Theor Appl Genet (2010) 120:943–953
DOI 10.1007/s00122-009-1223-y

ORIGINAL PAPER

A new SNP haplotype associated with blue disease resistance gene in cotton (*Gossypium hirsutum* L.)

David D. Fang · Jinhua Xiao · Paulo C. Canci · Roy G. Cantrell

Received: 27 July 2009 / Accepted: 12 November 2009 / Published online: 4 December 2009

Abstract Resistance to cotton blue disease (CBD) was evaluated in 364 $F_{2:3}$ families of three populations derived from resistant variety 'Delta Opal'. The CBD resistance in 'Delta Opal' was controlled by one single dominant gene designated *Cbd*. Two simple sequence repeat (SSR) markers were identified as linked to *Cbd* by bulked segregant analysis. *Cbd* resides at the telomere region of chromosome 10. SSR marker DC20027 was 0.75 cM away from *Cbd*. DC20027 marker fragments amplified from 3 diploid species and 13 cotton varieties whose CBD resistance was known were cloned and sequenced. One single nucleotide polymorphism (SNP) was identified at the 136th position by sequence alignment analysis. Screening SNP markers previously mapped on chromosome 10 identified an additional 3 SNP markers that were associated with *Cbd*. A strong association between a haplotype based on four SNP markers and *Cbd* was developed. This demonstrates one of the first examples in cotton where SNP markers were used to effectively tag a trait enabling marker-assisted selection for high levels of CBD resistance in breeding programs.

Communicated by I. Mackay.

D. D. Fang
Cotton Fiber Bioscience Research Unit, USDA-ARS-SRRC, 1100 Robert E. Lee Blvd., New Orleans, LA 70124, USA

J. Xiao · R. G. Cantrell (✉)
Monsanto Company, 800 North Lindbergh Blvd., St. Louis, MO 63167, USA
e-mail: roy.cantrell@monsanto.com

P. C. Canci
Monsanto Company, Av. Alexandre Ribeiro Guimaraes #620, Uberlandia, Minas Gerais 384085, Brazil

Introduction

Cotton (*Gossypium* spp.) is the most important fiber crop in the world. Diseases constitute one of the main challenges for sustainability of the cotton crop, especially in tropical and sub-tropical regions. One of the diseases with great economic importance is cotton blue disease (CBD). CBD was first described in the Central African Republic in 1949 and since then has been reported in regions of Africa, Asia, and the Americas (Cauquil and Vaissayre 1971; Cauquil 1977; Brown 2001; Correa et al. 2005; Junior et al. 2008). The causal agent was elusive for many years, but has recently been identified as a virus that belongs to the genus *Polerovirus* of the family *Luteoviridae* (Correa et al. 2005). CBD is transmitted by cotton aphids (*Aphis gossypii* Glover) in a persistent circulative manner. Symptoms include leaf rolling, vein yellowing, a moderate to severe stunting due to shortening of internodes, and dramatic dark green to bluish color of leaves from which the name "blue disease" originated (Brown 2001). In Brazil and other South American countries, CBD is a very serious problem for cotton production. This disease is capable of reducing productivity of susceptible varieties by up to 80% if cotton aphids are not properly controlled during the early growing season (Silva et al. 2008). Losses of up to 1,500 kg ha^{-1} of seed cotton due to CBD infection have been reported in Brazil (Freire 1998). Although insecticides can effectively control cotton aphids and consequently CBD, they are expensive, harmful to the environment, and do not provide season-long protection. The development and use of resistant variety offers the best management tool to control CBD. Commercial production in Brazil depends heavily on having highly resistant varieties available.

CBD resistance is present in upland cotton (*Gossypium hirsutum* L.). Royo et al. (2003) screened 283 upland

cotton germplasm (varieties) for CBD resistance under natural field infestation in Argentina. They found that a great majority of varieties bred in the USA are susceptible, but resistance exists in the germplasm from African countries as well as in new genetic materials derived from African germplasm. Recently, Junior et al. (2008) studied the inheritance of resistance to CBD using two crosses: 'CD401' (resistant)/'FM966' (susceptible), and 'Delta Opal' (resistant)/'FM966'. They determined that the CBD resistance in 'CD401' and 'Delta Opal' is controlled by one single dominant gene although they were not sure whether the same gene is present in both varieties, or each variety has a different gene.

Molecular markers provide efficient and powerful tools for constructing genomic maps and tagging genes of interest for marker-assisted selection. The number of molecular markers increased dramatically with the advent of polymerase chain reaction (PCR). In cotton, several genes controlling disease resistance traits, including root-knot nematode [*Meloidogyne incognita* (Kofoid & White) Chitwood] (Shen et al. 2006; Wang et al. 2006; Ynturi et al. 2006), reniform nematode (*Rotylenchulus reniformis* Linford & Oliveira) (Romano et al. 2009), verticillium wilt (*Verticillium dahliae* Kleb.) (Bolek et al. 2005), bacterial blight [*Xanthomonas axonopodis pv. malvacearum* (Smith) Dye] (Rungis et al. 2002; Xiao et al. 2010), black root rot (*Thielaviopsis basicola*) (Niu et al. 2008), and cotton leaf curl virus (Aslam et al. 2000) have been tagged by molecular markers. Currently, the most widespread PCR-based markers in cotton are simple sequence repeats (SSRs) (Blenda et al. 2006; Zhang et al. 2008). Genomic technology is improving to the point where genotyping is transitioning to single nucleotide polymorphism (SNP) markers (Rafalski 2002). SNP detection is not limited to gel- or capillary-based fragment size separation and thus can be fully automated (Eathington et al. 2007). SNP discovery can arise via conversion of existing molecular markers such as SSRs, mining EST sequence databases, or *de novo* sequencing and detection. The upland tetraploid cotton genome ($2n = 4X = 52$) is large and complex with an estimated DNA content of approximately 2,400 Mbp (1C) or 2.55 pg (1C) (Hendrix and Stewart 2005) that complicates *de novo* SNP discovery. The discovery and application of SNPs in plants is increasing, with recent expansion of SNP collections in soybean [*Glycine max* (L.) Merr.] (Choi et al. 2007) and wheat (*Triticum* spp.) (Akhunov et al. 2009). SNPs to date in cotton have focused primarily on sequence polymorphism in coding regions, such as transcription factor genes GhMyb8 and GhMyb10 (Hsu et al. 2008), Mt-Shsp sequence (Shaheen et al. 2009), and FlF1 sequence (Ahmad et al. 2007). However, to date few examples exist of SNP haplotypes being exploited to tag disease-resistant loci in plants.

