-use quality	CO_321:0000155	chr1A:4202215.0.4203588	TraesCS1A02G008000
TaPYL1-1B	1B	Drought tolerance	CO_321:0000131	chr1B:373628259.0.373629490	TraesCS1B02G206600
TOE-B1	1B	Flowering time	CO_321:0000007	chr1B:59192897.0.59197677	TraesCS1B02G076300
ELF3-B1	1B	Flowering time	CO_321:0000007	chr1B:685645287.0.685649392	TraesCS1B02G477400
TaFT3-B1	1B	Flowering time	CO_321:0000007	chr1B:581413558.0.581414952	TraesCS1B02G351100
TaNAAT1-B	1B	GFe/GZn	CO_321:0000224	chr1B:520925847.0.520929216	TraesCS1B02G300500
TaNAAT2-B	1B	GFe/GZn	CO_321:0000224	chr1B:520998902.0.521002315	TraesCS1B02G300600
Glu-B1-717	1B	Gluten/End-use quality	CO_321:0000153	chr1B:555765127.0.555766152	TraesCS1B02G329711
Glu-B3	1B	Gluten/End-use quality	CO_321:0000156	chr1B:5686611.0.5687693	TraesCS1B02G011700
AGP-L-1B	1B	Grain morphology	CO_321:0000040	chr1B:668129122.0.668132472	TraesCS1B02G449700
Elf3-D1	1D	Flowering time	CO_321:0000007	chr1D:493484553.0.493488588	TraesCS1D02G451200
Mot-D1	1D	Flowering time	CO_321:0000007	chr1D:492606158.0.492620025	TraesCS1D02G450200
TaNAAT1-D	1D	GFe/GZn	CO_321:0000224	chr1D:387796590.0.387800918	TraesCS1D02G289700
TaNAAT2-D	1D	GFe/GZn	CO_321:0000224	chr1D:387894784.0.387898194	TraesCS1D02G289800
Glu-D1	1D	Gluten/end-use quality	CO_321:0000154	chr1D:412160786.0.412163311	TraesCS1D02G317211
ZDS-A1	2A	Flour color	CO_321:0000214	chr2A:321150418.0.321156866	TraesCS2A02G238400
Ppd-A1	2A	Flowering time	CO_321:0000007	chr2A:36933684.0.36938202	TraesCS2A02G081900
TaVIT1-2A	2A	GFe	CO_321:0000222	chr2A:570192811.0.570195203	TraesCS2A02G336600
TaNAS1-A	2A	GFe/GZn	CO_321:0000224	chr2A:14976663.0.14978691	TraesCS2A02G033500
TaNAS3-A	2A	GFe/GZn	CO_321:0000224	chr2A:19162944.0.19164224	TraesCS2A02G049900
TaNAS9-A	2A	GFe/GZn	CO_321:0000224	chr2A:49221108.0.49222130	TraesCS2A02G095700
Sus2-2A	2A	Grain morphology	CO_321:0000040	chr2A:121141338.0.121145857	TraesCS2A02G168200
TaCwi-A1	2A	Grain morphology	CO_321:0000040	chr2A:508030243.0.508033950	TraesCS2A02G295400
TaCYP78A5	2A	Grain morphology	CO_321:0000040	chr2A:134273284.0.134275604	TraesCS2A02G175700
WFZP-A1	2A	Grain number	CO_321:0000391	chr2A:66848645.0.66849948	TraesCS2A02G116900
TaGS2-A1	2A	NUE	CO_321:0001671	chr2A:729293649.0.729297303	TraesCS2A02G500400
TaARF12	2A	Plant height	CO_321:0000020	chr2A:755768802.0.755776624	TraesCS2A02G547800
PPO-A1	2A	PPO activity	CO_321:0000214	chr2A:712187112.0.712189567	TraesCS2A02G468200
Ppo2-A1	2A	PPO activity	CO_321:0000214	chr2A:712344578.0.712346518	TraesCS2A02G468500
RMD-A1	2A	Root growth angle		chr2A:142707925.0.142709726	TraesCS2A02G182900
TaRSL4	2A	Root length		chr2A:162291365.0.162292945	TraesCS2A02G194200
Sdr-A1	2A	Seed dormancy/PHS	CO_321:0000081	chr2A:158452418.0.158453410	TraesCS2A02G191400
Ppd-B1	2B	Flowering time	CO_321:0000007	chrUn:293689186.0.293692375	TraesCSU02G196100
TaVIT1-2B	2B	GFe	CO_321:0000222	chr2B:492146188.0.492148400	TraesCS2B02G345300
TaNAS1-B	2B	GFe/GZn	CO_321:0000224	chr2B:23548049.0.23551608	TraesCS2B02G047100
TaNAS3-B	2B	GFe/GZn	CO_321:0000224	chr2B:29118956.0.29120236	TraesCS2B02G060800
TaNAS9-B	2B	GFe/GZn	CO_321:0000224	chr2B:72895029.0.72896639	TraesCS2B02G111100
TaSus2-2B	2B	Grain morphology	CO_321:0000040	chr2B:171030429.0.171034964	TraesCS2B02G194200
Tabas1	2B	Grain morphology	CO_321:0000040	chr2B:448904796.0.448907800	TraesCS2B02G313700
GNI	2B	Grain number	CO_321:0000391	chr2B:573974813.0.573975706	TraesCS2B02G405700
TaDA1-B	2B	Grain size	CO_321:0000040	chr2B:4646554.0.4654607	TraesCS2B02G007700
TaGS2-B1	2B	NUE	CO_321:0001671	chr2B:722629776.0.722634436	TraesCS2B02G528300
PPO-B1	2B	PPO activity	CO_321:0000214	chr2B:688478142.0.688480649	TraesCS2B02G491000
Ppo2-B1	2B	PPO activity	CO_321:0000214	chr2B:689764554.0.689766587	TraesCS2B02G491400
TaVSR-B1	2B	Root depth		chr2B:89554121.0.89558883	TraesCS2B02G122400
RMD-B1	2B	Root growth angle		chr2B:191742224.0.191744048	TraesCS2B02G209500
TaRSL4	2B	Root length		chr2B:197210852.0.197212507	TraesCS2B02G212700
Sdr-B1	2B	Seed dormancy/PHS	CO_321:0000081	chr2B:200572827.0.200573807	TraesCS2B02G215300
ZDS-D1	2D	Flour color	CO_321:0000214	chr2D:234144711.0.234150925	TraesCS2D02G236500
Ppd-D1	2D	Flowering time	CO_321:0000007	chr2D:33952224.0.33955766	TraesCS2D02G079600

(continued)

Table 9.2 (continued)

