BMC Bioinformatics

DATABASE Open Access

CerealsDB 2.0: an integrated resource for plant breeders and scientists

Paul A Wilkinson*†, Mark O Winfield†, Gary LA Barker, Alexandra M Allen, Amanda Burridge, Jane A Coghill and Keith J Edwards

Abstract

Background: Food security is an issue that has come under renewed scrutiny amidst concerns that substantial yield increases in cereal crops are required to feed the world's booming population. Wheat is of fundamental importance in this regard being one of the three most important crops for both human consumption and livestock feed; however, increase in crop yields have not kept pace with the demands of a growing world population. In order to address this issue, plant breeders require new molecular tools to help them identify genes for important agronomic traits that can be introduced into elite varieties. Studies of the genome using next-generation sequencing enable the identification of molecular markers such as single nucleotide polymorphisms that may be used by breeders to identify and follow genes when breeding new varieties. The development and application of next-generation sequencing technologies has made the characterisation of SNP markers in wheat relatively cheap and straightforward. There is a growing need for the widespread dissemination of this information to plant breeders.

Description: CerealsDB is an online resource containing a range of genomic datasets for wheat (*Triticum aestivum)* that will assist plant breeders and scientists to select the most appropriate markers for marker assisted selection. CerealsDB includes a database which currently contains in excess of 100,000 putative varietal SNPs, of which several thousand have been experimentally validated. In addition, CerealsDB contains databases for DArT markers and EST sequences, and links to a draft genome sequence for the wheat variety Chinese Spring.

Conclusion: CerealsDB is an open access website that is rapidly becoming an invaluable resource within the wheat research and plant breeding communities.

Keywords: Wheat, Single nucleotide polymorphisms, SNPs, Database

Background

The issue of food security (defined as the access to sufficient, safe, nutritious food to maintain a healthy and active life) has become a focus of concern for governments around the world and has moved steadily up the political agenda. The human population is expected to reach nine billion by 2050, and it has been estimated that cereal production will need to increase by 50% by this time if we are to meet the challenge of feeding ourselves and the animals upon which we are dependent [1]. The cereal crops wheat, maize and rice are the main source of nutrition for humans and domesticated animals [2]. Wheat, which accounts for approximately 30% of global cereal consumption [3], is of fundamental importance in this regard with an estimated 2010 harvest of 651 million metric tonnes [4]. Ensuring that yields of wheat increase to meet future needs has become an important focus in agricultural research. The remarkable increases in yield experienced during the period 1950 to 2000, which were based on traditional breeding techniques, have unfortunately been unable to keep up with the increased demand of a rapidly expanding world population. In addition, the relative increase in wheat yields has slowed towards the end of the twentieth century [5]. To help address this need plant breeders and scientists must embrace the modern molecular biological technologies that allow them a more targeted approach to plant improvement.

* Correspondence: Paul.Wilkinson@bristol.ac.uk
†Equal contributors
School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK

Single nucleotide polymorphisms (SNPs), which are single base pair changes in the DNA sequence, are the most common form of sequence variation between individuals of the same species. Because of their frequency in the genome, SNPs have become the marker of choice for marker assisted selection (MAS) of agronomic traits of importance such as disease and drought resistance [6]. In order to target genes at any position throughout the genome it is necessary to saturate the wheat genome with validated SNP markers that can be used in MAS. Through the use of next-generation sequencing (NGS), our laboratory is actively searching for SNPs in the wheat genome and developing PCR assays for these so that they can be used as molecular markers by plant breeders.

The development of NGS platforms and high-throughput genotyping methods has made it possible to rapidly characterise and validate thousands of putative SNP markers in wheat at relatively low cost [7,8]. A number of online databases already exist that link to SNP markers and other genomic resources, such as Gramene which is an online data resource for comparative genome analysis in the grasses [9] and GrainGenes, a database for genetic and genomic information about Triticeae species and their wild relatives [10]. Whilst these sites contain a wealth of information, they have been designed primarily for scientists. In order to maximise the utility and application of SNP markers for plant breeders, it is imperative that such information is made freely available and easily accessible to non-experts in molecular biology. This requirement for the dissemination of validated SNP marker data to plant breeders has provided the impetus for the development of the CerealsDB SNP database.

