Theor Appl Genet (2010) 120:1279–1287
DOI 10.1007/s00122-009-1255-3

ORIGINAL PAPER

Molecular mapping and candidate gene identification of the *Rf2* gene for pollen fertility restoration in sorghum [*Sorghum bicolor* (L.) Moench]

D. R. Jordan · Emma S. Mace · R. G. Henzell · P. E. Klein · R. R. Klein

Received: 13 September 2009 / Accepted: 15 December 2009 / Published online: 22 January 2010

Abstract The A1 cytoplasmic–nuclear male sterility system in sorghum is used almost exclusively for the production of commercial hybrid seed and thus, the dominant genes that restore male fertility in F_1 hybrids are of critical importance to commercial seed production. The genetics of fertility restoration in sorghum can appear complex, being controlled by at least two major genes with additional modifiers and additional gene–environment interaction. To elucidate the molecular processes controlling fertility restoration and to develop a marker screening system for this important trait, two sorghum recombinant inbred line populations were created by crossing a restorer and a non-restoring inbred line, with fertility phenotypes evaluated in hybrid combination with three unique cytoplasmic male sterile lines. In both populations, a single major gene segregated for restoration which was localized to chromosome SBI-02 at approximately 0.5 cM from microsatellite marker, *Xtxp304*. In the two populations we observed that approximately 85 and 87% of the phenotypic variation in seed set was associated with the major *Rf* gene on SBI-02. Some evidence for modifier genes was also observed since a continuum of partial restored fertility was exhibited by lines in both RIL populations. With the prior report (Klein et al. in Theor Appl Genet 111:994–1012, 2005) of the cloning of the major fertility restoration gene *Rf1* in sorghum, the major fertility restorer locus identified in this study was designated *Rf2*. A fine-mapping population was used to resolve the *Rf2* locus to a 236,219-bp region of chromosome SBI-02, which spanned ~31 predicted open reading frames including a pentatricopeptide repeat (PPR) gene family member. The PPR gene displayed high homology with rice *Rf1*. Progress towards the development of a marker-assisted screen for fertility restoration is discussed.

Communicated by Y. Xue.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1255-3) contains supplementary material, which is available to authorized users.

D. R. Jordan (✉) · E. S. Mace · R. G. Henzell
Queensland Primary Industries and Fisheries,
Hermitage Research Station, MS508,
Warwick, QLD 4370, Australia
e-mail: David.R.Jordan@deedi.qld.gov.au

E. S. Mace
e-mail: emma.mace@deedi.qld.gov.au

P. E. Klein
Department of Horticulture and Institute for Plant
Genomics and Biotechnology, Texas A&M University,
College Station, TX 77843-2123, USA

R. R. Klein
USDA-ARS, Southern Plains Agricultural Research Center,
College Station, TX 77845-9593, USA

Introduction

The exploitation of heterosis using F_1 hybrid cultivars is a major component of the strategy used by grain sorghum breeding programs in the developed world. This strategy relies exclusively on cytoplasmic–nuclear male sterility (CMS) to provide a cost-effective method for producing large quantities of hybrid seed. CMS in sorghum was first described by Stephens and Holland (1954) who observed that the interaction of "milo" or A1 cytoplasm and genes of "kafir" origin produced plants with male sterility and normal female fertility. Despite the identification of other CMS systems in sorghum (Schertz et al. 1989), the A1

CMS system is used almost exclusively for commercial sorghum hybrids.

A critical element of the success of the A1 CMS system has been the capacity to develop male parent lines (restorer or R lines) which carry dominant genes that restore male fertility in hybrid cultivars. Despite its importance to commercial sorghum production, the genetic architecture of fertility restoration is not well understood. Classical genetics studies indicate that fertility restoration in A1 cytoplasm is complex, but the majority of genotypic variation was reported to be controlled by 2–3 major genes with modifiers (Maunder and Pickett 1959; Miller and Pickett 1964; Erichsen and Ross 1963). Restoration is also influenced by environmental conditions, with cool conditions around flowering favouring sterility and high temperatures favouring fertility (Downs and Marshall 1971; Brooking 1976, 1979). The combination of moderately complex genetic control and environmental variation produces the continuum of male fertility that is observed in a range of A1 cytoplasm lines. At one end of this continuum are lines which exhibit high levels of sterility characterised by very small pale anthers which rarely produce viable pollen. Intermediate or partially fertile lines are characterised by larger darker coloured anthers that shed variable amounts of pollen depending on environmental conditions. At the other end of the continuum are fertile lines that are fully fertile and produce copious amounts of viable pollen under most environmental conditions. Segregation in breeding populations derived from crosses between fertile and sterile parents varies considerably from semi-qualitative to highly quantitative (Jordan, unpublished data).

The lack of detailed understanding of genetic control of fertility restoration places considerable constraints on sorghum breeders. For example, when developing new restorer lines using exotic germplasm or even crosses involving two restorer lines, the sole means by which a sorghum breeder can definitively determine the restoration status of a line is to cross it to a male sterile line and assess the fertility of the resulting F_1 hybrid. Further multi-environment testing may be required to ensure that acceptable restoration occurs under sub-optimal climatic conditions. When breeding new CMS (pollen-sterile, female) lines, the difficulties are even greater as any pollen production in female lines is unacceptable commercially.