Gene	Chr	Phenotype	Crop ontology	Position	Traes ID
TaVIT1-2D	2D	GFe	CO_321:0000222	chr2D:419781553.0.419783725	TraesCS2D02G326300
TaNAS1-D	2D	GFe/GZn	CO_321:0000224	chr2D:12870350.0.12873858	TraesCS2D02G033000
TaNAS3-D	2D	GFe/GZn	CO_321:0000224	chr2D:18168587.0.18170017	TraesCS2D02G049200
TaNAS9-D	2D	GFe/GZn	CO_321:0000224	chr2D:45799198.0.45800220	TraesCS2D02G094200
TaCYP78A5	2D	Grain morphology	CO_321:0000040	chr2B:181118653.0.181120839	TraesCS2B02G201900
WFZP-D1	2D	Grain number	CO_321:0000391	chr2D:67496011.0.67496898	TraesCS2D02G118200
TaDA1-D	2D	Grain size	CO_321:0000040	chr2D:8281359.0.8289277	TraesCS2D02G016900
TaGS2-D1	2D	NUE	CO_321:0001671	chr2D:595161545.0.595165983	TraesCS2D02G500600
Rht8	2D	Plant height	CO_321:0000020	chrUn:24893964.0.24897255	TraesCSU02G024900
PPO-D1	2D	PPO activity	CO_321:0000214	chr2D:572952347.0.572954307	TraesCS2D02G468200
Ppo2-D1	2D	PPO activity	CO_321:0000214	chr2D:573903210.0.573905141	TraesCS2D02G468600
RMD-D1	2D	Root growth angle		chr2D:134790880.0.134792691	TraesCS2D02G190700
TaRSL4	2D	Root length		chr2D:138754346.0.138756038	TraesCS2D02G193700
Lyce-A1	3A	End-use quality	CO_321:0000214	chr3A:370233784.0.370237786	TraesCS3A02G208800
TaGS5-A1	3A	Grain morphology	CO_321:0000040	chr3A:176555776.0.176559839	TraesCS3A02G212900LC
Pod-A1	3A	POD activity/quality	CO_321:0000214	chr3A:730397626.0.730398805	TraesCS3A02G510600
Tamyb10-A1	3A	Seed color/PHS	CO_321:0000037	chr3A:703905707.0.703905910	TraesCS3A02G631500LC
Phs1	3A	Seed dormancy/PHS	CO_321:0000081	chr3A:7294435.0.7297613	TraesCS3A02G006600
Lyce-B1	3B	End-use quality	CO_321:0000214	chr3B:377418979.0.377422751	TraesCS3B02G239100
Fhb1_His	3B	FHB resistance	CO_321:0000651	chr3B:8526628.0.8529572	TraesCS3B02G019900
TaNAS5-B	3B	GFe/GZn	CO_321:0000224	chr3B:40773361.0.40778748	TraesCS3B02G068500
Tamyb10-B1	3B	Seed color/PHS	CO_321:0000037	chr3B:757918298.0.757920082	TraesCS3B02G515900
Vp1B1	3B	Seed dormancy/PHS	CO_321:0000081	chr3B:693338001.0.693342761	TraesCS3B02G452200
COMT-3B	3B	WSC/drought	CO_321:0000131	chr3B:829391763.0.829392973	TraesCS3B02G612000
CKX-D1	3D	Grain morphology	CO_321:0000040	chr3D:106736525.0.106740667	TraesCS3D02G143500
Myb10-D1	3D	Seed color/PHS	CO_321:0000037	chr3D:570801243.0.570803210	TraesCS3D02G468400
ALPb-4A	4A	End-use quality	CO_321:0000070	chr4A:718033180.0.718034037	TraesCS4A02G453800
PRR73-A1	4A	Flowering time	CO_321:0000007	chr4A:119083489.0.119087436	TraesCS4A02G105300
TaDMAS1-A	4A	GFe/GZn	CO_321:0000224	chr4A:74150821.0.74153009	TraesCS4A02G074800
TaNAS6-A	4A	GFe/GZn	CO_321:0000224	Chr4A:148780629.0.148781781	TraesCS4A02G127900LC
TaCYP78A5	4A	Grain morphology	CO_321:0000040	chr2D:127258537.0.127260686	TraesCS2D02G183000
TaGS1-A1	4A	NUE	CO_321:0001671	chr4A:60668121.0.60671232	TraesCS4A02G063800
MOR1-A1	4A	Root length		chr4A:685380302.0.685381598	TraesCS4A02G415400
Lox-B1	4B	Flour color	CO_321:0000214	chr4B:27248262.0.27252524	TraesCS4B02G037700
PRR73-B1	4B	Flowering time	CO_321:0000007	chr4B:427491684.0.427496233	TraesCS4B02G198700
TaDMAS1-B	4B	GFe/GZn	CO_321:0000224	Chr4B:481847465.0.481849531	TraesCS4B02G400500LC
TaNAS6-B	4B	GFe/GZn	CO_321:0000224	Chr4B:402432887.0.402433879	TraesCS4B02G183900
TaGS1-B1	4B	NUE	CO_321:0001671	chr4B:499898695.0.499901767	TraesCS4B02G240900
Pds-B1	4B	PDS activity/quality	CO_321:0000214	chr4B:586575839.0.586580177	TraesCS4B02G300100
Rht-B1	4B	Plant height	CO_321:0000020	chr4B:30861382.0.30863247	TraesCS4B02G043100
TaERF73-D1	4B	Root depth		chr4D:467792044.0.467801204	TraesCS4D02G406100LC
MOR1-B1	4B	Root length		chr4B:605691920.0.605693239	TraesCS4B02G316200
TaDMAS1-D	4D	GFe/GZn	CO_321:0000224	chr4D:392726584.0.392728858	TraesCS4D02G232200
TaNAS6-D	4D	GFe/GZn	CO_321:0000224	chr4D:323095782.0.323098145	TraesCS4D02G184900
TaD14-4D	4D	Grain yield	CO_321:0000013	chr4D:428116830.0.428119151	TraesCS4D02G258000
TaGS1-D1	4D	NUE	CO_321:0001671	chr4D:403145655.0.403148815	TraesCS4D02G240700
Rht-D1	4D	Plant height	CO_321:0000020	chr4D:18781062.0.18782933	TraesCS4D02G040400
TaERF73-A1	4D	Root depth		chr4A:3351141.0.3352418	TraesCS4A02G003300LC
MOR1-D1	4D	Root length		chr4D:478997945.0.478999338	TraesCS4D02G312800
Lr67	4D	Rust resistance		chr4D:405770870.0.405775112	TraesCS4D02G243100
Dro1-A1	5A	Drought tolerance	CO_321:0000131	chr5A:428994186.0.428997632	TraesCS5A02G213300
Vrn-A1a	5A	Flowering time	CO_321:0000007	chr5A:587411824.0.587423240	TraesCS5A02G391700
TaNAS4-A	5A	GFe/GZn	CO_321:0000224	chr5A:705402044.0.705403372	TraesCS5A02G552400
TaDep1-A1	5A	Grain morphology	CO_321:0000040	chr5A:430486331.0.430493530	TraesCS5A02G215100

(continued)

Table 9.2 (continued)

Gene	Chr	Phenotype	Crop ontology	Position	Traes ID
TaGL3.3-5A	5A	Grain morphology	CO_321:0000979	chr5A:26440090.0.26449927	TraesCS5A02G030300
Egt2-A1	5A	Root growth angle		chr5A:151732800.0.151736140	TraesCS5A02G102000
Dro1-B1	5B	Drought tolerance	CO_321:0000131	chr5B:381041995.0.381044714	TraesCS5B02G210500
Vrn-B1b	5B	Flowering time	CO_321:0000007	chr5B:573803238.0.573815903	TraesCS5B02G396600
TaDep1-B1	5B	Grain morphology	CO_321:0000040	chr5B:378517204.0.378520796	TraesCS5B02G208700
TaGL3.3-5B	5B	Grain morphology	CO_321:0000979	chr5B:27830119.0.27840027	TraesCS5B02G029100
Egt2-B1	5B	Root growth angle		chr5B:304265954.0.304269177	TraesCS5B02G164200
Dro1-D1	5D	Drought tolerance	CO_321:0000131	chr5D:327631371.0.327634216	TraesCS5D02G218700
Vrn-D1	5D	Flowering time	CO_321:0000007	chr5D:467176608.0.467184463	TraesCS5D02G401500
TaDep1-D1	5D	Grain morphology	CO_321:0000040	chr5D:326126003.0.326129557	TraesCS5D02G216900
TaGL3.3-5D	5D	Grain morphology	CO_321:0000979	chr5D:37321983.0.37331860	TraesCS5D02G038500
Pina-D1	5D	Grain texture	CO_321:0000072	chr5D:3591495.0.3592002	TraesCS5D02G004100
Pinb-D1	5D	Grain texture	CO_321:0000072	chr5D:3609640.0.3610146	TraesCS5D02G004300
Egt2-D1	5D	Root growth angle		chr5D:131504758.0.131508027	TraesCS5D02G113600
TaNAS2-A	6A	GFe/GZn	CO_321:0000224	chr6A:158316641.0.158317931	TraesCS6A02G163100
TaNAS7-A2	6A	GFe/GZn	CO_321:0000224	chr6A:603249197.0.603250189	TraesCS6A02G386200
TaNAS7-A1	6A	GFe/GZn	CO_321:0000224	chr6A:60971892.0.60973259	TraesCS6A02G093000
TaGW2-6A	6A	Grain morphology	CO_321:0000980	chr6A:237734835.0.237759808	TraesCS6A02G189300
TaT6P	6A	Grain morphology	CO_321:0000040	chr6A:461145380.0.461147406	TraesCS6A02G248400
SPL21-6A	6A	Grain morphology	CO_321:0000040	chr6A:136541506.0.136544204	TraesCS6A02G152000
NAM-A1	6A	Grain protein	CO_321:0000073	chr6A:77098570.0.77100127	TraesCS6A02G108300
Kat-2A	6A	Grain weight	CO_321:0000025	chr6A:606969628.0.606973059	TraesCS6A02G392400
Rht-24	6A	Plant height	CO_321:0000020	chr6A:413732327.0.413735532	TraesCS6A02G221900
Rht24	6A	Plant height	CO_321:0000020	chr6A:432253559.0.432257969	TraesCS6A02G229500
TaNAS2-B	6B	GFe/GZn	CO_321:0000224	chr6B:212158654.0.212159706	TraesCS6B02G186000
TaNAS7-B	6B	GFe/GZn	CO_321:0000224	chr6B:694258986.0.694259978	TraesCS6B02G425200
SPL21-6B	6B	Grain morphology	CO_321:0000040	chr6B:200509075.0.200512019	TraesCS6B02G180300
GW2-6B	6B	Grain morphology	CO_321:0000040	chr6B:291761397.0.291778503	TraesCS6B02G215300
NAM-B1	6B	Grain protein	CO_321:0000073	chr6B:134662733.0.134665065	TraesCS6B02G207500LC
KAT-2B	6B	Grain weight	CO_321:0000025	chr6B:701871007.0.701874630	TraesCS6B01G432600
1fehw3	6B	WSC/Drought	CO_321:0000131	chr6B:57283367.0.57288151	TraesCS6B02G080700
TaNAS2-D2	6D	GFe/GZn	CO_321:0000224	chr6D:121579210.0.121580540	TraesCS6D02G148600
TaNAS2-D1	6D	GFe/GZn	CO_321:0000224	chr6D:121225536.0.121228339	TraesCS6D02G148200
TaNAS7-D	6D	GFe/GZn	CO_321:0000224	chr6D:456540490.0.456541773	TraesCS6D02G370800
SPL21-6D	6D	Grain morphology	CO_321:0000040	chr6D:111567638.0.111570051	TraesCS6D02G142100
TaGS1a	6D	Nitrogen use efficiency	CO_321:0001671	chr6D:386290812.0.386294394	TraesCS6D02G383600LC
Moc-A1	7A	Agronomic traits/drought	CO_321:0000131	chr7A:557553815.0.557555303	TraesCS7A02G382800
ALPa-7A	7A	End-use quality	CO_321:0000070	chr7A:15697493.0.15698020	TraesCS7A02G035500
ALPb-7A	7A	End-use quality	CO_321:0000070	chr7A:15639003.0.15639854	TraesCS7A02G035200
PSY-A1	7A	Flour color	CO_321:0000214	chr7A:729397558.0.729401208	TraesCS7A02G557300
TEF-7A	7A	Grain morphology	CO_321:0000040	chr7A:66228020.0.66229066	TraesCS7A02G108900
Sus1-7A1	7A	Grain morphology	CO_321:0000040	chr7A:115204109.0.115208145	TraesCS7A02G158900
TaGW7	7A	Grain morphology	CO_321:0000980	chr7A:205459137.0.205465028	TraesCS7A02G233600
SPL20-7A	7A	Grain morphology	CO_321:0000040	chr7A:685212680.0.685214713	TraesCS7A02G495000
AGP-S-7A	7A	Grain morphology	CO_321:0000040	chr7A:342609326.0.34261711	TraesCS7A02G287400
WAPO-A1	7A	Grain number	CO_321:0000391	chr7A:674081462.0.674082918	TraesCS7A02G481600
VRT-A2	7A	Grain number	CO_321:0000391	chr7A:128826237.0.128833021	TraesCS7A02G175200
FRK2-7A	7A	Starch synthesis/grain morphology	CO_321:0001674	chr7A:459209231.0.459211266	TraesCS7A02G319000
PSY-B1	7B	Flour color	CO_321:0000214	chr7B:739442503.0.739445446	TraesCS7B02G482000
TaSus1-7B	7B	Grain morphology	CO_321:0000040	chr7B:68344330.0.68348404	TraesCS7B02G063400
WAPO-B1	7B	Grain number	CO_321:0000391	chr7B:649950255.0.649951851	TraesCS7B02G384000
PIN-B2	7B	Grain texture	CO_321:0000072	chr7B:699388914.0.699389366	TraesCS7B02G431200
TaCOL-B5	7B	Grain yield	CO_321:0000013	chr7B:667070044.0.667071768	TraesCS7B02G400600