We have greatly extended the functionality of CerealsDB, which was originally constructed to store a dataset of 26,382 EST sequences [11], to include a number of searchable online databases relating to wheat genomics. Principally, the CerealsDB website is now aimed at those who wish to obtain information about SNP markers, for example, the sequence upon which they are based, the primers used for their PCR-based identification or SNP haplotype information of common UK varieties. The CerealsDB website also allows users to search the original CerealsDB EST sequences and the draft genome sequence of wheat variety Chinese Spring (assembled using Roche 454 technology) and to search for diversity array technology (DArT) markers that have been bin-mapped to chromosomes arms.

Construction and content

Implementation

CerealsDB was implemented using a MySQL relational database management system (RDMS) database (version 5.051b) running on a MacOSX server and hosted using the Apache web server (version 2.2) with Perl and PHP scripts used for all data retrieval and output. MySQL was selected for the database, as it is one of the leading open source industrial strength RDMS, and matches the quality and performance of other major proprietary databases. MySQL also finds advantage in multi-layered server design with independent modules and fast execution process. The Apache HTTP server was adopted due to its reliability, relative ease of configuration and wide language interface support system.

CerealsDB is a relational collection of entities (tables) with appropriate attributes (fields) to define the wheat SNPs. An entity-relationship (ER) model for the SNP database is displayed in Figure 1. At present, the SNP database is composed of 11 tables of which four share an identifying relationship with the 'Contig' table through their common primary key, 'SNP_id'. The 'Contig' table, which contains 111,442 records (representing putative SNP loci), consists of 16 attributes and includes information on those contigs that contain putative SNPs: e.g. source of the contig data, contig sequence and SNP position(s) within the sequence. The 'Contig' table also contains a field indicating whether the SNPs have been validated experimentally.

The 'Location' table consists of 3,813 entries and contains information with regard to chromosomal location and position in centimorgans (cM) of all the SNPs that have been mapped. The 'Primers' table, which has 6,337 records, contains the KASPar primer sequences (two forward and one reverse) for all the SNPs that have been experimentally assayed. The 'Haplotype' table includes 6,337 records and contains the SNP calls for all the validated SNPs against a panel of 101 wheat varieties. The 'blast_url2' table holds 37,837 entries that are linked to the contig table by the `NewContigName` attribute in a 1:1 identifying relationship. This table contains all hits reported when a BLASTX similarity search was performed for each contig versus the NCBI 'embryophyta plant' protein database [12]. Inclusion of this table allows users to search for SNPs on contigs that have a particular annotation (e.g., disease resistance). The 'Source' table cannot currently be queried by CerealsDB users and serves as an in-house resource. It contains 3,263 entries and holds information on the original source of a validated SNP and the location of the DNA on a sample plate stored in our laboratory.

The 'Homo_and_var_SNPs' (511,439 entries) and 'VarietalSNPS' tables (99,945 entries) are linked to one another by a 1:1 non-identifying relationship and are unlinked to any of the other tables. The 'Homo_and_var_SNPs' table contains information derived from a particular experiment designed to identify both homoeologous and varietal SNPs in eight wheat varieties (Alchemy, Avalon, Cadenza, Hereward, Rialto, Robigus, Savannah and Xi19): a homoeologous SNP is a sequence variation which exists between the different genomes (A, B and D) which make up the hexaploid wheat genome and a varietal SNP exists between wheat varieties and enables the differentiation of two varieties based upon that particular SNP location. The table, which contains 511,439 entries, contains a primary key which uniquely identifies the SNP and contig from which it has been derived, and the SNP calls, depth and value, for each locus for all eight varieties. This information has been published in Winfield et al. [13].

The 'VarietalSNPS' table contains information solely on those putative SNPs that have been classified as varietal; that is, SNPs that clearly distinguish one wheat variety from another. These varietal SNPs are of particular interest to breeders. This table is used principally for database searches used to populate the summary tables at the head of several web pages.