Previously Klein et al. (2001) mapped and subsequently cloned (Klein et al. 2005) a pollen fertility restoration gene designated *Rf1* to sorghum chromosome SBI-08. Sequence analysis of the *Rf1* loci identified it as a pentatricopeptide repeat (PPR) protein that cosegregated with the fertility restoration phenotype (Klein et al. 2005).

PPR proteins form part of a large multi-gene family that are thought to be RNA-binding proteins involved in post-transcriptional processes (RNA processing and translation) in mitochondria and chloroplasts (Lurin et al. 2004). To date, all the cloned *Rf* genes, except *Rf*2 of maize, are members of the PPR family (for review, see Saha et al. 2007). While the markers linked to *Rf*1 accurately predicted fertility restoration in progeny produced in crosses utilizing the R line (RTx432) used to map *Rf*1, the markers did not accurately classify all known restorers and non-restorers in elite germplasm (Klein, unpublished data). These observations have been attributed to the existence of at least two major non-allelic fertility restorer genes in commercial sorghum germplasm (Miller and Pickett 1964).

In this study we report on the fine mapping of a second major fertility restoration gene (*Rf*2) in sorghum and identify molecular markers that, when used in conjunction with *Rf1*-linked markers, could be used for marker-assisted breeding of new, commercially acceptable parental lines. We also present evidence of the potential identity of the *Rf*2 gene and discuss the genetic architecture of pollen fertility restoration in sorghum.

Materials and methods

Genetic stocks

Three populations were used in this study: two recombinant inbred line (RIL) populations were used for initial mapping of the gene, and a third BC_1F_2 population was used for subsequent fine mapping. Population 1 consisted of 285 F_5 RILs derived by single seed descent (SSD) from a bi-parental cross produced by crossing two inbred lines, R931945-2-2 and IS8525. R931945-2-2 is a restorer line derived from a complex cross developed by the Queensland Primary Industries and Fisheries (QPIF) sorghum breeding program while IS8525 is a non-restoring germplasm line of kafir origin.

Population 2 consisted of 233 F_4 RILs derived by SSD from a single F_1 plant that was produced by crossing two inbred lines, B923296 and SC170-6-8. B923296 is an elite B line from the QPIF sorghum breeding program while SC170-6-8 is a partially converted restorer line (Stephens et al. 1967) derived from the caudatum line IS12611. SC170-6-8 derivatives have been widely used in sorghum breeding programs worldwide (Doggett 1988). Crosses between SC170-6-8 and B lines in the QPIF sorghum breeding program have shown that SC170-6-8 contains a single restoration gene and does not appear to contain the partial fertility genes present in many other restorer lines (Jordan, unpublished data).

F_1 hybrids were produced by crossing each individual RIL from population 1 onto A1*F_B923171, an elite CMS line from the QPIF sorghum breeding program. In the case of population 2, F_1 hybrids were produced by crossing each

RIL with two elite CMS lines from the QPIF breeding program (A1*F_QL33 and A1*F_B923296). A1*F_B923296 and A1*F_B923171 are closely related lines derived from the same F_2 population that show similar agronomic and fertility responses. For population 2, seed was available for only 205 RIL hybrids using A1*F_QL33 as the female. These represented a subset of the 233 RILs used to produce hybrids with A1*F_B923296.

Population 3 was produced by crossing a fertile F_1 hybrid derived from a cross between one of the RILs from population 2 to the CMS line A1*F_B923296. A fertile plant from the resulting progeny was selfed to produce an F_2 population of 155 individuals.

Phenotypic data collection

The F_1 hybrids from population 1 were planted in 4-m, single-row plots in a randomised complete block design with two replicates. The trial was planted in February 2005 at Hermitage Research Station in Queensland, Australia. Two fertility-related traits were scored: (1) pollen production and (2) percent seed set on self-pollinated heads. Pollen production was determined by visual assessment of an average of five flowering heads in each row. On the basis of these quantitative scores genotypes were classified into three qualitative groups (fertile, sterile and segregating). Data from the two replicates were compared and genotypes with variable scores between reps were rechecked. Just prior to flowering, two heads in each row were covered with paper bags to ensure self-pollination. Approximately 40 days after flowering, the bags were removed and the individual panicles visually scored for percentage seed set. Temperatures in this experiment were low enough to influence sorghum fertility with minimum temperatures during the flowering period averaging 13°C and ranging between 6 and 17.5°C.

The F_1 hybrids from population 2 were planted in 4-m two-row plots in a randomised trial with two replicates planted in November 2004 at Hermitage Research Station. This trial consisted of 205 A1*F_QL33 hybrids and 233 A1*F_B923296 hybrids. The same two fertility-related traits were scored in this experiment with the exception that percent seed set on self-pollinated heads was only scored for two heads in a random sample of 100 plots. Again the genotypes were classified into qualitative groups based on the quantitative scores. Temperatures in this experiment were within the range normally experienced by commercial sorghum and unlikely to have had an impact on fertility.

The 155 F_2 individuals from population 3 were planted as spaced plants in a single row in October 2007 at the Hermitage Research Station. Each plant was bagged to prevent cross fertilization and the seed set was scored. Again the genotypes were classified into qualitative groups based on the quantitative scores. Temperatures in this experiment were within the range normally experienced by commercial sorghum and unlikely to have had an impact on fertility.