In this paper, we report the inheritance of resistance to CBD derived from 'Delta Opal', its chromosome location and linkage with SNP markers derived from an SSR and *de novo* SNP discovery.

Materials and methods

Plant materials

'Delta Opal' is a variety bred by Deltapine Australia Ltd. 'Delta Opal' is a common highly resistant source for CBD (Junior et al. 2008). Using 'Delta Opal' as female parent, three crosses, namely 'Delta Opal'/'DP388', 'Delta Opal'/'DP5305' and 'Delta Opal'/'SG747' were made in May 2000 in a greenhouse of Delta and Pine Land Company, Scott, Mississippi. Cotton varieties 'DP388', 'DP5305', and 'SG747' are highly susceptible to CBD. Each F_2 population derived from a single F_1 plant. These populations consisted of 253, 50, and 61 F_2 plants, respectively (Table 1). Populations were advanced to F_3 from each F_2 plant without selection. $F_{2:3}$ seeds were sent to Brazil in December 2001 for CBD evaluation.

CBD screening

CBD screening was conducted in Uberlandia, Brazil. Cotton aphids (*Aphis gossypii* Glover) were collected from field-grown plants with heavy CBD infection, and

Table 1 Populations used for genetic analysis of cotton blue disease resistance

Population	Parentage	Parental genotypes at the *Cbd* locus	No. observed[a]			Expected ratio	χ^2
			≤1.5 (1.17)	1.5–3.5 (2.16)	≥3.5 (4.67)		
1	Delta Opal/DP388	RR × rr	61	123	69	1:2:1	0.70 ns
2	Delta Opal/DP5305	RR × rr	10	31	9	1:2:1	2.92 ns
3	Delta Opal/SG747	RR × rr	13	27	21	1:2:1	2.90 ns
Total			84	181	99	1:2:1	1.25 ns

[a] Disease severity index (DSI) based on $F_{2:3}$ family. Mean DSI in parenthesis

ns Not significant at $P = 0.05$ level

maintained on seedlings of susceptible cotton material in laboratory conditions [28°C, and 50% relative humidity (RH)]. New seedlings were provided as aphid food source when CBD symptoms became visible on older plants so that aphid populations could be maintained at a high level. Trays with seedlings were placed on top of a table surrounded by an anti-aphid net.

Three hundred sixty-four $F_{2:3}$ families along with parental varieties and F_1 plants were evaluated for resistance or susceptibility to CBD. Eighteen plants were used to represent each $F_{2:3}$ family, F_1, or parents. According to Sedcole (1977), scoring 17 or more individuals will identify at least one susceptible plant in the progeny of a heterozygote in 99% of the tests. A total of more than 6,700 seedlings were screened. Seeds were sown in trays with 70 cells each in a growth chamber with temperature of 31°C and RH approximately 50%. Lights were maintained for 14 h per day. A completely randomized design was used for each experiment. Parental varieties were included in each experiment as controls. The inoculation took place at 10 days after planting, i.e., about 5 days after the emergence. At this time, seedlings displayed a well-developed first true leaf. Aphids that acquired virus were manually placed on seedlings, and left on seedlings for 10 days for virus transmission. Then aphids were eliminated through endosulfan CE insecticide pulverization (2.8 g L^{-1} active ingredient). Three weeks later, CBD symptom was scored for each individual plant as following: 1, no symptom; 2, normal color and slightly deformed leaf; 3, dark color and visibly deformed leaves; 4, thin bluish-green color and highly deformed leaves; 5, visible yellow veins, highly deformed and fragile leaves when caught by hand. Disease severity index (DSI) of $F_{2:3}$ families was used to determine the genotype at the CBD resistance locus for each one of F_2 plants. The DSI was calculated as following:

DSI = (no. of plants with score 1×1 + no. of plants with score 2×2 + no. of plants with score 3×3 + no. of plants with score 4×4 + no. of plants with score 5×5)/total no. of plants.

Bulked segregant analysis

Young leaves were collected from each individual F_2 plants of three populations. Total DNA was extracted from either fresh or frozen leaves using 2.0% hexadecyltrimethyl ammonium bromide according to Paterson et al. (1993). DNA was purified using Omega EZNA® DNA isolation column (Omega Bio-Tek, Norcross, GA). To rapidly identify DNA markers associated with CBD resistance, Bulked segregant analysis (BSA) was deployed as described by Michelmore et al. (1991). Only DNAs from F_2 plants of the population 'Delta Opal'/'DP388' were used to make bulks. For the CBD-resistant bulk, DNAs of ten F_2 plants whose F_3 families had disease severity indices ≤ 1.2 were pooled at equal ratio and diluted to 10 ng/μL. The susceptible bulk consisted of pooled DNA from 10 F_2 plants whose F_3 families had disease severity indices ≥ 4.0. SSR Primers that generated polymorphic patterns between bulks were retested using another two DNA bulks, each composed of five F_2 progeny individuals exclusive of those in the first two bulks. Reproducible polymorphisms were further tested using the 30 individual DNA samples that were included in the bulks. The markers linked to CBD resistance gene were analyzed on 364 individual F_2 progeny of all populations.

