



The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement

Emma Mace^{1,2} · David Innes³ · Colleen Hunt² · Xuemin Wang¹ · Yongfu Tao¹ · Jared Baxter² · Michael Hassall⁴ · Adrian Hathorn⁵ · David Jordan¹

Received: 27 August 2018 / Accepted: 11 October 2018 / Published online: 20 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Key message We describe the development and application of the Sorghum QTL Atlas, a high-resolution, open-access research platform to facilitate candidate gene identification across three cereal species, sorghum, maize and rice.

Abstract The mechanisms governing the genetic control of many quantitative traits are only poorly understood and have yet to be fully exploited. Over the last two decades, over a thousand QTL and GWAS studies have been published in the major cereal crops including sorghum, maize and rice. A large body of information has been generated on the genetic basis of quantitative traits, their genomic location, allelic effects and epistatic interactions. However, such QTL information has not been widely applied by cereal improvement programs and genetic researchers worldwide. In part this is due to the heterogeneous nature of QTL studies which leads QTL reliability variation from study to study. Using approaches to adjust the QTL confidence interval, this platform provides access to the most updated sorghum QTL information than any database available, spanning 23 years of research since 1995. The QTL database provides information on the predicted gene models underlying the QTL CI, across all sorghum genome assembly gene sets and maize and rice genome assemblies and also provides information on the diversity of the underlying genes and information on signatures of selection in sorghum. The resulting high-resolution, open-access research platform facilitates candidate gene identification across 3 cereal species, sorghum, maize and rice. Using a number of trait examples, we demonstrate the power and resolution of the resource to facilitate comparative genomics approaches to provide a bridge between genomics and applied breeding.

Communicated by Jessica Rutkoski.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00122-018-3212-5>) contains supplementary material, which is available to authorized users.

✉ Emma Mace
emma.mace@daf.qld.gov.au

- ¹ Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Warwick, QLD 4370, Australia
- ² Department of Agriculture and Fisheries, Hermitage Research Facility, Warwick, QLD 4370, Australia
- ³ Department of Agriculture and Fisheries, Ecosciences Precinct, Brisbane, QLD 4102, Australia
- ⁴ Department of Agriculture and Fisheries, Leslie Research Facility, Toowoomba, QLD 4350, Australia
- ⁵ Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St Lucia, Brisbane, QLD 4072, Australia

Introduction

Over the last two decades, around 150 studies identifying almost 6000 QTL for over 220 traits have been published in sorghum. These studies have produced a large body of information concerning the genetic basis of these traits including their genomic location, allelic effects and epistatic interactions. However, this information has yet to be fully utilised by sorghum improvement programs and sorghum genetic researchers worldwide. This is due in large part to the heterogeneous nature of the data, the challenges of projecting QTL detected in one population onto another population and the capacity of researchers to determine if QTL alleles are shared by unrelated lines. Many of these problems are the result of the current unstandardized analogue method of transmitting QTL information via papers in scientific journals. Currently, journals do not impose a specific format for QTL data presentation and hence QTL information is not readily, or efficiently, searchable due to the heterogeneous nature of the data presentation across studies. As a result, individual researchers must expend a large amount of effort

to compare QTL across studies, and generally, this can only be done in a limited way for a small number of studies.

The discovery of gene order conservation among the grasses provides the opportunity to further dissect the genetic control of traits in large genome species by investigating the underlying syntenic genes in small genome species. Sorghum, with its small genome size, provides an important template for studying the genetic control of quantitative traits of closely related large genome crops such as maize, in addition to being complementary to rice as a grass genome model (Draye et al. 2001). However, the difficulty of making use of this information in species closely related to sorghum, such as maize and rice, is even greater due to the lack of ready access to syntenic QTL and candidate gene details across species.

What is required is a global database of QTL in each crop which allows researchers to readily search for and compare QTL not only within a species but between species. Given the lack of data standards and the general heterogeneity of data sets, creating such a data resource is a nontrivial task. A major complicating factor for the direct utilisation of information from QTL studies among the grasses is the variability in the precision of initial QTL identification. This variability is caused by many factors including population size and type, precision of the phenotyping, statistical analysis methodology and marker order, coverage and density. As a result, projecting QTL and their confidence intervals identified across multiple studies on a single genetic map or physical map is a major challenge. Such projections from multiple experiments must take into account factors that affect the reliability of the initial study, including population size, QTL effect size and statistical confidence intervals for QTL locations. To date, two studies have collated existing QTL information for sorghum and positioned the QTL on a common map; Mace and Jordan (2011) projected 771 QTL relating to 161 unique traits from 44 studies onto the sorghum consensus map, and Zhang et al. (2013) positioned the flanking markers of 604 QTL from 35 studies onto the sorghum genome assembly (v1.4). The later study did not adjust the resulting QTL CI to take into account factors influencing the reliability of the initial study. Comprehensive QTL databases for maize and rice are also limited in terms of content relative to published studies (e.g. multi-trait databases: MaizeGDB; Andorf et al. 2010, 2016; RiceQTLPro; Kim et al. 2014; Q-TARO Yonemaru et al. 2010; Gramene; Ni et al. 2009; single-trait databases: root traits in rice Courtois et al. 2009; plant height in maize Wang et al. 2006; disease resistance in maize Zhao et al. 2015), and as such, it is very difficult to quantify the total number of QTL identified in rice and maize, in comparison to sorghum. A search in Google Scholar on key terms including crop name, QTL and GWAS, returns $2.7\times$ the number of hits for maize in comparison to sorghum and $3.6\times$ the

number of hits for rice in comparison with sorghum. A study published over 10 years ago reported that over 7000 QTL had been identified in rice (Zeng et al. 2006). At the same time in sorghum, ~660 QTL had been identified. However, currently the most comprehensive QTL database available for rice, through Gramene (Ni et al. 2009), contains 8646 QTL, indicating that not all of the QTL identified in the last decade have been included. The most comprehensive QTL database available for maize, MaizeGDB (Andorf et al. 2010, 2016), contains 2294 QTL, which is similarly unlikely to represent all available QTL. Additionally, these existing QTL databases for rice and maize have not used a comparable QTL CI prediction and projection to the approach taken for the sorghum QTL. In addition to the precision of the initial QTL identification, any comparative QTL analysis also depends on the quality of QTL projection onto a common framework map. The publication of the reference whole-genome sequence of sorghum (Paterson et al. 2009), rice (Goff et al. 2002; Yu et al. 2002) and maize (Schnable et al. 2009) provides a unified coordinate system for each crop. Additionally, the increasing, almost exclusive, use of sequence-based markers for genetic linkage and association mapping studies, including RFLPs, SSRs, DArTs and SNPs, facilitates accurate determination of the peak location of the QTL, or significant SNP, on the physical genome, permitting the development of more comprehensive QTL databases.

Increasingly, over the last 5 years, quantitative trait dissection has been undertaken using association mapping approaches in contrast to standard QTL genetic linkage mapping in biparental populations. For example, of the 43 GWAS studies published to date in sorghum, 41 have been published in the last 5 years. The information generated through GWAS studies presents new opportunities and challenges for genetic researchers. It provides researchers with opportunities to increase the mapping resolution due to the increased amount of recombination available, as well as to identify more allelic diversity than in a traditional biparental population. However, factors such as sample size, population structure, unexpected linkage disequilibrium (LD), small effect sizes and low allele frequency remain a challenge for association mapping approaches and can result in a higher frequency of false-positive associations than in traditional QTL mapping studies. The potential to compare both GWAS and QTL outputs could therefore provide validation opportunities to increase confidence in putative signals across studies.

The Nested Association Mapping (NAM) approach was recently described (Buckler et al. 2009) in order to combine the advantages and reduce the disadvantages of both genetic linkage QTL mapping and association mapping. However, to date only two NAM studies have been published in sorghum (Mace et al. 2013; Bouchet et al. 2017), and only one in rice (Fragoso et al. 2017) in comparison

with over 10 in maize (e.g. Buckler et al. 2009; Benson et al. 2015; Cook et al. 2012; Kump et al. 2011; Poland et al. 2011; Tian et al. 2011; Wallace et al. 2014; Zhang et al. 2015). It is unlikely that any single QTL, GWAS or NAM study will be able to detect all the QTL influencing a complex trait, due to the statistical power of the analyses and the noise inherent in the complex biological and environmental systems involved (Yamamoto et al. 2009). Therefore, users increasingly need to conduct meta-analyses across studies, and even across species, to more comprehensively dissect the genetic architecture of complex traits. Such an approach is increasingly being used to dissect complex traits in humans. For example, GWAS meta-analyses studying intelligence (Savage et al. 2018), height (Wood et al. 2014) and body mass index (Locke et al. 2015) have all used sample sizes of over a quarter of a million individuals, with a recent meta-analysis study combining data across previous studies, totalling almost 1 million individuals, to conduct a GWAS for both height and body mass index (Yengo et al. 2018). In contrast, large-scale association mapping studies are much more difficult in plants due in large part to the impact of genotype by environment interactions. In many cases, it is impossible to measure or sensibly compare traits in plants even when the plants are grown in the same environment due to the dominating influence of phenology (driven by differences in photoperiod or vernalisation). As a result, the largest GWAS in plants reported to date involving only 12,000 accessions (Bandillo et al. 2015). Due to the difficulties in conducting association mapping studies at sufficient scale in plants, tools to facilitate meta-analyses across studies are critical. The database described here provides researchers with a new meta-analysis tool to integrate data across multiple studies and types of studies (QTL, GWAS and NAM) and across species to compare positions across studies and to determine allelic relationships among QTL, allowing for the more comprehensive dissection of the genetic architecture of complex traits. The accumulation of QTL information will allow researchers to more effectively mine the QTL landscape (the “QTLome”; Salvi and Tuberosa 2015), with the potential to enhance the impact of such studies on crop improvement activities. The inclusion of genes and genomic regions controlling more qualitative and polygenic traits (e.g. major effect genes frequently used for selection by breeders) is also a critical element of the described resource to provide comprehensive information for both complex trait dissection and molecular breeding implementation strategies. Additionally, genome synteny between sorghum, rice and maize will facilitate knowledge transfer from related species and the identification of candidate genes across species that potentially underpin QTL regions, providing the opportunity to evaluate the extent to

which genomic features and specific genes and QTL play common roles across the grasses thereby accelerating the translation of diverse and heterogeneous QTL information into crop improvement progress.

