ORIGINAL ARTICLE

WheatQTLdb: a QTL database for wheat

Kalpana Singh[1] · Ritu Batra[1] · Shiveta Sharma[1] · Gautam Saripalli[1] · Tinku Gautam[1] · Rakhi Singh[1] ·
Sunita Pal[1] · Parveen Malik[1] · Manoj Kumar[1] · Irfat Jan[1] · Sahadev Singh[1] · Deepak Kumar[1] · Saksham Pundir[1] ·
Deepti Chaturvedi[1] · Anjali Verma[1] · Anshu Rani[1] · Anuj Kumar[1] · Hemant Sharma[1] · Jyoti Chaudhary[1] ·
Kuldeep Kumar[1] · Sourabh Kumar[1] · Vikas Kumar Singh[1] · Vivudh Pratap Singh[1] · Sachin Kumar[1] · Rahul Kumar[1] ·
Shailendra Singh Gaurav[1] · Shailendra Sharma[1] · Pradeep Kumar Sharma[1] · Harindra Singh Balyan[1] ·
Pushpendra Kumar Gupta[1]

Abstract

During the last three decades, QTL analysis in wheat has been conducted for a variety of individual traits, so that thousands of QTL along with the linked markers, their genetic positions and contribution to phenotypic variation (PV) for concerned traits are now known. However, no exhaustive database for wheat QTL is currently available at a single platform. Therefore, the present database was prepared which is an exhaustive information resource for wheat QTL data from the published literature till May, 2020. QTL data from both interval mapping and genome-wide association studies (GWAS) have been included for the following classes of traits: (i) morphological traits, (ii) N and P use efficiency, (iii) traits for biofortification (Fe, K, Se, and Zn contents), (iv) tolerance to abiotic stresses including drought, water logging, heat stress, pre-harvest sprouting and salinity, (v) resistance to biotic stresses including those due to bacterial, fungal, nematode and insects, (vi) quality traits, and (vii) a variety of physiological traits, (viii) developmental traits, and (ix) yield and its related traits. For the preparation of the database, literature was searched for data on QTL/marker-trait associations (MTAs), curated and then assembled in the form of WheatQTLdb. The available information on metaQTL, epistatic QTL and candidate genes, wherever available, is also included in the database. Information on QTL in this WheatQTLdb includes QTL names, traits, associated markers, parental genotypes, crosses/mapping populations, association mapping panels and other useful information. To our knowledge, WheatQTLdb prepared by us is the largest collection of QTL (11,552), epistatic QTL (107) and metaQTL (330) data for hexaploid wheat to be used by geneticists and plant breeders for further studies involving fine mapping, cloning, and marker-assisted selection (MAS) during wheat breeding.

Keywords WheatQTLdb · QTL database · Epistatic QTL · GWAS · MTA · metaQTL

Introduction

Wheat is a cereal grain crop, which is a worldwide staple food (Mauseth 2014) and accounts for approximately 30% of global cereal consumption (FAO 2002). It is grown on more land area than any other food crop (220.4 million hectares) (FAOSTAT 2014). World trade in wheat is greater than all other crops taken together (FAO 2019). The global wheat production for the year 2019–2020 is estimated to be 765.41 million tons (https://www.statista.com/topics/1668/wheat/), which is next to only maize. It is also estimated that the wheat production should increase by 50–60% of the present level to meet the demand and consumption in 2050. The grain quality of wheat also needs to improve to meet the challenges of nutritional security. This would require major breeding efforts and the sound knowledge of known QTL (quantitative trait loci) and marker-trait associations (MTAs) for all kinds of different traits will be of great help in wheat improvement. During the last three decades, QTL analysis in wheat has been conducted for a variety of individual traits, so that

