Theor Appl Genet (2011) 123:169–191
DOI 10.1007/s00122-011-1575-y

ORIGINAL PAPER

Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement

E. S. Mace · D. R. Jordan

Received: 5 August 2010 / Accepted: 18 March 2011 / Published online: 12 April 2011

Abstract A comprehensive analysis was conducted using 48 sorghum QTL studies published from 1995 to 2010 to make information from historical sorghum QTL experiments available in a form that could be more readily used by sorghum researchers and plant breeders. In total, 771 QTL relating to 161 unique traits from 44 studies were projected onto a sorghum consensus map. Confidence intervals (CI) of QTL were estimated so that valid comparisons could be made between studies. The method accounted for the number of lines used and the phenotypic variation explained by individual QTL from each study. In addition, estimated centimorgan (cM) locations were calculated for the predicted sorghum gene models identified in Phytozome (JGI GeneModels SBI v1.4) and compared with QTL distribution genome-wide, both on genetic linkage (cM) and physical (base-pair/bp) map scales. QTL and genes were distributed unevenly across the genome. Heterochromatic enrichment for QTL was observed, with approximately 22% of QTL either entirely or partially located in the heterochromatic regions. Heterochromatic gene enrichment was also observed based on their predicted cM locations on the sorghum consensus map, due to suppressed recombination in heterochromatic regions, in contrast to the euchromatic gene enrichment observed on the physical, sequence-based map. The finding of high gene density in recombination-poor regions, coupled with the association with increased QTL density, has implications for the development of more efficient breeding systems in sorghum to better exploit heterosis. The projected QTL information described, combined with the physical locations of sorghum sequence-based markers and predicted gene models, provides sorghum researchers with a useful resource for more detailed analysis of traits and development of efficient marker-assisted breeding strategies.

Communicated by E. Guiderdoni.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1575-y) contains supplementary material, which is available to authorized users.

E. S. Mace (✉)
Department of Employment, Economic Development and Innovation, Hermitage Research Station, 604 Yangan Road, Warwick, QLD 4370, Australia
e-mail: emma.mace@deedi.qld.gov.au

D. R. Jordan
Queensland Alliance for Agriculture and Food Innovation, Hermitage Research Station, 604 Yangan Road, Warwick, QLD 4370, Australia

Introduction

In recent decades, there has been a remarkable increase in the use of quantitative trait loci (QTL) mapping as a tool to uncover the genetic control of complex traits. Initially, a major motivation was a desire to locate useful genomic regions for use in marker-assisted selection (MAS) in breeding programmes. Increasingly, there has also been a desire to identify the underlying genes responsible for traits of interest. Following the publication of the first QTL studies in sorghum in 1995 (Lin et al. 1995; Paterson et al. 1995a, b; Pereira et al. 1995; Pereira and Lee 1995), just under 50 studies have identified QTL in sorghum for over 150 traits using over 30 unique mapping populations. Although sorghum is the world's fifth most important cereal crop based on tonnage (http://www.fao.org), investment in applied sorghum crop improvement is relatively low. This is in part because sorghum, in developing countries, is typically a staple crop of poor subsistence

farmers, while in developed countries it is grown as a relatively low-value animal feed grain in more marginal cropping areas. As a result, while many of the 50 published QTL mapping studies have the stated aim of identifying QTL and developing linked markers to facilitate molecular breeding, the actual use of MAS in sorghum breeding programmes to date has been minimal (e.g. Ejeta and Knoll 2007). The limited adoption of MAS in sorghum breeding programmes is, however, only partly due to the limited investment in sorghum crop improvement.

Another important factor that limits adoption of molecular breeding strategies is that, until recently, it has been very difficult and expensive for applied sorghum breeders and researchers to make use of existing QTL data. The key limitations have included; difficulty in comparing QTL locations based on different maps employed across studies, concerns about the accuracy of QTL location and size estimates due to the limited power of many QTL experiments, the need to validate QTL in different environments and genetic backgrounds, and the association of regions of the genome targeted for MAS with variation for other traits. A partial solution to these problems is to integrate the information available from multiple studies. The recent availability of the whole genome sequence of sorghum (Paterson et al. 2009) and the creation of a dense genetic linkage consensus map (Mace et al. 2009), combined with sequence data availability for legacy markers such as RFLPs used in early QTL mapping studies (http://cggc.agtec.uga.edu/) and the increased use of sequenced-based PCR (Ramu et al. 2010) and DArT (Mace et al. 2008) markers in more recent QTL studies, mean that we are now at a stage where it is possible to integrate information on QTL mapping studies and the location of major gene traits (Mace and Jordan 2010) together with the whole genome sequence. Thus, the resulting information is more accessible to applied plant breeding programmes. The effective use and integration of previous research data are of particular importance in a resource-poor crop like sorghum.

While it is theoretically possible to integrate the locations of published QTL and major genes to allow researchers and applied breeders to compare the results from different QTL studies and to study the congruency of QTL locations, it is not a trivial task. Comparison of the results of QTL studies is complicated, because QTL mapping experiments are generally heterogeneous; they involve different population types (RI, F_2, DH), different sample sizes, parents of different genetic origins, different environmental conditions, different methods of trait evaluation, different trait definitions, different markers and map construction approaches and different QTL detection algorithms. A recurring complication in the use of QTL data is that different parental combinations and/or experiments conducted in different environments often result in identification of partly or wholly non-overlapping sets of QTL (Rong et al. 2007). Such differences in the QTL landscape can be due to the environmental sensitivity of genes; however, a range of other factors, including lack of segregation in a particular cross, sampling error primarily due to population size and phenotypic evaluation, and errors in detection (e.g. the use of differing statistical thresholds to infer QTL), can also contribute to the lack of congruency between studies, both in terms of estimated genomic location and the magnitude of genetic effects. Previous simulations (e.g. Darvasi and Soller 1997) have demonstrated that the confidence interval of QTL location is inversely proportional to the population size and QTL effect. Large variation in the confidence intervals of QTL, even covering a whole chromosome, may occur when a QTL has a small genetic effect and a small population size is used. QTL detection can therefore be considered to be statistically biased both in the true number of QTL, which can be underestimated since only QTL with large effects are detected, and in the QTL effects, which can be overestimated since only significant effects are reported (Beavis 1994; Xu 2003). False-positive QTL can also be identified, because of the use of low threshold values and small population sizes, as discussed by Beavis (1994). Even though QTL mapping experiments must be considered with an awareness of these limitations, accumulated data from multiple studies have greatly improved our knowledge of the genetic control of some traits.

A number of methods have been developed to tackle the issues raised by heterogeneity among QTL studies (e.g. Goffinet and Gerber 2000; Arcade et al. 2004; Veyrieras et al. 2007). Integration of genetic maps and QTL locations by iterative projections onto a reference map is now widely used to position both markers and QTL on a single consensus map, and is referred to as meta-analysis. For sorghum researchers, this approach is simplified by the availability of a single, reference consensus map (Mace et al. 2009), with multiple anchor points to the whole genome sequence, allowing the integration of all sequence-based markers onto a common framework, which includes the majority of markers (RFLP and SSRs) used in historical QTL analyses. Marker order conservation across maps, for example as determined in sorghum by Feltus et al. (2006) and Mace et al. (2009), is another critical factor in integrating genetic map and QTL information across studies. In a recent study in wheat (Liu et al. 2009), it was found that the order of SSR markers associated with Fusarium head blight resistance QTL did not agree with the marker order of the ITMI consensus map. Such discrepancies make it very difficult to utilise markers linked to respective QTL from individual studies. The strategy to integrate QTL mapping information across different studies onto a single

reference map can be supplemented by approaches developed by Darvasi and Soller (1997) and adapted by Guo et al. (2006) to estimate the confidence intervals (CIs) of QTL. These approaches make use of existing published QTL information (genomic location, phenotypic variance explained, population type and size), to determine their most likely location and CI on a reference map. This process can be applied to merge a large number of genetic maps and QTL by projection and to refine the QTL positions based on a single reference map. The advantages of this method are not only to localise all markers and QTL against a common framework, but also to synthesise all the information related to a cluster of QTL by identifying consensus or meta-QTL (mQTL). The Darvasi and Soller (1997) approach is particularly useful in studies with limited numbers of QTL per trait per LG, as typically the Goffinet and Gerber (2000) methodology requires larger numbers (often >10) of observed QTL per trait per genomic region, to determine the most likely QTL distribution within a given genomic region.

For the foreseeable future, it is likely that sorghum breeding will involve a combination of conventional breeding and MAS. The development of effective, future breeding strategies will be enhanced by providing breeders with access to all available gene, marker and QTL information. For example, in addition to knowledge about the trait of interest, breeders need to be aware of the potential consequences of conventional or marker-assisted selection on other traits (Mace and Jordan 2010). In this study, we reviewed 48 QTL studies, which detected QTL corresponding to 161 traits. We projected these onto the consensus map and estimated 95% CIs of QTL, with consideration of the number of lines used and the phenotypic variation explained by individual QTL from each study. We also calculated the estimated cM locations for the predicted gene models identified in Phytozome (JGI GeneModels SBI v1.4; http://www.phytozome.net) and compared QTL distribution with gene density across the genome, both on a genetic linkage map (cM) and a physical map (base-pair/bp) scale.

Methodology

Collection of QTL studies

Data were collected from 48 publications by conducting a bibliographic review and using data stored in the public database Gramene (http://www.gramene.org). Four studies were identified with too few non-sequenced markers (AFLPs and RAPDs) in common to be included in this analysis: Tuinstra et al. (1996, 1997); Lijavetzky et al. (2000) and Natoli et al. (2002). On a study-by-study basis, details of the genetic linkage maps were collated (Table 1) including population type and size, total number of markers mapped, number of linkage groups identified, overall average marker density, and mapping function and QTL analysis methodology used. In 22 studies, common populations were evaluated for the same and/or different traits. In five studies, two populations were analysed. In total, 32 populations were studied across 44 publications.

Projection of QTL onto consensus map

A modified version of the consensus map developed by Mace et al. (2009) was used as the reference map for QTL projection. The consensus map was generated by integrating mapping data for 1,997 markers mapped to 2,029 unique loci across six component maps spanning 1,603.5 cM and with an average marker density of 1 marker/0.79 cM. Heterochromatic and euchromatic regions of the chromosomes were identified using the molecular cytogenetic maps of Kim et al. (2005) and genomic landscape of Paterson et al. (2009).

From the 44 studies, 771 individual QTL for 161 unique traits were included in the analysis. The 161 unique traits were classified into eight broad categories, for the purpose of facilitating analysis and reporting: grain, leaf, maturity, panicle, abiotic stress resistance, biotic stress resistance, stem composition and stem morphology. These are reported in the Supplementary Material (ESM, Table S1). Information was collected for each QTL or marker/trait association identified as follows: original marker interval, R^2 (% phenotypic variance explained), LOD value, direction of effect, published QTL symbol (where provided) and Gramene QTL accession ID. Marker nomenclature was standardised across maps and each genetic linkage map was aligned to the sorghum consensus map using Excel functions. A two-stage QTL projection strategy was then adopted.