Materials and methods

Sampling strategy

QTL and association mapping data for sorghum were collected from 146 publications by conducting a bibliographic review and using keyword searches (e.g. QTL, mapping, genome-wide association, major effect gene). On a study-by-study basis, details of the genetic linkage maps were collated where relevant including population type and size, total number of markers mapped, number of linkage groups identified, overall average marker density, mapping function used and analysis methodology used, in addition to the details of the QTL, or significant SNP, including significance level, R^2 value, flanking or most significant marker and allele effect.

QTL projection steps

From 146 studies, 5844 individual QTL for 223 unique traits were included in the analysis. The 223 unique traits were classified into seven broad categories, modified from Mace and Jordan (2011) for the purpose of facilitating analysis and reporting; leaf, maturity, panicle, abiotic stress resistance, biotic stress resistance, stem composition, stem morphology, and are reported in Supplementary Material (ESM, Table S1).

Each instance of a QTL, or significant SNP, for each trait identified in a single environment was recorded as a unique QTL even if it co-located with additional traits or it was detected in more than one environment. The locations of the QTL and significant marker–trait associations reported were aligned to six different maps: the sorghum consensus map (Mace et al. 2009), sorghum genome assemblies v1.0, v2.0 and v3.0 (Paterson et al. 2009; McCormick et al. 2018), the maize genome assembly (B73 agpv2; Schnable et al. 2009) and rice, *O. sativa* subsp. *japonica* (Release 7; Goff et al. 2002; Yu et al. 2002).

The consensus map reported in Mace and Jordan (2011) consisting of 3272 unique loci spanning 1603.5 cM, with 2335 sequence-mapped markers was used as the first reference map for QTL projection. As described previously (Mace and Jordan 2011), the map location with the highest test statistic on the chromosome in the individual studies was regarded as the estimated location of a QTL from a particular study. In the case where the sequence location of the QTL CI was provided, these were used. Otherwise, the confidence intervals (CI) for the projected QTL were

then estimated as described previously (Mace and Jordan 2011) based on the following formulae:

$$CI = 530/(NR^2) \text{ for } F_2 \text{ (described by Darvasi and Soller 1997)}$$

$$CI = 163/(NR^2) \text{ for RI (described by Guo et al. 2006)}$$

where N is the number of lines in the mapping population and R^2 is the proportion of phenotypic variation explained by the identified QTL.

The framework of sequence mapped markers on the consensus map, originally described in Mace and Jordan (2011), was updated to include the physical locations on all three sorghum genome assemblies and enabled the predicted locations of the CIs to be estimated across multiple sorghum genomes. In the case of significant SNPs reported in GWAS studies, the cM location on the consensus map was predicted using the same framework map and a small window of 1 cM (0.5 cM either side of the predicted SNP location) was used to estimate the CI, to accommodate the heterogeneous nature of GWAS studies which impacts upon the accuracy and resolution of the analysis output and to facilitate comparison across studies.

The predicted gene models within the CI across all three sorghum genomes were collated (Sbi1.4, Sbi2.1 and Sbi3.1.1 gene sets) and information on selection signatures on individual genes included based on Mace et al. (2013, 2014), which identified signatures of purifying and balancing selection based on whole-genome resequencing of 47 individuals categorised into three groups (1. Wild and Weedy genotypes, 2. Landraces, 3. Improved Inbreds).

The syntenic locations of all QTL/SNPs on the maize and rice genomes were predicted as follows: the mean location of the QTL or significant SNP CI on the first v1.0 release of the sorghum genome was used to predict the mean syntenic location in maize and rice based on publicly available syntenic gene sets (Lyons and Freeling 2008; Schnable et al. 2012). Publicly available framework genetic linkage maps of sequenced mapped markers (Andorf et al. 2010; Harushima et al. 1998) were then used to predict the cM location in maize and rice, respectively, and a small window of 1 cM (0.5 cM either side of the predicted SNP location) was used to estimate the CI, to facilitate comparison across studies. The physical locations of the 1 cM CI were then predicted using the framework genetic linkage maps of sequenced mapped markers for both maize and rice. The predicted gene models within the CI for both maize and rice were collated based on publicly available data set; for maize the most recent gene model set (Zm0001d.2.) available via maizegdb was used and for rice, Os-Nipponbare-Reference-IRGSP-1.0 as described in Kawahara et al. 2013).

Software and Implementation

MySQL (version 5.6.39) was used as the underlying relational database and WordPress as the content management system. All data and tools are hosted on a CentOS Enterprise server available through <http://aussorgm.org.au/sorghum-qtl-atlas>. WordPress template files written in PHP were used to develop the customised Web-based query interface. WordPress wraps these files to keep the theme consistent with the rest of the site and enforce login permissions.

Results

Database content

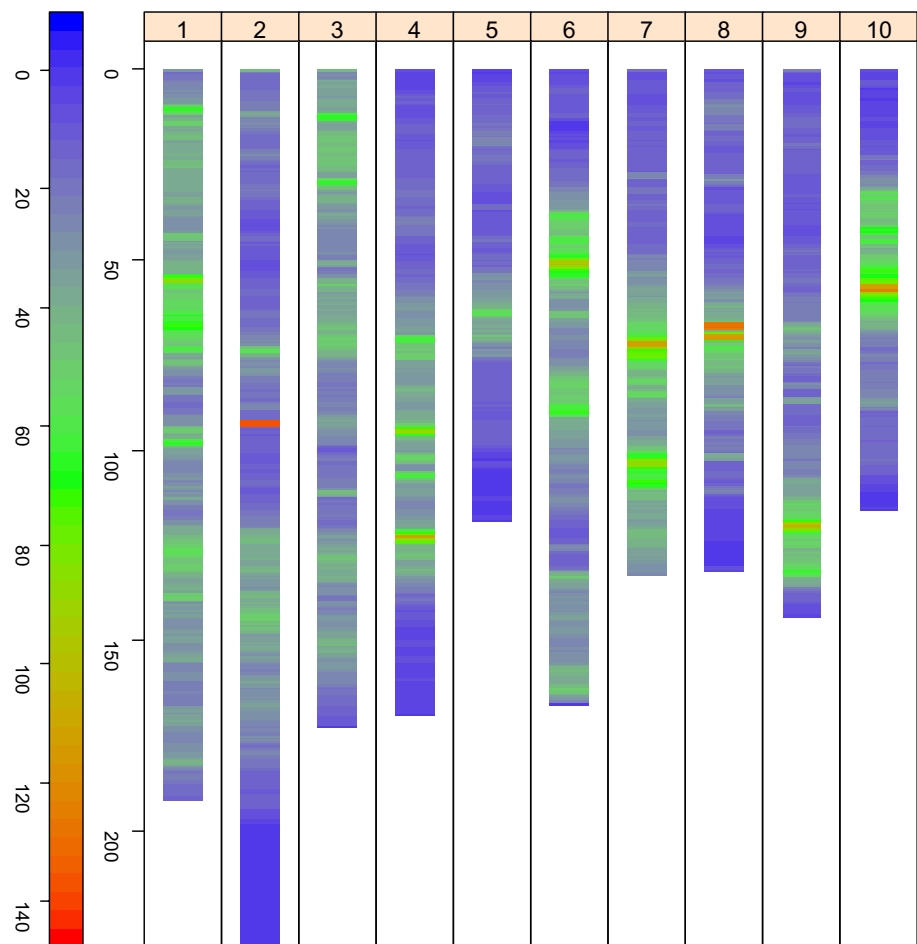
To date, sorghum QTL from over 140 studies are included that have utilised 82 unique biparental populations and 13 association mapping panels. From these 140 studies, 5843 individual QTL, or significant marker–trait associations, for 223 unique traits were projected onto four different sorghum reference maps, (1) the sorghum consensus map, (2) the sorghum genome assembly v1.0, (3) the sorghum genome assembly v2.0, (2) the sorghum genome assembly v3.0, in addition to the genome assemblies of both maize and rice.

The 5843 QTL were grouped into seven broad trait categories, 1753 related to the trait category Abiotic stress resistance, 1668 to Panicle, 753 to Stem morphology, 422 to stem composition, 407 to maturity, 394 to biotic stress resistance and 446 to the trait category Leaf (Table 1; Fig. 1).