1   Department of Genetics and Plant Breeding, Chaudhary
    Charan Singh University, Meerut, Uttar Pradesh 250004,
    India

thousands of QTL with their positions in wheat genome and their contribution to the phenotypic variation (PV) in the concerned traits are now known. Recent advances in high-throughput sequencing and genotyping technologies provided a large number of sequence-based markers for the discovery of marker-trait associations (MTAs) through QTL interval mapping and genome-wide association mapping studies (GWAS). As a result, enormous published literature on QTL analysis in wheat has become available. Although various web resources are available for plant genomics such as Gramene (Gupta et al. 2016), CerealsDB 3.0 (Wilkinson et al. 2016), MetaCrop 2.0 (Schreiber et al. 2012), PlantGDB (Dong et al. 2004), T3/Wheat (https://triticeaetoolbox.org/wheat/), and GrainGenes 2019 (Blake et al. 2019), but none is exclusively dedicated to wheat. Also, no available database provides information about all known QTL for wheat along with proper classification of traits to help researchers search the database according to their research interest. Currently, Gramene has only 23 QTL for *Triticum aestivum* (hexaploid wheat) and only 8 for *Triticum turgidum* (tetraploid wheat). CerealsDB 3.0 is wheat specific, while but does not include QTL data. MetaCrop 2.0 and PlantGDB are specific to plants; but none includes QTL data. GrainGenes include 372 QTL of *Triticum durum* only. Recently, Soriano and Alvaro (2019) published a wheat QTL database including 754 QTL dedicated to root only, but not available in the form of web-resource. URGI (https://urgi.versailles.inra.fr/) contains 749 wheat QTL and 19 metaQTL. None of the database include epistatic QTL. Thus, non-availability of an exhaustive database for wheat QTL prompted us to prepare a wheat QTL database described as WheatQTLdb, which is the subject of the current study.

WheatQTLdb is a manually curated QTL database for wheat that includes information about QTL identified through interval mapping and MTA identified using GWAS. For the purpose of this database, MTAs reported using GWAS have also been treated as QTL, since both identify genomic regions affecting the traits of interest. The available information on metaQTL, epistatic QTL, and candidate genes, wherever available, is also included in the database. Users could browse or query the database to find information regarding the genetic architecture of the different categories of traits of interest each of which include a variety of traits, which include the following (Table 1): (i) all morphological traits, (ii) N and P use efficiency, (iii) traits for biofortification (Fe, K, Se and Zn contents), (iv) tolerance to abiotic stresses including drought, water logging, heat, pre-harvest sprouting and salinity, (v) resistance to biotic stresses including not only bacterial, and fungal diseases, but also to infestation by nematodes and insects, (vi) other quality traits, (vii) physiological traits, (viii) developmental trait, and (ix) yield and its related traits. The users can

Table 1 Content of WheatQTLdb including QTL, metaQTL and epistatic QTL (with number of QTL) for different traits

Data	Traits	Number of QTL
QTL	Drought tolerance	1414
	Heat stress tolerance	453
	Pre-harvest sprouting tolerance	593
	Salt stress tolerance	687
	Water-logging tolerance	282
	Fungal resistance	3007
	Bacterial resistance	24
	Nematode resistance	85
	Insect resistance	111
	Iron content	312
	Zinc content	449
	Potassium content	745
	Selenium content	71
	Developmental traits	56
	Morphological traits	153
	Nitrogen use efficiency	1338
	Phosphorous use efficiency	185
	Physiological traits	391
	Quality traits	484
	Yield and its related traits	712
	Total	11,552
metaQTL	metaQTL	330
epistatic QTL	epistatic QTL	107

also find additional information about QTL including their genomic locations and associated markers (both are linked to genetic maps), crosses, and mapping populations used in each case and the links to original references. They can also download the whole data, if necessary. To our knowledge, WheatQTLdb is the largest web-resource having a collection of 11,552 QTL, 330 metaQTL and 107 epistatic QTL of wheat. The database should serve as a repository for the international wheat research community including plant breeders and geneticists for further studies involving fine mapping, cloning and marker assisted selection (MAS) during wheat breeding.

Materials and methods

Data sources and curation

In order to retrieve the published literature for WheatQTLdb, public literature libraries (including PubMed, ResearchGate, Agricola, BIOSIS Previews, CAB abstracts) were searched using the following keywords: wheat, bread wheat, *Triticum aestivum* and QTL, RQTL, GWAS, MTA, metaQTL, epistaticQTL along with the

topic of interest such as water logging tolerance, leaf rust resistance, pre-harvest sprouting tolerance, drought tolerance, etc. This search with topic of research interest was unique, since trait associated with QTL was used for search, to avoid highly heterogeneous textual descriptions of traits/phenotypes that was available in publications. Actual traits/phenotypes given in the publications were treated as parameters in WheatQTLdb; this was done to facilitate the user research community, since they would be interested in QTL related to one research topic at a time and retrieval of QTL of interest will be easier. Currently only hexaploid wheat (*Triticum aestivum*) QTL data is included in WheatQTLdb.