There are three stand-alone tables that are not linked to any other table in the database. The 'WheatVariety' table contains 101 records, one entry for each variety studied, and includes information on pedigree, winter/spring habit, and the original plant breeder for the 101 wheat varieties studied so far. The Images table (21 records) contains the URL to images of SNP maps of the 21 wheat chromosomes, and the table 'IUPAC_codes' holds 36 entries for translation of ambiguity codes.

The CerealsDB site has been extensively redesigned using cascading style-sheets (CSS) to standardise the format of the site, and improve the presentation, to provide a more intuitive browsing experience. Links are provided to the WheatBP educational resource and to other related sites that provide additional background information on wheat genomics.

Data sources

SNP data was included from previous studies [14] as well as from ongoing experiments. SNPs were validated

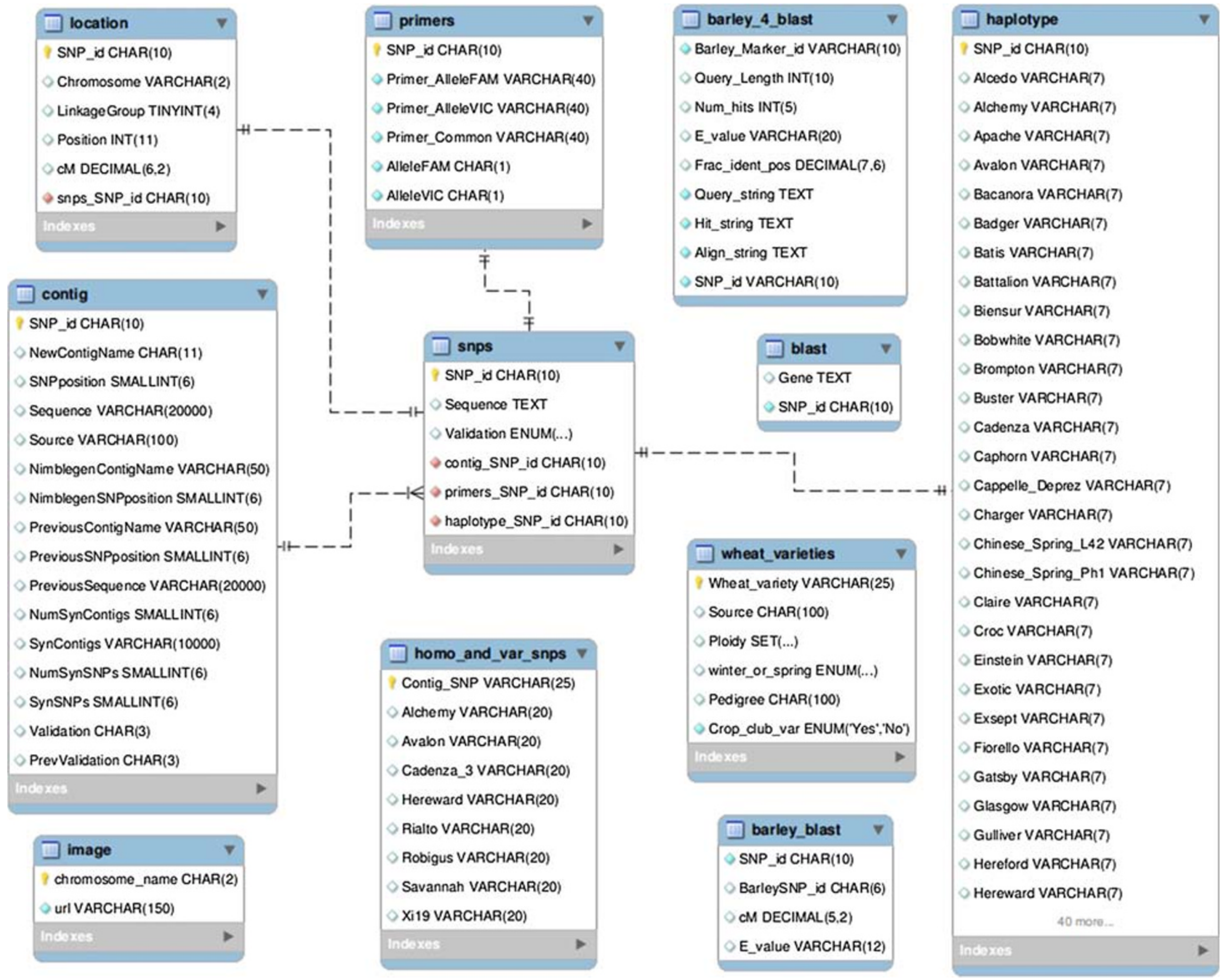


Figure 1 Entity-relationship (ER) model for the cerealsDB SNP database, generated with MySQL workbench. The schema displays the relationships between different tables within the SNP database.

using the KBioscience Competitive Allele-Specific PCR (KASPar) genotyping method which relies on the differential amplification of allele specific probes [15]. All SNP data were curated to remove redundancy and duplicate values. The contigs used in the database were generated from an assembly of wheat (5x Chinese Spring var.) cDNA sequences. Unique IDs were created for each contig using an 11-character code, a Bristol contig (BC) prefix for reference sequences followed by 9 digits (e.g., BC000000001). All unique SNPs were designated a 10-character code, beginning with a Bristol SNP (BS) prefix, followed by 8 digits (e.g., BS00000001). Cross-referencing and standardisation of file formats was performed using custom Perl scripts. The CerealsDB site also contains sequences from the draft assembly of the Chinese Spring genome, which is comprised of 5,321,847 contigs with a 5 x genome coverage. At this level of coverage, we would expect to have at least one read for > 95% of the genome.

