CrossMark

Molecular mapping of bunchy top disease resistance in *Gossypium hirsutum* L.

**Marc H. Ellis · Warwick N. Stiller · Tanya Phongkham · Walter A. Tate ·
Vanessa J. Gillespie · Washington J. Gapare · Qian-Hao Zhu ·
Danny J. Llewellyn · Iain W. Wilson**

Abstract Cotton bunchy top (CBT) is an aphid transmitted *Polerovirus* disease and a significant threat to the Australian cotton industry. Symptoms include stunted plant growth, and leaves often display pale green angular patterns at the leaf margins and dark green centers with a leathery texture. Resistance to CBT was evaluated in 206 F_2 plants and 76 $F_{2:3}$ families derived from the resistant cultivar 'Delta Opal' crossed to the susceptible cultivar 'Sicot 70', and in 25 other cultivars the majority susceptible to CBT. CBT resistance in 'Delta Opal' was shown to be controlled by a single dominant locus designated *Cbt*. A combination of AFLP and single nucleotide polymorphism markers located *Cbt* on chromosome A10, close to the mapped resistance locus in 'Delta Opal' to another *Polerovirus* disease of cotton; cotton blue disease. The markers identified flanking CBT resistance will provide useful tools for breeders for marker-assisted selection to alleviate the impact of this disease on cotton production.

Keywords Cotton bunchy top · Virus resistance · *Gossypium hirsutum* · *Polerovirus* · SNP markers

M. H. Ellis · T. Phongkham · W. A. Tate ·
V. J. Gillespie · W. J. Gapare · Q.-H. Zhu ·
D. J. Llewellyn · I. W. Wilson (✉)
CSIRO Agriculture, GPO Box 1600, Canberra,
ACT 2601, Australia
e-mail: iain.wilson@csiro.au

W. N. Stiller
CSIRO Agriculture, Locked Bag 59, Narrabri,
NSW 2390, Australia

M. H. Ellis
133 Route de Beauregard, 74540 Gruffy, France

Introduction

Cotton bunchy top (CBT) was first observed in Australian cotton fields in 1998–99 where it caused significant economic losses (Reddall et al. 2004). Symptoms include reductions in plant height, leaf surface area, petiole and internode length (Ali et al. 2007; Reddall et al. 2004) resulting in significant loss of fiber yield. Leaves also often display pale green angular patterns at the leaf margins and dark green centers with a leathery texture. The disease is spread by the cotton aphid (*Aphis gossypii* Glover) (Ali et al. 2007) that can acquire CBT within 5 min of feeding on infected plants, and viruliferous aphids can inoculate healthy plants after only an hour (Ali et al. 2007). The causal agent of CBT was recently identified as the cotton bunchy top virus (CBTV) belonging to the genus *Polerovirus* of the family *Luteoviridae* (Ellis et al. 2013). Viruses of the *Polerovirus* genus are aphid transmitted and contain non-enveloped single stranded

positive-sense RNA genomes that contain six open reading frames (ORF0 to ORF5) (Mayo and Ziegler-Graff 1996).

CBT symptoms and CBTV shares some similarities with another *Polerovirus* disease; cotton blue disease (CBD) and its causal agent cotton leafroll dwarf virus (CLRDV) (Pupim et al. 2008) that is widespread in Asia and South America. Partial genome sequence of CBTV revealed that its RNA-dependent-RNA polymerase (RdRP, ORF2) shared 90 % amino acid sequence identity to the RdRP of CLRDV, although other ORFs have lower homology. The majority of *Gossypium hirsutum* L. cultivars surveyed are susceptible to CBD, but resistant cultivars (mainly of African descent) have been identified (Royo et al. 2003), including the Australian cultivar 'Delta Opal' used in this study. Resistance to CBD has been shown to be controlled by a single dominant locus (Fang et al. 2010; Pupim et al. 2008).