Stage 1: individual QTL projected onto the consensus map

The location with the highest test statistics on the chromosome in the individual studies was regarded as the estimated location of a QTL from a particular study. This location was projected onto the consensus map based on flanking marker information in common between the individual study and the consensus map. In the case where markers flanking QTL were not included in the consensus map, their location on the consensus map was determined using a combined strategy of (1) projection based on common markers (Cone et al. 2002) and (2) sequence/in silico mapping of sequence-based markers analogous to e-PCR (Schuler 1998), following Mace and Jordan (2010). The latter strategy was made possible due to the availability of the primer sequence

Table 1 Details of the 44 QTL publications included in this study including population pedigree, cross type, generation (RI: recombinant inbred, BC_1: back-cross 1, F_2, F_3 or F_3), total number of loci mapped, total number of LGs (linkage groups) detected, total map length (cM), marker density (average distance between markers), mapping function used (K: Kosambi, H: Haldane, U: unknown/data not supplied) and analysis method used (SMA: single marker analysis, ANOVA: analysis of variance, IM: interval mapping, SIM: simple interval mapping, CIM: composite interval mapping, MQM: multiple QTL method; BSA: bulked segregant analysis)

Reference	Population pedigree	Cross type	Generation	Population size	No. of loci mapped	No. of LGs	Map length	Marker density	Mapping Function	Analysis method
Agrama et al. (2002)	GBIK/Redlan	Cultivated/cultivated	RI	93	113	12	1530	13.5398	K	SMA
Brown et al. (2006)	BTx623/IS3620C	Cultivated/cultivated	RI	137	396	10	1278.8	3.22929	K	CIM
Brown et al. (2008)	Association panel	Cultivated/wild		377	57	10	–	–	U	SMA
Chantereau et al. (2001)	IS2807/IS7680	Cultivated/cultivated	RI	85	129	12	878		H	CIM
Crasta et al. (1999)	B35/Tx430	Cultivated/cultivated	RI	96	128	14	1602		K	CIM
Deu et al. (2005)	Malisor 84-7/S34	Cultivated/cultivated	F_2	217	92	13	1160	12.6087	H	CIM
Feltus et al. (2006)	BTx623/IS3620C	Cultivated/cultivated	RI	137	145	10			K	IM
Feltus et al. (2006)	BTx623/*S. propinquum*	Cultivated/wild	F_2	370	96	10			K	IM
Hart et al. (2001)	BTx623/IS3620C	Cultivated/cultivated	RI	137	145	10	1278.8	8.81931	K	SMA
Haussmann et al. (2002)	IS9830/E36-1	Cultivated/cultivated	RI	226	128	10	1291.2	10.0875	H	CIM
Haussmann et al. (2002)	N13/E36-1	Cultivated/cultivated	RI	226	146	12	1438.1	9.85	H	CIM
Haussmann et al. (2004)	IS9830/E36-1	Cultivated/cultivated	RI	226	137	11	1498	10.9343	H	CIM
Haussmann et al. (2004)	N13/E36-1	Cultivated/cultivated	RI	226	157	11	1599	10.1847	H	CIM
Katsar et al. (2002)	BTx623/*S. propinquum*	Cultivated/wild	F_3	370	–	–	–	92%*	U	ANOVA
Katsar et al. (2002)	RTx430/PI550607	Cultivated/cultivated	F_3	195	–	–	–	36%*	U	ANOVA
Kebede et al. (2001)	SC56/Tx7000	Cultivated/cultivated	RI	125	144	10	1355	9.40972	K	CIM
Kim (2003)	BTx623/IS3620C	Cultivated/cultivated	RI	137	–	–	–	–	–	Physical mapping
Klein et al. (2001)	RTx430/Sureno	Cultivated/cultivated	RI	125	130	10	970	7.46154	K	SMA, IM
Knoll et al. (2008)	Shan Qui Red/SRN39	Cultivated/cultivated	RI	153	132	14	2128	16.1212	K	SIM, CIM
Lin et al. (1995)	BTx623/*S. propinquum*	Cultivated/wild	F_2	370	78	11	935	11.9872	K	–
Mohan et al. (2009)	296B/IS18551	Cultivated/cultivated	RI	168	96	–	–	–	K	CIM
Murray et al. (2008a)	Btx623/Rio	Cultivated/cultivated	RI	176	259	10	1836	7.0888	K	SMA, IM, CIM
Murray et al. (2008b)	Btx623/Rio	Cultivated/cultivated	RI	176	259	10	1836	7.0888	K	SMA, IM, CIM
Nagaraj et al. (2005)	96-4121/Redlan	Cultivated/cultivated	RI	88	60	13	603.5	10.0583	K	CIM
Nagy et al. (2007)	Shan Qui Red/Colby	Cultivated/cultivated	F_3	178						Physical mapping
Parh (2005)	R939145-2-2/IS8525	Cultivated/cultivated	RI	146	286	15	1599.1	5.59126	K	SMA, CIM
Parh et al. (2008)	R939145-2-2/IS8525	Cultivated/cultivated	RI	146	305	10	1625.2	5.32852	K	SMA, CIM
Paterson et al. (1995a)	BTx623/*S. propinquum*	Cultivated/wild	F_2	370	78	11	1020	13.0769	U	IM
Paterson et al. (1995b)	BTx623/*S. propinquum*	Cultivated/wild	F_2	370	78	11	1020	13.0769	U	IM
Paterson et al. (1995b)	BTx623/*S. propinquum*	Cultivated/wild	BC_1	378						
Pereira and Lee (1995)	CK60/PI229828	Cultivated/cultivated	F_2	152	111	10	1299	11.7027	H	SMA, IM
Pereira et al. (1995)	CK60/PI229828	Cultivated/cultivated	F_2	152	111	10	1299	11.7027	H	SMA, IM

Table 1 continued

Reference	Population pedigree	Cross type	Generation	Population size	No. of loci mapped	No. of LGs	Map length	Marker density	Mapping Function	Analysis method
Perumal et al. (2009)	SC748-5/BTx623	Cultivated/cultivated	F_2	146	98	–	–	–		Physical mapping
Rami et al. (1998)	IS2807/379	Cultivated/cultivated	RI	110	128	11	878	6.85938	H	SIM, CIM
Rami et al. (1998)	IS2807/249	Cultivated/cultivated	RI	90	151	11	977	6.4702	H	SIM, CIM
Ritter et al. (2008)	R9188/R9403463-2-1	Cultivated/cultivated	RI	184	228	16	1879.2	8.24211	K	SMA, CIM
Salas Fernandez et al. (2008)	KS115/Macia	Cultivated/cultivated	RI	351	112	11	1364.6	12.1839	K	SMA, CIM
Satish et al. (2009)	296B/IS18551	Cultivated/cultivated	RI	168	162	16	1143	7.05556	K	IM, MQM
Shiringani et al. (2010)	M71/SS79	Cultivated/cultivated	RI	188	157	11	1029	6.55414	K	CIM
Srinivas et al. (2009)	296B/IS18551	Cultivated/cultivated	RI	168	152	15	1098.7	7.22829	K	SMA, IM, MQM
Srinivasa Reddy et al. (2008)	IS22380/E36-1	Cultivated/cultivated	RI	93	85	10	650.3	7.65059	H	CIM
Subudhi et al. (2000)	B35/Tx7000	Cultivated/cultivated	RI	98	232	10	–	–	H	SIM, CIM
Tao et al. (1998)	QL39/QL41	Cultivated/cultivated	RI	160	166	21	1400	8.43373	U	SMA, IM
Tao et al. (2000)	QL39/QL41	Cultivated/cultivated	RI	160	311	10	~2750	8.84244	U	SMA, IM
Tao et al. (2003)	ICSV745/B890562	Cultivated/cultivated	RI	120	264	12	1472	5.57576	U	SMA, IM
Tuinstra et al. (1997)	Tx7078/B35	Cultivated/cultivated	RI	98	170	17	1645	9.67647	R	SMA
Winn et al. (2009)	P850029/Sureno	Cultivated/cultivated	F_4	277	8	1	61.5	7.6875	K	BSA
Wu and Huang (2008)	Westland A/PI550610	Cultivated/cultivated	F_2	277	118	10	1005	8.51695	K	IM, CIM
Wu et al. (2007)	Westland A/PI550610	Cultivated/cultivated	F_2	233	118	16	1000	8.47458	K	IM, CIM
Xu et al. (2000)	B35/Tx7000	Cultivated/cultivated	RI	98	145	10	837	5.77241	H	SIM

* Marker genome coverage as stated in text

information for the majority of SSRs (Ramu et al. 2010), probe sequence information for a subset of RFLP markers (http://cggc.agtec.uga.edu/) and clone sequences for the majority of DArT markers (Rami and Bouchet, pers. comm.) included on the consensus map, hence providing opportunities to anchor the consensus map to the whole genome sequence (Paterson et al. 2009). A reported QTL was not projected if its flanking markers in a particular study were inconsistent with the order on the consensus map and the whole genome sequence, if the chromosomal region containing the QTL was significantly different from the consensus map or if there were less than two flanking markers in common. In many cases, it was not possible to integrate all QTL identified based on maps established with a high number of AFLP (unless the AFLP markers had the txa prefix, detailed in Menz et al. 2002) or RAPD markers, due to the lack of common markers that allowed establishment of bridges between maps and due to lack of sequence information. This included QTL from Agrama et al. (2002); Crasta et al. (1999); Murray et al. (2008a, b); Ritter et al. (2008); Salas Fernandez et al. (2008); Subudhi et al. (2000) and Tuinstra et al. (1997).

The confidence intervals (CI) for the projected QTL were then estimated based on the following formulae:

$$\mathrm{CI} = 530/\left(NR^2\right) \text{ for } \mathrm{F}_2$$

(described by Darvasi and Soller 1997)

$$\mathrm{CI} = 163/\left(NR^2\right) \text{ for RI}$$

(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.

Stage 2: meta-QTL identified, trait by trait

Where estimated CIs of QTL for the same trait overlapped, those QTL were grouped into a meta-QTL (mQTL), following Liu et al. (2009) due to the limited number of QTL identified per trait per LG and per study. QTL for the same trait were classified as separate QTL if their CI had no region in common and mean QTL locations were equal or more than 15 cM away from each other. mQTL were identified across all traits and, additionally, a detailed meta-analysis was conducted for one selected trait, kernel weight.

Gene density comparisons

The physical (bp) locations of 35,854 predicted gene models (JGI GeneModels SBI v1.4; http://www.phytozome.net) were collated. To predict the most probable cM locations of all predicted gene models, a framework map was generated based on sequenced markers on the consensus map, preferentially selecting single copy, bridge markers or markers mapped across multiple component maps, evenly distributed across each chromosome. In total, 504 sequenced markers with known genetic linkage distances were selected, averaging 1 marker/3.1 cM or 1 marker/1.3 Mbp. The genetic linkage distance in cM was calculated for the predicted gene models by assuming a linear relationship between each adjacent marker pair. The number of genes per 0.5 cM across each chromosome was then calculated.

Generation of QTL heat map

The production of colour-coded QTL heat maps was performed in R (http://www.R-project.org), following the methods described in Norton et al. (2008) and Khowaja et al. (2009). The values used in the production of the QTL heat maps were the projected CIs around the LOD peak for each QTL. The QTL density at each location on the map was determined by integrating the overlapping CIs of the 771 QTL. The intensity was calculated by summing the overlapping CIs in 0.5 cM bins across each chromosome. The most probable physical (bp) locations of the CIs of the 771 QTL were also calculated, using the framework map of 504 sequenced markers with known genetic linkage distances, as described above. Colour-coded QTL heat maps based on the physical scale were then generated by summing the overlapping CIs in 0.5Mbp bins across each chromosome.

Results

Consensus map expansion

Following successive projections of markers flanking QTL from the genetic linkage maps detailed in the 44 studies, in addition to adding selected, known sequenced markers in critical genomic regions, the number of loci included on the genetic consensus map expanded from 2,029 to 3,272, including 562 SSR and STS, 577 RFLP, 2,079 DArT markers and 54 genes. Information on the physical location of the markers, based on the primer or probe sequence, was available for 2,335 (71.3%) overall (ESM, Table S2). These 2,335 sequenced markers showed a high degree of concordance in marker order between the sequence-based physical map and the consensus map. The overall marker density on the consensus map increased from 1 marker/0.79 cM to 1 marker/0.48 cM and the new map was used to project QTL from historical studies. Based on additional sequenced marker data, modifications were made to the telomeric end of the short arm of SBI-09 (ESM Table S2). The 2,335 sequenced markers mapped on the consensus

map were integrated with all publicly available sorghum SSR primer pairs, CISP (conserved introns scanning primers) markers and selected genes/gene products collated by Ramu et al. (2010), resulting in 8,504 markers ordered on the sequence-based physical map (bp) scale (1 marker/77 kb). The genetic linkage distances have also been detailed for the 2,335 sequenced markers from the consensus map, resulting in 27.5% of markers ordered on the sequence-based physical map (bp) scale also having genetic linkage distances, based on their consensus map location (ESM, Table S3).

QTL studies

Of the 44 studies included in this analysis, population size ranged from 85 to 378. The average distance between markers in each study ranged from 3.2 to 16.1 cM with an overall average distance between markers of 9.03 cM. Genetic linkage map sizes varied from 603.5 to 2,128 cM, and the total number of genetic linkage groups identified in whole genome studies ranged from 10 to 21. Almost all studies analysed the whole genome, except for Nagy et al. (2007) and Perumal et al. (2009), whose objectives were to fine map major effect QTL, and Kim (2003), whose objective was to estimate physical distances for the major effect gene, Ma_5.

The average number of QTL reported per study was 16.9, ranging from 1 to 76. The average number of QTL reported per trait per study was 3.4, close to that of 4 estimated by Kearsey and Farquhar (1998) and found in other meta-analysis studies, e.g. 4.6 reported by Chardon et al. (2004). The maximum number of QTL reported per trait in a single study was 29 (Haussmann et al. 2004; resistance to the parasitic weed striga).

