Theor Appl Genet (2012) 125:1525–1537
DOI 10.1007/s00122-012-1931-6

ORIGINAL PAPER

Development of candidate gene markers associated to common bacterial blight resistance in common bean

Chun Shi · Kangfu Yu · Weilong Xie · Gregory Perry · Alireza Navabi · K. Peter Pauls · Phillip N. Miklas · Deidré Fourie

Received: 21 February 2012 / Accepted: 28 June 2012 / Published online: 14 July 2012

Abstract Common bacterial blight (CBB), caused by *Xanthomonas axonopodis* pv. *phaseoli* (*Xap*), is a major yield-limiting factor of common bean (*Phaseolus vulgaris* L.) production around the world. Two major CBB-resistant quantitative trait loci (QTL), linked to the sequence characterized amplified region markers BC420 and SU91, are located at chromosomes 6 and 8, respectively. Using map-based cloning approach, four bacterial artificial chromosome (BAC) clones from the BC420-QTL locus and one BAC clone containing SU91 were sequenced by Roche 454 technique and subsequently assembled using merged assemblies from three different programs. Based on the quality of the assembly, only the sequences of BAC 32H6 and 4K7 were used for candidate gene marker (CGM) development and candidate gene (CG) selection. For the BC420-QTL locus, 21 novel genes were predicted in silico by FGENESH using *Medicago* gene model, whereas 16 genes were identified in the SU91-QTL locus. For each putative gene, one or more primer pairs were designed and tested in the contrasting near isogenic lines. Overall, six and nine polymorphic markers were found in the SU91- and BC420-QTL loci, respectively. Afterwards, association mapping was conducted in a breeding population of 395 dry bean lines to discover marker-trait associations. Two CGMs per each locus showed better association with CBB resistance than the BC420 and SU91 markers, which include BC420-CG10B and BC420-CG14 for BC420_QTL locus, and SU91-CG10 and SU91-CG11 for SU91_QTL locus. The strong associations between CBB resistance and the CGs 10 and 14 from BC420_QTL locus and the CGs 10 and 11 from SU91_QTL locus indicate that the genes 10 and 14 from the BC420 locus are potential CGs underlying the BC420_QTL locus, whereas the genes 10 and 11 from the SU91 locus are potential CGs underlying the SU91_QTL locus. The superiority of SU91-CG11 was further validated in a recombinant inbred line population Sanilac × OAC 09-3. Thus, co-dominant CGMs, BC420-CG14 and SU91-CG11, are recommended to replace BC420 and SU91 for marker-assisted selection of common bean with resistance to CBB.

Communicated by D. Lightfoot.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1931-6) contains supplementary material, which is available to authorized users.

C. Shi · K. Yu (✉) · A. Navabi
Greenhouse and Processing Crops Research Centre, Agriculture and Agri-Food Canada, Harrow, ON N0R 1G0, Canada
e-mail: yuk@agr.gc.ca

W. Xie · G. Perry · A. Navabi · K. Peter Pauls
Department of Plant Agriculture, University of Guelph, Crop Science Building, Guelph, ON N1G 2W1, Canada

P. N. Miklas
Vegetable and Forage Crop Production Unit, U.S. Department of Agriculture, Prosser, WA 99350, USA

D. Fourie
ARC Grain Crops Institute, Potchefstroom, Republic of South Africa

Introduction

Common bean (*Phaseolus vulgaris* L.) is the most important pulse crop for direct human consumption. Its nutritional composition includes complex carbohydrates (e.g. fibre, resistant starch, and oligosaccharides), vegetable protein, important vitamins and minerals like folate and iron, as well as antioxidants and only very small amounts of fat (McClean et al. 2008). In 2006, the bean industry was

valued at $1.2 billion and $180 million in USA and Canada, respectively (http://www.pulsecanada.com/). Common bacterial blight (CBB), incited by *Xanthomonas axonopodis* pv. *phaseoli* (*Xap*), is a major yield-limiting factor of common bean production around the world (Miklas et al. 2006). Since CBB resistance is a complex trait that is quantitatively inherited, host genetic resistance is the most effective and environmentally sound approach to control CBB (Miklas et al. 2006).