(continued)

Table 9.2 (continued)

Gene	Chr	Phenotype	Crop ontology	Position	Traes ID
PSY-D1	7D	Flour color	CO_321:0000214	chr7D:636766504.0.636770671	TraesCS7D02G553300
Vrn-D3	7D	Flowering time	CO_321:0000007	chr7D:68416507.0.68417532	TraesCS7D02G111600
GS3-D1	7D	Grain morphology	CO_321:0000040	chr7D:6483394.0.6485745	TraesCS7D02G015000
SPL20-7D	7D	Grain morphology	CO_321:0000040	chr7D:592816295.0.592819560	TraesCS7D02G482400
Lr34	7D	Rust resistance		chr7D:47412273.0.47424077	TraesCS7D02G080300
TaNAS4-D	UNK	GFe/GZn	CO_321:0000224	chrUn:108595828.0.108597155	TraesCSU02G125200
TaDA1-A	UNK	Grain size	CO_321:0000040	chrUn:11740231.0.11748045	TraesCSU02G007800
TaERF73-B1		Root depth		chr4B:585962983.0.585964402	TraesCS4B02G299500

nitrogen-dependent lesion mimic gene *Ndhrl1* (Li et al. 2016), powdery mildew resistance gene *Pm4b* (Wu et al. 2018), leaf senescence gene *els1* (Li et al. 2018), stripe rust resistance gene *Yr26* (Wu et al. 2018), *YrMM58, YrHY1* (Wang et al. 2018a, b), dwarfing gene *Rht12* (Sun et al. 2019), and *Pm61* (Hu et al. 2019). It is likely that this approach will get more attention because it replaces the genotyping of complete populations (Zou et al. 2016).

Very few genes in wheat have been discovered using the traditional map-based cloning approach, and most of the genes have been identified by comparative genomics between wheat and related grass species due to the high collinearity and genetic organization among grass genomes (Rasheed and Xia 2019; Chen et al. 2020). According to the recent literature search, almost 33 genes related to grain morphology have been isolated by homology-based cloning and functional markers have been developed for use in breeding (Table 9.1). Likewise, genes related to other morphological and phenological traits have been isolated including *TaPRR73* (Zhang et al. 2016) and *TaZIM-A1* (Liu et al. 2018) underpinning flowering time; *TaPPH-7A* (Wang et al. 2018a; b) underpinning morphological traits; *TaARF4* (Wang et al. 2019b) controlling root growth and plant height; and *TaSnRK2.9-5A* (Ur Rehman et al. 2019) controlling drought tolerance.

9.3.1 Functional Genomics and Map-based Cloning in Wheat

The continuous development of new genomic resources in wheat including new reference genomes, transcriptome resources, wheat TILLING mutants with exome sequencing data, and high-density SNP database are conduits for carrying out map-based cloning to discover new genes in wheat. A QTL for head length and spikelet number was identified and then fine mapped to an interval of 0.2 cM (Yao et al. 2019). The map-based cloning identified that *Head Length 2 (HL2)* is the designated gene controlling head length and spikelet number. Zhang et al. (2018) fine mapped a heading time gene, *TaHdm605,* in an EMS mutant line. Spike architecture is an important yield-related attribute, and three genes *TaTFL1-2D*, *TaHOX2-2B,* and *TaAGLG1-5A,* controlling spike architecture were discovered analyzing a large-scale transcriptome data of 90 wheat lines (Wang et al. 2017b). The effects of these genes were validated by the transgenic assays. Another approach used for discovery of gene was the screening of a yeast cDNA library constructed from a heat- and drought-tolerant wheat cv. HANXUAN 10. Using this approach, *TaPR-1-1*, for tolerance to abiotic stress tolerance, was identified which encodes the pathogenesis-related (PR) protein family (Wang et al. 2019a).

The development of male sterile lines is an important component of hybrid wheat breeding program. Two studies simultaneously cloned *Male Sterile 2* (*Ms2*) gene underpinning male sterility in wheat (Ni et al. 2017; Xia et al. 2016). The causal mutation was identified to be a terminal-repeat retrotransposon in miniature (TRIM) element in the promoter of *Ms2*. The TRIM element was involved in the gene activation and causes male sterility. Liu et al. (2019) cloned *TaSPL8* gene controlling leaf angle and is an important component of auxin and brassinosteroid pathways and associated with cell elongation. The knockout mutants of *TaSPL8* had erect leaves due to the loss of the lamina joint, compact architecture, and increased spike number. *Pm21* is a durable disease resistance gene derived from *Haynaldia villosa* confers resistance against powdery mildew, and currently wheat cultivars with *Pm21* are cultivated on 4 m ha in China (Cao et al. 2011). Two complementary studies cloned *Pm21* and identified that it encodes a typical CC-NBS-LRR protein involved in broad spectrum resistance to powdery mildew (He et al. 2019).

Fusarium head blight (FHB) is one of the most important yield and quality limiting factors in wheat globally. There are very few resources providing durable resistance to FHB in wheat including some landraces from China like SUMAI 3, which is known to carry *Fhb1* gene. Rawat et al. (2016) used multiple approaches including positional cloning, development of overexpression lines, and gene silencing to report that a pore-forming toxin-like (*PFT*) gene was the candidate for *Fhb1*. However, it was later found that several FHB susceptible cultivars also carry *PFT* and its candidacy was doubted. Two new studies further established that a histidine-rich calcium-binding (*TaHRC* or *His*) gene adjacent to *PFT* is the actual *Fhb1* and was identified as a susceptibility factor (Su et al. 2019). In contrast, Li et al. (2019) concluded that *Fhb1* is a gain-of-function gene and that the newly generated protein acts as a regulator of host immunity.

9.3.2 Functional Genes and Their Diagnostic Markers

All the above examples show the discovery of genes following different strategies and include various validation approaches. Once a gene is discovered and its phenotypic effect is validated, it becomes important to identify and select the favorable alleles of those genes in breeding using functional markers (FMs). FMs are referred to the PCR-based diagnostic markers designed to identify causal polymorphism underpinning phenotypic differences. FMs are routinely used in crop breeding programs to identify and select the desirable allelic variations of specific functional genes (Liu et al. 2012; Rasheed et al. 2017; Rasheed and Xia 2019; Rouse et al. 2019). As mentioned earlier, FMs due to their high diagnostic value are ideal markers for use in breeding to identify and pyramid different genes in marker-assisted recurrent selection. FMs are also used in genomic selection to improve selection accuracy. Rasheed et al. (2016) converted a collection of 72 FMs to kompetitive allele-specific PCR (KASP) formats for their use in high-throughput platforms. This effort currently now includes 157 KASP markers to diagnose alleles of traits of breeding interest. These KASP markers have been used by various breeding programs, and a recent estimate from citation indicated that currently more than 35 wheat breeding and genetic programs all over the world used these markers. For example, CIMMYT elite lines were tagged with *TaGS3-D1*, *TaTGW6,* and *TaSus1* genes using these KASP markers (Sehgal et al. 2019). Zhao et al. (2019) screened 1152 diverse global wheat germplasm lines with KASP markers of 47 functional genes underpinning a number of important traits of breeding interest (Zhao et al. 2019). Favorable alleles of more than 39 genes of breeding importance were also identified in East African wheat germplasm using the aforementioned KASP markers (Wamalwa et al. 2020).

Several commercial alternatives to the KASP master mix are now available which have made SNP genotyping more cost effective. Apart from these commercial alternates to the KASP technology, some open-source SNP genotyping methods are also available. Two examples are the development of semi-thermal asymmetric reverse PCR (STARP) (Long et al. 2017) and Amplifluor (Jatayev et al. 2017) methods which can be used with wide range of commercial master mix. Several SNP markers were converted to STARP format to further reducing the cost of genotyping (Wu et al. 2020).