QTL projections and genome-wide distribution

A total of 771 individual QTL from 44 studies relating to 161 unique traits from eight broad trait categories were projected onto the consensus map (ESM Figure S1). Among them, 169 QTL related to the trait category stem morphology, 128 to biotic stress resistance, 121 QTL to grain, 93 to abiotic stress resistance, 62 to maturity, 96 to stem composition, 56 to panicle and 46 to the trait category leaf (Table 2 and ESM Table S4). The original marker interval, R^2 (% phenotypic variance explained by the given trait), LOD value, direction of effect, published QTL symbol (if any) and Gramene QTL accession ID are detailed in the Supplementary Material (ESM, Table S1). The average size of the projected QTL was 11.5 cM, based on the estimated CIs. As much as 50% of QTLs had 95% CI less than 10 cM; 5.4% had 95% CIs greater than 25 cM and one QTL for green bug resistance (QGrBgR3_1) had a 95% CI greater than 100 cM. The large CI was due to small effect of the QTL as identified in the original study (R^2=1.2% in Wu and Huang 2008).

The QTL were distributed unevenly across the genome with SBI-01 containing almost 20% of the total QTL and SBI-05 containing only 3.9% of the QTL (Table 3). Assuming an even distribution of QTL midpoints across the genome, with an expected mean distance between QTL of 2.04 cM, eight of the ten chromosomes had either significantly more (SBI-01, SBI-07 and SBI-10) or significantly less (SBI-02, SBI-04, SBI-05, SBI-08 and SBI-09) QTL than expected (ESM Table S5). Chromosome length (from the genetic linkage map in cM) was not correlated with the number of total QTL. A suggestive association between the total number of QTL per chromosome and the total number of predicted gene models per chromosome was observed ($p = 0.063$).

Figure 1 shows the heat map generated for QTL distribution across the genome. There were five main regions of high QTL density (>20/0.5 cM): one on SBI-01 at approximately 70 cM, one on SBI-06 at approximately 84 cM, two on SBI-07, one at approximately 108 cM and the second at approximately 125 cM, and the final region of high QTL density was on SBI-10 at approximately 58 cM. Two of these regions (SBI-01 and SBI-10) overlapped wholly, or partially, with the heterochromatic regions of the chromosomes.

The SBI-01 high QTL density region contained QTL for tiller number, endosperm colour and carotenoid content, cold tolerance, ergot resistance, kernel weight, panicle architecture, total number of leaves, panicle length, stay-green, maturity, protein digestibility and shoot fly resistance The high QTL density region on SBI-06 contained QTL for kernel weight, grain yield, height, panicle weight, stay-green, ergot resistance, tillering, resistance to striga and shoot fly resistance. The first high QTL density region on SBI-07 contained QTL for height, leaf angle, kernel friability, stem yield, lodging, sugar-related traits, leaf hemi-cellulose, panicle architecture, tiller height and grain phosphorus. The second high-density QTL region on SBI-07 contained QTL for endosperm colour and carotenoid content, kernel weight, glume persistence, grain mould resistance, green bug resistance, height, leaf width, stay-green, stem cellulose and hemi-cellulose, sugar-related traits and tiller height The high-density QTL region on SBI-10 contained QTL for ergot resistance, yield, head and kernel weight, grain mould resistance, green bug resistance, height, maturity, panicle architecture, rhizomatousness, shoot fly resistance, stay-green, sugar-related traits and tiller number (ESM, Table S1 for publication details of specific QTL).

QTL clustering of different traits within trait categories was observed, in particular for the biotic stress resistance category, in which 65% of all QTL were located within

Table 2 Details of QTL identified per trait category

Trait category	Trait sub-category	Total	No. of publications	Publication details
Grain	Germination	3	1	Rami et al. (1998)
	Glume morphology	7	2	Feltus et al. (2006); Murray et al. (2008a)
	Grain composition	63	5	Klein et al. (2001); Murray et al. (2008a); Rami et al. (1998); Salas Fernandez et al. (2008); Winn et al. (2009)
	Kernel weight	35	8	Brown et al. (2006); Feltus et al. (2006); Murray et al. (2008a); Paterson et al. (1995a); Pereira et al. (1995); Rami et al. (1998); Srinivas et al. (2009); Tuinstra et al. (1997)
Leaf	Age-related leaf sensecence	4	1	Feltus et al. (2006)
	Leaf composition	9	1	Murray et al. (2008b)
	Leaf morphology	24	2	Feltus et al. (2006); Hart et al. (2001)
	Leaf yield	9	2	Murray et al. (2008b); Srinivas et al. (2009)
Maturity	Maturity	62	12	Brown et al. (2006); Chantereau et al. (2001); Crasta et al. (1999); Feltus et al. (2006); Hart et al. (2001); Kebede et al. (2001); Kim (2003); Lin et al. (1995); Parh (2005); Ritter et al. (2008); Shiringani et al. (2010); Srinivas et al. (2009)
Panicle	Grain and panicle yield	29	7	Feltus et al. (2006); Murray et al. (2008a, b); Rami et al. (1998); Ritter et al. (2008); Shiringani et al. (2010); Srinivas et al. (2009)
	Panicle architecture	39	5	Brown et al. (2006); Hart et al. (2001); Pereira et al. (1995); Rami et al. (1998); Srinivas et al. (2009)
Abiotic stress resistance	Cold tolerance	10	1	Knoll et al. (2008)
	Stay-green	83	7	Crasta et al. (1999); Haussmann et al. (2002); Kebede et al. (2001); Srinivas et al. (2009); Subudhi et al. (2000); Tao et al. (2000); Xu et al. (2000)
Biotic stress resistance	Anthracnose resistance	2	2	Klein et al. (2001); Perumal et al. (2009)
	Bacterial leaf stripe	1	1	Klein et al. (2001)
	Drechslera leaf blight resistance	1	1	Mohan et al. (2009)
	Ergot resistance	17	1	Parh et al. (2008)
	Grain mould resistance	9	2	Rami et al. (1998); Klein et al. (2001)
	Greenbug resistance	20	5	Agrama et al. (2002); Katsar et al. (2002); Nagaraj et al. (2005); Wu and Huang (2008); Wu et al. (2007)
	Head bug resistance	9	1	Deu et al. (2005)
	Leaf scorch	1	1	Feltus et al. (2006)
	Midge resistance	3	1	Tao et al. (2003)
	Milo disease resistance	1	1	Nagy et al. (2007)
	Rust resistance	4	1	Tao et al. (1998)
	Shoot fly resistance	25	1	Satish et al. (2009)
	Stalk rot resistance	3	1	Srinivasa Reddy et al. (2008)
	Striga resistance	29	1	Haussmann et al. (2004)
	Target leaf spot resistance	1	1	Mohan et al. (2009)
	Zonate leaf spot resistance	2	2	Klein et al. (2001); Mohan et al. (2009)
Stem composition	Stem biomass yield	9	2	Murray et al. (2008b); Ritter et al. (2008)
	Stem composition	17	1	Murray et al. (2008b)
	Sugar-related traits	70	3	Murray et al. (2008a); Ritter et al. (2008); Shiringani et al. (2010)

Table 2 continued

Trait category	Trait sub-category	Total	No. of publications	Publication details
Stem morphology	Height	101	15	Brown et al. (2006); Brown et al. (2008); Feltus et al. (2006); Hart et al. (2001); Kebede et al. (2001); Klein et al. (2001); Lin et al. (1995); Murray et al. (2008a); Parh (2005); Pereira and Lee (1995); Pereira et al. (1995); Rami et al. (1998); Ritter et al. (2008); Shiringani et al. (2010); Srinivas et al. (2009)
	Lodging tolerance	4	2	Kebede et al. (2001); Murray et al. (2008b)
	Regrowth	8	2	Murray et al. (2008b); Paterson et al. (1995b)
	Rhizomatousness	12	1	Paterson et al. (1995b)
	Stem morphology	10	3	Feltus et al. (2006); Murray et al. (2008a); Shiringani et al. (2010)
	Tillering	35	5	Feltus et al. (2006); Hart et al. (2001); Murray et al. (2008a); Paterson et al. (1995b); Shiringani et al. (2010)
Grand Total		771		

Table 3 The number of QTL per chromosome

LG	Total no. of QTL	% QTL/LG	Total length cM[a]	Total length bp[b]
SBI-01	139	18.0285	191.8	73,840,631
SBI-02	80	10.3761	229.6	77,932,606
SBI-03	90	11.6732	172.3	74,441,160
SBI-04	56	7.2633	169.4	68,034,345
SBI-05	30	3.8911	118.5	62,352,331
SBI-06	96	12.4514	166.4	62,208,784
SBI-07	113	14.6563	132.8	64,342,021
SBI-08	41	5.3178	131.9	55,460,251
SBI-09	47	6.096	143.6	59,635,592
SBI-10	79	10.2464	115.2	60,981,646
Totals	771		1571.5	659,229,367

[a] Based on the modified consensus map

[b] Based on SBI v1.4; http://www.phytozome.net

5 cM clusters of other biotic stress resistance QTL. On the short arm of SBI-01 at around 30 cM, the mean QTL location for resistance to head bug (Deu et al. 2005), shoot fly (Satish et al. 2009) and striga (Haussmann et al. 2004) were within 3.5 cM of each other. Similarly on the short arm of SBI-03 at around 33 cM, the mean QTL location for resistance to greenbug (Wu and Huang 2008), midge (Tao et al. 2003) and striga (Haussmann et al. 2004) were located within 3 cM of each other. On SBI-04, the mean location of four biotic stress resistance QTL were clustered within 7 cM of each other: ergot resistance (Parh et al. 2008), greenbug resistance (Nagaraj et al. 2005), grain mould resistance (Klein et al. 2001) and stalk rot resistance (Srinivasa Reddy et al. 2008). A further QTL for head bug resistance (Deu et al. 2005) was located within 5 cM of this cluster on SBI-04. SBI-06 contained the largest biotic stress resistance cluster with the mean location of 4 QTL for resistance to zonate leaf spot (Mohan et al. 2009), anthracnose (Klein et al. 2001), Drechslera leaf blight (Mohan et al. 2009) and bacterial leaf stripe (Klein et al. 2001) within 1 cM of each other. There were two further QTL for resistance to target leaf spot (Mohan et al. 2009) and stalk rot (Srinivasa Reddy et al. 2008) within 5 and 10 cM of this cluster, respectively.

Almost a quarter (22%) of QTL were either entirely or partially located in the heterochromatic regions across the genome (Table 4). Considering that on average, across all chromosomes, the heterochromatic region makes up just 4.1% of the total genetic linkage distance, the number of QTL in heterochromatin is significantly greater than would be expected if the QTL were distributed evenly over the genetic distance, even when considering only the QTL midpoint (chi-square = 58.45, 1 df, $p < 0.00001$; ESM Table S5). The proportion of QTL with their midpoint located in the heterochromatic regions was not distributed uniformly across chromosomes; four chromosomes (SBI-04, SBI-07, SBI-08 and SBI-10) had significantly more ($p < 0.00001$) QTL in the heterochromatic regions than expected, based on

genetic linkage distance (ESM Table S5); e.g. 24% of QTL on SBI-08 have their mean CI location in the heterochromatin region, which represents just 3.4% of the total genetic distance of the chromosome (ESM Figure S2A). In contrast, the heterochromatic region of SBI-03 contained the mean CI location of only 1 QTL, representing 1.1% of the total number of QTL on the chromosome.

In comparison with regions of high QTL density, eight regions of low QTL density, which contained no QTL for at least 10 cM, were observed across all chromosomes except for SBI-01; SBI-03 and SBI-07. SBI-04 contained two regions of low QTL density (from 18 to 28 cM and from 140 to 169 cM).

The locations of 35 major effect genes (Mace and Jordan 2010) were also plotted (ESM, Figure S1). Clustering of QTL around the major effect genes was observed, with just under 50% of all 771 QTL located with 5 cM of a major effect gene. In particular, there was significant clustering of QTL around 5 of the 35 major effect genes; 3 maturity genes (Ma_3, Ma_1 and Ma_4) and 2 height genes (dw_2 and dw_3).

On the long arm of SBI-01, Ma_3 was linked within 5 cM to 22 QTL including 4 QTL for maturity, 4 QTL for height, 3 QTL for yield, tillering, stay-green, rhizomatousness, grain composition, grain mould resistance, panicle architecture and green-bug resistance. On the long arm of SBI-06, $dw_2$and Ma_1 were within 5 cM of 36 QTL including 8 QTL for height, 3 for maturity, 2 for tillering, 2 for stem biomass yield, 5 for sugar-related traits, 2 for stay-green, 2 for leaf morphology, and 2 for panicle architecture, striga resistance and kernel weight. On the long arm of SBI-07, dw_3 was within 5 cM of 35 QTL including 11 for height, 5 for biomass yield, 5 for sugar-related traits, 6 for panicle architecture, 2 for green bug resistance and 1 QTL for lodging. This region corresponded to the first high-density QTL region on SBI-07, at around 108 cM, identified in Fig. 1. The gene Ma_4, located on the short arm of SBI-10, was within 5cM of 29 QTL, including QTL for 7 QTL for maturity; 5 QTL for height, grain and leaf yield, panicle architecture, endosperm colour and composition, greenbug resistance and stay-green (ESM, Table S1 for publication details of specific QTL).