SSR marker analysis

The DNAs of resistant and susceptible bulks, 'Delta Opal', 'DP388', and F_1 were analyzed with 4,247 SSR primer pairs. Primer sequences for the public markers (BNL, CIR, JESPR, CM, MGHES) can be obtained from Cotton Marker database (http://www.cottonmarker.org). The SSR primer sequences of Monsanto markers are listed in Xiao et al. (2009). Forward primers were fluorescent-labeled at 5′ end with 6-FAM (6-carboxyfluorescein), HEX (4, 7, 2′, 4′, 5, 7-hexachloro-carboxyfluorescein), or NED (7′, 8′-benzo-5-fluoro 2′, 4, 7,-trichloro-5-carboxyfluorescein). SSR primers were purchased from Sigma Genosys (Woodlands, Texas) or Applied Biosystems Inc. (Foster City, CA). Multiplex PCR was performed when conducting primer screening. Three pairs of primers with different dyes were multiplexed in each PCR reaction. After an SSR marker was putatively identified as linked to the trait, this marker was further analyzed using non-multiplex PCR. The 10-μL PCR reaction included 20 ng DNA, 2.5 μM each of the forward and reverse primers, 3.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 unit of DNA Taq polymerase (Promega Corporation, Madison, WI), and 1× reaction buffer without $MgCl_2$. Amplification conditions were 95°C for 3 min, followed by 34 cycles of 94°C for 45 s, 55°C for 45 s, and 72°C for 1 min, with a final step of 72°C for 10 min. Amplified PCR products were separated and measured on an automated capillary electrophoresis system ABI 3730 XL (Applied Biosystems Inc.). GeneScan-400 ROX® (Applied Biosystems Inc.) was used as an internal DNA size standard. The output was analyzed with GeneMapper 3.7 software (Applied Biosystems Inc.).

SNP marker discovery

To convert an SSR marker into a SNP, 13 cotton varieties with known reactions to CBD infection were used for cloning the target SSR marker fragments. The variety names, countries of origin, and marker genotypes are

listed in Table 4. In addition, three diploid species, i.e., *G. arboreum*, *G. herbaceum*, and *G. raimondii*, were also included. Genomic DNA was amplified using SSR primer DC20027 (forward 5′AATAAACCCTTCAGACAACA G3′, reverse 5′CTACCTAGTTTTGCATTATGT3′). PCR products were purified with Wizard® DNA Clean-up columns (Promega Corporation) before cloning. PCR products were directly cloned using TOPO® TA cloning kits from Invitrogen Company (Carlsbad, CA). Recombinant clones were screened by amplification of inserts in bacteria using primer DC20027 with 'Delta Opal' genomic DNA as control. The cloned marker fragments were sequenced in both directions in an automated ABI3730 DNA Analyzer using BigDye® terminator sequencing kits (Applied Biosystems Inc.). SNPs were identified by aligning all sequences using Clustalw2 (http://www.ebi.ac.uk/Tools/clustalw2) (Larkin et al. 2007). Once a SNP was identified, an end-point TaqMan® assay was developed to discriminate SNP alleles by properly designing primers and probes using design tools offered by Applied Biosystems Inc. (https://www2.appliedbiosystems.com/support/software/assaysbydesign). The 7-μL TaqMan® assay reaction contained 10 μM each of primers, 0.2 μM each of probes, 5 ng genomic DNA, and 1× TaqMan® universal PCR Master Mix (Applied Biosystems Inc.). PCR temperature profiles were 50°C 2 min, 95°C 10 min, followed by 40 cycles of 92°C 15 s, and 60°C 1 min. At the end of PCR, plates were scanned using ABI 7300 Real-Time PCR System, and SNP alleles plotted using software SDS 1.1 (Applied Biosystems Inc.).

The SSR marker DC20027 (GenBank Acc. No. MONCS1471) was previously mapped on cotton chromosome 10 based on linkages with public framework SSR markers (Xiao et al. 2009). In order to develop a SNP haplotype associated with the CBD resistance trait, SNP markers within 5 cM of SSR marker DC20027 on chromosome 10 in Monsanto's proprietary genetic map (unpublished) were screened between 'Delta Opal' and 'DP388'. Polymorphic SNP markers were tested among 253 F_2 progeny plants. An additional 3 SNP markers were identified as linked to the CBD resistance trait through this approach. The SNP attributes are presented in Table 5. All four SNP sequences found linked to the trait of interest were deposited in GenBank.

Linkage analysis

Segregation data for CBD resistance, SSR, and SNP markers from all segregating progeny were mapped using program JoinMap3.0 (Van Ooijen and Voorrips 2001) with LOD score ≥5.0. Chi-square tests were used to check segregation of markers and disease severity index against an expected 1:2:1 frequency.

Results

Inheritance of CBD resistance in 'Delta Opal'

During the course of this experiment, more than 50 plants of 'Delta Opal' or 'DP388' were evaluated for CBD resistance. For 'Delta Opal', great majority of plants (>90%) did not have any symptom (disease score = 1), and very few plants showing mild symptom (disease score = 2). Not a single 'Delta Opal' plant was ever scored as 3 or higher for CBD symptom. The overall disease severity index (DSI) of 'Delta Opal' was 1.1. On the contrary, for 'DP388', all plants developed severe disease symptom. Most plants had disease score of 5, and a few were scored as 4. No plants with disease score of 3 or lower were observed. The overall DSI of 'DP388' was 4.89. Similar to 'DP388', both 'DP5305' and 'SG747' had severe symptoms with DSI > 4.5. F_1 hybrids exhibited high resistance to CBD with DSI < 1.2. These results implied that the CBD resistance in 'Delta Opal' might be controlled by one single dominant gene.

If the CBD resistance in 'Delta Opal' is controlled by one single dominant gene, half of the F_2 plants would be heterozygotes, and their subsequent $F_{2.3}$ families would segregate for the CBD trait. Eighteen plants were used to represent each $F_{2.3}$ families when conducting CBD screening. We used DSI to reflect the segregation of CBD resistance in each $F_{2.3}$ family. The CBD segregation in 364 $F_{2.3}$ families of three populations is listed in Table 1. Table 2 listed the disease scores of individual plants from 2 parents and 12 $F_{2.3}$ families as examples to show the disease score distribution within a family. The DSI of these families ranged from 1.0 to 5.0. $F_{2.3}$ families with DSI ≤ 1.5 (mean 1.17) usually had less than 4 plants with mild symptoms, and no plants had disease score of 4 or higher. This group consisted of 84 families. The F_2 plants from which these $F_{2.3}$ families derived should be homozygous at the CBD resistance locus (Table 2). On the contrary, almost all plants had moderate-to-severe symptoms for the $F_{2.3}$ families with DSI ≥ 3.5 (mean 4.67). Ninety-nine families belonged to this category. Of 1782 plants tested, only 6 plants were scored as 1. These 6 plants might have escaped from CBD infection. The F_2 plants from which these $F_{2.3}$ families derived should be recessive at the CBD resistance locus. For the rest of 181 $F_{2.3}$ families (DSI range 1.5–3.5, and mean 2.16), both healthy and severely-infected plants were observed within each family, but healthy plants were more than infected ones. A great majority of these families had DSI lower than 2.5, and only 4 families had DSI greater than 3.0. When examining each family, it was found that the CBD resistance segregated in a 3:1 ration within a family if scores 2 and lower were considered as resistant (Table 2). The F_2 plants from which