9.4 Mining Gene Networks Using Database Resources

We have outlined many genome sequencing projects carried out to generate genome variation data in wheat populations (Table 9.1). The amount of genome sequencing data being generated in wheat can often hinder scientists from translating complex and sometimes contradictory information into biological understanding and discoveries. Apart from using the data to investigate the genetic diversity, population-level genomic variation data provides a valuable resources and great opportunities for identifying trait-related genes, designing markers, constructing gene trees, exploring the evolutionary history, and assisting the design of molecular breeding. Mining the relevant information from the extensive genome variation datasets is a time-consuming and error-prone process if the proper tools are not used to explore the genes in questions. New tools are indispensable to develop for explaining how genes and gene networks might be implicated in a complex trait or disease. Another limitation is that tapping large and complex genome variation datasets requires computational skills exceeding the abilities of the most crop breeders. In nutshell, the reuse of genomic variation data plays an important role in driving current plant science research. We have provided an overview of the various genome variation tools and resources for quick analysis of gene and gene networks (Table 9.3).

9.4.1 Gene–gene Synteny Using PRETZEL

In defining a genetic framework at the genome level, the reliance on similarity searches with transcripts and proteins is of primary importance, and in this context, features of genome structure such as sequence/gene repetition impact on the capacity to identify the correct gene for detailed analysis. Sequence alignments underpin all the studies. The capacity to visualize genome features such as uneven repetition between loci aligned between several genomes (Fig. 9.1) can anticipate complications when gene alignments are carried out without this prior knowledge.

PRETZEL (https://plantinformatics.io; Keeble-Gagnere et al. 2019) is an online, interactive, and real-time visualization tool for analyzing and integrating genetic and genomic datasets. In Fig. 9.1, the alignments of the fructosyltransferase genes at the fructan synthesis locus on 7AS for the wheat cv. LANCER, cv. CHINESE SPRING, and cv. MACE are shown as a complex example where the IWGSC 7A-LANCE 7A alignment of the array of GH32 genes is fully syntenic between gene models within the LACER and CS loci. In contrast, the IWGSC 7A-MACE 7A alignment is evidently ambiguous as a result of small genome rearrangements possibly due to assembly errors. The software PRETZEL enables any locus of interest to be analyzed and potential issues to be identified.

The variations in fructosyltransferases on chromosomes 7A, 4A, 7D, 6A, 6B, and 6D are candidate genes in QTL that characterize fructan content in wheat grain and thus relate to quality/nutritional attributes of the grain (Zhang et al 2008; Huynh et al 2012; Langridge and Fleury 2012). The component fructosyltransferases genes in the 4A and 7D loci showed good alignment across LANCE, CS, and MACE except for an inversion relative the CS in the MACE locus similar to that shown for the 7A locus (Fig. 9.1). The 6B and 6D loci carried the component fructosyltransferases genes, referred to as fructan

Table 9.3 Genomics database in wheat for genome-informed characterization of wheat genes

Name	URL	Description	Referece
GrainGene	https://wheat.pw.usda.gov/GG3/	A comprehensive resource for molecular and phenotypic information for Triticeae and Avena	Odell et al. (2017)
MASWheat	http://maswheat.ucdavis.edu/	Marker-assisted selection database for wheat	NA
expVIP	http://wheat-expression.com/	Wheat transcriptome resources for expression analysis	Borrill et al. (2016)
WheatExp	https://wheat.pw.usda.gov/WheatExp/	Homoeologue-specific database of gene expression profiles for polyploid wheat	Pearce et al. (2015)
Cerealsdb	http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php	Database for SNPs, genotyping arrays and sequences	NA
WheatIS	http://wheatis.org/	Wheat information system for wheat data, resources and bioinformatics tools	NA
OpenWildWheat	http://www.openwildwheat.org/	Sequencing resources of Ae. tauschii accessions	Gaurav et al. (2022)
IWGSC	http://www.wheatgenome.org/	Official website of IWGSC	NA
10+Wheat genomes	http://www.10wheatgenomes.com/	Wheat pan-genome resources	NA
Polymarker	http://polymarker.tgac.ac.uk/	SNP assay development tool	Ramirez-Gonzalez et al. (2015)
Triticeae tool box	https://triticeaetoolbox.org/wheat/	Repository of wheat data from wheat CAP	Blake et al. (2016)
Wheat Transcription factors	http://itak.feilab.net/	Database of wheat transcription factors	NA
TILLING	http://www.wheat-tilling.com/	Sequencing resource of CADENZA (6x) and KRONOS (4x) wheat TILLING population	Krasileva et al. (2017)
WGIN	http://www.wgin.org.uk/about.php	Wheat genetic improvement network	NA
URGI	http://wheat-urgi.versailles.inra.fr/	INRA-based resources for wheat sequence resources	NA
Gramene	http://www.gramene.org/	Open-source, integrated data resource for comparative functional genomics in crops and model plant species	Sun et al. (2022)
KnetMiner	http://knetminer.rothamsted.ac.uk/Triticum_aestivum/	Open-source software tools for integrating and visualizing large biological datasets	Hassani-Pak and Rawlings (2017)
Wheat SnpHub Portal	http://wheat.cau.edu.cn/Wheat_SnpHub_Portal/	A web interface to call variation data and map allele frequencies in global wheat populations based on exome capture and resequencing data	Wang et al. (2020)
Wheat Gmap	https://www.wheatgmap.org/	Bulk segregation analysis based on RNA or DNA sequencing data	Zhang et al. (2021)
WheatOmics	http://wheatomics.sdau.edu.cn/	Several wheat omics tools including blast, ID converter, sequence retriever, SNP marker	Ma et al. (2021)
WheatGene	http://wheatgene.agrinome.org	A Drupal-based interactive genome search database of wheat genomes and RNAseq	Garcia et al. (2021)

(continued)

Table 9.3 (continued)

Name	URL	Description	Referece
ggCOMP	http://wheat.cau.edu.cn/WheatCompDB/	A wheat resequencing database to enable unsupervised identification of pairwise germplasm resource-based identity by descent (gIBD) blocks	Yang et al. (2022)
ccnWHEAT	http://bioinformatics.cau.edu.cn/ccnWheat	A platform for searching and comparing specific functional co-expression networks, as well as identifying the related functions of the genes clustered therein	Li et al. (2022b)
TGT	http://wheat.cau.edu.cn/TGT/	A homology database, by integrating 12 Triticeae genomes and three outgroup model genomes and implemented versatile analysis and visualization functions	Chen et al. (2020)
Pretzel	https://plantinformatics.io/	An interactive, web-based environment for navigating multi-dimensional wheat datasets, including genetic maps and chromosome-scale physical assemblies	Keeble-Gagnere et al. (2019)
wheatQTL	http://wheatqtldb.net/	A QTL database of wheat	Singh et al. (2021)

1-exohydrolase (*1-FEH*) in Zhang et al (2008), and showed good alignment across LANCE, CS, and MACE. The 6A locus showed an inversion in MACE relative to CS and an absence of the locus in LANCER, consistent with the presence/absence polymorphism among wheat varieties for the 6A locus reported by Zhang et al. (2008).

In contrast to the locus carrying the fructosyltransferases, the wheat-*APO1* (*WAPO-A1*) locus on the long arm of 7A shows unambiguous alignments across the varieties examined (Fig. 9.2a, left-hand panel for entire chromosomes and right panel for the *WAPO1* locus region), and thus, the variation at the structural level that needs to be considered when gene functions are examined is not a significant factor. Interestingly, the h1 and h2 haplotypes at this locus (Fig. 9.2b) identified by Voss-Fels et al. (2018) using SNP variation in the genome sequence indicate striking sequence-level divergence in this *WAPO1* gene region that is not reflected at the gene–gene syntenic level shown in Fig. 9.2a.

The genome viewer in Fig. 9.2b is from DAWN (Watson-Haigh et al. 2018) and shows variation in SNP (colored drops) positions relative to the CHINESE SPRING refseq 2.1 as a reference and uses cv. LANCER and cv. MACE from the wheat 10Xgenome sequence dataset, and cv. XIAOYAN 54 and WESTONIA from Whole-Genome Shotgun (WGS) resequencing data (Watson-Haigh et al. 2018). In field trails, under rain-fed conditions, the SNP-based haplotype h2 was found to be significantly associated with increased grain yield compared to h1, conferring a 24% yield advantage relative to all other haplotypes, especially h1 which was the other prominent haplotype in the field trial (Voss-Fels et al. 2019).