The KASPar SNP database currently consists of data on 111,442 SNPs derived from 8 wheat varieties of which 99,945 are varietal SNPs (these are the SNPs that are of value to plant breeders) . Of these SNPs 4,986 have been validated (by screening against 101 varieties) and 3,813 mapped to a specific chromosomal location. Validated SNP entries contain information on location within a reference contig and primer sequences used for validation.

CerealsDB includes the original wheat EST database, containing 26,382 expressed sequence tags that were generated from hexaploid winter wheat (var. Mercia) cDNA and previously described by Wilson et al. [11]. These wheat ESTs can be searched by gene name or BLAST. There is also a database of DArT data that can be accessed through a clickable wheat ideogram; each chromosome image of the ideogram takes you to data related to that particular chromosome. DArT technology, which was developed as a hybridization-based tool for genotyping, incorporates many features of AFLP analysis combined with high throughput genetic analysis using a DNA microarray platform. We used deletion lines obtained from the Kansas Wheat Genetic and Genomic Resources Center (WGGRC) (http://www.k-state.edu/wgrc/Germplasm/Deletions/delindex.html) to assign DArT markers to chromosome arm specific bins. Of the 420 deletion lines held at the WGGRC, 74 lines were selected because they were the lines used in the NSF Wheat EST Genomics Project designed to assign wheat ESTs to chromosome bins. These lines possess homologous terminal deletions in single chromosome arms. These deletion lines (developed in the wheat variety Chinese Spring) provide powerful tools for the physical mapping of wheat chromosomes giving us the ability to map markers not only to specific homoeologous chromosomes but to assign them to bins of around 28 Mb [16] on specific chromosome arms. Whilst DArT markers are no longer used by plant breeders (having been superseded by SNP markers), this information is still relevant to the broader scientific research community.