CBT appears to be endemic to Australia and is currently managed by controlling its aphid vector with the use of insecticides, as most existing commercial cultivars are susceptible. Robust plant resistance is required for long term management of the disease. A source of CBT plant resistance in *G. hirsutum* has been identified in 'Delta Opal', but progress in introducing this trait into modern cultivars has been slowed by the difficulty of reliably assaying for the disease on large segregating breeding populations.

In this study the inheritance of resistance to CBT was analyzed in an F_2 and $F_{2:3}$ population from an intraspecific *G. hirsutum* cross between the resistant cultivar 'Delta Opal' and the susceptible 'Sicot 70'. CBT resistance in the population was consistent with the segregation of a single dominant gene, *Cbt* that was mapped to a 1.7 cM region on chromosome A10. The flanking markers identified in this study will provide useful tools for breeders for the marker-assisted selection of CBT resistant varieties to replace current difficult and labor-intensive phenotypic screening.

Materials and methods

Plant materials

A cross was made between a single plant each of the *G. hirsutum* cultivars 'Delta Opal' (CBT resistant) and

'Sicot 70' (CBT susceptible). After selfing, a population of 206 F_2 plants were evaluated for CBT resistance via infestation with viruliferous aphids (as outlined in CBT resistance assay). All F_2 plants were allowed to self-fertilize after the CBT assay, and F_3 seeds collected. For the F_2 assay the resistant control plants consisted of CSX180, a previously identified homozygous CBT resistant line from an earlier cross of 'Delta Opal' to 'Sicot 70'. For the $F_{2:3}$ assays the resistant control was a previously assayed breeding line designated CSX0895, a backcross derived line of CSX180 containing the genetically modified traits MON531 and MON15985. A panel of 25 mostly Australian varieties with known CBT resistance (W. Stiller pers communication) were selected for SNP genotyping using the CottonSNP63 K array (Table S1).

CBT resistance assay

CBT resistance assays were performed at CSIRO Agriculture, Narrabri NSW (Australia). Seeds were directly sown in 8L pots in a glasshouse with the temperature maintained at 32 °C during the day and 18 °C during the night. For the experiments on individual F_2 plants, three seeds per pot were sown and after establishment, thinned to one plant per pot. One resistant (either CSX180 for F_2 and CSX0895 for $F_{2:3}$ assays) and one susceptible control (parent 'Sicot 70') were included for every 22 test lines. When plants reached four true leaves, they were infested with aphids that had fed on cotton plants already expressing CBT symptoms. This was done by placing a minimum of 20 aphids on each test plant. The experiments were managed so that the plants were not lacking for nutrients or water. When present, aphid parasitoids were controlled by applying a solution of 0.3 g/L of the selective insecticide Spinosad. Four weeks after infestation, all aphids were killed with an application of 0.75 g/L Diafenthiuron. Plants were then monitored on a weekly basis for expression of CBT symptoms such as short internodes and mosaic like leaf patterning, and when symptoms were clearly visible on the susceptible controls, all plants were rated for symptoms on a weekly basis, generally over a 4 week period. Any plant with CBT like symptoms was considered susceptible to the disease. For subsequent $F_{2:3}$ experiments, 10 seeds were sown in each pot and thinned to five plants per pot after establishment. A

randomized complete block design was used, comprising one pot replicated four times, with a minimum of 16 plants assayed per $F_{2:3}$ family. In total, over all CBT F_2 and $F_{2:3}$ assays performed during this study, 271 susceptible control plants were assayed and CBT symptoms were found to develop on an average of 87 % per assay.

DNA extraction

Leaf tissue was collected for all plants in each trial, immediately frozen in liquid nitrogen and stored at −80 °C then freeze dried and stored with silica gel until DNA was extracted. Total genomic DNA was extracted using the DNeasy® Plant Kit (QIAGEN GmbH-Hilden-Germany) as indicated by the manufacturer.