Table 2 CBD scores of 12 $F_{2.3}$ families and their parental varieties

	Number of plants with disease score					Disease severity index	F_2 plant genotype at *Cbd* locus
	1	2	3	4	5		
Delta Opal	16	2	0	0	0	1.11	RR
DP388	0	0	0	2	16	4.89	rr
$F_{2.3}$ Family							
#047	11	3	0	2	2	1.94	Rr
#060	0	0	0	2	16	4.89	rr
#129	16	2	0	0	0	1.11	RR
#131	0	0	1	2	15	4.78	rr
#133	17	0	0	0	0	1.00	RR
#137	0	0	0	3	15	4.83	rr
#142	10	1	2	0	5	2.39	Rr
#144	0	0	0	0	18	5.00	rr
#154	12	0	1	0	5	2.22	Rr
#165	15	0	0	0	3	1.67	Rr
#215	8	6	0	0	4	2.22	Rr
#219	14	2	0	0	0	1.13	RR

these $F_{2.3}$ families derived should be heterozygous at the CBD resistance locus. Segregation of these 3 groups in all populations was consistent with 1:2:1 ratio (RR:Rr:rr), as expected if the CBD resistance in 'Delta Opal' was controlled by a single dominant gene. Our result is consistent with that from Junior et al. (2008). Although Junior et al. (2008) named this gene as *Rghv1* (*Resistance to Gossypium hirsutum Virus 1*), we suggest *Cbd* (Cotton blue disease) to better follow the genetic nomenclature rules established in cotton (Kohel 1973).

Identification of codominant SSR markers associated with *Cbd*

Of the 4,247 SSR markers screened, 265 (6.24%) were polymorphic between 'Delta Opal' and 'DP388'. However, only 3 markers (BNL1403, BNL3646, DC20027) showed polmorphism between two R and S bulks. When these 3 markers were tested in the second pair of bulks and 30 individuals comprising the bulks, only markers DC20027 and BNL3646 were reproducible and polymorphic. Subsequently, these two markers were analyzed on 364 F_2 progeny, F_1s, and parents. For the marker DC20027, three fragments, i.e. 182 bp, 200 bp and 202 bp, were observed (Table 3). The fragment 182 bp was present in all DNA samples. The fragments 200 bp and 202 bp were allelic, and linked to the susceptible and resistance alleles at *Cbd* locus, respectively. The genetic distance between DC20027 marker locus and *Cbd* is 0.75 cM (Fig. 2). Similarly, primer BNL3646 generated 3 fragments, i.e. 145 bp, 147 bp, and 155 bp. The fragments 145 bp and 147 bp were allelic, and linked to the susceptible and resistance alleles at *Cbd* locus, respectively at a distance of 1.65 cM. The fragment 155 bp was present in all DNA samples.

Table 3 DC20027 and BNL3646 marker fragments in parental varieties

	DC20027[a]		BNL3646[a]	
Delta Opal	182	202	147	155
DP388	182	200	145	155
DP5305	182	200	145	155
SG747	182	200	145	155

[a] DNA fragment size in bp

In separate projects, Fang (unpublished data) mapped more than 2,700 SSR marker loci in a *G. hirsutum*/*G. barbadense* population, and 379 SSR marker loci in 'Delta Opal'/'DP388' F_2 population. The loci DC20027_200 bp/202 bp and BNL3646_145 bp/147 bp were mapped at the telomere region of chromosome 10 in both maps. Thus, we conclude that the *Cbd* locus is located at the telomere region of chromosome 10.

Based on DC20027 marker genotypes, the F_2 plants were differentiated into three groups: 90 "AA" (202 bp only), 177 "AB" (200 bp & 202 bp), and 97 "BB" (200 bp only). This ratio fits 1:2:1 segregation. It is worth to mention that all 84 F_2 plants whose F_3 families had DSI $\leq$ 1.5 had marker genotype "AA", 97 of the 99 F_2 plants whose F_3 families had DSI $\geq$ 3.5 had "BB" genotype. The remaining two F_2 plants whose F_3 families had DSI $\geq$ 3.5 had "AA" and "AB" genotypes, respectively, and were considered as recombinants. Of the 181 F_2 plants whose F_3 families had DSI between 1.5 and 3.5, 176 had "AB" genotype, and the other 5 were "AA" type. These results clearly suggest that the genotypes of F_2 plants at *Cbd* locus can be determined using DSIs of $F_{2.3}$ families.

Development of SNP markers

Although SSR markers are PCR-based and codominant markers, they are not amenable to high throughput genotyping application in commercial breeding. In order to overcome this disadvantage, we converted the SSR marker DC20027 to a SNP marker. 'Delta Opal', 'DP388' and other 14 cotton genotypes (Table 4) were used to clone DC20027 marker fragments. The 182 bp fragment is present in *G. raimondii* (a diploid D_5 genome species) and all *G. hirsutum* varieties, but absent in *G. herbaceum* (a diploid A_1 genome) and *G. arboreum* (a diploid A_2 genome). This locus resides on Dt sub-genome chromosome 20, and is not associated with *Cbd*. As expected, both *G. arboreum* and *G. herbaceum* have 202 bp fragment.