Meta-QTL identification

The overall meta-QTL analysis conducted across all traits classified the 771 QTL into 371 unique QTL and 125 mQTL, representing 400 individual QTL (52% of the total). The mQTL consisted of a maximum of 21 individual QTL (for a height mQTL on SBI-07) to 2 individual QTL, with a mean of 3.2. Overall, the stem morphology trait category contained the highest number of mQTL (25) representing 107 individual QTL, averaging 4.28 individual QTL per mQTL, indicating a high degree of concordance across studies (Table 5 and ESM Tables S1 and S6). Trait categories biotic and abiotic stress resistances shared the next highest number of mQTL (20) averaging, 2.45 and 4 individual QTL per mQTL, respectively, again indicating a high degree of concordance across studies, particularly for the abiotic stress resistances (primarily stay-green) category. The mQTL identified within panicle and grain trait categories represented the lowest number of individual QTL per mQTL, 2 and 2.41, respectively. The trait category leaf contained no mQTL.

Consistent with the individual QTL distribution plots, the mQTL are widely and unevenly distributed (Fig. 2); as previously, assuming an even distribution of QTL midpoints across the genome, with an expected mean distance between the 125 mQTL and 371 unique QTL of 3.16 cM, five of the ten chromosomes had either significantly more (SBI-01 and SBI-07) or significantly less (SBI-04, SBI-05 and SBI-08) QTL than expected (ESM Table S5). Additionally, the proportion of individual QTL clustering into mQTL per trait varied across the genome (ESM Figure S1) with SBI-10 containing the lowest proportion (36.7%) of unique QTL and SBI-02 containing the highest proportion of unique QTL (62.5%).

Detailed meta-analysis was undertaken for the trait, kernel weight, to demonstrate the utility of this approach. Overall, 35 individual QTL for the kernel weight trait were identified in eight separate studies using eight different mapping populations. The total number of projected QTL per study varied from a maximum of 11 (Feltus et al. 2006; 7 identified in the BTx623/*S. propinquum* population and 4 identified in the BTx623/IS3620C population) to a minimum of 1 (Tuinstra et al. 1997 in the Tx7078/B35 population). Of the total of 35 QTL for kernel weight identified across all eight studies, 23 were in common across at least two studies and 12 were unique to a single study. Nine mQTL were identified from the 23 common QTL (Table 6 and Fig. 3), with mQTL representing between 2 and 4 individual QTL. The CI of mQTL ranged from 4.7 to 44.5 cM with an average of 20.7 cM and a median of 20 cM. The mQTL with the most precise CI was located on SBI-07 and consisted of two individual QTL identified by Rami et al. (1998) with large effect (mean R^2 = 33.3%). In contrast, the mQTL on SBI-08 had a very broad CI and was represented by four small effect individual QTL (mean R^2 = 0.8%). The individual QTL corresponding to the SBI-08 mQTL were all identified in a population with BTx623 as one parent and, where reported, BTx623 was identified as the source of the SBI-08 mQTL.

Relationships between QTL and gene density

We predicted probable cM locations for the 35,854 predicted gene models (JGI GeneModels list SBI v1.4) and plotted QTL density relative to gene density across each

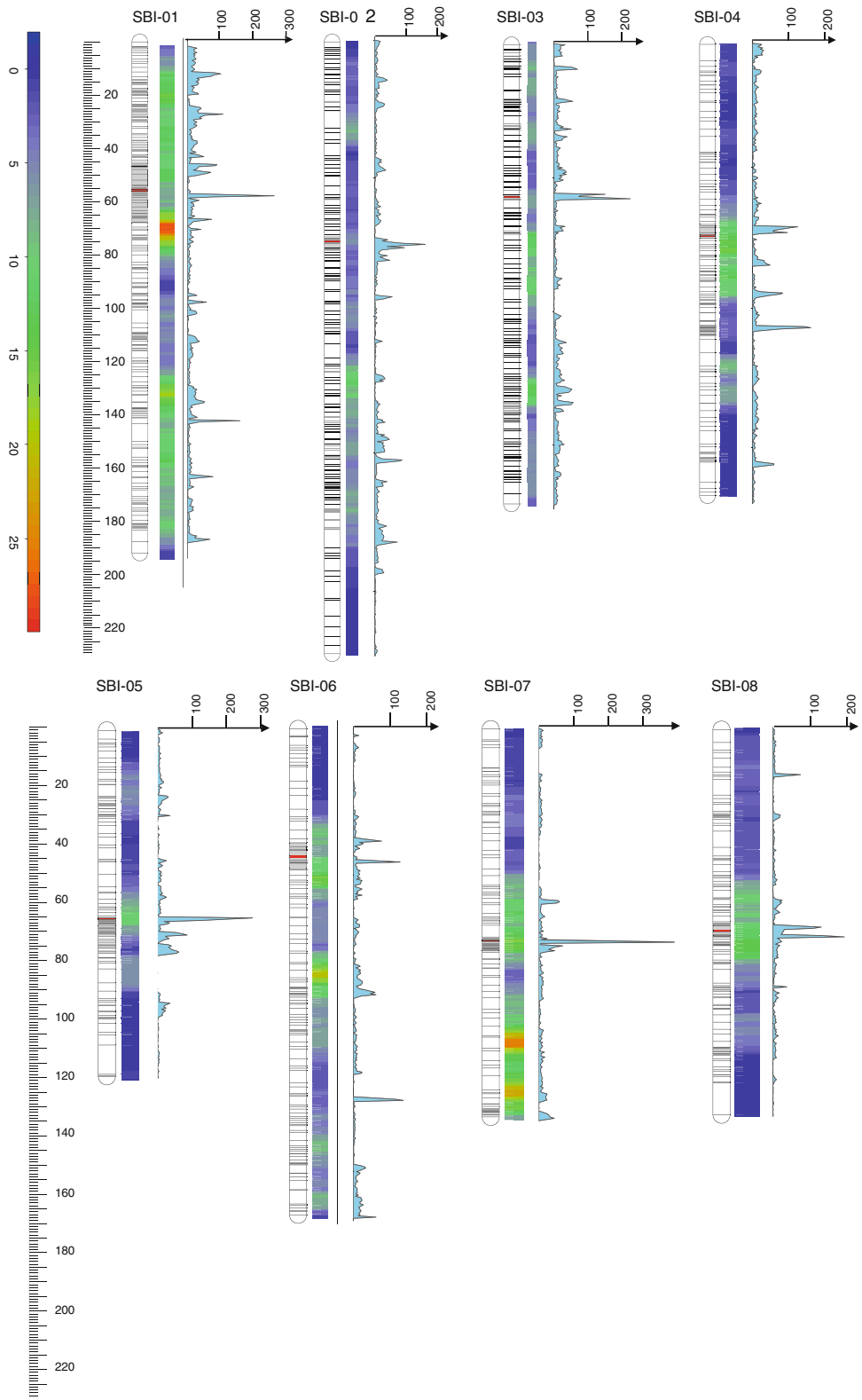


Fig. 1 QTL and gene density plots along the expanded sorghum consensus map. The 3272 loci of the consensus map are displayed schematically by *horizontal lines* across the *bars* representing the ten chromosomes. Heterochromatic regions on each chromosome are shaded *grey*, and the centromere in *red*. The *scale bar* to the *left* indicates the length of each chromosome in cM. The heat map to the *right* of each chromosome represents QTL density (number of QTL/0.5 cM) and the graph indicates gene density (number of genes/0.5 cM)

chromosome in cM (Fig. 1). Uneven gene distribution was observed, with gene enrichment in heterochromatic regions across all chromosomes. Mean gene density in the heterochromatic regions, across all chromosomes, was 118.3 genes/cM, compared to 19.3 genes/cM in the euchromatic regions. This was comparable to the pattern observed for QTL density, with a mean QTL density of 21.2 QTL/cM in heterochromatin, compared to 10.9 QTL/cM in

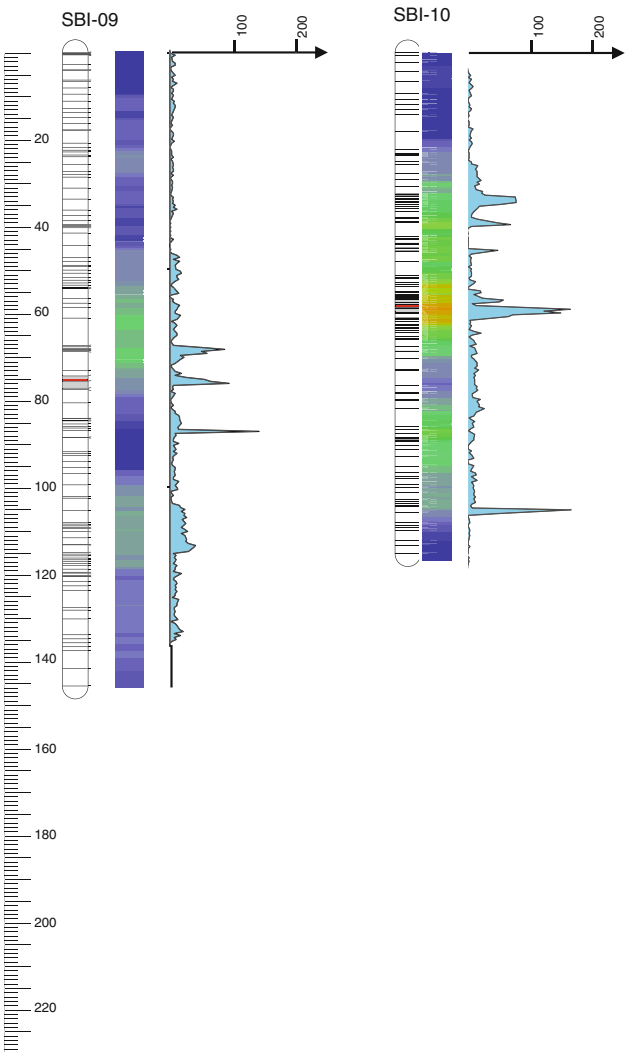


Fig. 1 continued

euchromatin. Ten regions of very high gene density (>150 genes/0.5 cM) were identified. The majority of these gene-rich regions were located in heterochromatin; however, three gene-rich regions were located in the euchromatic regions, on the long arms of SBI-01 (156 genes between 139.0 and 139.5 cM), SBI-04 (161 genes between 107.5 and 108.0 cM) and SBI-10 (167 genes between 103.5 and 104.0 cM). In the heterochromatic regions, 1 cM on average represented 8.46 Mb across all chromosomes. In comparison, in the euchromatic regions, 1 cM on average represented 0.22 Mbp.