AFLP genotyping

AFLP detection was performed using a protocol adapted from Vos et al. (1995), with minor modifications for cotton (Lopez-Lavalle et al. 2012). One microgram of genomic DNA was digested with one of three sets of enzyme combinations: *Pst*I and *Mse*I, *Eco*RI and MseI, or *Apa*I and *Mse*I to create different DNA libraries. Each library was amplified with 16 sets of primers containing two or three selective bases (Table S2). A molecular weight standard (ABI GeneScan-500 LIZ #4322682) was added at 1 % in Hi-Di Formamide (ABI #4311320) and separated on an ABI 3130xl capillary sequencer. Traces were analysed using ABI GeneMapper software v4.0 (Thermo Fisher Scientific, Waltham, MA).

AFLP analysis was performed on the parent lines and a subset of 40 $F_{2:3}$ lines (14 homozygous resistant, 7 heterozygotes, and 19 susceptible). Approximately 45 bands were scored per reaction, and in total over 2000 fragments were surveyed for linkage to *Cbt*. A single band of 213 bp (designated CL213) derived from *Pst*I-AT: *Mse*I-GAC amplification was found to be linked to CBT resistance. The AFLP fragment was scored as a co-dominant marker by using the height of nearby non-polymorphic peaks as a reference to identify heterozygous alleles based on peak height (Figure S1). To identify its sequence, the complexity of the amplified fragments was reduced by amplifying with an additional selective base on the *Pst*I-AT(A/C/G/T) primer. It was found that the correct nucleotide at

the third position was T and so the fragment was amplified using *Pst*I-ATT:*Mse*I-GAC and isolated based on size. The DNA was eluted from the gel in water, and re-amplified using the *Pst*I-ATT:MseI-GAC primers for Sanger DNA sequencing.

TaqMan genotyping

The SNP associated with markers NG0203481, NG0204310 and NG0211495 (DC20027) were genotyped using Taqman technology using primer sequences previously used to map CBD (Fang et al. 2010). The 7-μL TaqMan assay reaction contained 10 μM each of primers, 0.2 μM each of probes, 5 ng genomic DNA, and 1X TaqMan universal PCR Master Mix (Thermo Fisher Scientific, Waltham, MA). Amplification was carried out using either an ABI7900HT real-time PCR machine or an Eppendorf EP384 cycler, according to the cycling conditions recommended by Applied Biosystems, except the number of cycles was increased to 60. Pre-and post-PCR fluorescence were measured on an ABI7900HT and the results analyzed using the endpoint genotyping module of the ABI7900HT software package (Applied Biosystems Inc.).

Kompetitive Allele Specific PCR (KASP) genotyping

For KASP genotyping, DNA concentration of each sample was standardized to 15 ng/μL. The amplification was performed as for Zhu et al. (2015), in an 8 μL reaction, including 1 μL of DNA, 4 μL of $2 \times$ KASP master mix (LGC Group), 0.11 μL of primer mix (12 μM of each allele-specific primer and 30 μM of common primer) and 2.89 μL of H_2O, according to the manufacturer's instructions. Assays were performed in 384-well format. PCR cycling was performed on an Eppendorf Mastercycler ep384 using the following protocol: hotstart at 95 °C for 15 min, followed by ten touchdown cycles (95 °C for 20 s; touchdown 65–57 °C, 0.8 °C per cycle, 60 s), then followed by 31 cycles of amplification (94 °C for 20 s; 57 °C for 60 s). Plates were read on the ViiA7 Real-Time PCR System (Life Technologies) at ambient temperature and analyzed using the Applied Biosystems ViiA7 software (v1.2.2) (Thermo Fisher Scientific, Waltham, MA). Oligonucleotides used in KASP analyses are shown in Table S3.

Illumina CottonSNP63K array

For genotyping using the Illumina CottonSNP63K array (Illumina, San Diego CA), DNA was standardized at 50 ng/µL for each of the *G. hirsutum* accessions. DNA was processed according to Illumina protocols and hybridized to the CottonSNP63K array at CSIRO Agriculture (Brisbane, Australia) according to the manufacturer's instructions. Single-base extension was performed and the chips were scanned using the Illumina iScan (Illumina, San Diego CA). Image files were saved and analyzed using the GenomeStudio Genotyping Module (v 1.9.4, Illumina). Genotype calls for each SNP were based on the cluster file generated specifically for the CottonSNP63K array (Hulse-Kemp et al. 2015).