Wheat QTL reported till May, 2020 were curated from all available published research work and included in WheatQTLdb. For this purpose, the information about QTL that was extracted largely included the following details: (i) name of plant species, (ii) trait involved and parameter used, (iii) crosses and mapping population/germplasm (size of population) used, (iv) method of analysis, (v) QTL name, its assignment to individual chromosome and position (interval) on chromosome in cM/bp (hyperlinked to genetic map), (vi) type and name of associated markers, (vii) PVE/R^2 (if available), (viii) candidate genes (if known), and (ix) the link to original reference. Each of the available position of the markers linked to the QTL is hyperlinked with the constructed chromosome wise genetic maps of all the markers for QTL for a particular category of traits to facilitate the user. Another key utility feature is the inclusion of details of reference as hyperlink to the original publication of each QTL to facilitate the user to visit the same for the details of the experiments conducted for the detection of a particular QTL.

Same information was used for metaQTL also, except that for each metaQTL, the details of corresponding QTL are included. In case of epistatic QTL, the additional information included pairs of QTL involved in epistasis, with positions of these QTL within the genome.

Information about 11,552 QTL, 330 metaQTL, and 107 epistatic QTL were extracted and curated for the current WheatQTLdb. QTL for 22 categories are included, which are listed in Table 1 along with number of QTL in each category; most of the trait categories also served as traits, such as iron content, drought tolerance, and nitrogen (N) use efficiency, etc. Each QTL extracted from a published report was considered as unique entity in WheatQTLdb and it was left on the discretion of users to weigh the QTL extracted from more than one study to eliminate the issues related to tracking/merging confronted by the curators.

Implementation

WheatQTLdb webserver runs on a Linux server (Ubuntu OS version 19.10) having 256 GB RAM, 32 CPUs, and 6 TB hard disk. Specific features and activities involved in the preparation of WheatQTLdb web-resource and data retrieval are shown in Fig. 1 and included the following: (i) manually curated wheat QTL data that was utilized to create WheatQTLdb database and prepared in MySQL, a relational database management system (RDMS) (version 5.7) with 22 tables accommodating all data; (ii) WheatQTLdb web-resource was built using html, PHP, and Java; and it was hosted using Apache2 server (version 2.4.7). (iii) a request was created using PHP to be used from user's system to send it to WheatQTLdb webserver through internet; (iv) in response to this PHP request, query in MySQL was prepared