Utility and discussion

Web interface and user querying capabilities

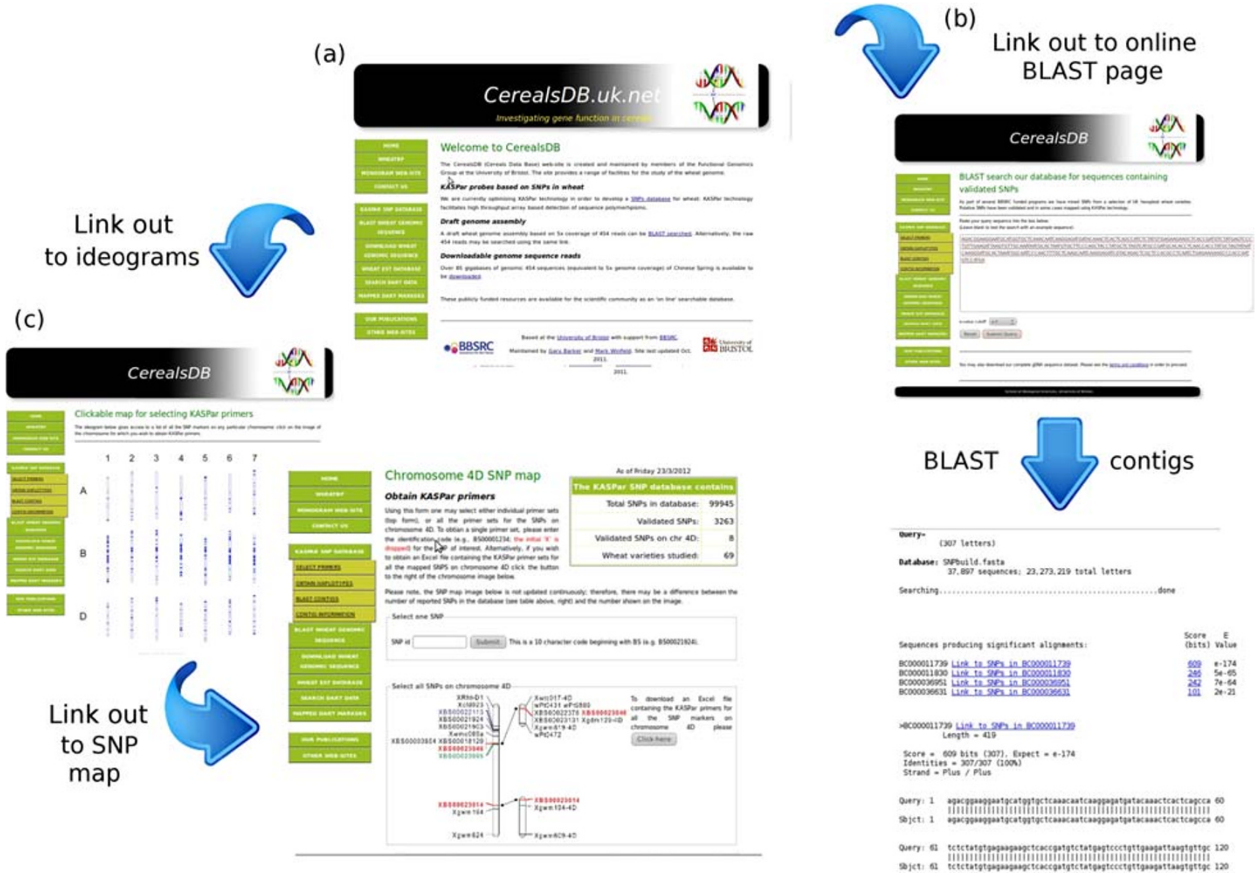
The CerealsDB website contains a range of datasets that are accessible via the menu bar and include the KASPar SNP database, a BLAST search page to query the Wheat genomic sequence, a download page for the draft wheat genome, a wheat EST database, a search page for wheat DArT data and a page describing the mapping of DArT markers using wheat deletion lines.

The KASPar SNP database allows SNP name and contig queries. Some queries (e.g. SNP name) are exact. Sequence queries are BLAST similarity based [17]. Some of the functionality of the CerealsDB site is displayed in Figure 2. All SNPs listed in CerealsDB have been assigned a unique ID and the output page for a query gives information about the gene name and locus. All validated SNP information can be searched by chromosome and sub-genome and the resulting data is downloadable as an Excel spreadsheet.

The interface is designed to be simple and provide quick access for plant breeders who may not possess an in depth knowledge of genetics. Chromosome information is accessed via a clickable image (ideogram). These are arrayed on the screen according to the chromosome and sub-genome to which they belong. The initial page for the KASPar SNP database contains statistics on the SNPs within the database and also links to bar graphs showing the distribution across the 21 wheat chromosomes of all mapped SNPs for the mapping populations Avalon x Cadenza and Rialto x Savannah. In addition, there are also graphs showing the relationship between the expected number of mapped SNPs to each of the three genomes (A, B and D), based on their relative lengths, and the actual number mapped.