Linkage analysis

JoinMap 4.0 software was used for linkage analysis and map construction (Stam 1993). Kosambi mapping function was used to convert recombination frequencies into map distances (centimorgan, cM) with LOD score ≥5.0. For the F_2 disease assays the CBT phenotypes were entered as a dominant trait in the initial F_2 genetic map. The phenotype data from 73 $F_{2:3}$ families was subsequently entered as a co-dominant trait for the final F_2 genetic map (see Table S4 for all $F_{2:3}$ genotype and phenotype data). The TM-1 (*G. hirsutum*) genome of Zhang et al. (2015) was used for DNA sequence comparisons and all genome nucleotide positions listed for markers are based on this publicly available sequence.

Phylogenetic analysis

Cluster analysis to reveal phylogenetic relationships between the 27 cultivars was conducted using R package *ape* and *hclust* function (R Development Core Team 2011) based on the simple matching distance of each pair of cultivars calculated using the same SNP set. The reliability of the dendrogram topology was confirmed with 1000 bootstraps with replacement.

Results

Inheritance of CBT resistance

A population consisting of 206 F_2 lines from the cross between Delta Opal and Sicot70 and eight replicates of

a CBT susceptible ('Sicot 70') and resistant (CS × 180) lines were tested for CBT resistance. Seven of the eight susceptible 'Sicot 70' parental plants displayed CBT related symptoms, whereas none of the resistant lines showed any sign of disease. This indicates false CBT negatives can occur in the assay. A total of 61 F_2 plants were identified as susceptible to CBT disease. This is consistent with a ratio of 3:1 (resistant: susceptible, $X^2 = 0.15$). All individual F_2 plants were selfed to generate $F_{2:3}$ seeds. To improve the accuracy of our CBT phenotype scores, approximately half (73 out of 145) of the F_2 plants that did not display CBT symptoms, were evaluated for CBT resistance using $F_{2:3}$ seeds. If the CBT resistance in 'Delta Opal' is controlled by one single dominant gene, heterozygotes to homozygous resistant families should be identified at a 2:1 ratio in these F_3 families. F_3 CBT assays identified 47 heterozygous to 24 homozygous resistant families ($X^2 = 1$, with two families found to be homozygous susceptible), consistent with a single dominant CBT resistance gene designated *Cbt*.

Genetic mapping of *Cbt*

AFLP analysis was performed on the parent lines and a subset of 40 $F_{2:3}$ lines (14 homozygous resistant, 7 heterozygotes, and 19 susceptible). Over 2000 AFLP fragments were surveyed for linkage to *Cbt* and only a single band of 213 bp amplified using *Pst*I-AT: *Mse*I-GAC primers (designated CL213) was found to be linked to the resistance locus in this population (2.5 cM). Genotyping CL213 on the full 206 F_2 lines determined that CL213 was 4.0 cM from *Cbt*.

The CL213 DNA fragment amplified by the primer combination *Pst*I-ATT and *Mse*I-GAC, was gel isolated and sequenced, and found to be 191 nt long when primer adaptor sequences were removed. Sequence analysis of region from the parent lines revealed that the AFLP polymorphism was due to an A to T transversion that abolished the *Pst*I restriction site in 'Sicot 70' (Figure S2). CL213 did not completely match any region of the draft TM-1 (*G. hirsutum*) genome sequence (Li et al. 2015; Zhang et al. 2015) but was very similar to a region located on chromosome A10 (~98,755,400). Fang et al. (2010), previously mapped the CBD resistance gene to a 0.8 cM region on Chr A10 flanked by SNP markers NG0204310 (Chr A10_ 99,881,319 nt) and