DNA extraction and marker analysis

Total genomic DNA was extracted from the three populations as described by DArT P/L (http://www.diversityarrays.com). Diversity Arrays Technology (DArT^{TM}) has recently been developed for sorghum and has been demonstrated to provide high-quality markers for the construction of medium-density genetic linkage maps (Mace et al. 2008). The technology was used in combination with other marker systems (AFLP^{TM} and SSR) to produce a map of population 1 (Mace et al. 2008).

Population 2 was screened with four SSRs, developed by Bhattramakki et al. (2000), located on SBI-02, *Xtxp50*, *Xtxp211*, *Xtxp297* and *Xtxp304*, and five SSRs developed for this study *(Xtxp616*, *Xtxp654*, Sb03QGM127, Sb03QGM128 and Sb03QGM130); (marker primer sequences detailed in Supplementary Table 1). The SSR markers developed by Bhattramakki et al. (2000) were previously screened in population 1 (Parh et al. 2008); the additional SSR markers developed by both TAMU-USDA and QPIF were subsequently added to increase marker density in the target region on population 1. PCR conditions were as previously detailed (Parh et al. 2008) and were visualised via capillary electrophoresis on a CEQ8800 (Beckman Coulter); between 0.25–1 μL of PCR product was mixed with approximately 25 μL of sample loading solution and 0.3 μL of CEQ 400 size standard (Beckman Coulter) prior to a 35-min run at 6 kV. Fragment sizes were subsequently determined using the Fragment Analysis software (Beckman Coulter).

Using the information obtained on the location of the trait locus provided by the first two experiments, an additional 11 markers spanning the region were developed. These SSR (with *Xtxp* prefix) and INDEL (with *Xtxi* prefix) markers were designated *Xtxp608, Xtxp609, Xtxp610, Xtxp611, Xtxp612, Xtxp613, Xtxp614, Xtxp615, Xtxp655, Xtxi51* and *Xtxi52* (see Supplementary Table 1). Population 3 was screened with a total of 16 markers using the PCR protocols previously outlined (Parh et al. 2008).

QTL analysis and genetic mapping

Linkage analysis was carried out using the multipoint mapping software (http://www.multipoint.com). The *Rf*2 marker was integrated into the maps generated for populations 1 and 2 as a phenotypic single gene marker. The RIL 'Selfing population' setting was selected, and the Kosambi mapping function and a maximum threshold Rf_s

value of 0.25 were used. Multipoint linkage analysis of loci within the cluster was then performed and marker order was further verified through re-sampling for quality control via jack-knifing (Mester et al. 2003). Only markers that could be ordered with a jack-knife value of 90% or greater were included as 'framework' markers, with any remaining markers causing unstable neighbourhoods being initially excluded from the map, including redundant markers mapping to the same location. Following a repeated multipoint linkage analysis with the reduced set of markers to achieve a stabilised neighbourhood, the previously excluded markers were attached by assigning them to the best intervals on the framework map.

The phenotypic data from the experiments involving populations 1 and 2 were used to classify individuals as fertile, sterile or heterozygous using a visual score of pollen production. Seed set data was used to confirm the classifications.

Composite interval mapping was conducted on pollen production and seed set using the computer program QTL Cartographer for Windows v2.0 (Wang et al. 2004). Empirical experiment-wise threshold values for significance ($P = 0.05$) were estimated from 1,000 permutations of the data for each trait (Churchill and Doerge 1994) as implemented by the program QTL Cartographer. The magnitude of the effect of significant QTL was estimated as the proportion of phenotypic variance explained (R^2). Single marker analysis of variance was conducted on the closely linked markers using QTL Cartographer and the phenotypic variance explained (R^2) was calculated.

Results

Segregation analysis

In both populations clear discontinuous distributions of fertility and seed set traits were observed. Table 1 shows the classification of the test-cross F_1 hybrids produced from the mapping experiments based on the RIL populations. In Population 1, 135 test-cross hybrids were classified as sterile, and of these lines 124 had no seed set while 11 test-cross hybrids had between 1 and 5% seed set. In the case of the 135 test-cross hybrids classified as fertile, 117 were rated as having 100 percent seed set while 18 test-cross hybrids were rated as having between 90 and 100% seed set. Fertility classifications for population 2 were consistent for test-cross hybrids produced from particular RILs on the two test-cross females (AQL33, A23296) examined. The seed set data for population 2 also strongly matched the fertility classification for the 100 random hybrids. Seed sets for the sterile A1*F_B923296 hybrids varied between 0 and 5%, while seed sets in the fertile A1*F_B923296 hybrids varied between 70 and 100%. Seed sets for the sterile A1*F_QL33 hybrids varied between 0 and 1%, while seed sets in the fertile A1*F_QL33 hybrids varied between 60 and 100%. A Chi-squared goodness of fit test for each of the three sets of test-cross hybrids conclusively supported the hypothesis of a single dominant gene.

Data from the QTL analysis of population 1 indicated the presence of a single QTL located between *Xtxp211* and sPb-8647 (LOD 39.6 and $R^2 = 84.5\%$). No other significant QTL were detected (LOD 3). The single marker analysis of population 2 was highly significant for the markers linked to the QTL in population 1 ($P < 00000.1$, $R^2 = 87\%$) and explained the majority of the phenotypic variance in the trait. However, the presence of lines which were classified as sterile which had low levels of seed set and lines classified as fertile lines with less than 100% seed set was suggestive of the presence of partial fertility genes in both RIL populations. If present it is likely that partial fertility genes were not detected as significant QTL due to their small effect size and the masking effect of the major gene. The large amount of variation explained by the single QTL provided high levels of confidence in our capacity to use qualitative classifications to map the trait as a single gene.