There are four main query roots for the KASPar SNP database:

1) "Select primers"; this can provide information on the primers for individual SNPs or for all the SNPs on a particular chromosome. Primers for SNPs on each chromosome can be selected by clicking on the relevant chromosome image of the ideogram, or by entering a specific SNP ID. Alternatively, all SNPs located on that particular chromosome can be downloaded as an Excel spreadsheet. Scrolling down the page reveals a map of the chromosome and the SNPs that are currently mapped to it.
2) "Obtain haplotypes"; select for information on the SNP alleles in the 101 wheat varieties against which



Welcome to CerealsDB

The CerealsDB (Cereals Data Base) web-site is created and maintained by members of the Functional Genomics Group at the University of Bristol. The site provides a range of facilities for the study of the wheat genome.

KASPar probes based on SNPs in wheat

We are currently optimising KASPar technology in order to develop a SNPs database for wheat. KASPar technology facilitates high throughput array based detection of sequence polymorphisms.

Draft genome assembly

A draft wheat genome assembly based on 5x coverage of 454 reads can be BLAST searched. Alternatively, the raw 454 reads may be searched using the same link.

Downloadable genome sequence reads

Over 85 gigabases of genomic 454 sequences (equivalent to 5x genome coverage) of Chinese Spring is available to be downloaded.

These publicly funded resources are available for the scientific community as an 'on line' searchable database.

Based at the University of Bristol with support from BBSRC.
Maintained by Gary Barker and Mark Winfield. Site last updated Oct. 2011.

BLAST search our database for sequences containing validated SNPs

Clickable map for selecting KASPar primers

Chromosome 4D SNP map

Obtain KASPar primers

Using this form one may select either individual primer sets (top form), or all the primer sets for the SNPs on chromosome 4D. To obtain a single primer set, please enter the identification code (e.g., BS00001234; the initial 'X' is dropped) for the SNP of interest. Alternatively, if you wish to obtain an Excel file containing the KASPar primer sets for all the mapped SNPS on chromosome 4D click the button to the right of the chromosome image below.

As of Friday 23/3/2012

The KASPar SNP database contains	
Total SNPs in database:	99945
Validated SNPs:	3263
Validated SNPs on chr 4D:	8
Wheat varieties studied:	69

Please note, the SNP map image below is not updated continuously; therefore, there may be a difference between the number of reported SNPs in the database (see table above, right) and the number shown on the image.

Select one SNP

SNP id [ ] Submit This is a 10 character code beginning with BS (e.g. BS00021924).

Select all SNPs on chromosome 4D

To download an Excel file containing the KASPar primers for all the SNP markers on chromosome 4D please Click here

```
Query=
         (307 letters)

Database: SNPbuild.fasta
           37,897 sequences; 23,273,219 total letters

Searching..................................................done

                                                            Score    E
Sequences producing significant alignments:                 (bits) Value

BC000011739 Link to SNPs in BC000011739                      609   e-174
BC000011830 Link to SNPs in BC000011830                      246   5e-65
BC000036951 Link to SNPs in BC000036951                      242   7e-64
BC000036631 Link to SNPs in BC000036631                      101   2e-21

>BC000011739 Link to SNPs in BC000011739
          Length = 419

 Score =  609 bits (307), Expect = e-174
 Identities = 307/307 (100%)
 Strand = Plus / Plus

Query: 1    agacggaaggaatgcatggtgctcaaacaatcaaggagatgatacaaactcactcagcca 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1    agacggaaggaatgcatggtgctcaaacaatcaaggagatgatacaaactcactcagcca 60

Query: 61   tctctatgtgagaagaagctcaccgatgtctatgagtccctgttgaagattaagtgttgc 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 61   tctctatgtgagaagaagctcaccgatgtctatgagtccctgttgaagattaagtgttgc 120
```