We also predicted bp locations for the coordinates of the CIs for the 771 QTL and plotted QTL density relative to gene density across each chromosome in bps. On the physical map (Fig. 4), as expected, euchromatic regions are enriched for genes, with an average gene density of 46.9 genes/0.5 Mbp across all chromosomes, as compared to 7.7 genes/0.5 Mbp in heterochromatin. In contrast, QTL density

Table 4 The number and frequency of QTL in heterochromatin

LG	Total no. of QTL	QTL with CI in heterochromatin	
		Total no. of QTL	QTL frequency
SBI-01	139	44	0.32
SBI-02	80	6	0.08
SBI-03	90	6	0.07
SBI-04	56	17	0.30
SBI-05	30	11	0.37
SBI-06	96	18	0.19
SBI-07	113	19	0.17
SBI-08	41	12	0.29
SBI-09	47	8	0.17
SBI-10	79	28	0.35

Table 5 The number of mQTL identified per trait category

Trait category and sub-category	No. of mQTL	No. of individual QTL comprising mQTL
Grain total	17	41
Glume morphology	1	2
Grain composition	7	16
Kernel weight	9	23
Maturity total	17	53
Maturity	17	53
Panicle total	7	14
Grain and panicle yield	3	6
Panicle architecture	4	8
Abiotic stress resistance total	20	80
Cold tolerance	3	7
Stay-green	17	73
Biotic stress resistance total	20	49
Ergot resistance	2	4
Grain mould resistance	1	2
Greenbug resistance	3	8
Head bug resistance	2	4
Shoot fly resistance	7	21
Striga resistance	5	10
Stem composition total	19	56
Stem biomass yield	2	4
Stem composition	2	4
Sugar-related traits	15	48
Stem morphology total	25	107
Height	16	85
Regrowth	1	2
Rhizomatousness	1	2
Tillering	7	18
Grand total	125	400

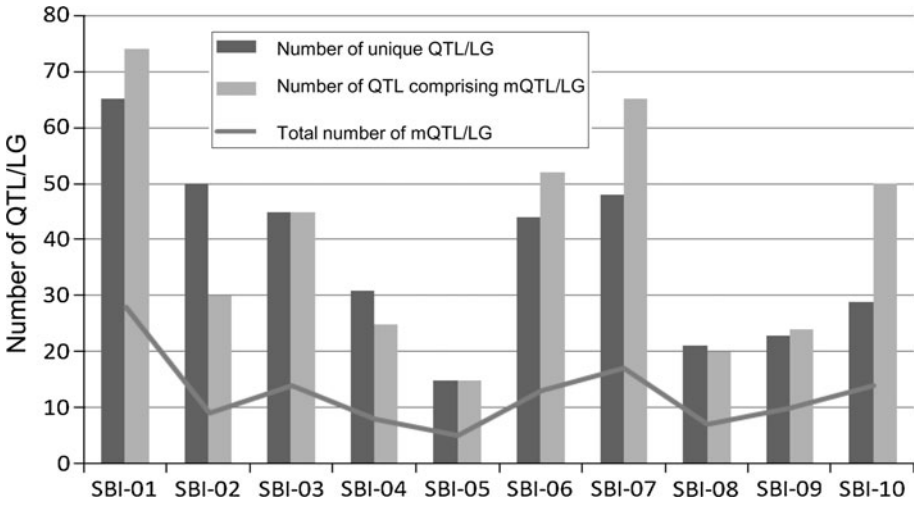


Fig. 2 Distribution of unique and meta-QTL across the genome

was higher in the heterochromatic regions across all chromosomes with a mean QTL density of 11 QTL/0.5 Mbp, as compared to 7.5 QTL/0.5 Mbp in euchromatin (ESM Figure S2B). Nine regions of very high (>80 genes/0.5 Mbp) gene density were identified across five chromosomes, all in the euchromatic regions: on SBI-01 (99 genes between 1.5 and 2.0 Mbp), SBI-02 (83 genes between 66.5 and 67.0 Mbp and 82 genes between 74.0 and 74.5 Mbp), SBI-03 (82 genes between 61.5 and 62.0 Mbp and 81 genes between 73.5 and 74.0 Mbp), SBI-06 (84 genes between 56.0 and 56.5 Mbp, 88 genes between 58.0 and 58.5 Mbp, 82 genes between 60.0 and 60.5 Mbp) and SBI-09 (80 genes between 56.0 and 56.5 Mbp). In comparison, QTL were enriched (>10 QTL/0.5 Mbp) in the heterochromatic regions of all chromosomes except SBI-02, SBI-03 and SBI-06.

Gene and QTL densities, on both the genetic linkage and physical scales, were compared for SBI-01 (Fig. 5). QTL density is represented both as a heat map and also as individual QTL plots to the left, on the genetic linkage scale, and to the right, on the physical scale. The individual plots of the QTL are colour coded for trait category as indicated. The non-linear relationship between genetic and physical distance in the euchromatic and heterochromatic regions, due to heterochromatic suppression of recombination, is shown through the contraction of the heterochromatic region on the genetic linkage distance scale. Similarly, the enrichment of both genes and QTL in the heterochromatin on the genetic linkage map, due to heterochromatic suppression of recombination, is clearly shown, in contrast to low gene density in heterochromatin on the physical scale.

Discussion

The primary objective of this study was to make information available from historical sorghum QTL experiments in a form that could be more readily used by sorghum researchers and plant breeders. To achieve this, we conducted a comprehensive comparative QTL study integrating, where possible, all of the QTL studies published in sorghum to date. We were able to project 771 QTL relating to 161 unique traits from 44 studies onto the sorghum consensus map and to compare to gene density, on both genetic linkage (cM) and physical (bp) map scales. In addition to the information on the projected QTL provided in ESM Table S1, the QTL information will also be available via Gramene (http://www.gramene.org) and the Molecular Marker Toolkit (Generation Challenge Programme 2009).

Comparative QTL mapping in sorghum

Any attempt to project QTL locations from multiple experiments onto a single genetic or physical map must take into account factors on which the reliability of the initial QTL study depends. The precision of initial QTL identification in QTL mapping studies is influenced by many factors including population size and type, precision of the phenotyping, QTL analysis methodology and marker order, coverage and density. These factors all impact on the principal parameters, which categorise QTL (position, CI, R^2 and LOD). For the sorghum studies integrated in this paper, this issue was further complicated by the fact that these parameters are not reported in a standardised manner across studies and in some cases are only partially reported or only available through graphical representations. To deal with this issue, we implemented the methodologies developed by Darvasi and Soller (1997) and adapted by Guo et al. (2006), which use parameters included in the majority of published QTL studies (location, R^2, population type and size) to determine the most likely location and CI of each QTL onto a single reference map. Using this method, the mean size of the estimated CI was 11.5 cM. In comparison, the mean CI for small effect ($R^2 < 10\%$) QTL, identified from populations with less than 150 individuals, was 15.9 cM, whereas the mean CI for large

Table 6 Characteristics of mQTL identified for kernel weight

mQTL_ID	LG	Position on consensus map (cM)	Flanking markers	No. of individual QTL comprising mQTL	Mean explained variance of QTL (R^2)%	Mean CI (cM)	Publications
QKWT_meta1.1	SBI-01	11.5	sPb-7015/sPb-7096	3	15.4	8.5	Murray et al. (2008a); Rami et al. (1998); Tuinstra et al. (1997)
QKWT_meta2.1	SBI-01	70.0	M187200/txp86	2	7.3	17.4	Feltus et al. (2006); Paterson et al. (1995a, b)
QKWT_meta1.2	SBI-02	160.0	M341504/txp296	2	4.8	17.4	Feltus et al. (2006); Paterson et al. (1995a)
QKWT_meta1.3	SBI-03	49.3	txs1053/txp33	2	11.0	12.2	Feltus et al. (2006); Rami et al. (1998)
QKWT_meta1.4	SBI-04	126.6	txp60/txs1901	2	15.9	7.5	Feltus et al. (2006); Brown et al. (2006)
QKWT_meta1.6	SBI-06	83.5	txs1206/txp274	3	10.4	11.0	Murray et al. (2008a); Feltus et al. (2006); Srinivas et al. (2009)
QKWT_meta1.7	SBI-07	73.5	gap342/sPb-5796	2	33.3	4.5	Rami et al. (1998)
QKWT_meta1.8	SBI-08	54.3	txp47/cdo459	4	10.8	17.1	Paterson et al. (1995a); Feltus et al. (2006); Murray et al. (2008a); Brown et al. (2006)
QKWT_meta1.10	SBI-10	63.5	sPb-6331/rz612	3	15.2	13.4	Feltus et al. (2006); Paterson et al. (1995a); Pereira et al. (1995)

effect ($R^2 > 25\%$) QTL, identified from populations with more than 150 individuals, was 3.9 cM.

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. In this study, the quality of the projection of each QTL location onto the consensus sorghum map was largely dependent on pair-wise comparison of common markers linked to the QTL and present on both the consensus map and the map generated for the original QTL experiment. Due to the link between the consensus map and the sequence-based physical map, the issue of too few common markers between studies could frequently be overcome, by retrospective in silico mapping of historical and current sequence-based molecular markers (primarily RFLPs and SSRs). To facilitate this activity, we further developed the existing consensus map of sorghum (Mace et al. 2009) and identified and added a further 1243 markers including an additional 888 DArTs, 229 SSRs, 81 RFLPs and 45 genes, via sequence mapping or marker projection. During this process, modifications were made to the telomeric end of the short arm of SBI-09 (ESM Table S2). The new expanded consensus map was integrated with over 7,000 collated sequence mapped markers recently described by Ramu et al. (2010). The use of these resources allowed the location of the majority of QTL to be identified by greatly increasing the probability of being able to place markers flanking QTL onto the consensus map. However, despite these resources it was not possible to project all of the QTL from the original studies onto the consensus map. This was the case when the order of markers in the genetic region containing the QTL was significantly different to the consensus map. In these regions, we compared the order of markers on the consensus and physical, sequence-based map and confirmed that the consensus map marker order was the same as that of the physical map. There are various possible reasons for marker order inconsistency across maps, as discussed in Mace et al. (2009). In other cases, it was not possible to project all of the QTL from the original study onto the consensus map due to there being too few markers in common. This was particularly problematic for studies that used maps established with a high numbers of AFLP or RAPD markers, due to the lack of sequenced or common markers that could be positioned on the consensus map. These two issues affected QTL from Agrama et al. (2002); Crasta et al. (1999); Murray et al. (2008a, b); Ritter et al. (2008); Salas Fernandez et al. (2008); Subudhi et al. (2000); and Tuinstra et al. (1997).

Use of comparative QTL data

One of the criticisms of many early QTL mapping experiments was that they had limited power due to small

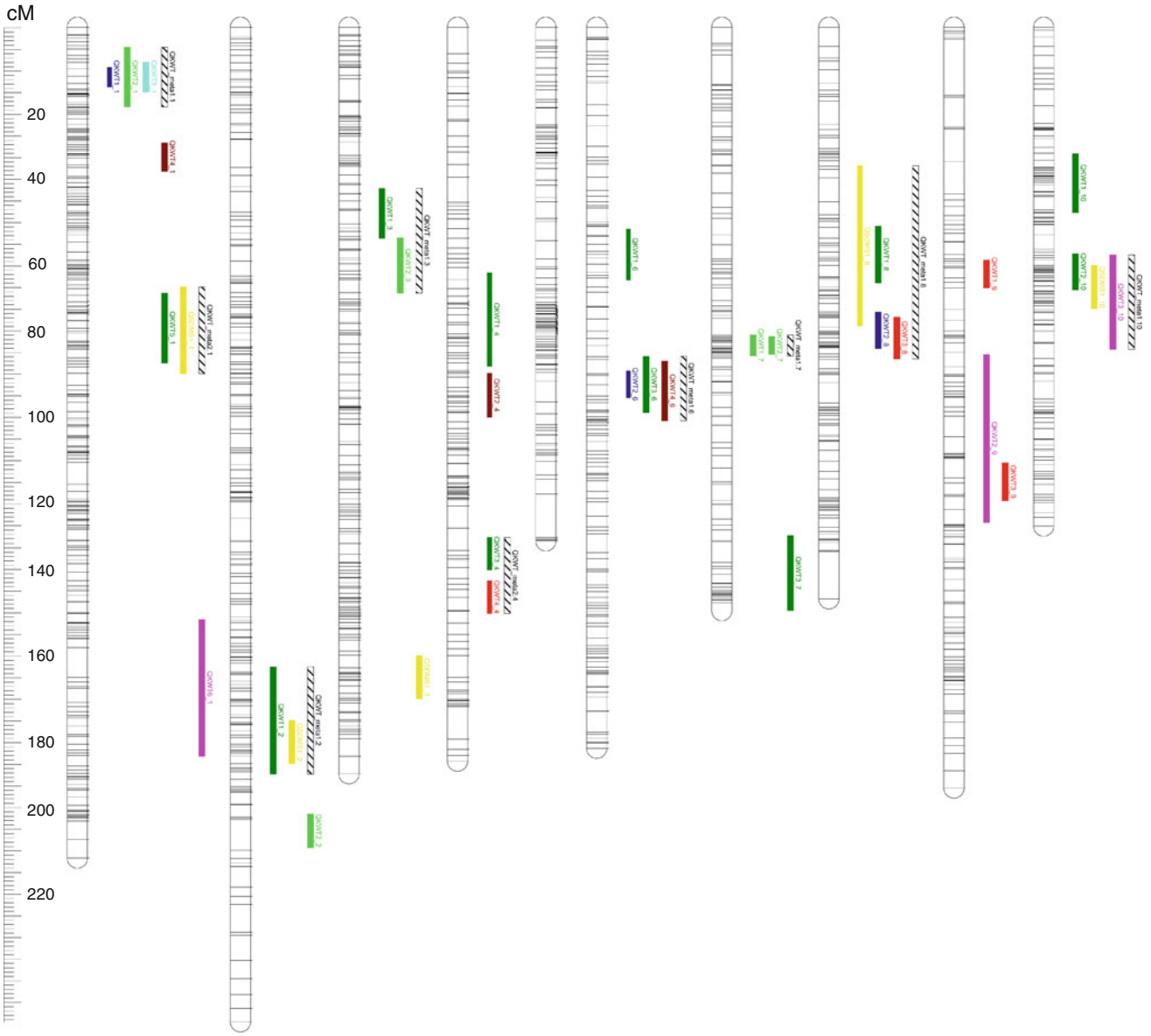


Fig. 3 QTL and mQTL for kernel weight identified from eight studies plotted against the consensus map

population sizes making them vulnerable to the detection of false QTL (Beavis 1994). The detection of QTL for the same trait in the same chromosome region in multiple independent studies provides strong evidence for the validity of that QTL, even when the size of the populations in the individual studies is low. The meta-QTL analysis we conducted for the trait kernel weight provides an example of the use of data from multiple studies to validate QTL. Tuinstra et al. (1997) detected a QTL for kernel weight on SBI-01 at ~12 cM using a population of 98 individuals. By current standards, it is probable that a QTL for a quantitative trait such as kernel weight, identified with a population size of just 98, would be reported as suggestive only. Subsequently, Rami et al. (1998) published a QTL in the same region, which explained 10% of the variation, based on a population of 110 individuals. Again, the number of individuals in that study was modest, but the fact that the region was associated with the trait in independent studies using unrelated genetic material provides strong evidence for the existence of a QTL for kernel weight in this region. Finally, Murray et al. (2008a) identified the same region as containing a QTL for kernel weight that explained 20% of the variation for kernel weight in a population of 176 individuals, providing very strong evidence that this region contains important genes for kernel weight that segregated in a range of genetic backgrounds and expressed in a range of environments. The kernel weight meta-analysis conducted also highlights the limitations of the analysis strategy adopted, compared with the meta-QTL