Genetic mapping

Regional linkage maps of chromosome SBI-02 for the *Rf2* RIL populations and a reference, consensus map (Mace et al. 2009) are presented in Fig. 1. The marker order spanning the targeted locus is consistent across two regional linkage maps and the high-density reference map

Table 1 Classification of test-cross hybrids from two RIL populations into fertility classes based on pollen production and seed set

Classification	Frequency population 1 Female A23171	%	Frequency population 2 Female AQL33	%	Frequency population 2 Female A23296	%
Male sterile	135	47.4	85	41	90	39
Heterozygous[a]	15	5.26	28	13	41	18
Fertile	135	47.4	92	45	102	43
Total	285		205		233	

[a] F_5 expectation of heterozygous individuals is 6.25% and the F_4 expectation of heterozygous individuals is 12.5%

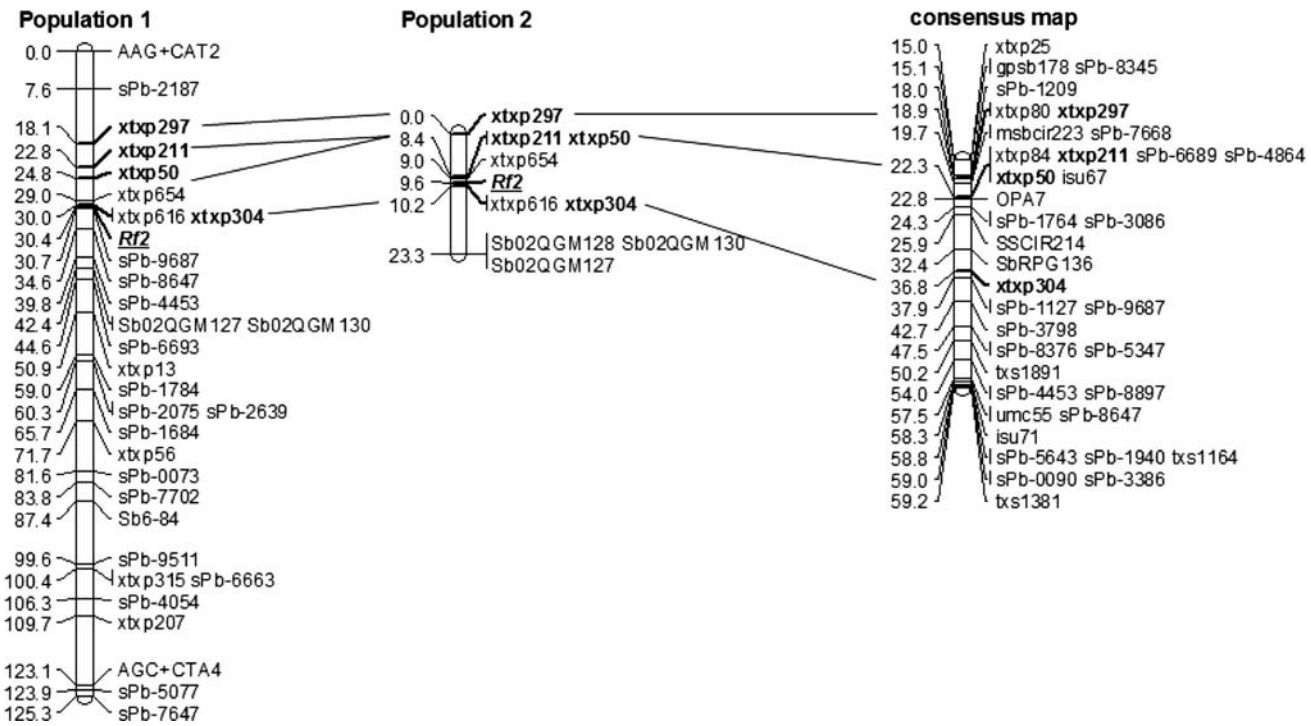

Fig. 1 Partial genetic linkage maps of sorghum chromosome SBI-02 for Populations *1*, *2* and a reference, *consensus* map (Mace et al. 2009) showing the location of a major fertility restoration gene. Markers in common across all 3 maps are in *bold*

of SBI-02 (see Supplemental Table 1 for details of the markers used). The results of the phenotypic segregation data supported a major gene model so the qualitative trait data was incorporated as a simple genetic marker in each population and designated *Rf*2. The *Rf2* gene maps to an almost identical genomic location in both RIL mapping populations suggesting a single locus (or closely linked genes) controlling the trait. In both populations, the *Rf2* gene was not included as a framework marker in the genetic linkage maps generated by Multipoint, as the marker order of the resulting maps, with the *Rf2* gene included, were found to be unreliable by the jack-knife re-sampling technique used, with a jack-knife value of less than 90%. This low jack-knife value indicated the likely presence of some individuals that were incorrectly classified for the *Rf2* gene within the populations. This being the case the inclusion of *Rf2* as a marker had the potential to perturb the order of markers in the framework map. To avoid this problem, the *Rf2* gene in both populations was attached to the respective framework maps by assigning it to the best intervals using the "Extend the linkage group" function of Multipoint. In population 1, the *Rf2* gene mapped between the co-located SSR markers *Xtxp616* and *Xtxp304* (SBI-02, ~5.668–5.703 Mbp) and DArT marker, sPb-9687. In population 2, the *Rf*2 gene mapped between *Xtxp654* and *Xtxp616* (SBI-02, ~5.432–5.668 Mbp). This discrepancy in position could be the result of real differences in map order in the two populations. However, the more likely explanation is that the variation in location is the result of classification errors in the fertility phenotype due to environmental effects thus changing the most likely position of the marker as indicated previously.