PRETZEL aims to solve alignment problems and structural changes in cultivar sequences by providing an interactive, online environment for data visualization and analysis which, when loaded with appropriately curated data, can enable researchers with no bioinformatics training to exploit the latest genomic resources

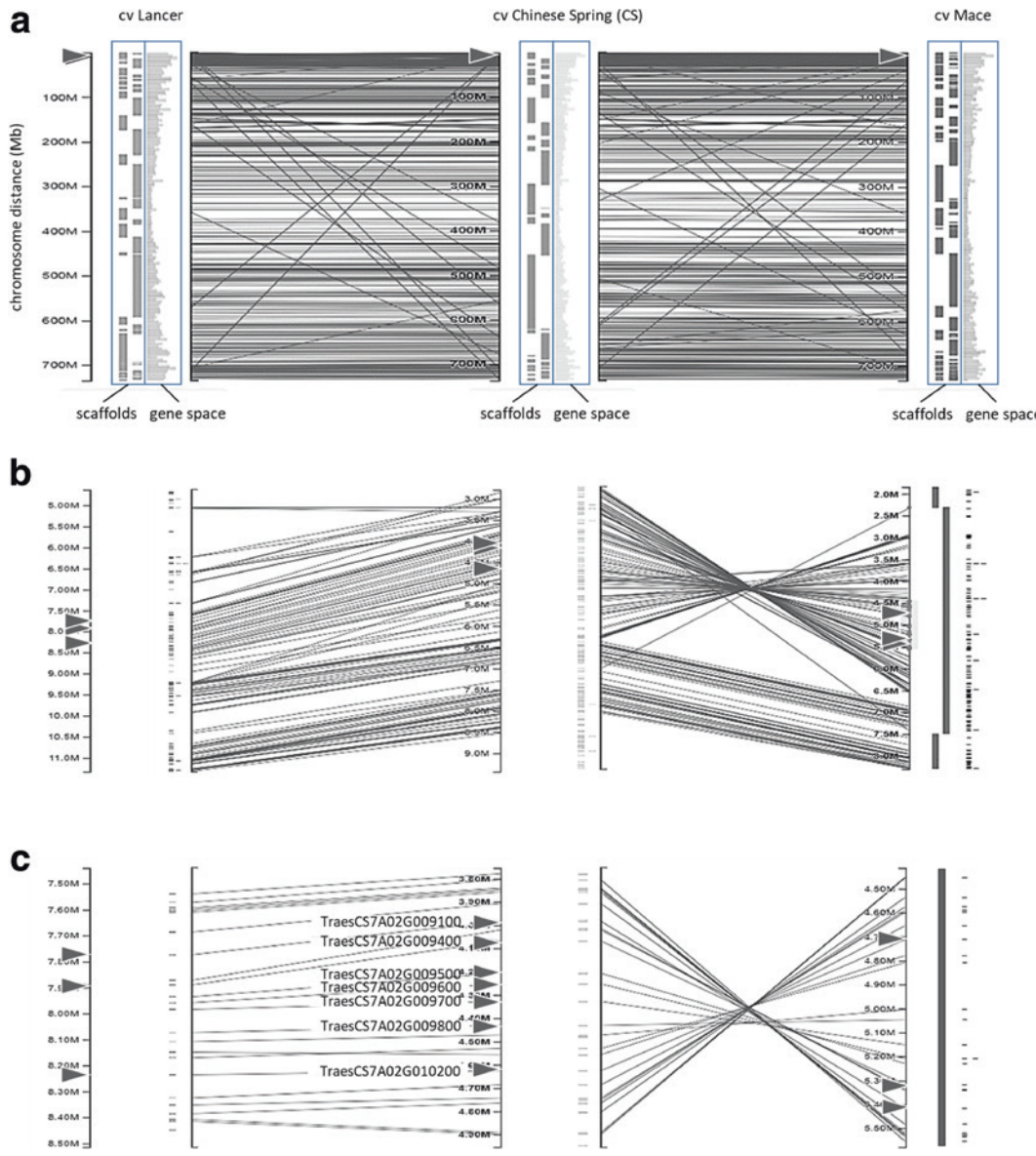

Fig. 9.1 Comparative analysis of 7AS fructan locus. In **a**, the arrows indicate the location of the locus within the entire chromosome, and **b** and **c** are the images resulting from ZOOMing into the locus. The marker genes *TraesCS7A02G009100*, *TraesCS7A02G009200* through *TraesCS7A02G010200* indicate the array of GH32 fructosyltransferases located at the locus in a ca 750 kb region (**c**). Scaffold columns to the right side of the PRETZEL maps are important for checking aberrations in colinearity (based on sequence similarity of 70% over 70% of the length of the sequence) as discussed in the text in terms of relating the boundaries of inverted regions to the boundaries of scaffolds in the assembly. In the region illustrated for MACE (**b**, **c**), the chance of the inverted region being an assembly error is reduced because the inversion is well within the respective scaffold

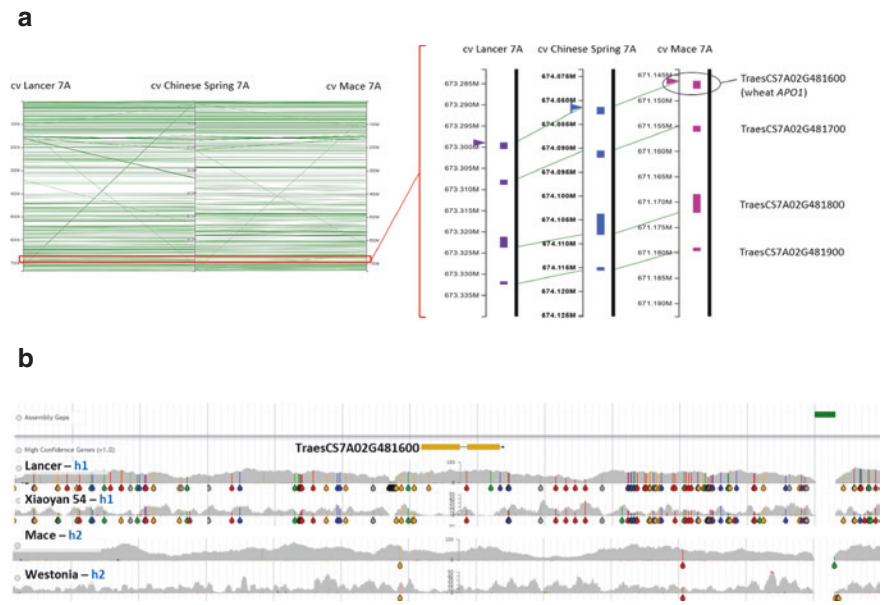

Fig. 9.2 **a** PRETZEL view of chr7A region (right panel) showing several genes including *WAPO-A1* (Voss-Fels et al. 2019; Kuzay et al. 2019, 2022) and structural changes in the *WAPO-A1* gene across three cvs. LANCER, CS, and MACE can be visualized with high-resolution (right panel). **b** is the genome viewer from DAWN (Watson-Haigh et al. 2018), and shows variation relative to the CHINESE SPRING refseq 2.1 as a reference and uses cv. LANCER and cv. MACE from the wheat 10Xgenome sequence dataset, and cv. XIAOYAN 54 and WESTONIA from Whole-Genome Shotgun (WGS) resequencing data

(Keeble-Gagnère et al. 2019). Apart from the visualization, PRETZEL can be used to retrieve the genome information (features including markers, genes, annotations, etc.) as dataset files of any selected chromosomal region for further downstream analysis.

9.4.2 Knowledge Graphs

Knowledge graphs (KG) are now extensively used to make search and information discovery more efficient. Knetminer is a data integration platform to visualize biological knowledge networks in an interactive web application (Hassani-Pak and Rawlings 2017). The data integration approach to build KGs has the ability to capture complex biological relationships between genes, traits, diseases, and many more information types derived from curated or predicted information sources. For example, *Rht24* is a new gene discovered associated with semi-dwarf phenotype in wheat and is present on chr6A. The Knetminer identified the gene network of *Rht24*, partially shown as Fig. 9.3 for clarity. The *Traes IDs* of both of the chr6B and chr6D homeologue are shown as interacting genes, and another gene, *TraesCS5B02G265400*, strongly interacts with *Rht24*. It can also be visualized that the gene interacts with bHLH27 transcription factor and physiologically influences the Gibberellin 20 pathway. Another feature is the identification of any stop/gain mutations in the CADENZA TILLING population, and mutant names and SNP positions can also be visualized.

The causal mutation of Rht24 on chr6A was identified in the exome capture data of the global hexaploid wheat collection (He et al. 2019). The target SNP was plotted for the frequency of wild-type and alternate SNP among global wheat accessions using SnpHub portal (Fig. 9.4).

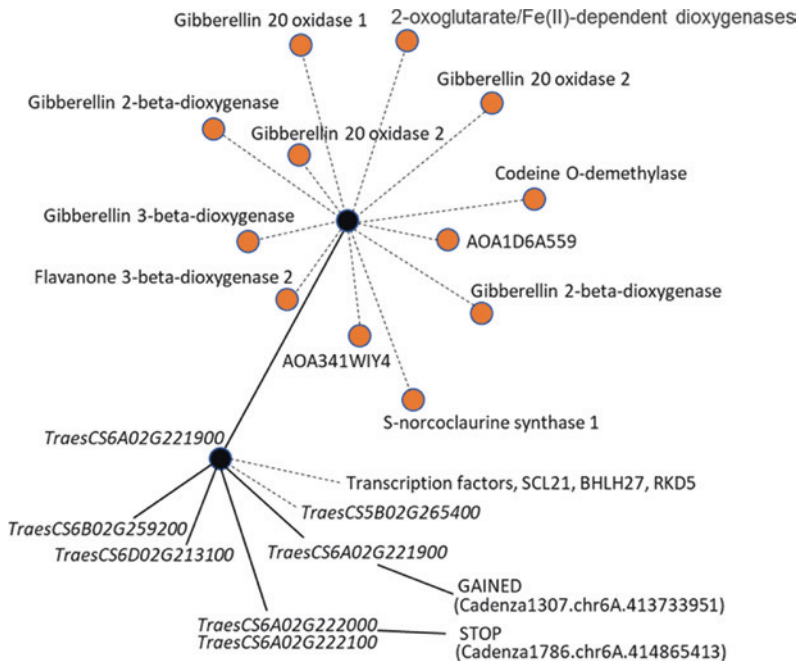

Fig. 9.3 KnetMiner network depicts connections with *Rht-24* on chr6A in wheat. This wheat reduced height gene, *Rht-24*, its homeologs on B- and D-genome along with other genes in cross-talk like TraesCS5B02G265400, associated transcription factors, and the mutations in the wheat TILLING population (e.g., two mutations in CADENZA TILLING population) can be visualized. Not all connections present in the KnetMiner network are depicted in the figure; only a subset is shown for clarity