Fig. 1 Layout depicting the steps involved in the preparation of WheatQTLdb web-resource and data retrieval steps
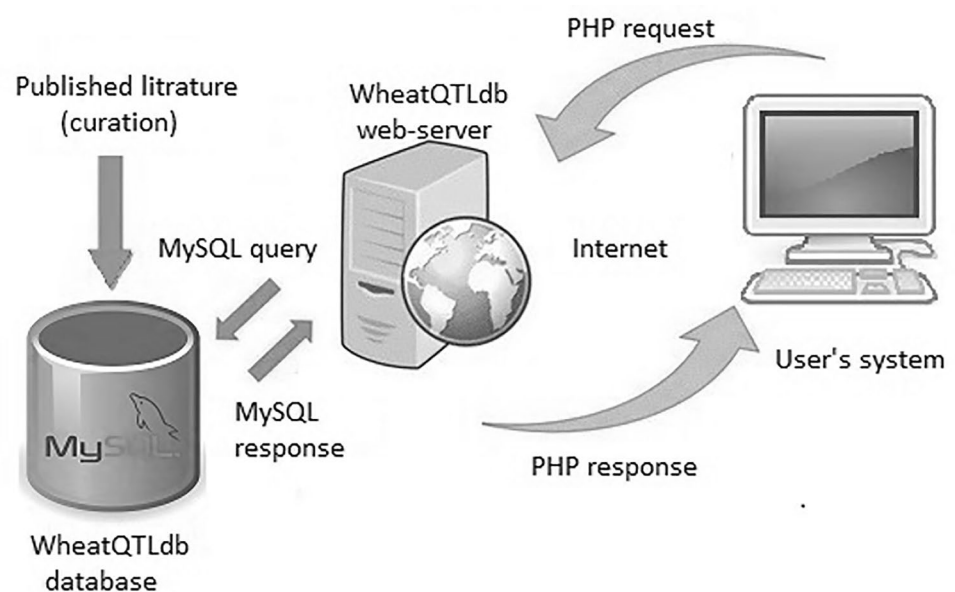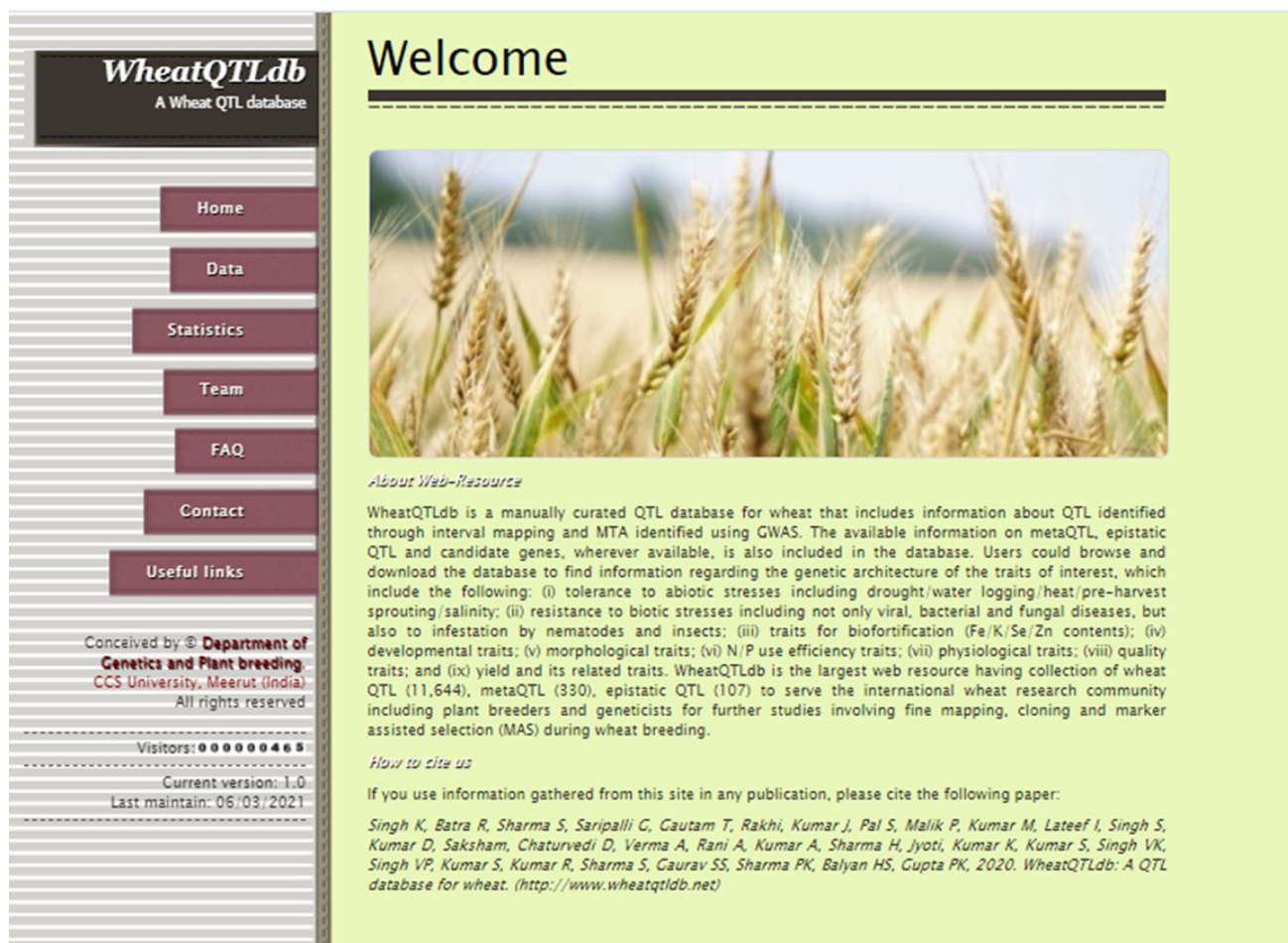
Fig. 2 The home page for WheatQTLdb with a vertical navigation bar at the left side that links to the separate parts of the site

to send it to WheatQTLdb database in MySQL; (v) a database response was also prepared in MySQL to send it to webserver; and (vi) a response was prepared in PHP to send it back to user's system through internet.