To further resolve the location of the *Rf2* locus, a third mapping population, consisting of 155 F_2 lines, was phenotyped and genotyped with additional markers spanning the locus. Of the 155 individuals genotyped, 17 contained a recombination event between *Xtxp297* and *Xtxp615* based on the three genotypic classes observed in this generation, however, 7 of these individuals contained a recombination event between heterozygous (Aa = fertile phenotype) and homozygous dominant loci (AA = fertile phenotype) in the targeted genomic region and were uninformative (without further progeny testing) due to the dominant nature of the *Rf*2 gene. Therefore, the ten remaining recombinants were used for fine mapping (Fig. 2). The genotypes of these individuals confirmed the location found for this gene in population 2, indicating that the gene was located between *Xtxp654* and *Xtxp616* rather than between *Xtxp616* and *Xtxp304* as predicted from the mapping in population 1. This region spans 236,219 bp and contains 31 predicted ORFs (Fig. 2; Table 2). This genomic region therefore averages one gene for every

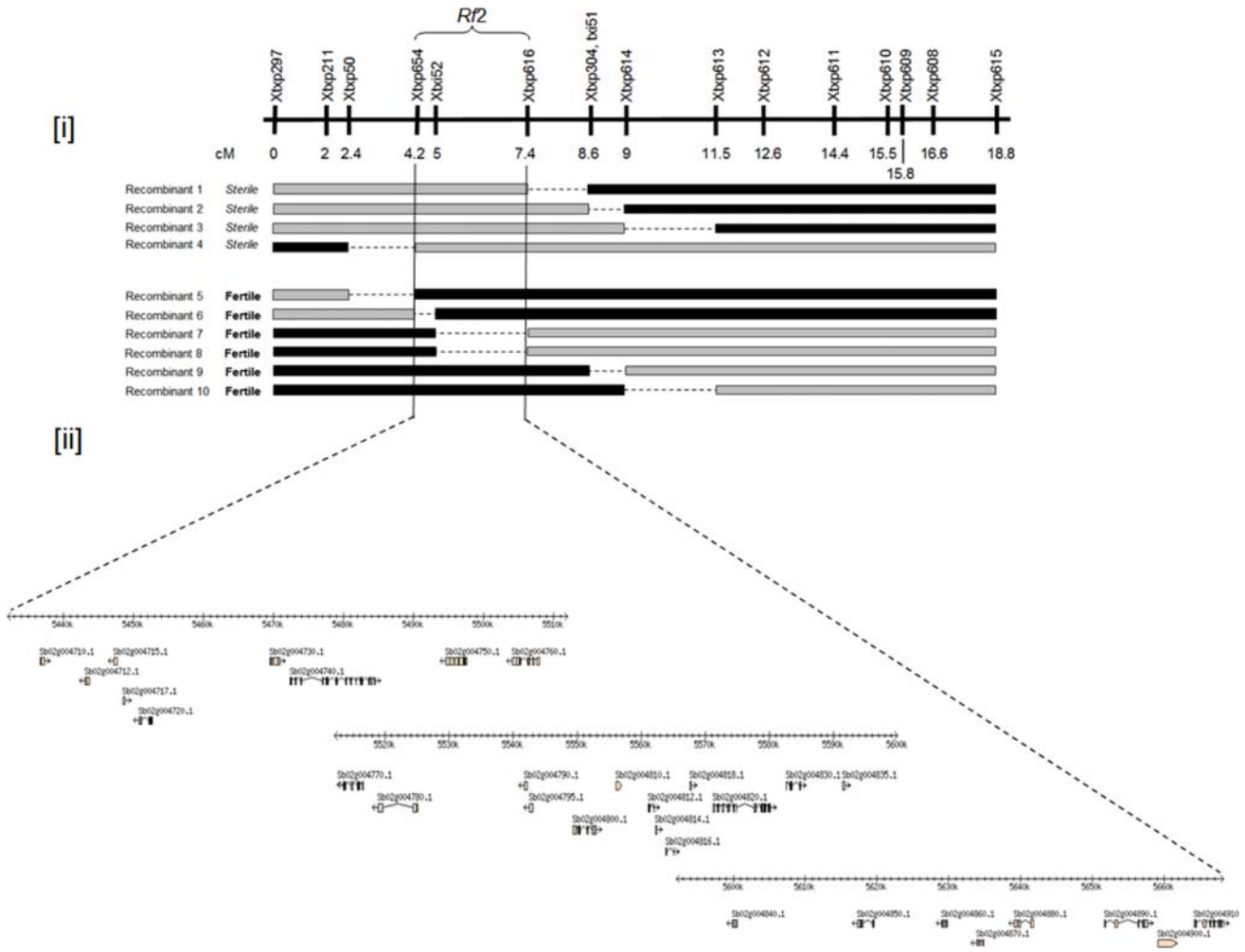

Fig. 2 A fine-scale, high-resolution genome map of the *Rf*2 locus on SBI-02. **i** Graphical genotypes of the ten recombinant individuals from population 3, based on fine mapping with 16 markers. Homozygous recessive alleles (aa) from parent 1 (source of the sterile allele) are *colour coded grey* and alleles heterozygous or homozygous dominant (Aa or AA) are *colour coded black*. *Dashed lines* indicate a recombination event between the markers. Phenotypic scores (sterile/fertile) for each line are detailed. The location of the *Rf*2 gene is indicated. **ii** Expanded view of the high-resolution physical map showing the positions of candidate genes, as identified on http://www.phytozome.net