Figure 2 The functionality of the KASPar SNP database is displayed with (a) the CerealsDB homepage, from which a range of pages are available for querying the SNP database. For example, contigs containing SNPs can be searched by BLAST (**b**) and SNP maps can be dynamically retrieved from the database by clicking on ideograms (**c**) for a particular chromosome on one of the sub genomes.

the SNPs have as yet been validated. Clicking the relevant chromosome image of the ideogram will take the user to information on that particular chromosome. Alternatively, haplotypes can be selected for particular varieties on specific sub-genomes and chromosomes. Once again, results can be downloaded as an Excel spreadsheet.

3) "Blast contigs"; allows the user to input a nucleotide sequence and perform similarity searches (using BLAST) against the Bristol contigs. If the search identifies a contig and it contains SNPs, then the user can display the positions of each SNP within that particular contig.
4) "Contig information"; facilitates the search for information about the contig on which a SNP is found by providing a SNP ID, or the contig name to see whether it contains any SNPs. If a SNP is found, the output displays the contig sequence, SNP position, information on mapping and any associated BLAST annotation (the original BLASTX reports are also included). A field is provided to search for contigs based on a specific term, for example 'disease resistance' or 'cold', and will return all contigs that contain this term in the contigs' associated BLASTX report.

The draft assembly of the gene-rich regions of the Chinese Spring genome, or the raw sequence reads can be searched using BLAST and a drop down field is available for setting the e-value cut-off.

Wheat ESTs can be searched by gene name, similarity search (via BLAST) or westdb cloneID. The wheat DArT markers can be searched by marker name or wheat line and the facility is available to download the DArT datasets in Excel format. Mapped DArT markers can also be viewed on their respective chromosome by clicking the appropriate chromosome image of the ideogram.

Finally, there is a Help/FAQ page for users, with answers to potential questions about the data in the database and other features on the website.

Conclusions

Substantial enhancement has been made to the original CerealsDB site through the inclusion of databases for

SNP and DArT markers as well as other genomic resources. In order to make CerealsDB comprehensible to non-scientists and to retain its relevance to scientific specialists, version 2.0 has improved both functionality and the user-friendliness of the web interface. As the SNP database grows in size and the number of KASPar validated SNPs increases, we expect traffic to the CerealsDB site to grow. The website has already attracted considerable interest from the plant-breeding and research communities and has recorded, on average, more than 5,600 unique visits per month over the past year (web traffic statistics generated by Webalizer version 2.23). Our laboratory is actively developing and validating SNP markers, and these are regularly added to the database. Future developments planned for the CerealsDB site include the linking of SNP data with phenotypic data and the integration of the KASPar SNP database with other cereal genomics databases. We aim to make CerealsDB an important, freely available resource by releasing all data into the public domain without restrictions. This will ensure that wheat breeders across the world have the data they require to create new varieties of wheat that will be needed to help feed the world beyond 2050.

Availability and requirements

Availability

CerealsDB can be accessed online from the following URL : http://www.cerealsdb.uk.net/CerealsDB/Documents/DOC_CerealsDB.php. The SNP database is publicly and freely accessible, requiring no registration and with no restrictions on use.

Technical requirements

It is recommended that one of the following browsers is used: Mozilla Firefox 3 on Linux, Mac OSX or Windows, Internet Explorer 8 on Windows, Safari 4 on Mac OsX or Windows, Chrome on Linux or Windows.

Abbreviations

BC: Bristol contig; BLAST: Basic local alignment search tool; BS: Bristol SNP; CSS: Cascading style sheets; DArT: Diversity array technology; ER: Entity-relationship; EST: Expressed sequence tag; FAO: Food and Agriculture Organisation of the United Nations; FAQ: Facts and questions; HTTP: Hypertext transfer protocol; ID: Identification; MAS: Marker-assisted selection; NCBI: National Center for Biotechnology Information; NGS: Next-generation sequencing; Perl: Practical extraction and report language; PHP: Hypertext pre-processor; RDMS: Relational database management system; SNP: Single nucleotide polymorphism; URL: Uniform resource locator.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PAW and MOW wrote the manuscript. MOW, GLAB & PAW built CerealsDB version 2.0, AMA and AB carried out the SNP validation experiments, JC carried out NGS sequencing and KJE conceived of the study, participated in its design and coordination, and helped to draft the manuscript. All authors have read and approved the final manuscript.