The two sequence characterized amplified region (SCAR) markers, SU91 and BC420, have been of particular interest to bean breeding programs for CBB resistance using marker-assisted selection (MAS) (Miklas et al. 2006). The tepary bean (*Phaseolus acutifolius* A. Gray) accession PI31944, via the inter-specific dry bean line XAN159, is considered to be the source of both BC420- and SU91-linked quantitative trait loci (QTL). Together with another tepary bean accession PI440795, they have provided the major source of resistance to CBB in Canadian bean breeding programs. The germplasm lines, HR45 and HR67, and the variety 'Apex' (HR199-4857) have obtained their resistance through crosses with XAN159 (Shi et al. 2011). BC420 was located on linkage group B6 (Miklas et al. 2006) and SU91 on linkage group B8 (Miklas et al. 2006). The different chromosome positions of the two SCAR markers make them attractive sources for introgressing independent CBB-resistance QTL into susceptible bean cultivars (Miklas et al. 2006). The presence of BC420 accounted for ~62 % of the phenotypic variation for resistance to CBB in a population derived from a cross between HR67 and a susceptible white bean W1744d (Yu et al. 2000). In the F_5 recombinant inbred lines (RILs) derived from a cross between HR67 and OAC95-4 (moderately susceptible in the greenhouse), the presence of BC420 accounted for ~63 % of the phenotypic variation for resistance to CBB (Yu et al. 2004). The presence of SU91 accounted for ~14 % of the phenotypic variation for resistance to CBB in an F_2 population derived from a cross between XAN159 (resistant) and Chase (moderately susceptible) (Vandemark et al. 2008), and its presence accounted for 17 % of the phenotypic variation for resistance in a BC2F1 population derived from the same cross (Vandemark et al. 2008). A recessive epistatic interaction has been observed between the two loci: (1) the expression of BC420 was epistatically suppressed by a homozygous recessive su91//su91 genotype; (2) SU91//SU91 and SU91//su91 genotypes conditioned an intermediate disease reaction when homozygous recessive for bc420//bc420; and (3) the highest level of disease resistance was conferred by genotypes with at least a single resistance allele at both QTL (BC420//-; SU91//-) (Vandemark et al. 2008). This result indicates that breeders may realize the greatest gains for resistance to CBB by selecting breeding materials that are fixed for both QTL.

Unfortunately, both BC420 and SU91 are the dominant markers. Dominant markers are less efficient for MAS because they cannot distinguish heterozygous QTL from homozygous QTL of a plant. So far, no co-dominant markers have been developed for the SU91 QTL, and could be used efficiently in MAS for CBB resistance. Moreover, both the SU91 and BC420 markers were derived at random from polymorphic sites in the genome. Genetic linkage between these markers and a target QTL allele can be broken by genetic recombination. In contrast, candidate gene markers (CGMs) are markers derived from polymorphisms within genes that should have complete linkage with the target loci (Andersen and Lubberstedt 2003). Physical mapping enables the efficient generation of new CGMs useful for co-dominant marker development. The BC420-QTL has recently been physically mapped using a bacterial artificial chromosome (BAC) library for HR45 (Liu et al. 2008). The minimum tiling path of the BAC contig contains six BAC clones spanning an estimated size of 750 kb covering this QTL region. BAC 4K7 is a clone containing the BC420 marker (Liu et al. 2010). To date, the SU91-QTL has not been physically mapped yet, and there is some ambiguity about its precise genetic mapping position. The current approach to BAC sequencing are the next-generation sequencing platforms, due to higher coverage and lower cost comparing to Sanger sequencing technique (Varshney et al. 2009). Since the bean genome had not been sequenced, the Roche 454 platform was selected, as the longer reads generated (currently about 400 bases) are more amenable to de novo assembly and annotation (Varshney et al. 2009). Moreover, different assembly software packages use different assembly algorithms; therefore, combining optimal assemblies from different programs can give a more credible final product (Kumar and Blaxter 2010).

Physical mapping provides a glimpse of local gene content useful for searching candidate genes underlying respective QTL. Annotation of plant genomic sequences can be separated into structural and functional annotation (Ouyang et al. 2009). Structural annotation is dependent on sensitive, specific computational programs and deep experimental evidence to identify gene features within genomic DNA, while functional annotation is highly dependent on sequence similarity to other known genes or proteins. Coupling structural and functional annotation across genomes in a comparative manner promotes a more accurate annotation, i.e. candidate gene catalogue (Ouyang et al. 2009). Near isogenic lines (NILs) harbouring BC420- and/or SU91-QTL have been developed through crossing XAN159 to Teebus (susceptible) (Vandemark et al. 2008). NILs are premium plant materials for screening candidate genes as in the absence of other segregating QTL, the target QTL become the major genetic source of genetic