~7.6 and 80 kb per cM. Of the 31 genes in the region, Sb02g004810 was annotated as a member of the pentatricopeptide repeat gene family, which showed high homology to the *Rf1* gene in rice (LOC_Os10g35240.2; e-value of e-119). Further fine-mapping efforts are presently underway to refine this locus. Nevertheless, the markers delimiting the sorghum *Rf2* locus, in conjunction with markers spanning the sorghum *Rf1* locus (Klein et al. 2005), are well suited for marker-assisted selection for these two major fertility restoration genes.

Discussion

The restoration of male fertility in F_1 hybrids is essential to the existence of the hybrid seed industry in sorghum. Early classical genetic studies indicated that restoration in A1 cytoplasm is controlled by at least two major genes with modifiers (Maunder and Pickett 1959; Miller and Pickett 1964; Erichsen and Ross 1963). The interaction of this genetic architecture and the environment, particularly temperature around flowering (Downs and Marshall 1971; Brooking 1976, 1979), results in a continuum of fertility phenotypes. Despite its commercial importance and the wealth of recent mapping studies in sorghum, the genetic architecture of the trait is not well understood. Since the advent of molecular marker technology, only a single fertility restoration gene has been mapped in sorghum. This gene designated as *Rf1* was mapped to SBI-08 by Klein et al. (2001) and subsequently cloned (Klein et al. 2005).

The major restorer gene described and mapped in this study has the same function as *Rf*1 but maps to

Table 2 List of genes and their putative function in the 236 kb region between markers *Xtxp654* and *Xtxp616* on SBI-02

Gene/marker name	LG	Start position (bp)	End position (bp)	Gene length (bp)	Rice BLAST Hit[a]	Description	e-value
Xtxp654	2	5,432,190	5,432,681	491			
Sb02g004710.1	2	5,436,792	5,437,305	513	LOC_Os03g16760.1	Protein phosphatase 2C	9.7e-41
Sb02g004712.1	2	5,443,142	5,443,768	626	LOC_Os10g07310.1	Retrotransposon protein	3.3e-30
Sb02g004715.1	2	5,447,246	5,447,788	542	LOC_Os08g37040.1	Gibberellin receptor GID1L2	1.5e-39
Sb02g004717.1	2	5,448,545	5,448,944	399	LOC_Os02g56830.1	Transposon protein	5.3e-30
Sb02g004720.1	2	5,450,800	5,452,778	1,978	LOC_Os07g08660.1	40S ribosomal protein S15	8.9e-76
Sb02g004730.1	2	5,469,565	5,470,916	1,351	LOC_Os03g58470.1	Retrotransposon protein, putative	2.3e-52
Sb02g004740.1	2	5,472,453	5,484,548	12,095	LOC_Os07g08729.1	ATP-dependent DNA helicase 2 subunit 1	1.6e-289
Sb02g004750.1	2	5,494,644	5,497,606	2,962	LOC_Os07g08750.1	STE_PAK_Ste20_Slob_Wnk.1	1.4e-155
Sb02g004760.1	2	5,504,147	5,508,081	3,934	LOC_Os07g08760.1	START domain containing protein	1.8e-189
Sb02g004770.1	2	5,509,289	5,516,060	6,771	LOC_Os01g46580.1	Actin-related protein 2/3 complex subunit 2	4.1e-158
Sb02g004780.1	2	5,518,411	5,524,571	6,160	LOC_Os07g08770.1	YGGT family protein	6.1e-68
Sb02g004790.1	2	5,541,268	5,541,753	485	LOC_Os05g23550.1	Hypothetical protein	2.9e-11
Sb02g004795.1	2	5,542,053	5,542,559	506	LOC_Os05g24790.1	Transposon protein, putative	1.5e-27
Sb02g004800.1	2	5,548,907	5,552,510	3,603	LOC_Os07g08790.1	Targeting protein for Xklp2, putative	1.7e-134
Sb02g004810.1	2	5,555,548	5,556,498	950	LOC_Os10g35240.2	Rf1, mitochondrial precursor, putative	8.9e-76
Sb02g004812.1	2	5,560,672	5,561,545	873	LOC_Os07g08810.1	Expressed protein	6.0e-14
Sb02g004814.1	2	5,561,786	5,561,914	128	LOC_Os07g08790.1	Targeting protein for Xklp2	5.1e-07
Sb02g004816.1	2	5,563,381	5,564,647	1,266	LOC_Os07g08790.1	Targeting protein for Xklp2	2.1e-26
Sb02g004818.1	2	5,567,157	5,567,369	212	LOC_Os07g08790.1	Targeting protein for Xklp2	1.1e-16
Sb02g004820.1	2	5,570,757	5,579,770	9,013	LOC_Os07g08810.1	Expressed protein	4.2e-236
Sb02g004830.1	2	5,582,201	5,584,358	2,157	LOC_Os07g08820.1	C-Myc-binding protein,	9.2e-42
Sb02g004835.1	2	5,591,003	5,591,359	356	LOC_Os07g35050.1	OsFBX237—F-box domain containing protein	3.6e-20
Sb02g004840.1	2	5,599,852	5,600,552	700	LOC_Os03g52410.1	Expressed protein	1.4e-13
Sb02g004850.1	2	5,617,307	5,619,542	2,235	LOC_Os07g08840.1	Thioredoxin, putative	1.1e-40
Sb02g004860.1	2	5,628,983	5,629,764	781	LOC_Os07g08840.1	Thioredoxin, putative	5.9e-31
Sb02g004870.1	2	5,633,851	5,634,778	927	LOC_Os07g08840.1	Thioredoxin, putative	7.2e-26
Sb02g004880.1	2	5,639,123	5,641,822	2,699	LOC_Os06g20630.1	SAM dependent carboxyl methyltransferase, putative	1.3e-108
Sb02g004890.1	2	5,651,661	5,657,700	6,039	LOC_Os07g08880.1	ES43 protein, putative	6.1e-116
Sb02g004900.1	2	5,659,100	5,661,781	2,681	LOC_Os07g08890.1	Disease resistance protein RPM1, putative	0.0
Sb02g004910.1	2	5,664,206	5,671,661	7,455	LOC_Os07g08950.1	FAD-linked oxidoreductase protein, putative	5.2e-245
Sb02g004910.2	2	5,664,206	5,671,661	7,455	LOC_Os07g08950.1	FAD-linked oxidoreductase protein, putative	5.2e-245
Xtxp616	2	5,668,140	5,668,763	623			