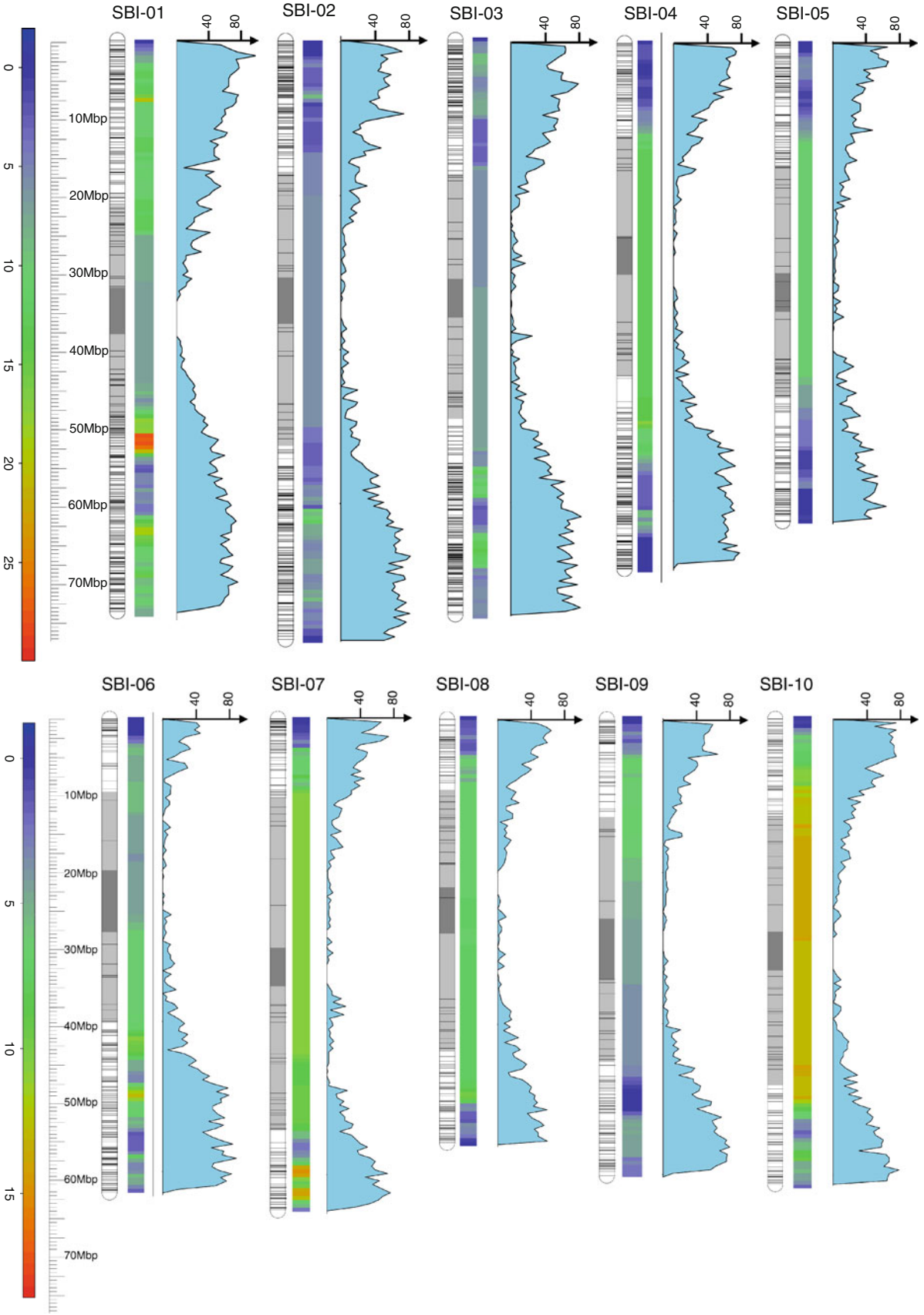
approach proposed by Goffinet and Gerber (2000) and refined by Veyrieras et al. (2007); specifically the loss of resolution when merging component QTL into an mQTL. In four of the nine mQTL identified for kernel weight (QKWT_meta1.3, QKWT_meta2.4, QKWT_meta1.6 and QKWT_meta1.8), the extent of the mQTL is greater than for any of the component QTL individually. In contrast, the meta-analysis approaches described by Goffinet and Gerber (2000) and implemented in BioMercator (Arcade et al. 2004) and meta-QTL (Veyrieras et al. 2007) aim to obtain more accurate estimates of QTL positions and to increase the resolution in the QTL region. In order to determine the most likely QTL distribution within a given genomic region, the approach proposed by Goffinet and Gerber (2000) requires a QTL data set of between 10 to 40 QTL lying within a genome region no longer than 200 cM, identified for the same trait across independent studies. Of the 161 unique traits identified in this study, only 11 traits with at least 10 identified QTL were identified across a minimum of two studies. Of these 11 traits, only 1 (plant height) had over 10 QTL on a single LG. Hence, due to the limited number of QTL per trait per LG used in this study, the meta-analysis approaches described by Goffinet and Gerber (2000) were not suitable. Instead, the meta-analysis strategy adopted here provides strong evidence for the association between a particular genomic region and selected traits, when the multiple QTL for the same traits are identified in independent studies. Additionally, the frequency with which particular QTL are detected in multiple studies provides information about their likely frequency in breeding populations, while variation in effect sizes provides some information about the likely range of alternative alleles that may be present (e.g. Griffiths et al. 2009). Similarly, the accuracy of the original QTL mapping can be confirmed in cases where the underlying gene responsible for the QTL is known, e.g. dw_3 (sb07g0232730) with a physical start location of 58,610,896 bp and a predicted cM location of approximately 104 cM on SBI-07. In this case, 7 QTL (Pereira et al. 1995; Pereira and Lee 1995; Feltus et al. 2006; Murray et al. 2008a, b; Srinivas et al. 2009) with LOD scores ranging from 2.97 to 8.4, R^2 values ranging from 9 to 29% and peak positions ranging from 102.5 to 109.3 cM were combined into an mQTL covering the gene.

The integration of QTL for 161 traits provides the opportunity to identify other traits that may be affected either positively or negatively by selection for a particular trait or QTL. For example, the 18 QTL for grain and panicle yield were found to co-localise with QTL for height, maturity, tillering, kernel weight, stay-green and resistances to ergot, shoot fly and striga. Such co-localisations are either the result of genetic linkage between QTL for the same trait or due to pleiotropic effects of a QTL on multiple traits. Pleiotropic effects are likely to contribute to a trait-like yield, which is determined by the cumulative effect of many processes. Such co-localisations are commonly observed in other crops including wheat (Quarrie et al. 2006), rice (Shomura et al. 2008), barley (Waller et al. 2005), soybean (Yuan et al. 2002) and brassicas (Shi et al. 2009). However, genetic linkage between co-locating QTL is also likely to be important. Where two traits are measured in the same study but do not show co-locating QTL, it provides some evidence for the association observed in the meta-analysis being the result of linkage rather than pleiotropy. Since selection is usually carried out in breeding populations that are often unrelated to mapping populations, the co-location of QTL serves as an indicator that caution may be required before aggressive selection is employed, particularly marker-assisted selection. In such cases, an understanding of linkage disequilibrium between QTL alleles for important traits or knowledge of pleiotropic effects of particular QTL is needed.

Fig. 4 QTL and gene density plots along the sequence-based physical map. The 2335 sequence-mapped loci are displayed schematically by *horizontal lines* across the *bars* representing the ten chromosomes. Heterochromatic regions on each chromosome are shaded *grey*, with the centromere shaded *dark grey*. The *scale bar* to the *left* indicates the length of each chromosome in bp. The heat map to the *right* of each chromosome represents QTL density (number of QTL/0.5 Mbp) and the graph indicates gene density (number of genes/0.5 Mbp)

Gene and QTL density as affected by physical and genetic distance

It is well known that genetic distance and physical distance, while correlated, do not vary consistently along chromosomes. With the advent of whole genome sequencing projects in a range of crop species, including sorghum, it is now possible to accurately determine gene order and align physical and genetic distances. With the exception of errors in determining marker order or chromosomal rearrangements, the relationship between order in genetic and physical maps should be consistent. In sorghum, where chromosomal rearrangements appear rare, the relationship between physical and genetic order is highly consistent (Mace et al. 2009). In contrast, the relationship between genetic and physical distances is not as consistent as the relationship between gene and marker order on the physical and genetic linkage scales. This inconsistency is mainly due to the variation in the average number of recombination events in different regions of the genome (e.g. Akhunov et al. 2003; Kao et al. 2006). In sorghum, areas of low recombination are observed in the heterochromatic regions surrounding the centromeres, while recombination is usually high towards the ends of the chromosome arms (Kim et al. 2005). Gene density, on a physical basis, is also known to vary across the genome with gene-rich regions generally


SBI-01
SBI-02
SBI-03
SBI-04
SBI-05
SBI-06
SBI-07
SBI-08
SBI-09
SBI-10
0
5
10
15
20
25
10Mbp
20Mbp
30Mbp
40Mbp
50Mbp
60Mbp
70Mbp
40
80