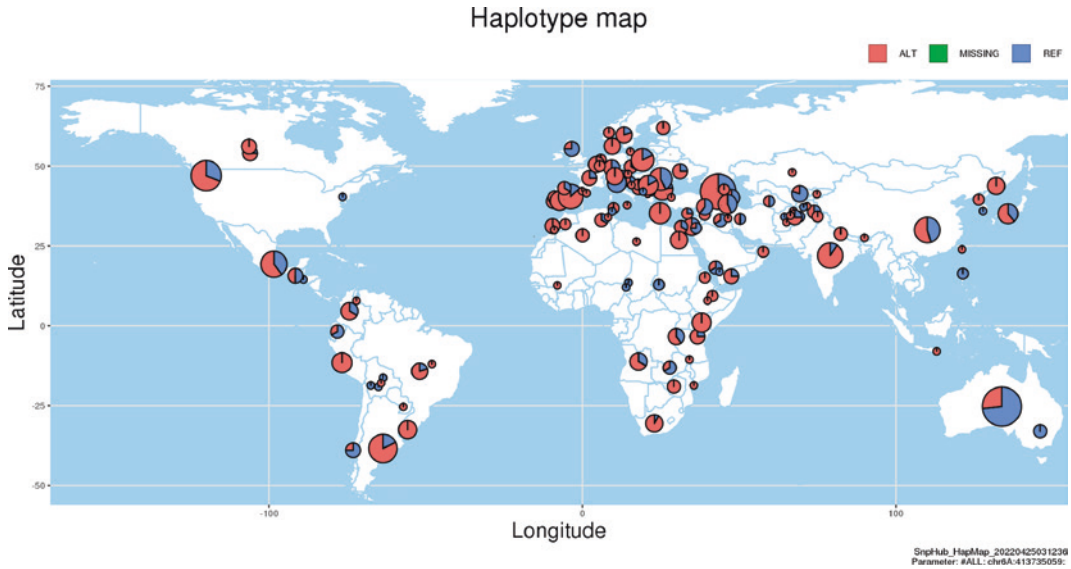

Fig. 9.4 SnpHub-based global haplotype map of non-synonymous mutation in Rht-24 is plotted based on the global exome sequencing data. In pie chart, the red proportion represents the frequency of wild-type mutation, while the blue proportion represents the frequency of non-synonymous mutation associated with reduced height

9.4.3 SnpHub Portal for Global Overview of Functional Gene Frequencies

SnpHub portal is a convenient way to identify mutations in the wheat genomes and then plotting the frequency of the SNPs country-wise in global what population (Wang et al. 2020). It is a Shing/R-based platform for mining and visualizing large genome variation data in wheat. Genome variation data in terms of .vcf files and genome annotation files can be accessed by a chromosomal interval of specific gene (*Traes ID*) to visualize genomic variation in heatmap, phylogenetic trees, haplotype networks, and haplotype geographic maps.

Apart from these platforms, several other platforms can be interactively used to mine useful genome variation and gene expression analysis (Table 9.3). The exVIP is an excellent resource for gene expression studies across various tissues and various experiments where the expression of certain genes can be visualized as heatmaps or as datafiles for further analysis. Similarly, WheatOmics (Ma et al. 2021) provides several features for analysis of genes including JBrowse with distinct track of several SNP genotyping and exome sequencing resources, *TraesID* converter, and sequence retriever. Last but not least, a wheat QTL database has been released recently which is an important resource to align QTL information with the IWGSC reference sequence (Singh et al. 2021).

9.5 Conclusion and Prospects

The complete annotation of functional genes in wheat is a challenge at multiple levels. For example, a first important intrinsic feature to impact annotation is the fragmentation level at the level of the number of exons per gene. As a CDS is fragmented into several exons, the difficulty to predict the correct intron/exon structure increases. In a detailed analysis of the wheat genome space by Choulet et al., (this volume, Chap. 4) it was emphasized that an important intrinsic feature of eukaryote gene structure that impacts on annotation is the fragmentation level at the level of the number of exons per gene. Choulet et al., (Chap. 4) noted that as a CDS is fragmented into several exons, the difficulty in predicting the correct intron/exon structure increases, although in wheat, (RefSeq Annotation v2.1) the average number of exons per CDS is only 4, and some genes (up to 10%) can have up to 17 exons. In this chapter, we have assigned genes and QTL to the reference genome and utilized available annotations to significantly improve the value of the outputs as reference documentation to be used in wheat breeding. The alignment of traits to annotated genes in the reference genome provides their position and *TraesIDs* to define a framework for establishing more informative markers for selecting lines to be deployed in crosses as well as for tracking targeted traits in segregating progeny from crosses.

Integration of a range of datasets has been emphasized in this chapter in order to deal with the complexity of the wheat genome and generating robust associations between genome haplotypes and agronomic traits for selecting parents for crossing and accurately tracking progeny from crosses. Since only 17% of genes are single copies, most key agronomic traits are likely to be the product of gene network interactions involving genes/gene families distributed across the chromosomes of the A-, B-, and D-subgenomes and genome signatures (haplotypes).

The sequencing data generated from cultivated and wild wheats, natural and breeding populations, and mutants is enabling the discovery of genes underpinning important traits of breeding interest. This information is useful to further develop and deploy the diagnostic markers for use in wheat breeding. The wheat genome variation is very complex for downstream analysis; therefore, the data analytics platforms have been developed to visualize genome variations and expression in heatmaps, haplotype and geographic maps, and gene

networks. We have provided an elucidated of gene frameworks discovered so far in wheat, and these need to be integrated with the thousands of QTL that have been discovered in wheat in different mapping populations and with many different marker platforms. The integration of wheat QTL information with genome visualization platforms for better understanding of gene networks and trait discovery is a key challenge.

Acknowledgements The authors are grateful for the perceptive comments provided by Dr Yuri Shavrukov (Flinders University, SthAustralia).

References

Avni R, Nave M, Barad O, Baruch K, Twardziok SO et al (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357:93–97

Avni R, Lux T, Minz-Dub A, Millet E, Sela H, Distelfeld A et al (2022) Genome sequences of three Aegilops species of the section Sitopsis reveal phylogenetic relationships and provide resources for wheat improvement. Plant J 110:179–192

Balfourier F, Bouchet S, Robert S, De Oliveira R, Rimbert H, Kitt J, Choulet F, Paux E (2019) Worldwide phylogeography and history of wheat genetic diversity. Sci Adv 5:eaav0536

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in Maize. Crop Sci 47:1082–1090

Bevan MW, Uauy C, Wulff BB, Zhou J, Krasileva K, Clark MD (2017) Genomic innovation for crop improvement. Nature 543:346–354

Blake VC, Birkett C, Matthews DE, Hane DL, Bradbury P, Jannink J-L (2016) The triticeae toolbox: combining phenotype and genotype data to advance small-grains breeding. Plant Genome 9:2. https://doi.org/10.3835/plantgenome2014.12.0099

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

Browning BL, Yu Z (2009) Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. Am J Hum Genet 85:847–861

Cao A, Xing L, Wang X, Yang X, Wang W, Sun Y, Qian C, Ni J, Chen Y, Liu D, Wang X, Chen P (2011) Serine/threonine kinase gene Stpk-V, a key member of powdery mildew resistance gene Pm21, confers powdery mildew resistance in wheat. Proc Natl Acad Sci U S A 108:7727–7732

Chen H, Jiao C, Wang Y, Wang Y, Tian C, Yu H, Wang J, Wang X, Lu F, Fu X, Xue Y, Jiang W, Ling H, Lu H, Jiao Y (2019) Comparative population genomics of bread wheat (Triticum aestivum) reveals its cultivation and breeding history in China. bioRxiv:519587

Chen Y, Song W, Xie X, Wang Z, Guan P, Peng H, Jiao Y, Ni Z, Sun Q, Guo W (2020) A collinearity-incorporating homology inference strategy for connecting emerging assemblies in the Triticeae tribe as a pilot practice in the plant pangenomic era. Mol Plant 13:1694–1708

Driscoll CJ, Jensen N (1964) Chromosomes associated with waxlessness, awnedness and time of maturity of common wheat. Can J Genet Cytol 6:324–333

Finno CJ, Aleman M, Higgins RJ, Madigan JE, Bannascha DL (2014) Risk of false positive genetic associations in complex traits with underlying population structure: a case study. Vet J 202:543–549

Gao F, Wen W, Liu J, Rasheed A, Yin G, Xia X, Wu X, He Z (2015) Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese Spring. Front Plant Sci 6:1099

Garcia DF, Wang Z, Guan J, Yin L, Geng S, Li A, Mao L (2021) WheatGene: a genomics database for common wheat and its related species. Crop J 9:1486–1491

Gaurav K, Arora S, Silva P, Sánchez-Martín J, Horsnell R et al (2022) Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. Nat Biotechnol 40:422–431

Halloran GM, Boydell CW (1967) Wheat chromosomes with genes for vernalization response. Can J Genet Cytol 9:632–639

Hao Z, Geng M, Hao Y, Zhang Y, Zhang L, Wen S, Wang R, Liu G (2019) Screening for differential expression of genes for resistance to Sitodiplosis mosellana in bread wheat via BSR-seq analysis. Theor Appl Genet 132:3201–3221

Hao C, Jiao C, Hou J, Li T, Liu H, Wang Y, Zheng J, Liu H, Bi Z, Xu F, Zhao J, Ma L, Wang Y, Majeed U, Liu X, Appels R, Maccaferri M, Tuberosa R, Lu H, Zhang X (2020) Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. Mol Plant 13:1733–1751