[a] Sbi1.4 gene sequences were downloaded from http://www.phytozome.net and subjected to BLASTX analysis to the genes in the MSU Osa1 rice pseudomolecules at http://rice.plantbiology.msu.edu/blast.shtml. The top rice gene and corresponding e-values are shown

chromosome SBI-02. As a result we have designated it as *Rf2*. The *Rf2* locus was mapped in two different RIL populations and the locus was further refined in a F_2 population. In all three populations (and both CMS lines in the case of population 2) two clear non-overlapping restoration classes were observed that fitted a single gene model. The phenotypes were scored as simple genetic markers and the underlying gene was found to map to almost identical genomic locations on SBI-02 in the two RIL populations. Fine mapping in the F_2 population delimited the *Rf2* locus to a 236 kb region of the sorghum genome on SBI-02 that is predicted to encode 31 genes (Table 2). Within A1 cytoplasm of sorghum, there are no obvious differences in the restoration phenotypes for the *Rf1* and *Rf2* genes. Sequence analysis of the *Rf1* trait loci identified a pentatricopeptide repeat (PPR) protein that cosegregates with the

fertility restoration phenotype (Klein et al. 2005). To date, all the cloned *Rf* genes, except *Rf*2 of maize, are members of the PPR family. Lurin et al. (2004) identified hundreds of PPR genes in the rice and Arabidopsis genomes, while the human genome encoded for only six putative PPR proteins. This expansion of the PPR gene family in higher plants and the recent transposition of the *Rf1* gene into sorghum chromosome SBI-08 (Klein et al. 2005) suggest that several PPR genes with a similar function may reside across the sorghum genome. Based on this evidence it seems probable that *Rf2* also encodes a PPR gene.

The region identified in this study contains a single member of the PPR gene family (Sb02g004810), which shows high homology to the *Rf1* protein of rice (LOC_Os10g35240.2; e-value of e-119). While its similarity to *Rf1* from rice and its inclusion in the PPR family make Sb02g004810 a strong candidate, additional fine mapping along with further experimental observations are necessary to confirm the identity of the *Rf2* gene. Further efforts to delimit the locus by developing new markers were unsuccessful. This region, on a gross level, appears highly homologous between the parents of the mapping populations. Further refinement of the locus may require mapping in a different populations where this region is not nearly identical in state as it is in the current parental lines.

In the first population, the donor of the restoration allele was an elite line from the QPIF sorghum breeding program (R931945-2-2). The donor of the restoration allele in the second and third populations was SC170-6-8. This line is a partially converted line from the Sorghum Conversion Program (Stephens et al. 1967). SC170 and its derivatives have been widely used in sorghum breeding programs worldwide. Progeny of SC170 include the elite restorer line RTx430 which has been used extensively in commercial hybrids in the USA and elsewhere (Miller, personal communication).