Seven QTL hot spot regions were identified with a QTL density > 100 QTL/0.5 cM, on chromosomes SBI-02, SBI-04, SBI-06, SBI-07, SBI-08, SBI-09 and SBI-10. In three of these hot spot regions, a single-trait category accounted for $> 70\%$ of the QTL in the hot spot region, e.g. for the QTL hot spot region on SBI-02, over 90% of the QTL in this region were associated with the panicle trait category, and in particular grain fat and protein content, tannin content and kernel hardness. In contrast, the QTL hot spot on SBI-07 contained high numbers of QTL from multiple trait categories, and in particular abiotic resistance (e.g. stem biomass yield and leaf yield under cold temperatures), stem composition (e.g. vegetative yield and sugar yield) and panicle related traits (e.g. panicle architecture). Three cold-spot regions were identified with a QTL density of ≤ 1 QTL/0.5 cM, on the distal ends of chromosomes SBI-02, SBI-05 and SBI-08. Clustering of QTL within a trait category was observed most frequently for Biotic Resistance traits; e.g. of the 36% of the sorghum genome implicated in quantitative disease response (QDR) to fungal pathogens, half of the total QDR genomic space consisted of co-localising QTL for the same trait. As noted previously (Mace et al. 2014), the majority ($> 90\%$)

Table 1 Number of sorghum QTL per chromosome and per trait category

Chromosome	Leaf	Maturity	Panicle	Resistance abiotic	Resistance biotic	Stem composition	Stem morphology	Grand total
SBI-01	107	63	276	246	50	58	102	902
SBI-02	23	45	319	174	48	36	34	679
SBI-03	56	34	159	240	39	53	83	664
SBI-04	41	29	284	125	34	32	55	600
SBI-05	20	19	70	103	21	15	31	279
SBI-06	48	84	177	145	43	73	147	717
SBI-07	43	19	126	127	23	58	115	511
SBI-08	57	29	87	269	22	18	39	521
SBI-09	21	40	70	162	76	57	99	525
SBI-10	30	45	100	162	38	22	48	445
Grand total	446	407	1668	1753	394	422	753	5843

Fig. 1 QTL density plots (number of QTL/0.5 cM) along the sorghum genome. The scale bar to the left indicates the length of each chromosome in cM based on the sorghum consensus map

of co-localising QTL conditioned resistance to multiple diseases and pests with a hot spot region on SBI-04 where the mid-points of 25 QTL for resistance to head bug, shoot fly, green bug, rust, anthracnose, stalk rot and grain mould resistance map mapped within 20 cM on SBI-04.

In total, 9325 syntenic locations were predicted for 5745 sorghum QTL in maize, and 6057 syntenic locations were predicted for 5813 sorghum QTL in rice (ESM Figures S1 and S2).

Web-based tools

The database homepage contains background information and links to the relevant literature about the development of the database content. The homepage allows users to select different options for querying the database (ESM Figure S3).

- Search via trait
- Search via genomic location
- Search via study
- Search via syntenic location
- Advanced search

Search via trait allows users to select search terms which can be entered at the level of trait category, trait subcategory or trait description. For example, to search for QTL associated with grain weight users can either enter the search term “Grain weight” in Trait Description, or select the trait category “Panicle” or Trait subcategory “Grain & Panicle Yield”, and search in the Trait Description options to select grain weight-related traits.

Search via Genomic Location allows users to search for a specified physical location on any of the three available sorghum genome assemblies in addition to the genetic linkage scale on the sorghum consensus map. The location can be specified as a set distance from a target location or with a defined start and end location.

Search via Study allows users to search for QTL identified in specific populations, in specific publications (accessible through author names, publication title or keywords, the journal name, year of publication), the original published QTL identifier (e.g. stg A, Sb-HT9.1, qGY-5), flanking markers detailed in the original publication (e.g. txp302, Dsenhsbm64, S1_10629136), and to limit the search to only QTL, GWAS or major effect gene studies.

Search via Syntenic Location allows user to QTL in sorghum based on the syntenic location in either maize or rice. Users can specify the location in either the maize genome assembly (B73 agpv2; Schnable et al. 2009) or the rice genome assembly, *O. sativa* subsp. *japonica* (Release 7; Goff et al. 2002; Yu et al. 2002), or provide a candidate gene ID from either species (e.g. GRMZM2G108663; LOC_Os01g01080.1).

Additionally, the advanced search option allows users to combine elements from all four searches above (via trait, genomic location, study and syntenic location).

The search output is displayed in a downloadable summary QTL results table that can be further interrogated by clicking on any individual row, which provides a link to the QTL details page. The QTL details page provides trait hierarchy information, general information (the population used for the QTL identification, the significance value and measure used, e.g. $\min \log_{10}P$, LOD, p value (\pm adjustments made

for FDR/Bonferroni correction, as detailed in each study), the additive effect, a hyperlink to the publication page), information about the location of the QTL. The location information includes the physical position of the QTL CI on all three genome assemblies, with a hyperlink to the list of predicted gene models under each CI, in addition to information on the genetic linkage (cM) coordinates on the sorghum consensus map. It also includes information on the flanking markers and information on the predicted syntenic location in maize and rice, with hyperlinks to a downloadable list of predicted gene models in the rice genome underlying the QTL CI, in addition to links to the genome browsers (for rice at MSU (<http://rice.plantbiology.msu.edu/>; and for maize at maizegdb <https://gbrowse.maizegdb.org/>).

Application examples

Do QTL identified for a selected trait in one population or environment corresponds to those identified in other populations or environments?

Clustering of QTL within a trait category was observed most frequently for Biotic Resistance traits. In particular, half of all the QTL associated with quantitative disease response (QDR) to fungal pathogens co-localised, often for the same trait. However, clustering of resistance responses across pests and diseases was also observed by Mace et al. (2014) with a hot spot region on SBI-04 where the mid-points of 34 QTL for resistance to head bug, shoot fly, green bug, rust, anthracnose, stalk rot, ergot, Southern root-knot nematode and grain mould resistance map mapped within 20 cM on SBI-04. As speculated previously, this could be due to single gene effects, whereby the resistance gene and QTL are allelic, or by the effects of clusters of genes. A hot spot region on SBI-09 was also identified where the mid-points of 43 QTL for resistance to, rust, rice weevil, striga, stalk rot and midge resistance co-located within 20 cM on the long arm of SBI-09. In both of these hot spot examples, resistance traits of importance to other species were identified, e.g. southern root-knot nematode (*Meloidogyne incognita* race 3) resistance for maize, and rice weevil (*Sitophilus oryzae*) resistance for rice, providing opportunities to investigate the syntenic regions in maize and rice for evidence of association with biotic resistance traits.

For a selected chromosomal region where a QTL has been detected, what other traits have been associated with the same region?

In many crop species, intense selection is practiced in early generations for major effect genes such as morphological traits, major gene resistance or phenology. This strong selection will have strong effects on the frequency of closely

linked genes in the genome. These impacts can be heightened where major effect genes are linked. Typically, breeders have little or no idea of the potential impact of this selection on other genes or QTL. To illustrate this issue, Mace and Jordan (2010) highlighted the impact on QTL allele frequencies using a simulation study on a 100 cM region on the long arm of SBI-03 containing both stay-green QTL and major effect genes for grain colour and the presence of awns. Different selection strategies were investigated for the impact on recovering the beneficial stay-green QTL alleles (Fig. 2), and the study demonstrated that strong selection pressure applied for red grain colour in early generations resulted in significantly lower frequencies of the beneficial stay-green QTL alleles.

In the public sorghum breeding program in Australia, breeders practiced strong selection for the dominant allele of grain colour (R) and by doing so effectively reduced the frequency of the beneficial alleles for stay-green 2 to very low levels despite the stay-green trait being a major selection target of the breeding program (unpublished data). In the light of our knowledge of the association between the major effect genes for grain colour and awns and the two stay-green QTL in this genomic region (*stg1* and *stg2*), alternative and more effective molecular breeding strategies can be identified and deployed, as described previously (Mace and Jordan 2010).

This example on SBI-03 has potential impact for closely related species. The *stg2* QTL in sorghum is syntenic with the maize QTL stay-green 3 (*qstgr3* in MaizeGDB) and has the potential to provide insights into unforeseen consequences, due to linkage drag, of selection decisions based on genes or QTL that are closely linked. Simulation studies such as the one described here may help determine whether closely linked traits are due to allelism or pleiotropic effects and underline the potential risks and opportunities available when breeders and researchers have access to high-resolution data for specific genomic regions of interest.

Do QTL locations identified in one species correspond to QTL or candidate genes detected in corresponding regions of other related species?

The QTL atlas facilitates the identification and prioritisation of candidate genes through the identification of all predicted gene models underlying each QTL CI, not only across the three sorghum genome assemblies but also in maize and rice. Inclusion of the Gene Ontology (GO) terms and the sequence diversity statistics generated previously (Mace et al. 2013) for all predicted gene models, in addition to selection signatures, provides opportunities to prioritise candidate genes based on gene sequence variability and over-representation of GO terms or specific gene families.