NG0211495 (DC20027, Chr A10_ 99,894,811 nt). Since these markers are in close proximity to the likely location of CL213, three of the SNP markers developed by Fang et al. (2010) for the mapping of *Cbd*, (another close SNP marker NG0203671, was not polymorphic in our population) and four SNP markers (GhCSIRO_2461, GhCSIRO2551, GhCSIRO_2637 and GhCSIRO_2637, Table S3), previously identified between 'Delta Opal' and 'Sicot 70' (Zhu et al. 2014) that were in the same region, were mapped onto the F_2 population (Table S4). All markers were found to be tightly linked to *Cbt*, confirming its location on chromosome A10 (Fig. 1). Initial analysis of the marker data indicated that three F_2 lines (#120, #182 and #208) that had not displayed CBT symptoms in the F_2 assay, were 'Sicot 70'-like in genotype across the entire interval. $F_{2:3}$ families of those lines were assayed for CBT resistance and found to be homozygous susceptible, indicating the original F_2 plants were false negatives for symptom development. Updating the CBT resistance status of these three lines allowed *Cbt* to be mapped to a 1.7 cM region flanked by NG0203481 (A10_ nt 98,715,931) and GhCSIRO_2637 (A10_nt 99,614,176) on Chr A10 (Fig. 1). Based on the draft *G. hirsutum* genome from Zhang et al. (2015) there are 25 annotated genes

(Gh_A10G2062 to Gh_A10G2086) in this region including 11 putative disease resistance like NB-LRR genes (Table S5).

Targeted association analysis

The recent development and availability of the Cotton SNP63K array enabled SNP analysis of a range of *G. hirsutum* cultivars through targeted association analysis, as was recently used to narrow the genetic interval underlying the Okra leaf shape in cotton and identify the causal gene (Zhu et al. 2015). Twenty seven cultivars of mainly Australian origin with known CBT phenotype (4 resistant and 23 susceptible, Table S1) were genotyped using the CottonSNP63K array (Hulse-Kemp et al. 2015). Phylogenic analysis indicated that genetic distances (Nei 1978) among the cultivars were generally small, averaging 0.137, and ranging from 0.007 between TamcotSP37 and TamcotSP37H, to 0.251 between SicotF1 and TamcotSP37H (Figure S3). The cultivars could be separated into three major clusters with the four CBT resistant lines distributed across two clusters (Figure S3). Close examination of the region around Chr A10 where the *Cbt* resistance trait had been initially mapped, indicated forty SNPs that were

Fig. 1 Genetic map of *Cbt* compared to the genetic map of the *Cbd* locus (Fang et al. 2010) on chromosome A10. *Arrows* indicate markers in common between the two genetic maps. All distances in cM
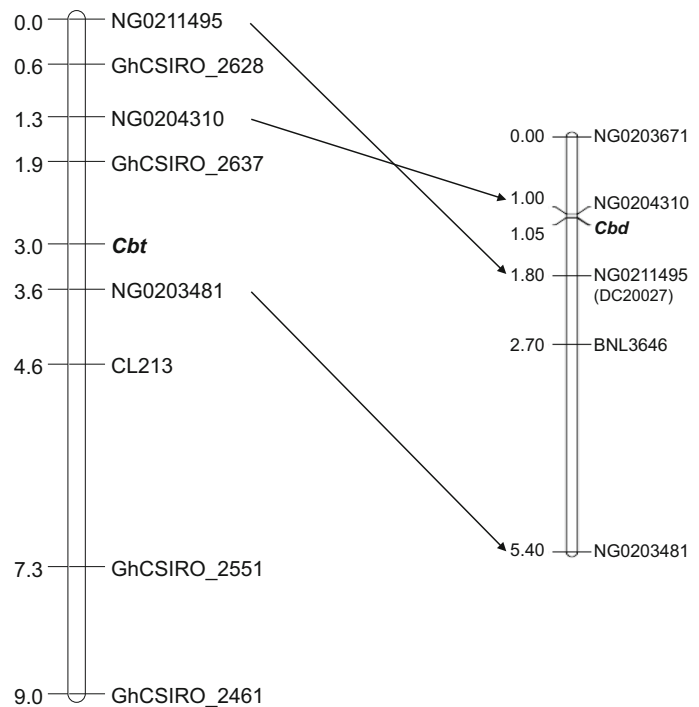
completely associated with either the resistant or susceptible phenotype, and these spanned a region of approximately 1,220,000 nts, from A10 nt 98,718,917 to nt 99,938,987, based on the *G. hirsutum* genome from Zhang et al. (2015) (Table 1, Table S6). This region is slightly larger, but overlaps the genetically mapped region spanned by the markers NG0204310 (A10_ nt 98715931) and GhCSIRO_2637 (A10_nt 99614176), establishing that targeted SNP association analysis in cotton cultivars can provide an alternative means of mapping traits in cotton, even from a relatively small number of cultivars.