Hassani-Pak K, Rawlings C (2017) Knowledge discovery in biological databases for revealing candidate genes linked to complex phenotypes. J Integr Bioinform 14. https://doi.org/10.1515/jib-2016-0002

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

Hu J, Li J, Wu P, Li Y, Qiu D, Qu Y, Xie J, Zhang H, Yang L, Fu T, Yu Y, Li M, Liu H, Zhu T, Zhou Y, Liu

Z, Li H (2019) Development of SNP, KASP, and SSR markers by BSR-Seq technology for saturation of genetic linkage map and efficient detection of wheat powdery mildew resistance gene *Pm61*. Int J Mol Sci 20:750

Huynh B-L, Mather DE, Schreiber AW, Toubia J, Baumann U, Shoaei Z, Stein N, Ariyadasa R, Stangoulis JCR, Edwards J, Shirley N, Langridge P, Fleury D (2012) Clusters of genes encoding fructan biosynthesizing enzymes in wheat and barley. Plant Mol Biol 80:299–314

IWGSC (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:eaar7191

Jatayev S, Kurishbayev A, Zotova L et al (2017) Advantages of Amplifluor-like SNP markers over KASP in plant genotyping. BMC Plant Biol 17:254

Jayakodi M, Schreiber M, Stein N, Mascher M (2021) Building pan-genome infrastructures for crop plants and their use in association genetics. DNA Res 28:dsaa030

Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, Toledo FH, Govindan V, Mondal S, Kumar U, Bhavani S, Singh PK, Randhawa MS, He X, Guzman C, Dreisigacker S, Rouse MN, Jin Y, Pérez-Rodríguez P, Montesinos-López OA, Singh D, Mokhlesur Rahman M, Marza F, Singh RP (2019) Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. Nature 51:1530–1539

Keeble-Gagnère G, Isdale D, Suchecki R, Kruger A, Lomas K, Carroll D, Li S, Whan A, Hayden M, Tibbits J (2019) bioRxiv. https://doi.org/10.1101/517953

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

Kuzay S, Xu Y, Zhang J, Katz A, Pearce S, Su Z, Fraser M, Anderson JA, Brown-Guedira G, DeWitt N, Haugrud AP, Faris JD, Akhunov E, Bai G, Dubcovsky J (2019) Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. Theor Appl Genetics 132:2689–2705

Kuzay S, Lin H, Li C, Chen S, Woods DP, Zhang J, Lan T, von Korff M, Dubcovsky J (2022) WAPO-A1 is the causal gene of the 7AL QTL for spikelet number per spike in wheat. PLoS Genet 18(1):e1009747

Langridge P, Fleury D (2012) Clusters of genes encoding fructan biosynthesizing enzymes in wheat and barley. Plant Mol Biol 80:299–314

Law CN (1966) The location of genetic factors affecting a quantitative character in wheat. Genetics 53:487–498

Li L, Shi X, Zheng F et al (2016) A novel nitrogen-dependent gene associates with the lesion mimic trait in wheat. Theor Appl Genet 129:2075–2084. https://doi.org/10.1007/s00122-016-2758-3

Li H, Rasheed A, Hickey L, He Z (2018) Fast-forwarding genetic gain. Trends Plant Sci 23:184–186

Li G, Zhou J, Jia H et al (2019) Mutation of a histidine-rich calcium-binding-protein gene in wheat confers resistance to Fusarium head blight. Nat Genet 51:1106–1112. https://doi.org/10.1038/s41588-019-0426-7

Li A, Hao C, Wang Z, Geng S, Jia M, Wang F, Han X, Kong X, Yin L, Tao S, Deng Z, Liao R, Sun G, Wang K, Ye X, Jiao C, Lu H, Zhou Y, Liu D, Fu X, Zhang X, Mao L (2022a) Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. Mol Plant 15:504–519

Li Z, Hu Y, Ma X, Da L, She J, Liu Y, Yi X, Cao Y, Xu W, Jiao Y, Su Z (2022b) ccnWheat: A database for comparing co-expression networks analysis of allohexaploid wheat and its progenitors. bioRxiv:2022b.2001.2017.476536

Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, Yu Y, Du H, Qi M, Li Y, Lu H, Yu H, Cui Y, Wang N, Chen C, Wu H, Zhao Y, Zhang J, Li Y, Zhou W, Zhang B, Hu W, van Eijk MJT, Tang J, Witsenboer HMA, Zhao S, Li Z, Zhang A, Wang D, Liang C (2018) Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. Nature 557:424–428

Liu YN, He ZH, Appels R, Xia XC (2012) Functional markers in wheat: current status and future prospects. Theor Appl Genet 125:1–10

Liu J, He Z, Rasheed A et al (2017) Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). BMC Plant Biol 17:220 https://doi.org/10.1186/s12870-017-1167-3

Liu H, Li T, Wang Y, Zheng J, Li H, Hao C, Zhang Z (2018) TaZIM-A1 negatively regulates flowering time in common wheat (Triticum aestivum L.). J Integr Plant Biol 61:359–376. https://doi.org/10.1111/jipb.12720

Liu K, Cao J, Yu K, Liu X, Gao Y, Chen Q, Zhang W, Peng H, Du J, Xin M, Hu Z, Guo W, Rossi V, Ni Z, Sun Q, Yao Y (2019) Wheat TaSPL8 modulates leaf angle through auxin and brassinosteroid signaling. Plant Physiol 18:179–194. https://doi.org/10.1104/pp.19.00248

Long YM, Chao WS, Ma GJ et al (2017) (2017) An innovative SNP genotyping method adapting to multiple platforms and throughputs. Theor Appl Genet 130:597–607. https://doi.org/10.1007/s00122-016-2838-4

Lopes MS, Reynolds MP, Manes Y, Singh RP, Crossa J, Braun HJ (2012) Genetic yield gains and changes in associated traits of CIMMYT spring bread wheat in a "Historic" set representing 30 years of breeding. Crop Sci 52:1123–1131

Ma S, Wang M, Wu J, Guo W, Chen Y, Li G, Wang Y, Shi W, Xia G, Fu D, Kang Z, Ni F (2021) WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat. Mol Plant 14:1965–1968

McIntosh R (1973) A catalogue of gene symbols for wheat. In: Missouri C (ed) Proceedings of 4th International Wheat Genetics Symposium

Ni F, Qi J, Hao Q et al (2017) Wheat Ms2 encodes for an orphan protein that confers male sterility in grass species. Nat Commun 8. https://doi.org/10.1038/ncomms15121

Odell SG, Lazo GR, Woodhouse MR, Hane DL, Sen TZ (2017) The art of curation at a biological database: principles and application. Curr Plant Biol 11–12:2–11

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

Pont C, Leroy T, Seidel M, Tondelli A, Duchemin W et al (2019) Tracing the ancestry of modern bread wheats. Nat Genet 51:905–911

Ramirez-Gonzalez RH, Uauy C, Caccamo M (2015) PolyMarker: a fast polyploid primer design pipeline. Bioinformatics 31:2038–2039

Rasheed A, Xia X (2019) From markers to genome-based breeding in wheat. Theor Appl Genet 132:767–784

Rasheed A, Wen W, Gao F, Zhai S, Jin H, Liu J, Guo Q, Zhang Y, Dreisigacker S, Xia X, He Z (2016) Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. Theor Appl Genet 2016:1843–1860

Rasheed A, Hao Y, Xia XC, Khan A, Xu Y, Varshney RK, He ZH (2017) Crop breeding chips and genotyping platforms: progress, challenges and perspectives. Mol Plant 10:1047–1064

Rawat N, Pumphrey M, Liu S et al (2016) Wheat Fhb1 encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. Nat Genet 48:1576–1580. https://doi.org/10.1038/ng.3706

Rouse MN, Jin Y, Pérez-Rodríguez P, Montesinos-López OA, Singh D, Mokhlesur Rahman M, Marza F, Singh RP (2019) Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. Nature 51:1530–1539

Sears ER (1954) The aneuploids of common wheat. University of Missouri, College of Agriculture, Agricultural Experiment Station

Sears ER, Sears LMS (1978) The telocentric chromosomes of common wheat. In: 5th international wheat genetics symposium, pp 389–407

Sehgal D, Mondal S, Guzman C, Garcia Barrios G, Franco C, Singh R, Dreisigacker S (2019) Validation of candidate gene-based markers and identification of novel loci for thousand-grain weight in spring bread wheat. Front Plant Sci 10:1189

Sehgal D, Mondal S, Crespo-Herrera, Velu G, Juliana P, Huerta-Espino J, Shrestha S, Poland J, Singh R, Dreisigacker S (2020) Haplotype-based, genome-wide association study reveals stable genomic regions for grain yield in CIMMYT spring bread wheat. Front Genet 11. https://doi.org/10.3389/fgene.2020.589490

Shepherd K (1968) Chromosomal control of endosperm proteins in wheat and rye. In: Proceedings of 3rd International Wheat Genetics Symposium Australian Acadamic Science, pp 86–96

Shi W, Hao C, Zhang Y,Cheng J, Zhang Z, Liu J,Yi X, Cheng X, Sun D, Xu, Zhang X, Cheng S, Guo P, Guo J (2017) A combined association mapping and linkage analysis of kernel number per spike in common wheat (Triticum aestivum L.). Front. Plant Sci. Sec. Plant Breed 18. https://doi.org/10.3389/fpls.2017.01412

Singh K, Batra R, Sharma S et al (2021) WheatQTLdb: a QTL database for wheat. Mol Genet Genomics 296:1051–1056. https://doi.org/10.1007/s00438-021-01796-9