Acknowledgements and funding

We are grateful to the Biotechnology and Biological Sciences Research Council, UK, for providing the funding for this work (awards BB/I003207/1, BB/I017496/1, BB/F010370/1). We are also grateful to the Lady Emily Smyth Agricultural Research Station for providing the funds for next-generation sequencing.

Received: 11 May 2012 Accepted: 9 August 2012
Published: 3 September 2012

References

1. Foresight: *The Future of Food and Farming*. The Government Office for Science: In Final Project Report. London; 2011.
2. Shewry PR: **Wheat.** *J Exp Bot* 2009, **60**:1537–1553.
3. FAO: *World Agriculture: towards 2015/2030. Summary Report*. Rome: Food and Agriculture Organization of the United Nations; 2002.
4. FAOSTAT database [http://faostat.fao.org/].
5. Alston JM, Babcock BA, Pardey PG: *The Shifting Patterns of Agricultural Production and Productivity Worldwide*. Ames: Iowa State University; 2010. CARD-MATRIC on-line volume.
6. Gupta PK, Rustgi S, Mir RR: **Array-based high-throughput DNA markers for crop improvement.** *Heredity* 2008, **101**(1):5–18.
7. Berkman PJ, Lai K, Lorenc MT, Edwards D: **Next generation sequencing applications for wheat crop improvement.** *American J Bot* 2012, **99**(2):365–371. 2012.
8. Paux E, Sourdille P, Mackay I, Feuillet C: **Sequence-based marker development in wheat: Advances and applications to breeding.** *Biotechnol Advances* 2011, **2011**:2011–2011.
9. Youens-Clark K, Buckler E, Casstevens T, Chen C, Declerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, Lu J, McCouch SR, Ren L, Spooner W, Stein JC, Thomason J, Wei S, Ware D: **Gramene database in 2010: updates and extensions.** *Nucleic Acids Res* 2011, **39**:D1085–D1094.
10. Carollo V, Matthews DE, Lazo GR, Blake TK, Hummel DD, Lui N, Hane DL, Anderson OD: **GrainGenes 2.0. An improved resource for the small-grains community.** *Plant Physiol* 2005, **139**:643–651.
11. Wilson ID, Barker GLA, Beswick RW, Shepherd SK, Lu C, Coghill JA, Edwards D, Owen P, Lyons R, Parker JS, Lenton JR, Holdsworth MJ, Shewry PR, Edwards KJ: **A transcriptomics resource for wheat functional genomics.** *Plant Biotechnol J* 2004, **2**:495–506.
12. Pruitt KD, Tatusova T, Brown GR, Maglott DR: **NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy.** *Nucleic Acids Res* 2012, D130–D135.
13. Winfield MO, Wilkinson PA, Allen AM, Barker GLA, Coghill JA, Burridge A, Hall A, Brenchley RC, D'Amore R, Hall N, Bevan MW, Richmond T, Gerhardt D, Jeffrey A, Jeddeloh JA, Edwards KJ: **Targeted re-sequencing of the allohexaploid wheat exome.** *Plant Biotechnol J* 2012 Aug, **10**(6):733–742.
14. Allen AM, Barker GL, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ: **Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.).** *Plant Biotechnol J* 2011, **9**:1086–1099.
15. Cuppen E: **Genotyping by allele-specific amplification (KASPar).** *Cold Spring Harb Protocols* 2007, pdb.prot4841.
16. Qi LL, Echalier B, Friebe B, Gill BS: **Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs.** *Functional Integrative Genomics* 2003, **3**:39–55.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.

doi:10.1186/1471-2105-13-219
Cite this article as: Wilkinson *et al.*: **CerealsDB 2.0: an integrated resource for plant breeders and scientists.** *BMC Bioinformatics* 2012 **13**:219.