Table 4 Cotton varieties used to clone DC20027 marker fragments associated with *Cbd*

#	Variety[a]	Country of origin	CBD resistance	DC20027 marker fragment	
				Size (bp)[b]	Linked to *Cbd* allele
1	*G. raimondii*	D genome	Unknown	182	NA
2	Delta Opal	Australia	R	182	NA
3	DP388	USA	S	182	NA
4	PM183	USA	S	198	r
5	DP388	USA	S	200	r
6	DP90	USA	S	200	r
7	DP5305	USA	S	200	r
8	IAC21	Brazil	S	200	r
9	SG747	USA	S	200	r
10	Delta Opal	Australia	R	202	R
11	Sicala 32	Australia	R	202	R
12	Reba 50	Central Africa Republic	R	202	R
13	Pora	Argentina	R	202	R
14	CD401	Brazil	R	202	R
15	Guazuncho	Argentina	R	202	R
16	Albar AF884	Zimbabwe	R	202	R
17	*G. arboreum*	A genome	R	202	R
18	*G. herbaceum*	A genome	R	202	R

[a] *G. arboreum* and *G. herbaceum* are diploid A genome specie, *G. raimondii* is a diploid D genome species. All others are Upland cotton (*G. hirsutum*) varieties

[b] 182 bp fragment is from Dt subgenome

Besides 200 bp and 202 bp fragments, a 198 bp fragment was observed in 'PM183'. Allele 198 bp is associated with the susceptible allele of *Cbd* locus. The CBD resistance gene in 'CD401' should be the same as that in 'Delta Opal' because both had 202 bp fragments with almost identical sequences (Fig. 1).

DC20027 amplicons had two microsatellite motifs (Fig. 1), i.e. TA (between 69th and 82nd positions) and GT (between 85th and 112th positions). The number of repeats resulted in the size differences among fragments. After aligning all sequences, we identified two SNPs. The first was an "A/C" SNP at the 39th position, and the second was an "A/T" SNP at the 136th position (Fig. 1). For the 39th position SNP, the 182 bp and 202 bp fragments had nucleotide A, while the 198 bp and 200 bp fragments had nucleotide C. Because the 182 bp fragment was present in all upland cotton samples, the 39th position SNP had no value in determining the genotypes at *Cbd* locus. However, for the 136th position SNP, all 202 bp fragments had nucleotide A, and fragments 200 bp, 198 bp and 182 bp had nucleotide T. A TaqMan® assay to discriminate the SNP alleles was successfully designed and the SNP designated NG0211495. The primer and probe sequences are listed in Table 5. The SNP was analyzed on 364 F_2 and completely matching DC20027 SSR marker genotypes.

Monsanto Company has developed a proprietary high density cotton genetic map that contains about 7,000 SSR and SNP loci. Because the SSR marker DC20027_202 bp and *Cbd* are mapped on chromosome 10, 15 SNP markers that are within 5 cM of SSR marker DC20027 on this chromosome were first screened between 'Delta Opal' and 'DP388'. Three of them were polymorphic between two parents, and were analyzed among the 253 F_2 progeny of "Delta Opal'/'DP388'. These 3 SNP markers, i.e., NG0203671, NG0204310, NG0203481, were found as tightly linked to *Cbd* (Fig. 2). The primer and probe sequences of all 4 SNP markers are listed in Table 5. The optimum situation for MAS is to select based on marker haplotype which should include SNPs flanking the gene of interest as in this case. The haplotype 'CC-CC-AA-TT' denotes resistance, while 'TT-TT-TT-CC' denotes susceptibility.

Discussion

'Delta Opal' is highly resistant but not immune to CBD as very mild disease symptom was occasionally observed on its plants. In the current research, we used DSIs of $F_{2.3}$ families to determine the *Cbd* genotypes of F_2 plants. The principle is similar to calculating the R:S ratio within a population. However, DSI can let us avoid determining resistant *vs* susceptible at the beginning, and still gives us a quantification results. Furthermore, these results are quite accurate. For example, for the 181 $F_{2.3}$ families that segregated for CBD resistance, the mean DSI was 2.16. In a