1. Maturity

As proof of concept, we compared the predicted syntenic location of the major effect sorghum maturity locus, *Ma3* (Childs et al. 1997), known to be encoded by PHYB, with the location of the orthologous PHYB genes in maize and rice (Fig. 3a). The predicted syntenic location of the CI for *Ma3* was found to co-locate with PHYB in maize (GRMZM2G124532) and rice (LOC_Os03g19590). Further to this, a syntenous maturity QTL identified in both sorghum and maize NAM resources (Mace et al. 2013; Buckler et al. 2009), which also co-located with a candidate gene, was investigated to compare the predicted syntenic location in maize with the reported significant marker–trait association from Buckler et al. (2009). The significant marker PZA00402.1 associated with days to anthesis in maize (Zm Chr 3 at 223,179,497 bp) co-located within the predicted syntenic CI location (Zm Chr3 222,942,597–223,384,631 bp), in addition to containing the orthologous photoperiod candidate gene identified in sorghum, CHLOROPHYLL A/B BINDING PROTEIN (CAB) (Fig. 3b).

2. Fertility restoration

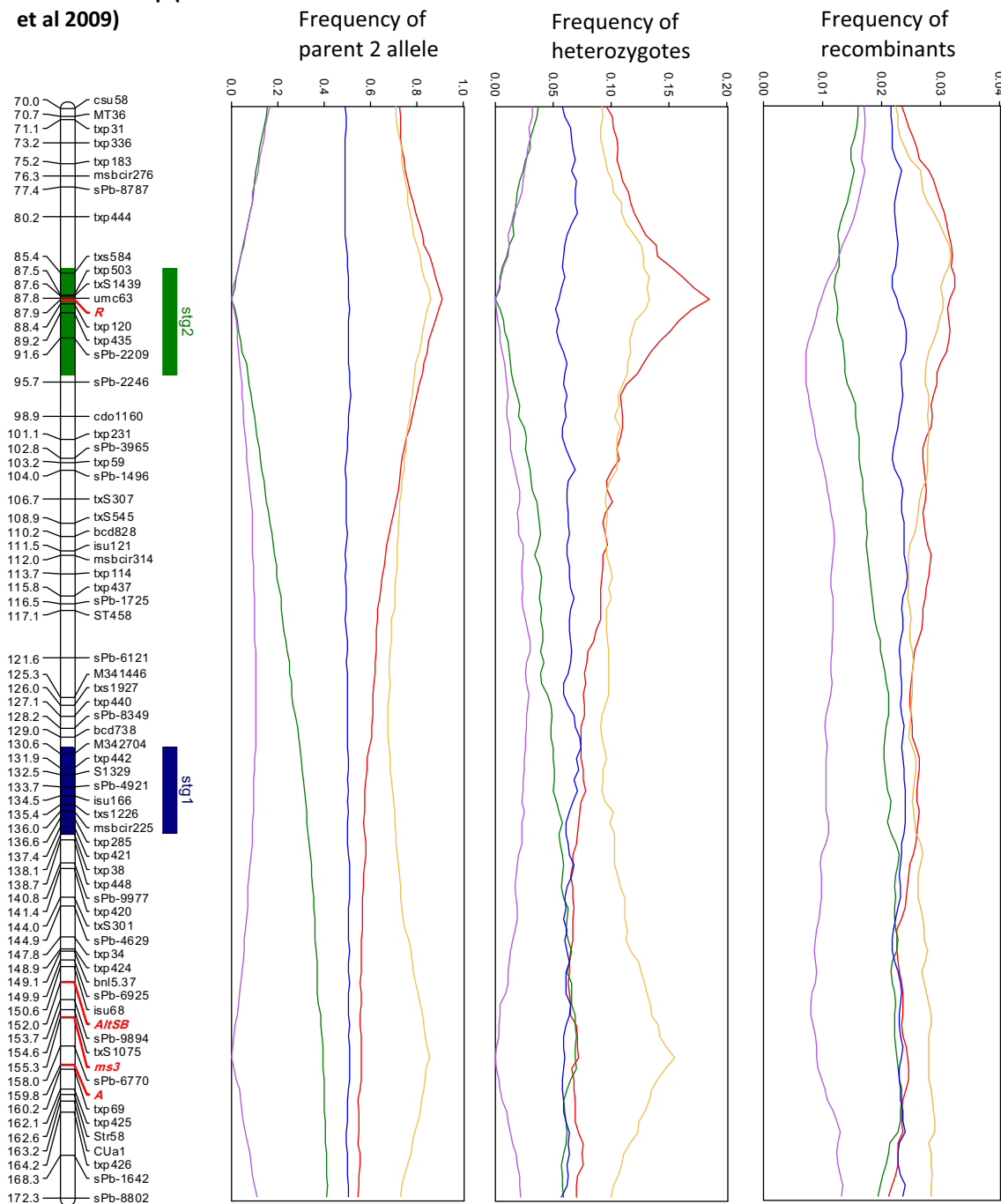
The genomic space in sorghum associated with fertility restoration QTL was found to be significantly enriched for PPR-encoding genes (X^2 p value 0.00035), with only 1 of the 17 QTL described in the literature without a PPR gene within the QTL CI. The predicted syntenic locations of the sorghum fertility restoration QTL in maize and rice were also found to co-locate with PPR genes, e.g. the CI of the *Rf5* locus in sorghum (Jordan et al. 2011) co-locates with 8 PPR genes and the predicted syntenic locations in maize and rice co-locate with a single PPR gene (Fig. 4). The PPR gene within the maize CI has recently been identified as *Emp11* (Empty pericarp11) shown to be essential for proper seed development in maize (Ren et al. 2017). Users are also able to use the maize gene identifier for *Emp11* (GRMZM2G353301) as the input search parameters and interrogate all the sorghum QTL that have been identified in the predicted syntenic location in sorghum.

Can a meta-analysis across studies improve the resolution of the QTL/GWAS output for selected traits?

A meta-analysis across studies can provide opportunities for increasing the accuracy of the QTL location, and hence the underlying candidate gene. As proof of concept, we looked at the QTL for height that have been identified within 4 Mb of *dw3*, the causative gene (Multani et al. 2003). In total, 70 QTL for height have been identified from 26 studies. The average distance of the QTL mid-point from *dw3* was

Results from simulation study

Consensus map (Mace et al 2009)



◀**Fig. 2** Genetic region on SBI-03 with two major effect stay-green QTL highlighted in addition to major effect genes for grain colour (R) and awns (A), together with the results from a simulation study to investigate the impact on stay-green QTL frequency across five different selection strategies. Figure modified with permission from Mace and Jordan (2010) (color figure online)

1.128 Mb across the 70 QTL (Fig. 5); however, by collating and averaging all of the existing height QTL data across studies, a target region approximately 250 Kb from the causative gene can be identified increasing the resolution 4.5-fold (Fig. S4).

Discussion

The Sorghum QTL Atlas is a high-resolution, open-access research platform to facilitate candidate gene identification across three cereal species, sorghum, maize and rice. The platform provides users with multiple QTL search options and resource links, both to external sites, and to downloadable data, as shown in Figure S3. The versatility of the search options, the comprehensiveness of the database content and the simplicity of the user interface provide a valuable bridge between genomics and applied breeding. Such a resource for searching for and comparing QTL locations across studies, traits and species is more efficient than the current model for sharing QTL information through journal publications which do not impose data format standards and requires significant time for individual researchers to mine the available literature. The time required and inefficiency of the current model of QTL information sharing is reflected in the limited number of QTL and/or GWAS studies that effectively review the published QTL literature in a species, with even fewer including cross-species comparisons. The described QTL Atlas can leverage analyses conducted in different populations and species to provide increased power for complex trait dissection. The inclusion of predicted syntenic locations of sorghum QTL CI in two closely related species, maize and rice, in the QTL Atlas, in addition to all predicted gene models within the syntenic QTL CI further strengthens the application of genomics and QTL data to powerful genomics-assisted methods for cereal crop improvement. Together the atlas enables user to address a range of key questions including:

- Do QTL identified for a selected trait in one population or environment correspond to those identified in other populations or environments?
- For a selected chromosomal region where a QTL has been detected, what other traits have been associated with the same region?

- Do QTL locations identified in one species correspond to QTL or candidate genes detected in corresponding regions of other related species?
- Can a meta-analysis across studies improve the resolution of the QTL/GWAS output for selected traits?

The first three of these questions were identified as critical for maximising the utility of QTL data over 20 years ago, before either comprehensive data sets or structured databases were available (e.g. Byrne et al. 1995; Bigwood 1997). The questions are still highly relevant today, as acknowledged by more recent studies describing QTL resources (e.g. Lawrence et al. 2007) and enable users to develop new hypotheses for QTL in addition to leveraging the power of comparative genomics to identify candidate genes for complex traits.

Dynamic approaches characterise the identification of significant marker–trait associations

Cataloguing, standardising, summarising and integrating information across studies are very time consuming. Further work is required by global crop research communities to standardise trait ontologies across existing databases and future studies which would allow for automated data collection, integration and meta-analysis, including adopting standardised approaches for identifying and reporting significant marker–trait association and for controlling the rate of false discoveries (FDR). However, the platform presented in this paper is unique in integrating data in a standardised format across traditional QTL studies and association mapping studies in plants. To date only a very limited number of databases have attempted to do this, with most progress being made in animal research, e.g. Animal QTLdb which was recently updated to include GWAS studies in addition to QTL data, and e-QTL (Hu et al. 2013), and is described by the authors as an animal QTL/association database which serves as a bridge between genotypes (genes) and phenotypes (traits). The comprehensive and effective synthesis across studies, however, is critical for maximising value and potential impact of the QTL information in crop improvement activities. Any synthesis must therefore provide opportunities to assess the reliability of the component studies, including the significance thresholds used. The QTL search tools enable users to compare QTL across studies; users can select either specific traits to compare within a population or select the same trait to compare across populations and studies. This enables the identification of QTL clusters for key traits, e.g. QDR traits, and enables users to identify common versus population–specific traits.