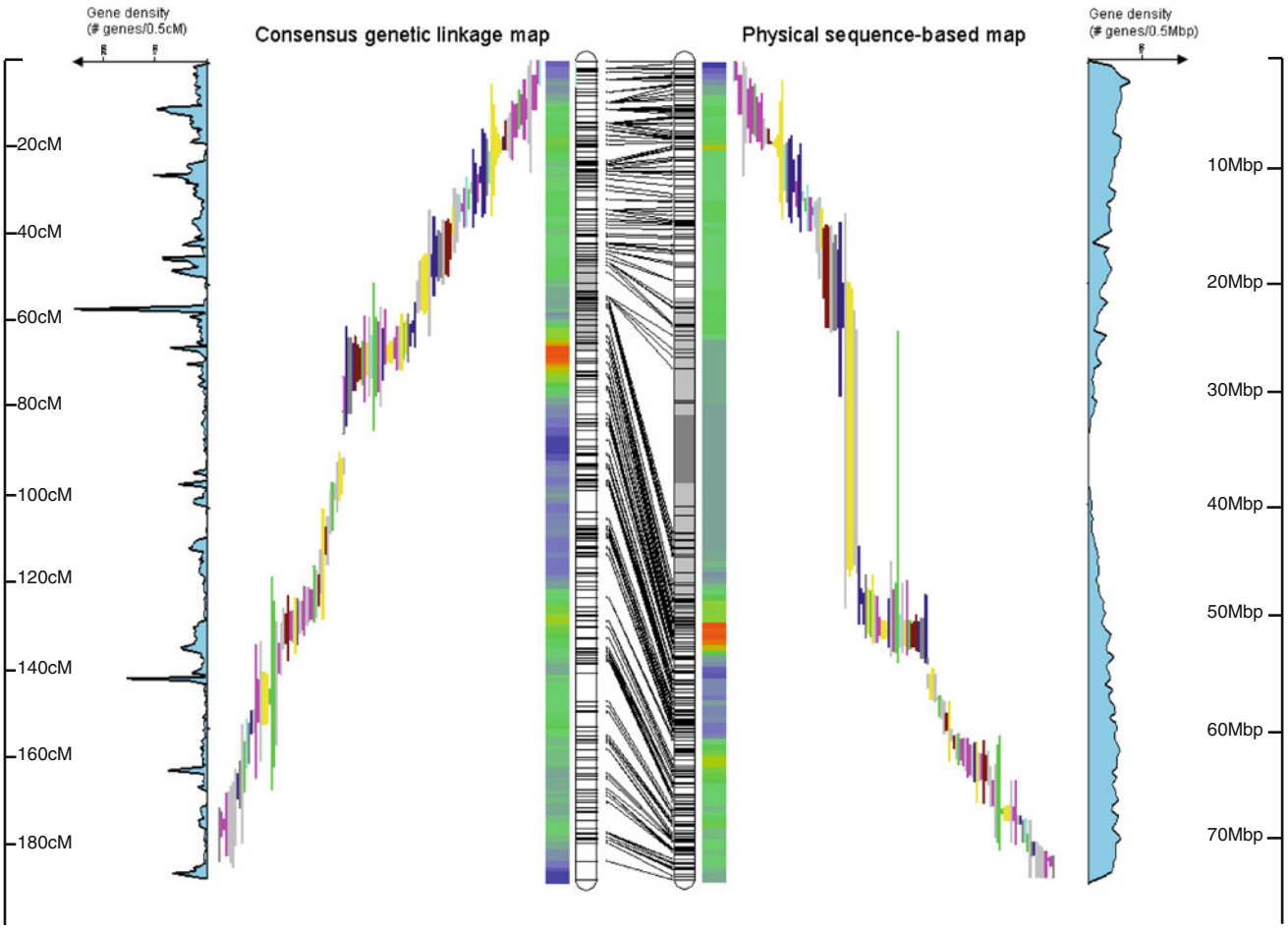


Fig. 5 QTL and gene density comparisons on SBI-01 on the consensus genetic linkage map (# QTL/0.5 cM, on the *left-hand side*) and sequence-based physical map (# QTL/0.5Mbp, on the *right-hand side*). QTL density is shown both as a heat map adjacent to each map and as projected CI plots, colour coded as indicated below. Gene density is plotted on the graphs as the # genes/0.5 cM, on the *far left*, and as # genes/0.5Mbp, on the *far right*. QTL colour coded by trait category as follows: grain *red*; leaf *dark blue*; maturity *yellow*; panicle *pink*; abiotic stress resistance *bright green*; biotic stress resistance *brown*; stem composition *bright blue*; stem morphology *dark green*

occurring in the euchromatic regions on the chromosome arms, while regions located in the heterochromatin closer to the centromere have relatively lower gene densities (Paterson et al. 2009).

For plant breeding applications genetic distance, rather than physical distance, remains the critical factor governing breeders' capacity to manipulate genes that reside on the same chromosome. This variation in genetic and physical distance has a range of implications for the use of the sorghum whole genome sequence information by plant breeders. The full benefits of information on molecular markers, QTL positions and the whole genome sequence information will only come when this information is integrated. In this study, the use of the consensus map, linked directly to the sequence-based physical map through large numbers of sequence-based markers, enabled us to compare both QTL and gene density on two different scales, genetic distance (cM) and physical distance (bp) scale. For such a comparison to be broadly applicable, the genetic distances on the consensus map must be representative of the genetic distances between common marker pairs in populations of interest to plant breeders. In constructing the consensus map (Mace et al. 2009), we compared distances across six individual component maps and found them to be broadly consistent. The six populations used to construct this map also encompassed a significant proportion of sorghum diversity, enhancing the utility of the gene and QTL density plots against the consensus map.

Following the integration of gene and QTL density on the genetic linkage (cM) scale, regions of high gene and QTL density were observed, somewhat counter-intuitively, around the centromeres of seven chromosomes (SBI-01, SBI-04, SBI-05, SBI-06, SBI-07, SBI-08 and SBI-10). When plotted against physical distances, the regions around the centromeres have low gene density compared with the chromosome arms. The heterochromatic gene

enrichment, as observed on the genetic linkage scale, was the result of suppressed recombination in heterochromatic regions. In the heterochromatic regions, 1 cM on average represented approximately 40 times the physical distance (8.46 Mbp) than 1 cM in euchromatin (0.22 Mbp). Overall, approximately 20% of the predicted gene models are located in the heterochromatic region (Paterson et al. 2009). Similarly, 22% of the QTL were either entirely or partially located in the heterochromatic regions. Considering that on average, across all chromosomes, the heterochromatic region makes up ~4.1% of the total genetic linkage distance, the number of QTL in heterochromatin is much greater than would be expected based on an even genome distribution ($p < 0.00001$ based on CI peak only). The QTL contained in these regions were detected in multiple studies and contained a mixture of related and unrelated traits. While some of these QTL can be due to pleoitropic effects of a single gene, in many cases there is no biological support for pleiotropy suggesting that linkage is the primary cause of the increased QTL density. Three gene-rich (>150 genes/0.5 cM) regions were located in the euchromatic regions on SBI-01, SBI-04 and SBI-10. In all cases, there was a suggestion that these regions were also associated with high increased QTL frequency. Detailed investigation of the individual maps used to construct the consensus map (Mace et al. 2009) indicates that gene-rich areas have similar genetic size on the component maps suggesting that these are not artefacts resulting from errors in estimation of genetic distance.

In other cases, regions of the genome were observed with increased QTL density, but no obvious increase in gene density. It is likely that this increase in QTL density is the result of the non-random selection of the traits investigated (e.g height and maturity measured in many studies) and the impact of major genes with pleiotropic effects across a number of traits. For example, there was a region of high QTL density surrounding the location of the major height gene dw_3, which contains a number of plant height QTL. The height gene, dw_3, is also known to have pleiotropic effects on a number of other traits including tiller number and grain size (Casady 1965). QTL for these traits were located nearby, as were QTL for biomass yield, lodging, panicle architecture, sugar-related traits and green bug resistance. This contributed to the high QTL density observed in this region, and the overall significantly higher than expected number of QTL on SBI-07 (chi-square = 35.13, 1 *df*, $p < 0.00001$).

Implications for sorghum genetic improvement

The presence of regions of high gene density in the heterochromatic regions around the centromeres and the associated higher QTL density when plotted against a genetic linkage map have a number of important implications for sorghum breeders.

In the developed world, sorghum breeders exploit the phenomena of heterosis or hybrid vigour through the use of F_1 cultivars produced using a cytoplasmic male sterility system. These F_1 hybrids are generally higher yielding and more vigorous than their inbred parents. Estimates of mid-parent heterosis for grain yield range from 6 to 54% (Kirby and Atkins 1968; Liang and Walter 1968; Wenzel 1988). A variety of theories have been advanced to describe potential genetic causes of this phenomenon. These include the dominance (Davenport 1908; Bruce 1910; Keeble and Pellew 1910), overdominance (East 1908; Shull 1908) and pseudo-overdominance (Jones 1917) theories. The dominance theory proposes that independent sets of slightly deleterious alleles accumulate throughout the genome during inbreeding of parental lines. Hybridization produces F_1 progeny with complementary superior alleles. The overdominance theory proposes that allelic interactions at a single heterozygous locus result in a synergistic effect on performance that surpasses both homozygous parents. The pseudo-overdominance theory is a variation of the dominance theory, where favourable alleles of multiple genes are linked in repulsion phase such that hybridization brings together complementary superior alleles, which are difficult to accumulate within a single individual. While it is likely that all of these mechanisms contribute to heterosis, there is growing consensus that dominance (and pseudo-overdominance) is important. In sorghum, alleles with favourable effects on fitness tend to show dominance in the heterozygous condition over unfavourable alleles. Jordan et al. (2003) observed that parental inbred diversity, measured using molecular markers distributed over the genome, explained approximately 18% of the variation in yield in a set of 161 F_1 hybrids. The observation of high gene density in recombination-poor regions in the current study, coupled with the association with increased QTL density, creates conditions favourable for pseudo-overdominance to occur, as also observed recently in maize (Gore et al. 2009). Under such conditions, it is likely that, on average, parents that are genetically diverse in these regions of low recombination will produce hybrids that express higher levels of hybrid vigour. This has potential implications for the development of more efficient breeding systems for sorghum and warrants further investigations. For example, it is possible that heterotic groups in sorghum could be enhanced by ensuring that complementary regions are accumulated in male and female germplasm pools resulting in maximal expression of heterosis in F_1 hybrids.

For many applications, genetic distances based on evenly distributed molecular markers are used as a measure of the likelihood that individuals contain similar genes (e.g. Dean et al. 1999) and for a variety of other purposes,

including estimating the polygenic background effects in association mapping studies (e.g. Brown et al. 2008). The results of this study suggest that, at least in sorghum, a strategy to weight markers targeting gene and QTL-rich regions of the genome may be more effective in estimating true genetic, and therefore phenotypic, similarity.

Conclusion

In this study, an analysis of QTL location of 771 QTL across 44 studies and 161 traits was conducted based on the adjusted confidence intervals of QTL location, projected onto the sorghum consensus map. The consensus map was expanded through the addition of 1,243 additional markers, increasing the overall marker density from 1 marker/0.79 cM to 1 marker/0.48 cM. Additionally, 125 mQTL were identified representing 400 individual QTL across all trait categories, with the exception of the leaf trait category. The use of the consensus map, linked directly to the sequence-based physical map through 2,335 sequence-based markers, enabled us to compare QTL on two different scales, genetic distance (cM) and physical distance (bp) scale. This integrated a wealth of information, and in particular gene density, associated with the recently completed whole genome sequence for sorghum (Paterson et al. 2009). For practical plant breeding applications, genetic distance rather than physical distance, remains the critical factor governing the capacity for breeders to manipulate genes that reside on the same chromosome. However, the direct use of the sorghum whole genome sequence information by plant breeders is complicated by the non-linear relationship between genetic and physical distance. Access to the full benefits of the recent expansion in genomics resources in sorghum can be enhanced by integrating information on molecular markers, QTL positions, gene density and the whole genome sequence. The projected QTL information described, combined with resources such as the physical locations of all the publicly available sorghum sequence-based markers collated by Ramu et al. (2010), and predicted gene models as detailed on Phytozome, provides sorghum researchers with a useful resource for more detailed analyses of the genetic control of target traits and development of more effective marker-assisted breeding strategies. The inclusion of both genetic (cM) and physical (bp) co-ordinates of CIs in future QTL studies would contribute to the ongoing utility of such a resource.

Acknowledgments The authors would like to acknowledge Colleen Hunt for her assistance in generating the heat maps, and Jerry Franckowiak, Mandy Christopher, Tracey Shatte, Ian Godwin and Peter Prentis for their critical review of the manuscript. We would also like to thank the anonymous reviewers for their comments and suggestions. We also acknowledge the Australian Grains Research and Development Cooperation (GRDC; http://www.grdc.com.au) for contributing a part of the funding for this research.