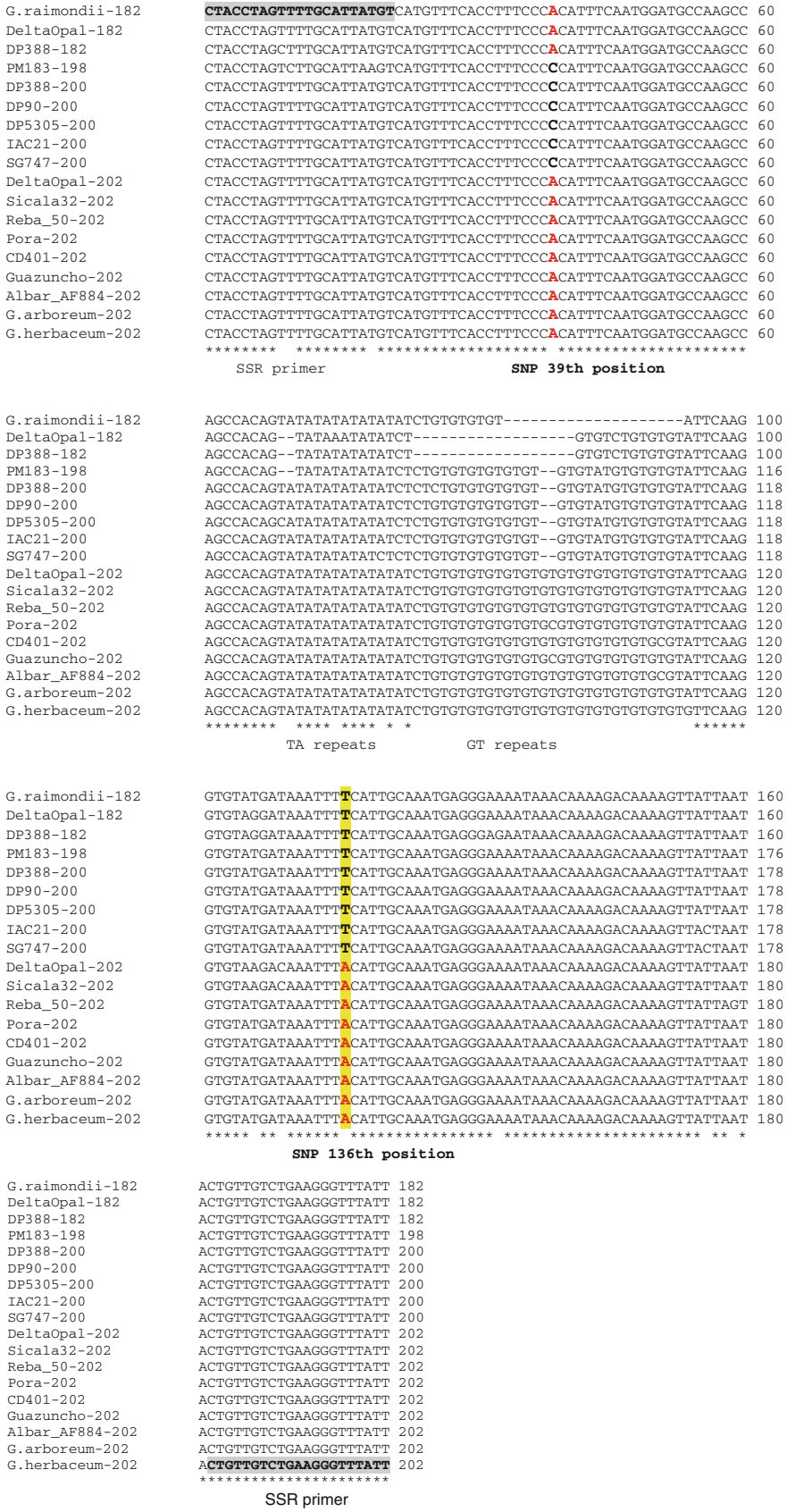

Fig. 1 Sequence alignment of DC20027 marker fragments from 13 cotton varieties and 3 diploid *Gossypium* species

Table 5 Attributes of SNP markers associated with *Cbd*

	SNP marker			
	NG0203671	NG0204310	NG0211495	NG0203481
GenBank Acc. No.	MONCS3001	MONCS3002	MONCS1474	MONCS3003
SNP position[a]	243	463	136	150
"R" allele	C	C	A	T
"S" allele	T	T	T	C
Forward primer[b]	TGTGACCTTAAGACAGCCTAAAACC	CCCCTGTTTACGAGGCTATCTATTCT	TGTGTGTGTGTGTGTGTGTGT	CTCCGCGTGCTTACTTTTTTTA AAGAAA
Reverse primer[b]	GCAAAAATCTACACGTGTTGAAGCT	GGTTTGGCCCAGTGACTAGAAG	CCCTTCAGACAACAGTATT AATAACTTTTGT	CATCAACCAAACATTGC AGCTCTAT
"R" allele probe[b]	TCCTACACAAAACTC	CTAGAATATATACATGAAATGAA	TGCAATGAAAATTT	ATTTTTGAACTGATAAATT
"S" allele probe[b]	AATCCTACATAAAACTC	CTAGAATATATACATAAAATGAA	ATTTGCAATGTAAATTT	TTTTGAGCTGATAAATT

[a] SNP position within consensus target sequence

[b] 5′–3′ orientation

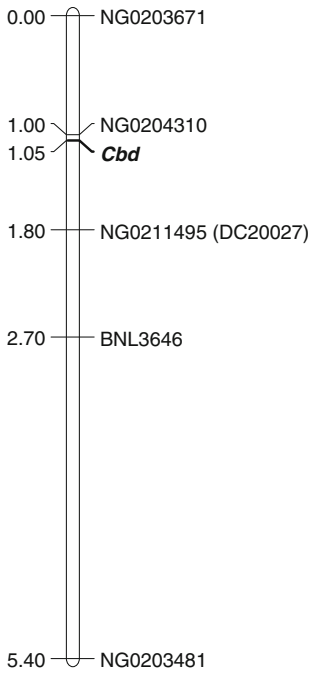

Fig. 2 Genetic map of *Cbd*, SSR, and SNP markers on chromosome 10

perfect situation, the DSI for a segregating $F_{2:3}$ family is 2.0 $[(13.5\times1 + 4.5\times5)/18 = 2.0]$. Our results clearly show the advantage of using DSI of an $F_{2:3}$ family to determine the F_2 plant genotype at the *Cbd* locus.

Upland cotton might have originated from a single event of hybridization between a diploid A genome and a D genome species about 1–2 million years ago (Wendel et al. 1992). Moreover, long history of domestication and selection by human has significantly reduced the genetic diversity within Upland cotton. This was reported by many researchers using different types of molecular markers (Wendel et al. 1992; Iqbal et al. 2001; Liu and Myers 2002; Rungis et al. 2005). For example, Rungis et al. (2005) detected only 5–7% polymorphism between any two upland cotton varieties after analyzing 216 genomic SSR markers. The diversity detected by EST SSR markers is even lower, and only about 2.46% (Lin et al. 2009). Our research showed that only 6.24% of SSR markers were polymorphic between 'Delta Opal' and 'DP388' after testing more than 4,200 SSR markers. Analysis of SNP markers around *Cbd* region revealed higher (20%) level of diversity between 'Delta Opal' and 'DP388'. However, this higher polymorphic rate was biased because *Cbd* region was previously known as polymorphic between the two parents based on the marker DC20027. The low level of genetic variation combined with larger genome size and allotetraploid nature has hindered the development of molecular markers associated with the traits of interest in cotton. However, because of this limited diversity, once

identified when using an intraspecific population, a marker is usually very close (<2 cM) to the trait of the interest as evidenced by the present research.

The BSA method developed by Michelmore et al. (1991) has been widely used to rapidly identify markers linked to the gene of interest. BSA is particularly useful for cases for which no near-isogenic lines exist like the research reported here. In their paper, Michelmore et al. (1991) calculated that the probability of an unlinked locus being polymorphic between bulks of 10 individuals is 2×10^{-6}. However, a high frequency of false positives was observed in the present experiment. Although 3 primers gave polymorphic patterns between the first bulk pair, only two of them revealed polymorphism in the second bulk pair and among progeny plants. This translates into 33.3% false positive rate. In a separate experiment identifying markers linked to the bacterial blight resistance gene, we also found a high rate of false positives (Xiao et al. 2010). A high rate of false positives in BSA has been observed by other researchers (Haley et al. 1993; Young and Kelly 1996). A repeatability of PCR-based markers might be one cause of false positives, especially when multiplex PCR was employed as the case in this research. In the current research, we did not observe artifacts related to multiplex PCR after the marker BNL1403 was analyzed using non-multiplex PCR. However, though not often, we did observe some artifacts related to multiplex PCR in our other research. For example, if BNL2662 and BNL3279 were multiplexed, a new fragment 161 bp will appear in some cotton genotypes (Fang, unpublished). PCR products may vary with different DNA polymerases, $MgCl_2$ concentrations, and many other factors. To reduce false positives and amplification work, it is helpful to construct two different bulk pairs with one of them being used as retest bulks, and to include one parent contributing the gene of interest as a control. Nevertheless, successful identification of markers closely linked to *Cbd* illustrates the power of the BSA.