Combining QTL from previous studies facilitates the identification of genomic regions that are enriched for QTL, also known as QTL clusters, across traits. In most cases, it

Fig. 3 Comparison of days to flowering QTL in sorghum (in dark green) with predicted syntenic CI QTL locations (in light green) and syntenic gene models across species (represented by horizontal black bars), with known candidate genes highlighted in red. **a** Major effect maturity locus Ma3 in sorghum, and causative gene PHYB, and the predicted syntenic QTL CI locations in maize and rice, and orthologous PHYB genes. **b** Days to flowering QTL identified in the sorghum NAM population and underlying candidate genes and predicted QTL CI in comparison to the significant marker identified in Buckler et al. (2009) (represented as a black bar to the left of the maize chromosome), and the underlying candidate genes (color figure online)

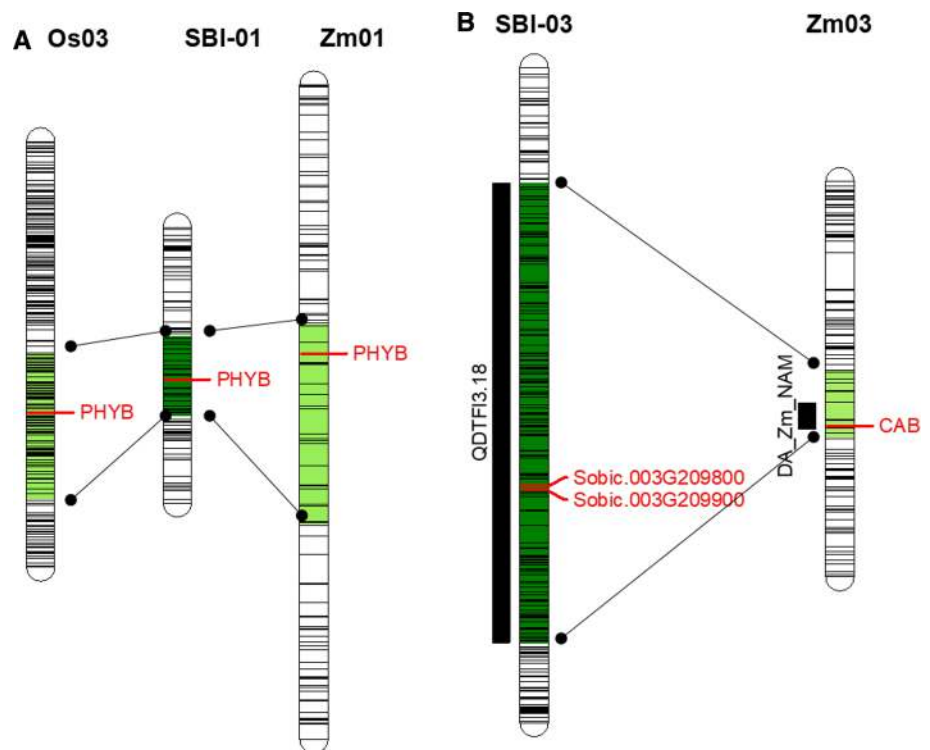
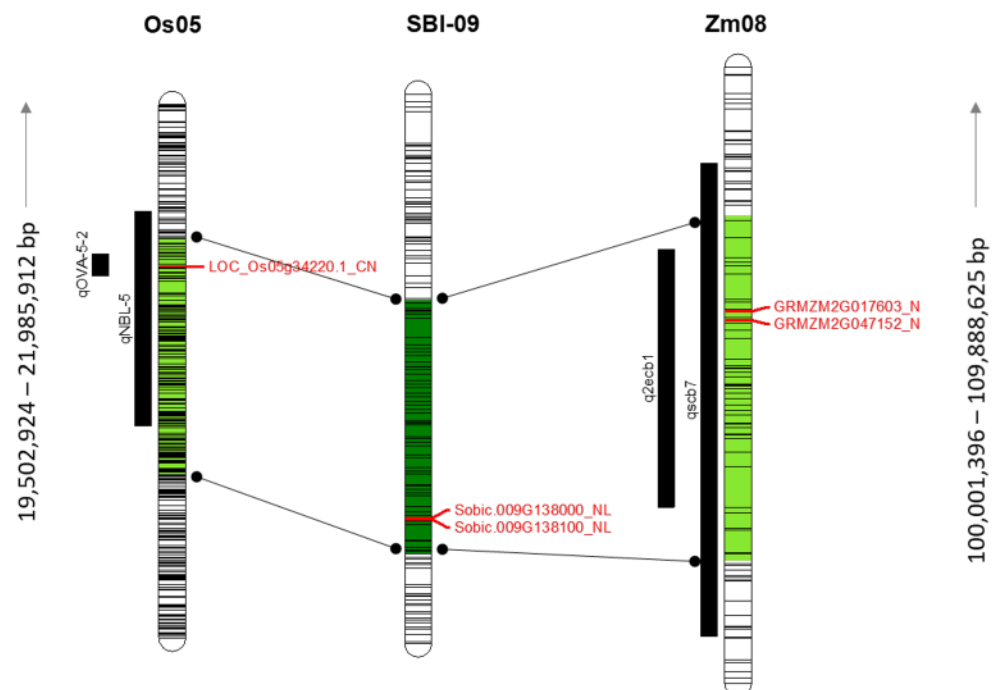


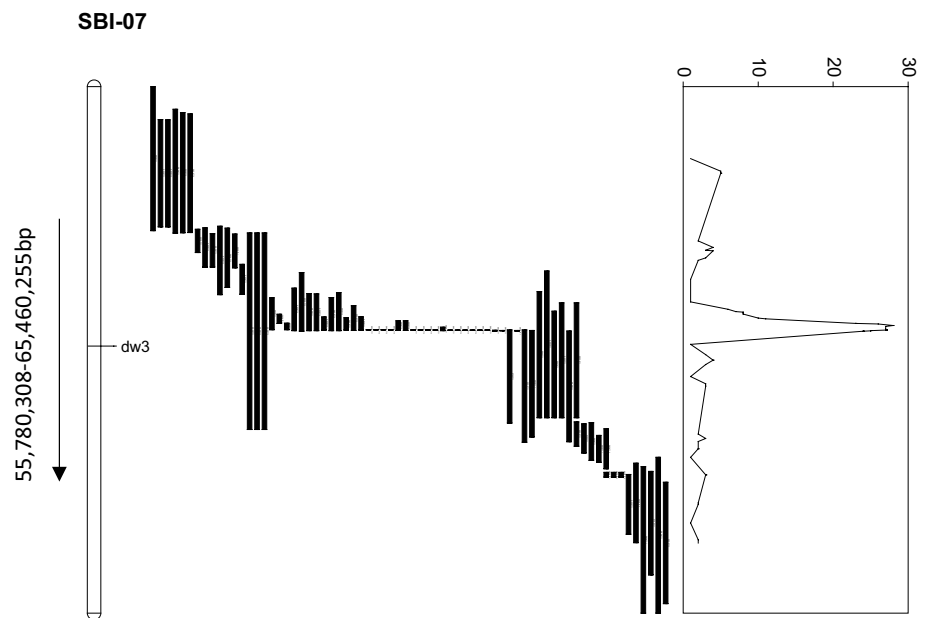
Fig. 4 The CI of the Rf5 locus in sorghum (Jordan et al. 2011) co-locates with 8 PPR genes in sorghum, highlighted in red, and the predicted syntenic locations in maize and rice co-locate with a single PPR gene (color figure online)



has not yet been determined whether QTL clusters associated with multiple traits represent the pleiotropic effects of a single gene or close linkage of different genes affecting different traits. However, the synthesis of standardised, multi-trait, multi-study information in platforms such as this will provide further insights into potential pleiotropic effects

which could explain physiological or functional interactions between traits. For example, the chromosomal segment on SBI-02 in sorghum containing the B2 gene controlling the presence of the high-tannin testa layer in the sorghum grain has been associated with QTL for grain quality (Rami et al. 1998), grain mold resistance (Upadhyaya et al. 2013), grain

Fig. 5 Distribution of 70 height QTL within 4 Mb of dw3 on SBI-07



composition (Shakoor et al. 2016; Rhodes et al. 2017) and cold tolerance (Knoll et al. 2008) and it is likely that the high tannin content in the sorghum contributes to grain mold resistance; (2) the chromosomal segment on SBI-07 containing the dw3 gene controlling plant height co-locates and likely has a pleiotropic effect on QTL for stem juice yield (Guan et al. 2011; Murray et al. 2008a), panicle exertion (Zhao et al. 2016), biomass (Guan et al. 2011), panicle architecture (Pereira et al. 1995; Brown et al. 2006) and lodging (Murray et al. 2008b). Observations of major genes' pleiotropic effects across crops have been made, e.g. the gene in *Arabidopsis* that controls cuticular wax biosynthesis, CUT1, was found to have a pleiotropic effect resulting in conditional male sterility, which Millar et al. (1999) determined was caused by the absence of waxes in the trypine layer of the pollen grain disrupting pollen-pistil interactions. A database that integrates QTL information across traits is critical in order to identify potential pleiotropic interactions. However, to date comprehensive, up-to-date multi-trait databases are very rare for crop species. Even single-trait database availability is limited, e.g. root traits in rice, Courtois et al. 2009; plant height in maize, Wang et al. 2006; disease resistance in maize, Zhao et al. 2015. Additionally, such single-trait databases do not provide the opportunity to look for potential pleiotropic interactions across traits. Knowledge of QTL clusters can also be valuable for exploiting heterosis in hybrid breeding. Breeders can target complementary beneficial QTL allele accumulation in male and female germplasm pools separately, that can be deployed in hybrid combination resulting in maximal expression of heterosis in F_1 hybrids.