Restoration genes are under strong differential selection in hybrid breeding programs with selection for complete restoration of fertility in A_1 cytoplasm by male parents and complete sterility in A_1 cytoplasm in female parents. In both RIL hybrid populations, variation in seed set was observed in both the fertile and sterile classes. This is consistent with the presence of modifier or partial fertility genes as indicated by the classical genetic studies contributing to full restoration (Miller and Pickett 1964). In both RIL populations the lowest seed sets observed in fertile individuals were still within the range observed in Australian commercial hybrids grown in the same (data not shown). The maximum seed sets in the lines classified as sterile were greater than those that would be acceptable in commercial CMS lines. This observation of partial fertility is typical of the situation encountered by breeders who make crosses between B lines and exotic germplasm or R lines. In some cases this partial fertility is only expressed under particular environmental conditions and may pass unnoticed for some generations. Fertility in female parents in seed production blocks can result in considerable commercial losses and commercial breeders are very cautious about making such crosses. As a result it is rare for crosses between B and R (or exotic) lines to be made. This constraint has markedly restricted genetic diversity observed of the A/B line germplasm pool as observed in a number of diversity studies (e.g. Menz et al. 2004; Ahnert et al. 1996; Mace et al. 2008). In contrast there is strong selection in the case of male parent lines for combinations of restorer genes that give complete restoration of A1 cytoplasm in the heterozygous state under a range of environmental conditions. Again failure to restore fertility can be commercially expensive and crosses between non-restorers and restorers aimed at developing new male parents are rarely made. In both types of parents, but particularly female parents, these constraints restrict access to genetic variation limiting potential genetic gain for quantitative traits. Another limitation imposed by strong phenotypic selection for major genes are the constraints on a sorghum breeder's capacity to access genetic variation for other traits that may be controlled by genes linked to these regions.

The identification of the major restorer genes *Rf1* and *Rf2* is an essential step toward developing a detailed understanding of the genetic and physiological architecture of the trait and identifying the genes for partial fertility. Knowledge of their location will permit more effective strategies for developing male parents by allowing breeders to identify the most effective restorer alleles and to ensure that they are present in their elite material if necessary by marker-assisted selection. Knowledge of the location of restoration genes relative to QTLs for other traits can be used to improve the efficiency of selection for particular combinations. In both male and female parents, markers could be used to identify recombinants between these genes where favourable genes are in repulsion phase or to maintain coupling phase linkage where this is desired.

The location and putative sequence of *Rf2* will be helpful in unravelling the complex genetic and cellular control of pollen fertility restoration providing some of the knowledge and molecular tools to manipulate this commercially important trait.

The next critical step in understanding this trait is to attempt to identify markers for the partial fertility genes. If identified, these markers could be used in combination with markers linked to *Rf1* and *Rf2* to provide breeders with the molecular screening tools needed to more effectively broaden the genetic base of female parents B lines by permitting diverse germplasm to be efficiently used to develop acceptable B lines.

Acknowledgments We thank the Australian Grains Research and Development Cooperation (GRDC; http://www.grdc.com.au), USDA-ARS and Queensland Primary Industries and Fisheries providing financial support for this research.

References

Ahnert D, Lee M, Austin DF, Livini C, Woodman WL, Openshaw SJ, Smith JSC, Porter K, Dalton G (1996) Genetic diversity among elite sorghum inbred lines assessed with DNA markers and pedigree information. Crop Sci 36:1385–1392

Bhattramakki D, Dong J, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

Brooking IR (1976) Male sterility in *Sorghum bicolor* (L.) Moench induced by low night temperature. I. Timing of the stage of sensitivity. Aust J Plant Physiol 3:589–596

Brooking IR (1979) Male sterility in *Sorghum bicolor* (L.) Moench induced by low night temperature. II. Genotypic differences in sensitivity. Aust J Plant Physiol 6:143–147

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Doggett H (1988) Sorghum, 2nd edn. Longman Scientific and Technical co-published with John Wiley and Sons, New York

Downs RW, Marshall DR (1971) Low temperature induced male sterility in grain sorghum. Aust J Agric Res Animal Husb 11:352–356

Erichsen AW, Ross JG (1963) Irregularities at microsporogenesis in colchicine-induced male-sterile mutants in *Sorghum vulgare* Pers. Crop Sci 3:481–483

Klein RR, Klein PE, Chhabra AK, Dong J, Pammi S, Childs KL, Mullet JE, Rooney WL, Schertz KF (2001) Molecular mapping of the *Rf1* gene for pollen fertility restoration in sorghum (*Sorghum bicolor* L.). Theor Appl Genet 102:1206–1212

Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF (2005) Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theor Appl Genet 111:994–1012

Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16:2089–2103

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 J-F, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Sakrewski 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

Maunder B, Pickett RC (1959) The genetic inheritance of cytoplasmic-genetic male sterility in grain sorghum. Agron J 51:47–49

Menz MA, Klein RR, Unruh NC, Rooney WL, Klein PE, Mullet JE (2004) Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci 44:1236–1244

Mester D, Ronin Y, Minkov D, Nevo E, Korol A (2003) Constructing large-scale genetic maps using an evolutionary strategy algorithm. Genetics 165:2269–2282

Miller DA, Pickett RC (1964) Inheritance of partial male-fertility in *Sorghum vulgare* Pers. Crop Sci 4:1–4

Parh D, Jordan DR, Aitken E, Mace ES, Junai P, McIntyre C, Godwin I (2008) QTL analysis of ergot resistance in sorghum. Theor Appl Genet 117:369–382

Saha D, Prasad AM, Srinivasan (2007) Pentatricopeptide repeat proteins and their emerging role in plants. Plant Physiol Biochem 45:521–534

Schertz KF, Sotomayor-Rios A, Torres-Cardona S (1989) Cytoplasmic-nuclear male sterility: opportunities in breeding and genetics. Proc Grain Sorghum Res and Utility Conf 16:175–186

Stephens JC, Holland RF (1954) Cytoplasmic male sterility for hybrid sorghum seed production. Agron J 46:20–23

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghum to early combine types. Crop Sci 7:396

Wang S, Basten CJ, Zeng Z-B (2004) Windows QTL Cartographer 2.0. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)