Upland cotton is an allotetraploid with 26 pairs of chromosomes. It has two sub-genomes, At (chromosomes 1–13) and Dt (chromosomes 14–26). Due to sequence homology between these two sub-genomes, many SSR primers revealed duplicate loci present in both sub-genomes (Guo et al. 2007; Nguyen et al. 2004). In the present research, SSR primers DC20027 and BNL3646 revealed two loci each. The loci DC20027_200 bp/202 bp and BNL3646_145 bp/147 bp were mapped at the telomere region of chromosome 10, and *Cbd* locus is residing on this chromosome as well. Two diploid A genome species *G. arboreum* and *G. herbaceum* have these marker loci and *Cbd* gene (Table 4 and Fig. 1). As expected, the diploid D genome species *G. raimondii* does not have these loci. Recently, we analyzed the marker DC20027 in *G. thurberi* (D1), *G. armourianum* (D2), *G. aridum* (D4), and *G. trilobum* (D8). None of these D species has the target locus. Chromosome 10 is homoeologous to chromosome 20 (Guo et al. 2007). Because the marker loci DC20027_182 bp and BNL3646_155 bp are homozygous in the parental varieties used in this study and in our other mapping populations, we were not able to map them. However, Guo et al. (2007) mapped the locus BNL3646_155 bp to the telomere region of chromosome 20. Thus, it is safe to suggest that locus DC20027_182 bp may also reside on chromosome 20. The marker loci on chromosome 20 are not associated with *Cbd*.

Breeders had been trying to breed CBD resistant varieties since it was discovered in 1949. Little resistance was found in upland cotton, however, *G. arboreum* showed strong resistance to CBD and other viral diseases such as cotton leaf curl virus (CLCuV) (Nateshan et al. 1996). Interestingly, CLCuV was also first discovered in Africa in 1912 (Brown 2001). In order to introgress pest resistance traits into upland cotton from *G. arboreum*, trispecies hybrid cotton lines that derived from crosses of *G. hirsutum*, *G. arboreum* and *G. raimondii* (HAR) was developed by P. Kammacher in the Ivory Coast in 1970s (Innes 1983). Due to its resistance to viral and bacterial diseases, these HAR hybrid lines had been widely used in breeding programs in African countries. The CBD resistance gene present in many African cotton varieties might come from HAR hybrid lines with *G. arboreum* as the primary resistance source. In 1980s, the African germplasm was introduced to South American countries especially Brazil and Argentina to combat CBD and bacterial blight (Royo et al. 2003). Likewise, breeders in Pakistan and India introduced the African germplasm to fight against CLCuV. Because there have been little viral diseases in the USA, the African germplasm was not extensively used in breeding programs. We screened hundreds of cotton germplasm (varieties) collected from 25 countries with DC20027 SSR and SNP marker NG0211495 (data not shown). Almost all germplasm from African countries were predicted to be RR or Rr at *Cbd* locus based on the markers. A great majority of germplasm that carried resistant allele at *Cbd* locus were from Africa, South America, or Southern Asia. Almost all varieties from North America, Europe, and China are susceptible to CBD based on the SSR marker DC20027. Most varieties in Australia are predicted to be susceptible to CBD.

Due to difficulties to efficiently distinguish between genome-specific polymorphism and locus-specific polymorphism, SNP discovery in cotton is difficult and lags behind other row crops such as soybean and maize. Less than 300 SNP markers have been reported in the public domain for cotton. To the best of our knowledge, no major agronomic or disease trait in cotton has been tagged directly using SNP markers. This trend is likely to continue

until a large number of SNP markers are available to the cotton community. In the present research, we developed a strategy of first localizing the *Cbd* gene using SSR markers, then enriching the target region with SNP markers based on *de novo* mapping, and eventually developing a robust and reliable SNP haplotype associated with *Cbd*. Marker-assisted selection based on a haplotype has obvious advantage over using a single marker. It will dramatically improve selection accuracy because a haplotype consists of at least two marker loci. If the SNP markers flank the target gene as the case in this study, it almost can achieve 100% accuracy.

Acknowledgment We greatly thank two anonymous reviewers who made excellent suggestions for revising the manuscript.

References

Ahmad S, Zhang T, Noor-Ul-Islam, Shaheen T, Mehboob-Ur-Rahman (2007) Identifying genetic variation in *Gossypium* based on single nucleotide polymorphism. Pak J Bot 39:1245–1250

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet. doi:10.1007/s00122-009-1059-5

Aslam M, Jiang C, Wright R, Paterson AH (2000) Identification of molecular markers linked to leaf curl virus disease resistance in cotton. J Sci I R Iran 11:277–280

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape J-M, Yu J, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D (2006) CMD: a Cotton Microsatellite Database resource for *Gossypium* genomics. BMC Genomics 7:132

Bolek Y, El-Zik KM, Pepper AE, Bell AA, Magill CW, Thaxton PM, Reddy OUK (2005) Mapping of verticillium wilt resistance genes in cotton. Plant Sci 168:1581–1590

Brown JK (2001) Cotton blue disease. In: Kirkpatrick TL, Rothrock CS (eds) Compedium of cotton diseases, 2nd edn. APS Press, St. Paul, pp 50–51

Cauquil J (1977) Etudes sur une maladie d'origin virale du cotonnier: La maladie bleue. Coton Fibres Trop 32:259–278

Cauquil J, Vaissayre M (1971) La "maladie bleue" du cotonnier en Afrique: transmission de cotonnier à cotonnier par *Aphis Gossypii* Glover. Coton Fibres Trop 6:463–466

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685–696

Correa RL, Silva TF, Simoes-Araujo JL, Barroso PAV, Vidal MS, Vaslin MFS (2005) Molecular characterization of a virus from the family Luteoviridae associated with cotton blue disease. Arch Virol 150:1357–1367