The synthesis and interpretation of QTL information that is enabled by the sorghum QTL atlas platform potentially allows sorghum breeders and researchers to focus their

efforts on high-confidence genomic regions with the highest percentage of phenotypic variances/large effect sizes, as well as providing opportunities to investigate potential unintended consequences that could arise if a particular genomic region is used in selection decisions for crop improvement. This is very important as reliance on individual studies, and individual populations, can be problematic, due to both the limited genetic diversity, and allelic variation, studied in a single population and the possible inclusion of false positives in results. For example, the majority of sorghum QTL and GWAS publications (118) only studied a single population, with the remaining 22 publications studying between 2 and 5 populations. In addition to the use of low statistical threshold values, false-positive QTL can also be identified because of small population sizes, as discussed by Beavis (1994). In the sorghum studies included in this resource, the population sizes ranged from 70 up to 400, with an overall increase in population size in QTL mapping studies in the last two decades from an average of 98 in 1997 up to 214 in 2018. Association mapping methodologies have been used increasingly for quantitative trait dissection over the last 5 years. This provides an opportunity to sample a broader range of allelic variation. Additionally in the sorghum literature, the average size of the association panels was almost twice that of the mapping populations, with the panel sizes ranging from 107 to 2214. The two largest panels used to date are the two sorghum NAM panels developed (Mace et al. 2013; Bouchet et al. 2017). Marker density increased in both QTL and GWAS studies in the last 3 years with the increasing use of GBS technology, with QTL mapping studies published after 2012 having an average marker density of over 6300, in comparison with just 170 markers per study pre-2012. GWAS studies had an average marker density

tenfold greater than QTL mapping studies (80,000 markers per study). GWAS studies also reported, on average, double the number of significant marker–trait associations reported per study in comparison to QTL studies (82 per GWAS study vs 28 in QTL studies). The average number of QTL reported per trait per study was 5.8, slightly higher than reported for other meta-analysis studies, e.g. 4.6 reported by Chardon et al. (2004).

In order to manage the challenges raised by heterogeneity among QTL studies, meta-analysis methods have been developed (e.g. Goffinet and Gerber 2000; Arcade et al. 2004; Veyrieras et al. 2007). These enable the integration of multiple diverse genetic linkage maps and QTL onto a single reference map through iterative projections. However, such approaches typically require large numbers of QTL per trait and also per genomic region in order to estimate the most likely location of a QTL in a specific region, and to date the approaches also do not accommodate GWAS output. As GWAS output represents over half (3200) of the compiled QTL Atlas content, a meta-analysis approach would not capture the extent of the genomic regions associated with a specific trait and hence would not contribute to increasing the resolution in the QTL region or increasing the value of the data included in this resource. We therefore decided it was unsuitable for inclusion in the analysis of the QTL and GWAS outputs for this resource. This highlights the need, however, for analysis methods that allow GWAS and QTL output to be integrated in order to increase the power of meta-analysis approaches.

Leveraging the power of multiple studies for complex trait dissection

One of the first steps taken when interpreting the results of a QTL or GWAS study is to determine which genes fall either within the QTL CI or with a pre-defined LD window around the SNP peaks. This generally involves manually going through output files to identify the coordinates of the SNPs or CIs, and then using these to filter large lists of genes. This process can be prone to errors, particularly when dealing with large output files from GWAS studies, as well as the difficulties of dealing with multiple peaks with complex traits. The QTL atlas provides not only all of the predicted gene models underlying the CIs and within a 1-cM LD window around significant GWAS peaks, but also the predicted gene models within a 1 cM window of the syntenic location in both maize and rice. The extent of the physical genomic region captured by the 1-cM window varies throughout the genome in all 3 species, influenced highly by recombination rate variation across the genome. For example, in the heterochromatic regions in sorghum, 1 cM is on average represented by approximately forty-times the physical distance (8.46 Mbp) than 1 cM in the euchromatin (0.22

Mbp). The use of a standardised 1-cM window, rather than a standardised physical window defined by base-pair coordinates, has attempted to take into account the variation in recombination frequencies which influence linkage disequilibrium throughout the genome. However users should be aware that in some contexts, a 1-cM window may be either too large (e.g. in a large diversity panel) or too small (in a structured population) and in these situations, the predicted gene models reported may need to be extended or reduced.

The multi-species nature of this platform is novel. Very few attempts have been made to integrate QTL information across plant species. A comparative QTL database has been reported for Saccharinae grasses (Zhang et al. 2013), however, at the time of publication of this study, it was not available. TropGeneDB has recently been developed containing data on molecular markers, QTL, maps, genotypes and phenotypes for 10 tropical crops (Hamelin et al. 2013); however, the database does not contain functionality for across species QTL comparisons. A mammalian multi-species QTL database has been developed for mice, rat and human QTL for the analysis of underlying candidate genes (Star et al. 2006), in addition to a broader Animal QTLdb (Hu et al. 2013), including QTL information from cattle, chicken, pigs, sheep and rainbow trout, for comparison of QTL both across and within species studies.

One of the major goals of QTL and GWAS studies is to identify the causative gene responsible for the phenotypic variation observed for the particular trait. The integration of results from different mapping studies within and across species can facilitate the identification of overlapping genomic regions which, when combined with information on diversity and selection of all the predicted gene models in the region in addition to orthologous candidate genes in closely related species, provides an opportunity to further mine the wealth of mapping data available for quantitative trait dissection. A meta-analysis across studies can also provide opportunities for increasing the accuracy of the QTL location, and hence the underlying candidate gene, as indicated with the 4.5-fold increase in resolution around the height gene, *dw3* on SBI-07, through a meta-analysis involving 70 height QTL from 26 studies. This approach is increasingly being used in human GWAS studies, e.g. a meta-analysis combining data on educational attainment from multiple sources and studies has now reached over 1 million subjects making it the largest GWAS sample ever assembled, providing a very powerful predictive instrument (Martin 2018). In plants the influence of the environment, which drives the high levels of context dependency observed in plant phenotypes, means that a meta-analysis approach will be critical to achieve more powerful mapping studies, with high resolution and precision. Comparative genomics also has a critical role to play to facilitate causative gene identification, with evidence from classical genetics in maize suggesting that the vast majority

of visible plant phenotypes result from genes conserved at syntenic locations across the grasses, or in other words syntenic conservation between species has been shown to be associated with function (Schnable 2015). However, varying linkage equilibrium signatures genome-wide impacts on candidate gene identification and prioritisation within a given QTL CI. Previous studies have estimated that LD largely decays in sorghum between 10 and 50 kb, depending on the germplasm sets studied (Mace et al. 2013; Morris et al. 2013; Hamblin et al. 2005; Wang et al. 2013). However, LD patterns differ significantly throughout the genome and for any given QTL CI, the extent of surrounding LD blocks should be considered. Comparative genomics, however, can provide opportunities to exploit low LD regions from related species to narrow down the list of candidate genes. For example, a recent study (Mace et al. 2013) comparing the maturity QTL identified in the sorghum and maize NAM populations identified examples where the QTL identified in one species occurred in a region of low recombination and the syntenic QTL in the other species occurred in a region of high recombination, providing higher resolution for candidate gene identification. Figure 6 highlights an example of four mature QTL identified in sorghum in a region of high recombination and a single syntenic maturity QTL identified in maize (Buckler et al. 2009) in a region of low recombination, demonstrating how information from one species, in this example sorghum, can be used to inform the other, in this case maize.

The number of studies reporting QTL and GWAS in maize and rice has grown rapidly and total several fold than that available in sorghum. However, due to the limited available of comprehensive multi-trait QTL databases for maize and rice, with adjusted CIs taking into account factors influencing the reliability of the original study, we have not attempted to incorporate the maize and rice QTL information into the QTL Atlas. Users can instead search using the QTL locations identified for selected traits in maize and rice to find the QTL identified in sorghum in the syntenic locations. However, it is critical that global databases of QTL and GWAS outputs are established for each species to enable researchers to readily search for and compare QTL within and across species in the future.