References

Agrama HA, Widle GE, Reese JC, Campbell LR, Tuinstra MR (2002) Genetic mapping of QTLs associated with greenbug resistance and tolerance in *Sorghum bicolor*. Theor Appl Genet 104:1373–1378

Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, Miftahudin Gustafson JP, Lazo G, Chao SM, Anderson OD, Linkiewicz AM, Dubcovsky J, La Rota M, Sorrells ME, Zhang DS, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng JH, Lapitan NLV, Gonzalez-Hernandeiz JL, Anderson JA, Choi DW, Close TJ, Dilbirligi M, Gill KS, Walker-Simmons MK, Steber C, McGuire PE, Qualset CO, Dvorak J (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13:753–763

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

Beavis WD (1994) The power and deceit of QTL experiments: Lessons from comparative QTL studies. In: Proceedings of 49th Ann Corn Sorghum Res Conf Am Seed Trade Ass, Washington, DC, pp 250–266

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

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:629–637

Bruce AB (1910) The mendelian theory of heredity and the augmentation of vigour. Science 32:627–628

Casady AJ (1965) Effect of a single height Dw gene of Sorghum on grain yield, grain yield components, and test weight. Crop Sci 5:385–388

Chantereau J, Trouche G, Rami JF, Deu M, Barro C, Grivet L (2001) RFLP mapping of QTLs for photoperiod response in tropical sorghum. Euphytica 120:183–194

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

Cone KC, McMullen MD, Bi IV, Davis GL, Yim YS, Gardiner JM, Polacco ML, Sanchez-Villeda H, Fang ZW, Schroeder SG, Havermann SA, Bowers JE, Paterson AH, Soderlund CA, Engler FW, Wing RA, Coe EH (2002) Genetic, physical and informatics resources for maize. On the road to an integrated map. Plant Physiol 130:1598–1605

Crasta OR, Xu WW, Nguyen HT, Rosenow DT, Mullet J (1999) Mapping of post flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 262:579–588

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

Davenport C (1908) Degeneration, albinism and inbreeding. Science 28:454–455

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S (1999) Genetic redundancy and diversity among 'orange' accessions in the US National Sorghum Collection as assessed with simple sequence repeat (SSR) markers. Crop Sci 39:1215–1221

Deu M, Ratnadass A, Hamada MA, Noyer JL, Diabate M, Chantereau J (2005) Quantitative trait loci for head-bug resistance in sorghum. Afr J Biotechnol 4:247–250

East E (1908) Inbreeding in corn. Rep Connecticut Agric Exp Stat 1907–1908, pp 419–428

Ejeta G, Knoll J (2007) Marker-assisted selection in sorghum. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement. Springer, Berlin, pp 187–205

Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, Klein PE, Brown PJ, Paterson AH (2006) Alignment of genetic maps and QTLs between inter- and intraspecific sorghum populations. Theor Appl Genet 112:1295–1305

Generation Challenge Programme (2009). The GCP Molecular Marker Toolkit. Material developed by Veerle Van Damme for the Generation Challenge Programme and available online at http://www.generationcp.org/sp5_toolboxhomepage_april_09

Goffinet B, Gerber S (2000) Quantitative trait loci: a meta-analysis. Genetics 155:463–473

Gore MA, Chia J-M, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Griffiths S, Simmonds J, Leverington M, Wang Y, Fish L, Sayers L, Alibert L, Orford S, Wingen L, Herry L, Faure S, Laurie D, Bilham L, Snape J (2009) Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. Theor Appl Genet 119:383–395

Guo B, Sleper 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

Hart GE, Schertz KF, Peng Y, Syed NH (2001) Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet 103:1232–1242

Haussmann BIG, Mahalakshmi V, Reddy BVS, Seetharama N, Hash CT, Geiger HH (2002) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106:133–142

Haussmann BIG, Hess DE, Omanya GO, Folkertsma RT, Reddy BVS, Kayentao M, Welz HG, Geiger HH (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. Theor Appl Genet 109:1005–1016

Jones D (1917) Dominance of linked factors as a means of accounting for heterosis. Proc Natl Acad Sci USA 3:310–312

Jordan DR, Tao Y, Godwin ID, Henzell RG, Cooper M, McIntyre CL (2003) Prediction of hybrid performance in grain sorghum using RFLP markers. Theor Appl Genet 106:559–567

Kao FI, Cheng YY, Chow TY, Chen HH, Liu SM, Cheng CH, Chung MC (2006) An integrated map of *Oryza sativa* L. chromosome 5. Theor Appl Genet 112:891–902

Katsar CS, Paterson AH, Teetes GL, Peterson GC (2002) Molecular analysis of Sorghum resistance to the greenbug (Homoptera:Aphididae). J Econ Entomol 95:448–457

Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants: where are we now? Heredity 80:137–142

Kebede H, Subadhi PK, Rosenow DT, Nguyen HT (2001) Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 103:266–276

Keeble F, Pellew C (1910) The mode of inheritance of stature and time of flowering in peas (*Pisum sativum*). J Genet 1:47–56

Khowaja FS, Norton GJ, Courtois B, Price AH (2009) Improved resolution in the position of drought-related QTLs in a single mapping population of rice by meta-analysis. BMC Genomics 10

Kim J (2003) Genomic Analysis of sorghum by fluorescence in situ hybridization. PhD thesis. Texas A&M University

Kim JS, Islam-Faridi MN, Klein PE, Stelly DM, Price HJ, Klein RR, Mullet JE (2005) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics 171:1963–1976

Kirby JS, Atkins RE (1968) Heterotic response for vegetative and mature plant characters in grain sorghum, *Sorghum bicolor* L.) Moench. Crop Sci 8:335–339

Klein RR, Rodriguez-Herrera R, Schlueter JA, Klein PE, Yu ZH, Rooney WL (2001) Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum. Theor Appl Genet 102:307–319

Knoll J, Gunaratna N, Ejeta G (2008) QTL analysis of early-season cold tolerance in sorghum. Theor Appl Genet 116:577–587

Liang GHL, Walter TL (1968) Heritability estimates and gene effects for agronomic traits in grain sorghum, *Sorghum vulgare* Pers. Crop Sci 8:77–81

Lijavetzky D, Martinez MC, Carrari F, Hopp HE (2000) QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. Euphytica 112:125–135

Lin Y-R, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411

Liu SY, Hall MD, Griffey CA, McKendry AL (2009) Meta-analysis of QTL associated with fusarium head blight resistance in wheat. Crop Sci 49:1955–1968

Mace E, Jordan D (2010) Location of major effect genes in sorghum (*Sorghum bicolor* (l.) Moench). Theor Appl Genet. doi: 10.1007/s00122-010-1392-8

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9:26

Mace ES, Rami JF, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran 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

Menz MA, Klein RR, Mullet J, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP(R), RFLP and SSR markers. Plant Mol Biol 48:483–499

Mohan SM, Madhusudhana R, Mathur K, Howarth CJ, Srinivas G, Satish K, Reddy RN, Seetharama N (2009) Co-localization of quantitative trait loci for foliar disease resistance in sorghum. Plant Breed 128:532–535

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

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

Nagaraj N, Reese JC, Tuinstra MR, Smith CM, Amand PS, Kirkham MB, Kofoid KD, Campbell LR, Wilde GE (2005) Molecular mapping of sorghum genes expressing tolerance to damage by greenbug (Homoptera: Aphididae). J Econ Entomol 98:595–602

Nagy ED, Lee TC, Ramakrishna W, Xu ZJ, Klein PE, SanMiguel P, Cheng CP, Li JL, Devos KM, Schertz K, Dunkle L, Bennetzen JL (2007) Fine mapping of the Pc locus of *Sorghum bicolor*, a

gene controlling the reaction to a fungal pathogen and its host-selective toxin. Theor Appl Genet 114:961–970

Natoli A, Gorni C, Chegdani F, Marsan PA, Colombi C, Lorenzoni C, Marocco A (2002) Identification of QTLs associated with sweet sorghum quality. Maydica 47:311–322

Norton GJ, Aitkenhead MJ, Khowaja FS, Whalley WR, Price AH (2008) A bioinformatic and transcriptomic approach to identifying positional candidate genes without fine mapping: an example using rice root-growth QTLs. Genomics 92:344–352

Parh D (2005) DNA-based markers for ergot resistance in sorghum. PhD thesis. University of Queensland, Brisbane

Parh DK, Jordan DR, Aitken EAB, Mace ES, Jun-ai P, McIntyre CL, Godwin ID (2008) QTL analysis of ergot resistance in sorghum. Theor Appl Genet 117:369–382

Paterson AH, Lin Y-R, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S-C, Stansel JW, Irvine JE (1995a) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Paterson AH, Schertz KF, Lin Y-R, Liu S-C, Chang Y-L (1995b) The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. Proc Natl Acad Sci USA 92:6127–6131

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang HB, Wang XY, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang LF, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90:380–388

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

Perumal R, Menz MA, Mehta PJ, Katile S, Gutierrez-Rojas LA, Klein RR, Klein PE, Prom LK, Schlueter JA, Rooney WL, Magill CW (2009) Molecular mapping of Cg1, a gene for resistance to anthracnose (*Colletotrichum sublineolum*) in sorghum. Euphytica 165:597–606

Quarrie SA, Quarrie SP, Radosevic R, Rancic D, Kaminska A, Barnes JD, Leverington M, Ceoloni C, Dodig D (2006) Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. J Exp Bot 57:2627–2637

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

Ramu P, Deshpande SP, Senthivel S, Jayashree B, Billot C, Deu M, Ananda Reddy L, Hash CT (2010) In silico mapping of important genes and markers available in the public domain for efficient sorghum breeding. Mol Breed. doi:10.1007/s11032-009-9365-9

Ritter KB, Jordan DR, Chapman SC, Godwin ID, Mace ES, McIntyre CL (2008) Identification of QTL for sugar-related traits in a sweet x grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Mol Breed 22:367–384

Rong J, Feltus EA, Waghmare VN, Pierce GJ, Chee PW, Draye X, Saranga Y, Wright RJ, Wilkins TA, May OL, Smith CW, Gannaway JR, Wendel JR, Paterson AH (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577–2588

Salas Fernandez MG, Hamblin MT, Li L, Rooney WL, Tuinstra MP, Kresovich S (2008) Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. Crop Sci 48:1732–1743

Satish K, Srinivas G, Madhusudhana R, Padmaja PG, Nagaraj Reddy R, Murali Mohan S, Seetharama N (2009) Identification of quantitative trait loci (QTL) for resistance to shoot fly in sorghum [*Sorghum bicolor* (L.) Monech]. Theor Appl Genet 119:1425–1439

Schuler GD (1998) Electronic PCR: bridging the gap between genome mapping and genome sequencing. Trends Biotechnol 16:456–459

Shi JQ, Li RY, Qiu D, Jiang CC, Long Y, Morgan C, Bancroft I, Zhao JY, Meng JL (2009) Unraveling the Complex Trait of Crop Yield With Quantitative Trait Loci Mapping in *Brassica napus*. Genetics 182:851–861

Shiringani AL, Frisch M, Friedt W (2010) Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. Theor Appl Genet 121:323–336

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nature Genet 40:1023–1028

Shull G (1908) The composition of a field of maize. Ann Breed Assoc Rep 4:296–301

Srinivas G, Satish K, Madhusudhana R, Nagaraja Reddy R, Murali Mohan S, Seetharama N (2009) Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. Theor Appl Genet 118:1439–1454

Srinivasa Reddy P, Fakrudin B, Rajkumar, Punnuri SM, Arun SS, Kuruvinashetti MS, Das IK, Seetharama N (2008) Molecular mapping of genomic regions harboring QTLs for stalk rot resistance in sorghum. Euphytica 159:191–198

Subudhi PK, Rosenow DT, Nguyen HT (2000) Quantitative trait loci for the stay green trait in sorghum (*Sorghum bicolor* L. Moench): consistency across genetic backgrounds and environments. Theor Appl Genet 101:733–741

Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998) Identification of genomic regions for rust resistance in sorghum. Euphytica 103:287–292

Tao YZ, Henzell RG, Jordan DR, Butler DG, Kelly AM, McIntyre CL (2000) Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments. Theor Appl Genet 100:1225–1232

Tao YZ, Hardy A, Drenth J, Henzell RG, Franzmann BA, Jordan DR, Butler DG, McIntyre CL (2003) Identifications of two different mechanisms for sorghum midge resistance through QTL mapping. Theor Appl Genet 107:116–122

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1996) Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum. Crop Sci 36:1337–1344

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1997) Genetic analysis of post-flowering drought tolerance and components of grain development in *Sorghum bicolor* (L.) Moench. Mol Breed 3:439–448

Veyrieras JB, Goffinet B, Charcosset A (2007) MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. BMC Bioinformatics 8:49

Waller F, Achatz B, Baltruschat H, Fodor J, Becker K, Fischer M, Heier T, Huckelhoven R, Neumann C, von Wettstein D, Franken P, Kogel KH (2005) The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease

resistance, and higher yield. Proc Natl Acad Sci USA 102:13386–13391

Wenzel WG (1988) Relationships between yield, base temperature and drought resistance in grain sorghum. S Afr J Plant Soil 5:150–154

Winn J, Mason R, Robbina A, Rooney W, Hays D (2009) QTL mapping of a high protein digestibility trait in *Sorghum bicolor*. Int J Plant Genomics. doi:10.1155/2009/471853

Wu YQ, Huang YH (2008) Molecular mapping of QTLs for resistance to the greenbug *Schizaphis graminum* (Rondani) in *Sorghum bicolor* (Moench). Theor Appl Genet 117:117–124

Wu YQ, Huang YH, Porter DR, Tauer CG, Hollaway L (2007) Identification of a major quantitative trait locus conditioning resistance to greenbug Biotype E in sorghum PI550610 using simple sequence repeat markers. J Econ Entomol 100: 1672–1678

Xu SZ (2003) Theoretical basis of the Beavis effect. Genetics 165:2259–2268

Xu WW, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43:461–469

Yuan J, Njiti VN, Meksem K, Iqbal MJ, Triwitayakorn K, Kassem MA, Davis GT, Schmidt ME, Lightfoot DA (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Sci 42:271–277