Discussion

Aphid transmitted diseases such as CBT are very difficult and time consuming to assay. Despite the assay being performed under optimal conditions for aphid growth there were a significant number of plants that initially appeared without CBT symptoms, but were later found to be susceptible (false negatives for symptom development). This was most clearly observed in the infection rate of our susceptible controls, as on average only 87 % of control plants (271 in total) displayed obvious symptoms in each set of assays. False negatives are most likely due to the failure of aphids to become infected when feeding on infected host, or failing to thrive and pass on the virus to their new host. The inability to reliably identify susceptible individuals makes it essential to screen multiple individuals from the same family in order to confidently determine CBT resistance, but this makes large scale screening of genotypes in the field or glasshouse prohibitive.

CBT resistance segregated as a single dominant locus and genetically mapped to a 1.7 cM region on chromosome A10. The use of the close flanking markers NG0204310 and GhCSIRO_2637 in breeding will enable rapid selection of CBT resistant germplasm with high accuracy, and have already been adopted in the CSIRO breeding program. As these markers do not underlie the causal gene for resistance (perfect markers), disease assay validation would still be required, but only on the reduced number of lines pre-selected using those markers. Based on the published *G. hirsutum* genome as a reference (Zhang et al. 2015), the flanking markers NG0204310 and GhCSIRO_2637 are estimated to be 898,000 nt apart,

although the genome is only a draft sequence and so physical distances are only a guide. Targeted SNP association analysis on 27 cultivars with known CBT phenotype, located the CBT resistance to a slightly larger region 1,220,070 nt, that almost completely overlaps the genetically mapped region (overlap of 895,259 nt). This indicates that as long as a trait is genetically simple and the gene underlying the trait in the population surveyed is the same, targeted association analysis on a relatively small number of cultivars can provide a quick method for finding linked markers for a trait that will be amenable to marker-assisted breeding. The linked SNPs we have identified could be used to survey for CBT resistance from CottonSNP63K data in other cotton varieties, possibly identifying other backgrounds with CBT resistance that could be useful for breeding.

It has been previously noted that CBT resistant cultivars, such as 'Sicala 3-2', 'Delta Opal', 'Guazuncho II', 'Reba' and others that share similar pedigrees, are also resistance to CBD (Ellis et al. 2013). Genetic mapping of CBT resistance confirms that the gene is near the CBD resistance gene mapped by Fang et al. (2010) using three populations of 'Delta Opal' crossed to the susceptible cultivars 'DP388', 'DP5305' and 'SG747'. Our mapping suggests that *Cbt* is located in an adjacent region but is not likely to be the same gene conferring CBD resistance (Fig. 1). However, the genetic distance and marker order between the flanking *Cbd* markers NG0204310 and NG0211495 (DC20027) are different in our population compared to Fang et al. (2010) possibly due a chromosomal rearrangement, making direct comparisons difficult (Fig. 1). The difficulties in reliably assaying for either CBD or CBT resistance means that we cannot rule out the possibility that the same gene is responsible for resistance to both viruses without testing our population for resistance to CBD, which is difficult as this disease does not occur in Australia. It will require identification of the underlying resistance gene before this matter can be resolved. Resistance to multiple viral diseases by the same resistance gene is not uncommon, for example, the tomato *Sw-5* gene confers broad spectrum resistance against tomato spotted wilt virus, and two related viruses; groundnut ringspot virus and tomato chlorotic spot virus (Brommonschenkel et al. 2000). The high identity between some of the viral proteins of CBTV and CLRDV could mean that a similar viral product is detected by the host