Future developments

The goal is to keep the QTL Atlas current and useful to the cereal community by providing regular updates with QTL and significant marker–trait association from succeeding QTL and GWAS studies, in addition to updating QTL projection to the most recent genome assembly releases across species. In addition to search functions, the sorghum QTL Atlas also contains a contact form to enable users to alert us about new publications and to

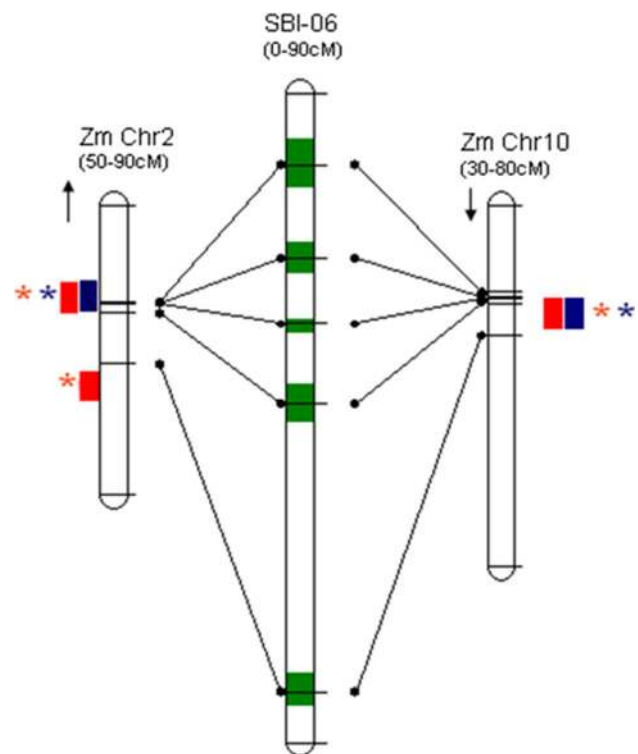


Fig. 6 Sorghum/maize flowering time QTL synteny between sorghum SBI-06 and maize chromosomes 2 and 10. Sorghum QTL identified in Mace et al. (2013) indicated as green segments on sorghum chromosome. Projected locations of the mid-point of the sorghum QTL onto the syntenic maize chromosomes indicated by lines between chromosomes. Locations of the maize DA (blue) and DS (red) significant markers identified in Buckler et al. (2009) indicated. Co-ordinates of segments of maize chromosomes detailed (based on maize NAM map) and orientation of maize chromosomes, relative to sorghum chromosome, indicated by arrows. Asterisks indicate maize QTL with mid-points less than 10 cM from the location of the sorghum QTL mid-point. Figure modified with permission from Mace et al. (2013) (color figure online)

ensure that it is integrated as rapidly as possible into the atlas. Additionally we will investigate the use of new tools, such as QTLTableMiner++ (Singh et al. 2018), in addition to in-house customised programming scripts, to automate data collation from diverse sources. Direct data submission has not been provided to ensure consistency in data standards and in QTL CI prediction methodologies. Maintaining accurate up-to-date content will be critical for meta-analysis studies that will increase the power and precision for candidate gene identification. This is particularly important in plant species, where genetic by environmental (G×E) interactions drives high context dependencies in plant phenotypes, and where the number of individuals included in mapping studies are many folds less than in human studies, e.g. the largest GWAS study in plants to date used 12,000 individuals in comparison to over 1 million individuals in humans. Meta-analyses in plants will

be critical to achieving higher power and precision for mapping complex traits.

There are also future opportunities to identify causal loci within candidate genes by linking to available whole-genome resequencing data across multiple genomes and multiple species. With the increase in sequencing studies within and across species, researchers will have access not only to high-resolution sequence data genome-wide but also to estimated haplotype information across many thousands of individuals (e.g. using the Practical Haplotype Graph approach; Buckler et al. 2018). Linking haplotype data to QTL and/or GWAS studies will provide significant power to improve association mapping studies going forward. This will provide powerful approaches for causal loci identification within QTL CI through SNP identification between parental lines, haplotype construction across diverse lines, and Genomic Evolutionary Rate Profiling Scores (GERP) analysis across species to identify deleterious mutations. Comparable QTL and GWAS databases across species will be a critical step for maximising the power of comparative genomics approaches to complex trait dissection and deployment in crop improvement activities.

Author Contribution statement EM collated the QTL and GWAS data and wrote the manuscript, XW and YT projected QTL locations and co-wrote the manuscript, CH and AH undertook statistical analysis of the data, DI, JB and MH developed the web application and DJ provided the original concept and co-wrote the manuscript.

Acknowledgements We would like to acknowledge funding support for this activity from the University of Queensland and the Department of Agriculture and Fisheries.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Andorf CM, Lawrence CJ, Harper LC, Schaeffer ML, Campbell DA, Sen TZ (2010) The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. *Bioinformatics* 26:434–436. <https://doi.org/10.1093/bioinformatics/btp556>

Andorf CM, Cannon EK, Portwood JL, Gardiner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalasu VV, Huerta M, Cho KT, Wimalanathan K, Richter JD, Mauch ED, Rao BS, Birkett SM, Sen TZ, Lawrence-Dill CJ (2016) MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Res* 44:1195–1201. <https://doi.org/10.1093/nar/gkv1007>

Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J (2004) BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* 14:2324–2326. <https://doi.org/10.1093/bioinformatics/bth230>

Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, Lorenz A (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome* 8:3. <https://doi.org/10.3835/plantgenome2015.04.0024>

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Wilkinson DB (ed) *Proceedings 49th annual corn and sorghum research conference*, American Seed Trade Association, Chicago, IL, pp 250–266

Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ (2015) Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. *PLOS Genet* 11(3):e1005045. <https://doi.org/10.1371/journal.pgen.1005045>

Bigwood DW (1997) Compilation and distribution of data on complex traits. In: Paterson AH (ed) *Molecular dissection of complex traits*. CRC Press, New York, pp 175–184

Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, Tuinstra M, Morris GP (2017) Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics* 206:573–585. <https://doi.org/10.1534/genetics.116.198499>

Brown PJ, Klein PE, Bortiri E, Acharya CB, Rooney WL, Kresovich S (2006) Inheritance of inflorescence architecture in sorghum. *Theor Appl Genet* 113:931–942. <https://doi.org/10.1007/s00122-006-0352-9>

Buckler ES et al (2018) Practical haplotype graph. <https://bitbucket.org/bucklerlab/practicalhaplotypegraph/overview>. Accessed 20 Aug 2018

Buckler ES et al (2009) The genetic architecture of maize flowering time. *Science* 325:714–718. <https://doi.org/10.1126/science.1174276>

Byrne P, Berlyn M, Coe E, Davis G, Polacco M, Hancock D, Letovsky S (1995) Reporting and accessing QTL information in USDA's Maize Genome Database. *J Agric Genomics* 1:1–11

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* 168:2169–2185. <https://doi.org/10.1534/genetics.104.032375>

Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE (1997) The sorghum photoperiod sensitivity gene, *Ma3*, encodes a Phytochrome B. *Plant Physiol* 113:611–619

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol* 158:824–834. <https://doi.org/10.1104/pp.111.185033>

Courtois B, Ahmadi N, Khowaja F, Price AH, Rami JF, Frouin J, Hamelin C, Ruiz M (2009) Rice root genetic architecture: meta-analysis from a drought QTL database. *Rice* 2:9028. <https://doi.org/10.1007/s12284-009-9028-9>

Darvasi A, Soller M (1997) A simple method to calculate resolving power and confidence interval of QTL map location. *Behav Genet* 27:125–132

Draye X, Lin YR, Qian XY, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren SX, Wing RA, Paterson AH (2001) Toward integration of comparative genetic, physical, diversity, and cytomechanical maps for grasses and grains, using the sorghum genome as a foundation. *Plant Physiol* 125:1325–1341. <https://doi.org/10.1104/pp.125.3.1325>

Fragoso CA, Moreno M, Wang Z, Heffelfinger C, LArbelaez LJ, Aguirre JA, Franco N, Romero LE, Labadie K, Zhao H, Dellaporta SL, Lorieux M (2017) Genetic architecture of a rice nested association mapping population. *G3: Genes Genomes Genet* 7:1913–1926. <https://doi.org/10.1534/g3.117.041608>