Web interface

Web interface of WheatQTLdb provides vertical navigation bar in the left part of website to link with other parts of the site (Fig. 2), which include statistics, data, team, FAQ and contact parts. Figure 3 shows the recommended browse map for data. In data part, three options were provided. On data page, epistatic QTL and metaQTL buttons directly hyperlinked to separate web page provided with multiple options in drop-down. QTL button provides a drop-down option list to opt; selecting a option from drop down list would open a separate web page provided with multiple options for a particular trait category/ sub-category in dropdown. On submitting the selected option, QTL data in tabular form appears on same page. User could search this tabular data and download in comma separated value (.csv) file. Statistics

part was prepared with Java scripts to show the data statistics in pi-charts. The contact form provides facility to any user, who wants to add his data to WheatQTLdb or like to ask any query; though few probable queries and their answers are provided in FAQ part. Few useful links to other databases were also provided along with links to other sites to share the link of WheatQTLdb with other interested users to facilitate the user.

Utility of WheatQTLdb

During the last 30 years, due to intensive research involving QTL analysis and GWAS in wheat, thousands of QTL are now known. Yet currently available wheat QTL data domains are presenting a small set of data without complete information about the trait category for users to find QTL data according to their research problem (e.g., all QTL associated with drought tolerance). To address this issue, we developed a QTL web-resource that enhances user's ability to utilize more efficiently the previously published wheat QTL data. Curation of wheat QTL data according to

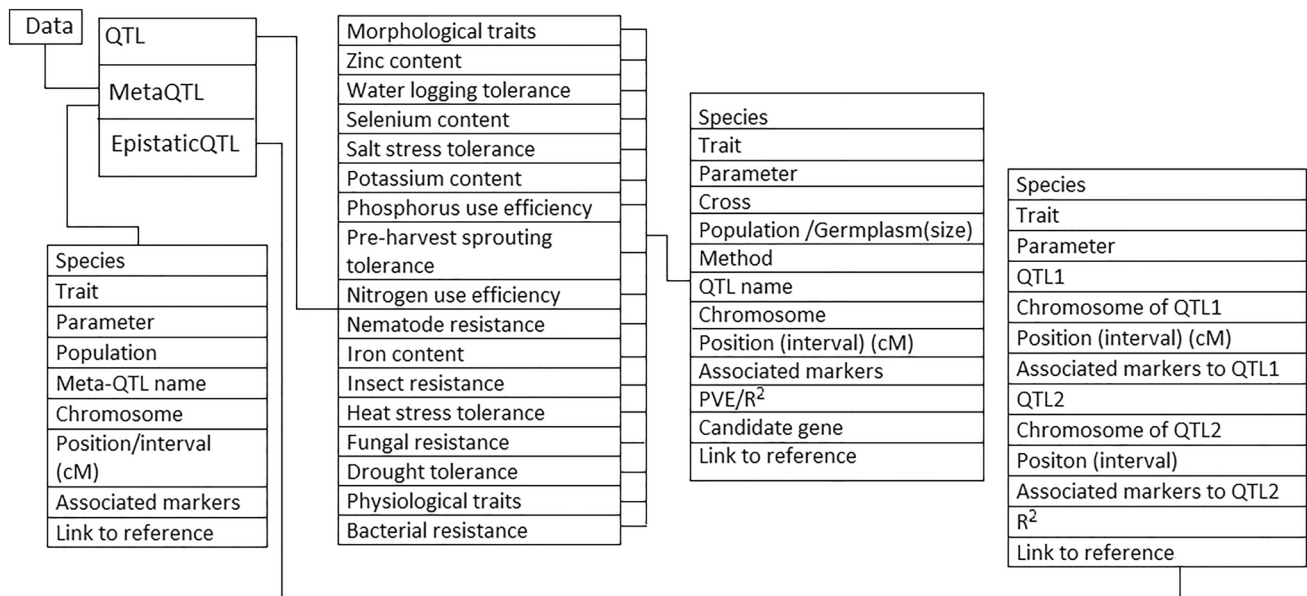| Data	QTL	Morphological traits		Species		Species