Table 1 SNPs associated with CBT resistance

SNP ID	Chr A10 nt[a]	Guazuncho II[b]	Delta Opal	CD401	Sicala3-2	Sicala 40[c]	Sicot 71	CS 50	CS 8S	Sicot 53	Siokra 1-4	Siokra V-15	Namcala	Sicot 70	Siokra V-16	Sicala 43	Siokra V-18	Sicala 45	Sicala V-1	Sicot 80	Sicala V-2	Sicot 189	Sicot F-1	DP90	Sicot 41	Sicot72	Tamcot SP37	Tamcot SP37H
i51659Gb	98718917	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i46697Gh	98719018	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i17634Gh	98721475	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i17635Gh	98721572	C	C	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i17636Gh	98722143	T	T	T	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i12290Gh	98722856	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i12301Gh	98763482	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i49403Gh	99334483	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i46125Gh	99373906	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i30671Gh	99412502	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i12305Gh	99476258	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i17649Gh	99478067	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i12307Gh	99521724	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i17643Gh	99522388	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i46124Gh	99526061	C	C	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i49130Gh	99580654	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i12309Gh	99613820	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i17651Gh	99614107	G	G	G	-	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i12310Gh	99640885	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i52307Gb	99657483	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i12311Gh	99708344	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i45223Gh	99732087	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i49933Gh	99753152	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i27989Gh	99761579	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i46070Gh	99763501	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i12317Gh	99877422	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i17657Gh	99877535	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i17658Gh	99877603	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i17659Gh	99877654	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i12318Gh	99877834	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i12319Gh	99878110	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i12321Gh	99879387	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i12323Gh	99884701	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i12327Gh	99892096	C	C	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
i00530Gh	99893165	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i12328Gh	99893165	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
i12329Gh	99894024	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
i12330Gh	99934309	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i12331Gh	99937131	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
i12332Gh	99938987	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T

[a] SNP position based on Zhang et al. (2015) *G. hirsutum* genome sequence

[b] Cultivars in *red* are CBT resistant [a]SNP position based on Zhang et al. (2015) *G. hirsutum* genome sequence

[c] Cultivars in black are CBT susceptible

defence systems of resistant cultivars upon infection by either pathogen.

To date no plant resistance gene has been cloned that acts against *Poleroviruses.* The majority of single dominant anti-viral resistance traits in plants have been found to encode nucleotide binding leucine rich repeat proteins (NB-LRR) (de Ronde et al. 2014). Analysis of the *G. hirsutum* genome sequence between

NG0204310 and GhCSIRO_2637 reveals eleven putative NB-LRR genes (A10G2067, A10G2068, A102070, A102072, A102073, A102074, A102076, A102077, A102078, A102080 and A102081) that could be potential candidates for *Cbt* (Table S5). Identification of the *Cbt* gene will require either fine mapping of the region and/or testing of each of the eleven potential NB-LRR candidates isolated from 'Delta Opal' transformed into a susceptible variety.

In summary, long term management of CBT in Australia requires the transfer of host plant resistance against CBTV to all current and future commercial cotton varieties. Phenotypical assays for CBT resistance have proven to be difficult and somewhat unreliable and as yet no new CSIRO CBT resistant germplasm has been released in Australia. The SNP markers identified in this study will significantly improve the speed and accuracy of screening for CBT resistant germplasm and that should lead to a more rapid deployment of resistant varieties in the future.