- Goff SA et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296:92–100. <https://doi.org/10.1126/science.1068275>
- Goffinet B, Gerber S (2000) Quantitative trait loci: a meta-analysis. *Genetics* 155:463–473
- Guan Y, Wang H-L, Qin L, Zhang H-W, Yang Y-B, Gao F-J, Li R-Y, Wang H-G (2011) QTL mapping of bio-energy related traits in *Sorghum*. *Euphytica* 182:431–440. <https://doi.org/10.1007/s10681-011-0528-5>
- Guo B, Slepner DA, Lu P, Shannon JG, Nguyen HT, Arelli PR (2006) QTLs associated with resistance to soybean cyst nematode in soybean: meta-analysis of QTL locations. *Crop Sci* 46:595–602. <https://doi.org/10.2135/cropsci2005.04-0036-2>
- Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S (2005) Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. *Genetics* 171:1247–1256. <https://doi.org/10.1534/genetics.105.041566>
- Hamelin C, Sempere G, Jouffe V, Ruiz M (2013) TropGeneDB, the multi-tropical crop information system updated and extended. *Nucleic Acids Res* 41:172–1175. <https://doi.org/10.1093/nar/gks1105>
- Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiji H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* 148:479–494
- Hu ZL, Park CA, Wu XL, Reecy JM (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res* 41:871–879. <https://doi.org/10.1093/nar/gks1150>
- Jordan DR, Klein RR, Sakreowski K, Henzell RG, Klein PE, Mace ES (2011) Mapping and characterization of Rf5: A new loci conditioning pollen fertility restoration in A1 and A2 cytoplasm in sorghum (*Sorghum bicolor* (L.) Moench). *Theor Appl Genet* 123:383–396. <https://doi.org/10.1007/s00122-011-1591-y>
- Kawahara Y et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. <https://doi.org/10.1186/1939-8433-6-4>
- Kim C-K, Seol Y-J, Lee D-J, Lee J-H, Lee T-H, Park D-S (2014) RiceQTLPro: an integrated database for quantitative trait loci marker mapping in rice plant. *Bioinformatics* 10:664–666. <https://doi.org/10.6026/97320630010664>
- Knoll J, Gunaratna N, Ejeta G (2008) QTL analysis of early-season cold tolerance in sorghum. *Theor Appl Genet* 116:577–587. <https://doi.org/10.1007/s00122-007-0692-0>
- Kump KL, Bradbury PJ, Wissner RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat Genet* 43:163–168. <https://doi.org/10.1038/ng.747>
- Lawrence CJ, Schaeffer ML, Seigfried TE, Campbell DA, LHarper LC (2007) MaizeGDB's new data types, resources and activities. *Nucleic Acids Res* 35:895–900. <https://doi.org/10.1093/nar/gkl1048>
- Locke AE et al (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518:197–206. <https://doi.org/10.1038/nature14177>
- Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J* 53:661–673. <https://doi.org/10.1111/j.1365-3113.2007.03326.x>
- Mace ES, Jordan DR (2010) Location of major effect genes in sorghum [*Sorghum bicolor* (L.) Moench]. *Theor. Appl Genet.* 121:1339–1356. <https://doi.org/10.1007/s00122-010-1392-8>
- Mace ES, Jordan DR (2011) Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals non-random distribution of QTL and of gene rich regions with significant implications for crop improvement. *Theor Appl Genet* 123:169–191. <https://doi.org/10.1007/s00122-011-1575-y>
- Mace ES, Rami J-F, Bouchet S, Klein PP, Klein RE, Kilian A, Wenzl P, Xia L, Sakreowski K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. *BMC Plant Biol* 9:13. <https://doi.org/10.1186/1471-2229-9-13>
- Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frère C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole genome resequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat Commun* 4:2320. <https://doi.org/10.1038/NCOMMS3320>
- Mace ES, Tai SS, Innes DJ, Godwin ID, Hu WS, Campbell BC, Gilding EK, Cruickshank A, Prentis PJ, Wang J, Jordan DR (2014) The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC Plant Biol* 14:253. <https://doi.org/10.1186/s12870-014-0253-z>
- Martin N (2018) Getting to the genetic and environmental roots of educational inequality. *npj Sci Learn* 3:4. <https://doi.org/10.1038/s41539-018-0023-z>
- McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, Mattison A, Morishige DT, Grimwood J, Schmutz J, Mullet JE (2018) The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* 93:338–354. <https://doi.org/10.1111/tpl.13781>
- Millar AA, Clemens S, Zachgo S, Giblin EM, Taylor DC, Kunst L (1999) CUT1, an Arabidopsis gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell* 11:825–838
- Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci USA* 110:453–458. <https://doi.org/10.1073/pnas.1215985110>
- Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS (2003) Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. *Science* 302:81–84. <https://doi.org/10.1126/science.1086072>
- Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S (2008a) Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Sci* 48:2165–2179. <https://doi.org/10.2135/cropsci2008.01.0016>
- Murray SC, Rooney WL, Mitchell SE, Sharma A, Klein PE, Mullet JE, Kresovich S (2008b) Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. *Crop Sci* 48:2180–2193. <https://doi.org/10.2135/cropsci2008.01.0068>
- Ni J, Pujar A, Youens-Clark K, Yap I, Jaiswal P, Tecle I, Tung CW, Ren L, Spooner W, Wei X, Avraham S, Ware D, Stein L, McCouch S (2009) Gramene QTL database: development, content and applications. *Database*. <https://doi.org/10.1093/database/bap005>
- Paterson AH et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556. <https://doi.org/10.1038/nature07723>

- Pereira MG, Ahnert D, Lee M, Klier K (1995) Genetic-mapping of quantitative trait loci for panicle characteristics and seed weight in sorghum. *Braz J Genet* 18:249–257
- Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc Natl Acad Sci* 108:6893–6898. <https://doi.org/10.1073/pnas.1010894108>
- Rami JF, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, Blanchard P, Hamon P (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). *Theor Appl Genet* 97:605–616
- Ren X, Pan Z, Zhao H, Zhao J, Cai M, Li J, Zhang Z, Qiu F (2017) EMPTY PERICARP1 serves as a factor for splicing of mitochondrial *nad1* intron and is required to ensure proper seed development in maize. *J. Exp Bot* 68:4571–4581. <https://doi.org/10.1093/jxb/erx212>
- Rhodes DH, Hoffmann L, Rooney WL, Herlad TJ, Bean S, Boyles R, Brenton ZW, Kresovich S (2017) Genetic architecture of kernel composition in global sorghum germplasm. *BMC Genomics* 18:15. <https://doi.org/10.1186/s12864-016-3403-x>
- Salvi S, Tuberosa R (2015) The crop QTLome comes of age. *Curr Opin Biotechnol* 32:179–185. <https://doi.org/10.1016/j.copbio.2015.01.001>
- Savage JE et al (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 50:912–919. <https://doi.org/10.1038/s41588-018-0152-6>
- Schnable JC (2015) Genome evolution in maize: from genomes back to genes. *Annu Rev Plant Biol* 66:329–343. <https://doi.org/10.1146/annurev-arplant-043014-115604>
- Schnable PS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115. <https://doi.org/10.1126/science.1178534>
- Schnable JS, Freeling M, Lyson E (2012) Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol Evol* 4:265–277. <https://doi.org/10.1093/gbe/evs009>
- Shakoor N, Ziegler G, Dilkes BP, Brenton Z, Boyles R, Connolly EL, Kresovich S, Ivan R, Baxter IR (2016) Genetic determinants of seed element composition. *Plant Physiol* 170:1989–1998. <https://doi.org/10.1104/pp.15.01971>
- Singh G, Kuzniar A, van Mulligan EM, Gavai A, Bachem CW, Visser RGF, Finkers R (2018) QTLTableMiner++: semantic mining of QTL tables in scientific articles. *BMC Bioinform* 19:183. <https://doi.org/10.1186/s12859-018-2165-7>
- Star KV, Song Q, Zhu A, Böttinger EP (2006) QTL MatchMaker: a multi-species quantitative trait loci (QTL) database and query system for annotation of genes and QTL. *Nucleic Acids Res* 34:586–589. <https://doi.org/10.1093/nar/gkj027>
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43:159–162. <https://doi.org/10.1038/ng.746>
- Upadhyaya HD, Wang YH, Sharma R, Sharma S (2013) SNP markers linked to leaf rust and grain mold resistance in sorghum. *Mol Breed* 32:451–462. <https://doi.org/10.1007/s11032-013-9883-3>
- Veyrieras JB, Goffinet B, Charcosset A (2007) MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinform* 8:49. <https://doi.org/10.1186/1471-2105-8-49>
- Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet* 10:e1004845. <https://doi.org/10.1371/journal.pgen.1004845>
- Wang Y, Yao J, Zhang Z, Zheng Y (2006) The comparative analysis based on maize integrated QTL map and meta-analysis of plant height QTLs. *Chin Sci Bull* 51:2219. <https://doi.org/10.1007/s11434-006-2119-8>
- Wang YH, Upadhyaya HD, Burrell AM, Sahraeian SME, Klein RR, Klein PE (2013) Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant, *Sorghum bicolor*. *G3: Genes Genomes Genet* 3:783–793. <https://doi.org/10.1534/g3.112.004861>
- Wood AR et al (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 46:1173–1186. <https://doi.org/10.1038/ng.3097>
- Yamamoto T, Yonemaru J, Yano Y (2009) Towards the understanding of complex traits in rice: substantially or superficially? *DNA Res* 16:141–154. <https://doi.org/10.1093/dnares/dsp006>
- Yengo L et al (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *bioRxiv* <https://doi.org/10.1101/274654>
- Yonemaru J, Yamamoto T, Fukuoka S, Uga Y, Hori K, Yano M (2010) Q-TARO: QTL annotation rice online database. *Rice* 3:194. <https://doi.org/10.1007/s12284-010-9041-z>
- Yu J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296:79–92. <https://doi.org/10.1126/science.1068037>
- Zeng H, Luo L, Zhang W, Zhou J, Li Z, Liu H, Zhu T, Feng X, Zhong Y (2006) PlantQTL-GE: a database system for identifying candidate genes in rice and Arabidopsis by gene expression and QTL information. *Nucleic Acids Res* 35:879–882. <https://doi.org/10.1093/nar/gkl814>
- Zhang D, Guo H, Kim C, Lee TH, Li J, Robertson J, Wang X, Wang Z, Paterson A (2013) CSGRqtl, a comparative QTL database for Saccharinae grasses. *Plant Physiol* 161:594–599. <https://doi.org/10.1104/pp.112.206870>
- Zhang N, Gibon Y, Wallace JG, Lepak N, Li P, Dedow L, Chen C, So YS, Kremling K, Bradbury PJ, Brutnell T, Stitt M, Buckler ES (2015) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol* 168:575–583. <https://doi.org/10.1104/pp.15.00025>
- Zhao L, Liu HJ, Zhang CX, Wang QY, Li XH (2015) Meta-analysis of constitutive QTLs for disease resistance in maize and its synteny conservation in the rice genome. *Genet Mol Res* 14:961–970. <https://doi.org/10.4238/2015.February.3.3>
- Zhao J, Mantilla Perez MB, Hu J, Salas Fernandez MG (2016) Genome-wide association study for nine plant architecture traits in sorghum. *Plant Genome*. <https://doi.org/10.3835/plantgenome2015.06.0044>