Su Z, Bernardo A, Tian B et al (2019) A deletion mutation in TaHRC confers Fhb1 resistance to Fusarium head blight in wheat. Nat Genet 51:1099–1105. https://doi.org/10.1038/s41588-019-0425-8

Sun J, Zhan K, Chu J, Zhang A (2018) A wheat dominant dwarfing line with Rht12, which reduces stem cell length and affects gibberellic acid synthesis, is a 5AL terminal deletion line (2019). Plant J 97:887–900. https://doi.org/10.1111/tpj.14168

Sun L, Yang W, Li Y, Shan Q, Ye X, Wang D, Yu K, Lu W, Xin P, Zhong X, Pei Z, Guo X, Liu D, Tello-Ruiz MK, Jaiswal P, Ware D (2022) Gramene: a resource for comparative analysis of plants genomes and pathways. In: Edwards D (ed) Plant bioinformatics: methods and protocols. Springer, US, New York, NY, pp 101–131

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327:818–822

Ur Rehman S, Wang J, Chang X et al (2019) A wheat protein kinase gene TaSnRK2.9-5A associated with yield contributing traits. Theor Appl Genet 132:907–919

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Voss-Fels KP, Robinson H, Mudge SR, Richard C, Newman S, Wittkop B, Stahl A, Friedt W, Frisch M, Gabur I, Miller-Cooper A (2018) *VERNALIZATION1* modulates root system architecture in wheat and barley. Mol Plant 11(1):226–229

Voss-Fels KP, Keeble-Gagnère G, Hickey LT, Tibbits J, Nagornyy S, Hayden MJ, Pasam RK, Kant S, Friedt W, Snowdon RJ, Appels R, Wittkop B (2019) High-resolution mapping of rachis nodes per rachis, a critical determinant of grain yield components in wheat. Theor Appl Genet 132:2707–2719

Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J et al (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature 588:277–283

Wamalwa M, Tadesse Z, Muthui L et al (2020) Allelic diversity study of functional genes in East Africa bread wheat highlights opportunities for genetic improvement. Mol Breed 40:104. https://doi.org/10.1007/s11032-020-01185-x

Wang Y, Xie, Zhang H et al (2017a) Mapping stripe rust resistance gene YrZH22 in Chinese wheat cultivar Zhoumai 22 by bulked segregant RNA-Seq (BSR-Seq) and comparative genomics analyses. Theor Appl Genet 130:2191–2201. https://doi.org/10.1007/s00122-017-2950-0

Wang Y, Yu H, Tian C, Sajjad M, Gao C, Tong Y Wang X, Jiao Y (2017b) Transcriptome association identifies regulators of wheat spike architecture. Plant Physiol 175:746–757. https://doi.org/10.1104/pp.17.00694

Wang H, Wang S, Chang X et al (2018a) Identification of TaPPH-7A haplotypes and development of a molecular marker associated with important agronomic traits in common wheat. BMC Plant Biol 19:296. https://doi.org/10.1186/s12870-019-1901-0

Wang Y, Zhang H, Xie J, Guo B, Chen Y, Zhang H, Lu P, Wu Q, Li M, Zhang D, Guo G, Yang J, Zhang P, Zhang Y, Wang X, Zhao H, Cao T, Liu Z (2018b) Mapping stripe rust resistance genes by BSR-Seq: YrMM58 and YrHY1 on chromosome 2AS in Chinese wheat lines Mengmai 58 and Huaiyang 1 are Yr17. Crop J 6:91–98

Wang J, Mao X, Wang R, Jing R et al (2019a) Identification of wheat stress-responding genes and TaPR-1-1 function by screening a cDNA yeast library prepared following abiotic stress. Sci Rep 9. https://doi.org/10.1038/s41598-018-37859-y

Wang J, Wang R, Mao X, Li L Chang X, Zhang X, Jing R (2019b) TaARF4 genes are linked to root growth and plant height in wheat. Ann Bot 124:903–915. https://doi.org/10.1093/aob/mcy218

Wang W, Wang Z, Li X, Ni Z, Hu Z, Xin M, Peng H, Yao Y, Sun Q, Guo W (2020) SnpHub: an easy-to-set-up web server framework for exploring large-scale genomic variation data in the post-genomic era with applications in wheat. GigaScience 9:giaa060

Wang X et al (2021) Genome-wide association study identifies QTL for thousand grain weight in winter wheat under normal- and late-sown stressed environments. Theor Appl Genet 134:143

Watson-Haigh NS, Suchecki R, Kalashyan E et al (2018) DAWN: a resource for yielding insights into the diversity among wheat genomes. BMC Genomics 19:941. https://doi.org/10.1186/s12864-018-5228-2

Wu P, Xie J, Hu J, Qiu D, Liu Z, Li J, Li MM, Zhang H, Yang L, Zhou Y, Zhang Z, Li H (2018) Development of molecular markers linked to powdery mildew resistance gene Pm4b by combining SNP discovery from transcriptome sequencing data with bulked segregant analysis (BSR-Seq) in wheat. Front Plant Sci 9. https://doi.org/10.3389/fpls.2018.00095

Wu Y, Li M, He Z, Dreisigacker S, Wen W, Jin H, Zhai S, Li F, Gao F, Liu J, Wang R (2020) Development and validation of high-throughput and low-cost STARP assays for genes underpinning economically important traits in wheat. Theor Appl Genet 133:2431–2450

Xia C, Zhang L, Zou C, Gu Y, Duan J, Zhao G, Wu J, Liu Y, Fang X, Gao L, Jiao Y, Sun J, Pan Y, Liu X, Jia J, Kong X (2016) A TRIM insertion in the promoter of Ms2 causes male sterility in wheat. Nat Commun 8:15407. https://doi.org/10.1038/ncomms15407

Yang Z, Wang Z, Wang W, Xie X, Chai L, Wang X, Feng X, Li J, Peng H, Su Z, You M, Yao Y, Xin M, Hu Z, Liu J, Liang R, Ni Z, Sun Q, Guo W (2022) ggComp enables dissection of germplasm resources and construction of a multiscale germplasm network in wheat. Plant Physiol 188:1950–1965

Yao H, Xie Q, Xue S, Ma Z et al (2019) HL2 on chromosome 7D of wheat (Triticum aestivum L.) regulates both head length and spikelet number. Theor Appl Genet 132. https://doi.org/10.1007/s00122-019-03315-2

Zhai S, He Z, Wen W et al (2016) Genome-wide linkage mapping of flour color-related traits and polyphenol oxidase activity in common wheat. Theor Appl Genet 129:377–394. https://doi.org/10.1007/s00122-015-2634-6

Zhang J, Huang S, Fosu-Nyarko J, Dell B, McNeil M, Waters I, Moolhuijzen P, Conocono E, Appels R (2008) The genome structure of the 1-FEH genes in wheat (Triticum aestivum L): new markers to track stem carbohydrates and grain filling QTLs in breeding. Mol Breed 22:339–351

Zhang W, Zhao G, Gao L, Kong X, Guo Z, Wu B, Jia J (2016). Functional studies of heading date-related gene TaPRR73, a paralog of Ppd1 in common wheat. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.00772

Zhang X, Liu G, Zhang L Xia C, Zhao T, Jia J Liu X Kong X (2018). Fine mapping of a novel heading date gene, TaHdm605, in hexaploid wheat. Front Plant Sci. https://doi.org/10.3389/fpls.2018.01059

Zhang L, Dong C, Chen Z, Gui L, Chen C, Li D, Xie Z, Zhang Q, Zhang X, Xia C, Liu X, Kong X, Wang J (2021) WheatGmap: a comprehensive platform for wheat gene mapping and genomic studies. Mol Plant 14:187–190

Zhao G, Zou C, Li K, Wang K, Li T, Gao L, Zhang X, Wang H, Yang Z, Liu X, Jiang W, Mao L, Kong X, Jiao Y, Jia J (2017) The *Aegilops tauschii* genome reveals multiple impacts of transposons. Nat Plants 3:946–955

Zhao J, Wang Z, Liu H et al (2019) Global status of 47 major wheat loci controlling yield, quality, adaptation and stress resistance selected over the last century. BMC Plant Biol 19:5. https://doi.org/10.1186/s12870-018-1612-y

Zhou Y, Chen Z, Cheng M, Chen J, Zhu T, Wang R, Liu Y, Qi P, Chen G, Jiang Q, Wei Y, Luo M-C, Nevo E, Allaby RG, Liu D, Wang J, Dvorák J, Zheng Y (2018) Uncovering the dispersion history, adaptive evolution and selection of wheat in China. Plant Biotechnol J 16:280–291

Zhou Y, Zhao X, Li Y, Xu J, Bi A, Kang L, Xu D, Chen H, Wang Y, Wang Y-g, Liu S, Jiao C, Lu H, Wang J, Yin C, Jiao Y, Lu F (2020) Triticum population sequencing provides insights into wheat adaptation. Nat Genet 52:1412–1422

Zhou Y, Bai S, Li H, Sun G, Zhang D, Ma F, Zhao X, Nie F, Li J, Chen L, Lv L, Zhu L, Fan R, Ge Y, Shaheen A, Guo G, Zhang Z, Ma J, Liang H, Qiu X, Hu J, Sun T, Hou J, Xu H, Xue S, Jiang W, Huang J, Li S, Zou C, Song C-P (2021) Introgressing the *Aegilops tauschii* genome into wheat as a basis for cereal improvement. Nat Plants 7:774–786

Zou C, Wang P, Xu Y (2016) Bulked sample analysis in genetics, genomics and crop improvement. Plant Biotechnol J 14:1941–1955