References

Ali A, Reddall A, Roberts J, Wilson LJ, Rezaian MA (2007) Cytopathology, mode of aphid transmission and search for the causal agent of cotton bunchy top disease. J Phytopathol 155(4):220–227. doi:10.1111/j.1439-0434.2007.01220.x

Brommonschenkel SH, Frary A, Frary A, Tanksley SD (2000) The broad-spectrum tospovirus resistance gene Sw-5 of tomato is a homolog of the root-knot nematode resistance gene Mi. Mol Plant-Microbe Interact 13(10):1130–1138

de Ronde D, Butterbach P, Kormelink R (2014) Dominant resistance against plant viruses. Front Plant Sci 5:307

Ellis MH, Silva TF, Stiller WN, Wilson LJ, Vaslin MFS, Sharman M, Llewellyn DJ (2013) Identification of a new Polerovirus (family Luteoviridae) associated with cotton bunchy top disease in Australia. Australas Plant Pathol 42(3):261–269. doi:10.1007/s13313-012-0177-8

Fang DD, Xiao JH, Canci PC, Cantrell RG (2010) A new SNP haplotype associated with blue disease resistance gene in cotton (Gossypium hirsutum L.). Theor Appl Genet 120(5):943–953

Hulse-Kemp AM, Lemm J, Plieske J, Ashrafi H et al (2015) Development of a 63 K SNP Array for Cotton and high-density mapping of intraspecific and interspecific populations of Gossypium spp. G3-genes genomes. Genetics 5(6):1187–1209. doi:10.1534/g3.115.018416

Li FG, Fan GY, Lu CR et al (2015) Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nat Biotechnol 33(5):524

Lopez-Lavalle LAB, Gillespie VJ, Tate WA, Ellis MH, Stiller WN, Llewellyn DL, Wilson IW (2012) Molecular mapping of a new source of Fusarium wilt resistance in tetraploid cotton (Gossypium hirsutum L.). Mol Breed 30(2):1181–1191. doi:10.1007/s11032-012-9705-z

Mayo MA, ZieglerGraff V (1996) Molecular biology of luteoviruses. Adv Virus Res 46(46):413–460

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89(3):583–590

Pupim O, Schuster I, Pinto RB, Pires E, Belot JL, Silvie P, Chitarra LG, Hoffmann LV, Barroso P (2008) Inheritance of resistance to cotton blue disease. Pesqu Agropecu Bras 43(5):661–665

Reddall A, Ali A, Able JA, Stonor J, Tesoriero L, Wright PR, Rezaian MA, Wilson LJ (2004) Cotton bunchy top: an aphid and graft transmitted cotton disease. Australas Plant Pathol 33(2):197–202. doi:10.1071/AP03094

Royo O, Erazzu L, Bonacic I, Poisson J, Montenegro A (2003) Screening of cotton germplasm for ''blue disease'' under natural field infestation. In: Zambolim L (ed) Proceedings of the world cotton research conference-3. Cape Town, pp 305–316

Stam P (1993) Construction of integrated genetic-linkage maps by means of a new computer package—joinmap. Plant J 3(5):739–744

Team RDC (2011) R: a language and environment for statistical computing. the R Foundation for Statistical Computing. Vienna. http://www.R-project.org

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) Aflp—a new technique for DNA-Fingerprinting. Nucleic Acids Res 23(21):4407–4414

Zhang TZ, Hu Y, Jiang WK et al (2015) Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol 33(5):531

Zhu QH, Spriggs A, Taylor JM, Llewellyn D, Wilson I (2014) Transcriptome and complexity-reduced, DNA-based identification of intraspecies single-nucleotide polymorphisms in the polyploid Gossypium hirsutum L. G3-genes genomes. Genetics 4(10):1893–1905. doi:10.1534/g3.114.012542

Zhu Q-H, Zhang J, Liu D, Stiller W, Liu D, Zhang Z, Llewellyn D, Wilson I (2015) Integrated mapping and characterization of the gene underlying the okra leaf trait in Gossypium hirsutum L. J Exp Bot. doi:10.1093/jxb/erv494