Fig. 3 Recommended browse map for data in WheatQTLdb

research topics and link to the reference are the user-friendly options provided in WheatQTLdb web-resource. Another important utility of WheatQTLdb is that a good number of wheat metaQTL and epistatic QTL are included. Probably, this is the first web-resource, where epistatic QTL are also included along with the largest collection of metaQTL for wheat. WheatQTLdb will facilitate the plant breeders and research community for functional annotation of the wheat genome involving following purposes: (i) to locate the positions of gene/genes responsible for the traits of interest within in the genome; (ii) to facilitate marker-assisted selection (MAS); (iii) to conduct fine mapping and map-based cloning; (iv) to study linkage and epistatic relationships among genes and QTL for traits of interest; (v) to facilitate identification of individual genes responsible for a specific quantitative trait by narrowing the search; and (vi) to provide information about the location and relative importance of individual markers.

Currently WheatQTLdb includes only curated data from published literature for hexaploid wheat. We would like to expand it by integrating other data domains and also the data for diploid and tetraploid wheats.

Conclusion

WheatQTLdb is a wheat QTL database with large data sets for QTL, metaQTL, and epistatic QTL to serve the wheat research community including geneticists and plant breeders, so that one can access the information related to QTL from extensive published literature in a user-friendly manner. It will be extended further in future to include information on QTL reported in diploid and tetraploid wheats. The database includes data that was available till May 2020. After the acceptance of the manuscript for publication, additional data for a large number of QTL (>10,000) has been collected. Therefore, a manuscript to be named as supplement to this database is being prepared, which will soon be submitted for publication.

Author contributions KS prepared the database and first draft of the manuscript. SK, RK, SSG, SS, PKS, HSB, and PKG conceived and planned the study and also finalised the manuscript. All the remaining authors collected and tabulated the data for different traits. GS, AK, and KK also jointly prepared the chromosome maps showing the markers linked to QTLs for different traits.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.
Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Reference

Blake VC, Woodhouse MR, Lazo GR, Odell SG, Wight CP et al (2019) Grain Genes: centralized small grain resources and digital platform for geneticists and breeders. Database 2019:baz065

Dong Q, Schlueter SD, Brendel V (2014) Plant GDB, plant genome database and analysis tools. Nucleic Acids Res 32(Database issue):D354–D359

FAO (2002) FAO: World Agriculture: towards 2015/2030. Summary Report. Food and Agriculture Organization of the United Nations, Rome

FAO (2019) World food situation: FAO cereal supply and demand brief. Rome: United Nations, Food and Agriculture Organization. Retrieved 14 Dec 2016.

FAOSTAT (2014) Crops/World Total/Wheat/Area Harvested/2014 (pick list). United Nations, Food and Agriculture Organization, Statistics Division (FAOSTAT). Archived from the original on 6 Sept 2015. Retrieved 8 Dec 2016.

Gupta P, Naithani S, Tello-Ruiz MK et al (2016) Gramene database: navigating plant comparative genomics resources. Curr Plant Biol 7–8:10–15

Mauseth JD (2014) Perhaps the simplest of fruits are those of grasses (all cereals such as corn and wheat)… These fruits are caryopses. Botany. Jones & Bartlett Publishers, Burlington, p 223

Schreiber F, Colmsee C, Czauderna T et al (2012) MetaCrop 2.0: managing and exploring information about crop plant metabolism. Nucleic Acids Res 40(Database issue):D1173–D1177

Soriano JM, Alvaro F (2019) Discovering consensus genomic regions in wheat for root-related traits by QTL meta-analysis. Sci Rep 9:10537

Wilkinson PA, Winfield MO, Barker GLA et al (2016) CerealsDB 3.0: expansion of resources and data integration. BMC Bioinform 17:256

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.