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S154–S163

Freire EC (1998) Doenca azul tem solucao. Cultivar, Pelotas 1:64–65

Guo WZ, Cai CP, Wang CB, Han ZG, Song XL, Wang K, Niu XW, Wang C, Lu KY, Shi B, Zhang TZ (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. Genetics 176:527–541

Haley SD, Miklas PN, Stavely JR, Byrum J, Kelly JD (1993) Identification of RAPD markers linked to a major rust resistance gene block in common bean. Theor Appl Genet 86:505–512

Hendrix B, Stewart JM (2005) Estimation of the nuclear DNA content of *Gossypium* species. Ann Bot (London) 95:789–797

Hsu CY, An C, Saha S, Ma DP, Jenkins JN, Scheffler B, Stelly DM (2008) Molecular and SNP characterization of two genome specific transcription factor genes GhMyb8 and GhMyb10 in cotton species. Euphytica 159:259–273

Innes NL (1983) Bacterial blight of cotton. Biol Rev 58:157–176

Iqbal MJ, Reddy OUK, El-Zik KM, Pepper AE (2001) A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. Theor Appl Genet 103:547–554

Junior OP, Schuster I, Pinto RB, Pires E, Belot JL, Silvie P, Chitarra LG, Hoffmann LV, Barroso P (2008) Inheritance of resistance to cotton blue disease. Pesq Agropec Bras Brasilia 43(5):661–665

Kohel RJ (1973) Genetic nomenclature in cotton. J Hered 64:291–295

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2. Bioinformatics 23:2947–2948

Lin ZX, Zhang YX, Zhang XL, Guo XP (2009) A high-density integrative linkage map for *Gossypium hirsutum*. Euphytica 166:35–45

Liu HJ, Myers GO (2002) Genetic relationships and discrimination of ten influential Upland cotton varieties using RAPD markers. Theor Appl Genet 105:325–331

Michelmore RW, Paran I, Kessell RV (1991) Identification of markers linked to disease-resistance genes bybulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. PNAS 88:9828–9832

Nateshan H, Muniyappa V, Swanson MM, Harrison BD (1996) Host range, vector relations and serological relationships of cotton leaf curl virus from southern India. Ann Appl Biol 128:233–244

Nguyen TB, Giband M, Brottier P, Risterucci AM, Lacape JM (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor Appl Genet 109:167–175

Niu C, Lister HE, Nguyen B, Wheeler TA, Wright RJ (2008) Resistance to *Thielaviopsis basicola* in the cultivated A genome cotton. Theor Appl Genet 117:1313–1323

Paterson AH, Brubaker CL, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Report 11:122–127

Rafalski JA (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Romano GB, Sacks EJ, Stetina RS, Robinson AF, Fang DD, Gutierrez OA, Scheffler JA (2009) Identification and genomic location of a reniform nematode (*Rotylenchulus reniformis*) resistance locus (Ren^{ari}) introgressed from *Gossypium aridum* into upland cotton (*G. hirsutum*). Theor Appl Genet 120:139–150

Royo OM, Erazzu L, Bonacic I, Poisson J, Montenegro A (2003) Screening of cotton germplasm for "blue disease" under natural field infestation. In: Swanepoel A (ed) Proceedings of the world cotton research conference-3, Cape Town, South Africa. Agricultural Research Council-Institute for Industrial Crops (Publisher), pp 305–316

Rungis D, Llewellyn D, Dennis ES, Lyon BR (2002) Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium hirsutum* L.) cultivar. Aust J Agri Res 53:551–560

Rungis D, Llewellyn D, Dennis ES, Lyon BR (2005) Simple sequence repeat (SSR) markers reveal low levels of polymorphism

between cotton (*Gossypium hirsutum* L.) cultivars. Aust J Agric Res 56:301–307

Sedcole JR (1977) Number of plants necessary to recover a trait. Crop Sci 17:667–668

Shaheen T, Asif M, Zafar Y, Mehboob-Ur-Rahman (2009) Single nucleotide polymorphism analysis of Mt-Shsp gene of *Gossypium arboreum* and its relationship with other diploid cotton genomes, *G. hirsutum* and *Arabidopsis thaliana*. Pak J Bot 41:177–183

Shen X, Van Becelaere G, Kumar P, Davis RF, May LO, Chee P (2006) QTL mapping for resistance to root-knot nematodes in the M-120 RNR Upland cotton line (*Gossypium hirsutum* L.) of the Auburn 623 RNR source. Theor Appl Genet 113:1539–1549

Silva TF, Correa RL, Castilho Y, Silvie P, Belot JL, Vaslin MFS (2008) Widespread distribution and a new recombinant species of Brazilian virus associated with cotton blue disease. Virol J 5:123–135

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Wang C, Ulloa M, Roberts PA (2006) Identification and mapping of microsatellite markers linked to a root-knot nematode resistancegene (*rkn1*) in Acala NemX cotton (*Gossypium hirsutum* L.). Theor Appl Genet 112:770–777

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot 79:1291–1310

Xiao J, Wu K, Fang DD, Stelly DM, Yu J, Cantrell RG (2009) New SSR markers for use in cotton (*Gossypium* spp.) improvement. J Cotton Sci 13:75–157

Xiao J, Fang DD, Hendrix B, Bhatti M, Cantrell RG (2010) A SNP haplotype associated with a gene resistant to *Xanthomonas campestris pv. malvacearum* (Race 18) in upland cotton (*Gossypium hirsutum* L.). Mol Breed (in press)

Ynturi P, Jenkins JN, McCarty JC Jr, Gutierrez OA, Saha S (2006) Association of root-knot nematode resistance genes with simple sequence repeat markers on two chromosomes in cotton. Crop Sci 46:2670–2674

Young RA, Kelly JD (1996) RAPD markers flanking the *Are* gene for anthracnose resistance in common bean. J Am Soc Hortic Sci 121:37–41

Zhang HB, Li Y, Wang B, Chee PW (2008) Recent advances in cotton genomics. Int J Plant Genomics. Hindawai Publishing Corporation, 20 pp. doi:10.115/2008/742304