resistance (Salvi and Tuberosa 2005). Since common bean is reputed to be a species recalcitrant to transformation (Seo et al. 2006), the validation of promising candidate genes remains a challenge in common bean. As an alternative to direct proof, candidate genes can be indirectly functionally approved through association mapping (i.e. by identifying, within a set of genotypes such as germplasm accessions and cultivated varieties, a statistical association between allelic variants at marker or candidate loci and the mean of the analysed trait) (Andersen and Lubberstedt 2003). Shi et al. (2011) successfully applied genome-wide association mapping in a bean breeding population to detect the markers associated to CBB resistance. Fourteen markers were significant for both days after inoculation (DAI) and five markers were highly significant ($P \leq 0.001$), including BC420 and SU91.

The objectives of our study were to (1) identify a target BAC clone harbouring SU91 marker, (2) sequence five BAC clones at the BC420 locus and assemble two of them using a combination of three different assembly algorithms, (3) structurally and functionally annotate BAC sequences, (4) develop and screen CGMs using contrasting NILs, (5) apply association study to evaluate CGMs and validate candidate genes, and (6) validate SU91 CGMs in a bi-parental population.

Materials and methods

Plant materials

A pair of NILs, BBSS and bbss, are BC6-derived F_2 (BC6:F_2) plants from a cross between XAN159 × Teebus (Vandemark et al. 2008). A population of 392 advanced bean breeding lines were used in association study (Fig. 1). The population includes genotypes from different gene-pool origins and breeding materials from different stages of variety development in the AAFC-University of Guelph Bean Breeding Program. These included 130 lines in the advanced yield trials and 262 lines in the preliminary yield trials. The population represents the range of genetic diversity within the breeding program in Ontario, Canada. Moreover, a population of 90 $F_{4:5}$ RILs derived from a cross between Sanilac (CBB susceptible) and OAC 09-3 (CBB resistant) was used to confirm the association of the new markers with CBB resistance.

BAC sequencing, assembly, and annotation

Bacterial artificial chromosomes previously found to contain the regions of interest were sequenced using the 454 platform (Roche Canada, Laval, Quebec) in Genome Quebec (BAC 4K7 and 59I10) (Montreal, Quebec), Beckman Coulter Genomics (BAC 32H6) (Beverly, Massachusetts), and Plant Biotechnology Institute (BAC 52C11 and 79B3) (Saskatoon, Saskatchewan). BAC assembly used three programs, Newbler (Roche Canada, Laval, Quebec), CLC Genomics Workbench (Katrinebjerg, Denmark), and Geneious (Auckland, New Zealand). Initial assembly by Newbler was received from the BAC sequencing service provider and was treated as the primary assembly. Afterwards, secondary assembly was conducted by CLC Genomics Workbench using the "high-throughput de novo assembly" module on the trimmed 454 reads. Results from the 454 trimmed reads, the primary and secondary assemblies were integrated into Geneious to create final assembly. The sequences of BAC 4K7 (JX000234) and 32H6 (JX 000235) were submitted to the National Centre for Biotechnology Information (NCBI) database. Gene predictions were based on ab initio method by FGENESH software using *Medicago* gene model (http://www.softberry.com). Sequences were subjected to BlastN similarity search analysis against the NCBI, non-human, non-mouse ESTs (est_others) database and annotated according to their homology with known EST sequences.

Phenotypic evaluation

Plant materials for the association studies were field evaluated in 2009 at the CBB nursery in Harrow, Ontario, Canada. The experimental design was a randomized complete block with two replications. Each experimental unit consisted of a single 0.5 foot long row with a 2 foot row spacing. Artificial inoculation was carried out using fresh bacterial inoculum, prepared by mixing equal amount of two fuscans isolates 12 and 118, and two non-fuscans isolates 18 and 98 with a concentration of 10^8 CFU/ml. These four strains are endemic in Ontario, Canada, and were isolated from the field (Park and Dhanvantari 1987). Plots were mechanically inoculated at the unifoliolate growth stage using a high-pressure sprayer at the constant pressure of 250 psi. Two CBB ratings were taken at 14 and 21 DAI using a 0–5 scale for disease severity ratings, based on a visual estimate of the percentage of CBB symptoms on the total leaf area, where 0 = no symptoms, 1 = <10 %, 2 = 11–30 %, 3 = 31–50 %, 4 = 51–80 %, and 5 = more than 80 % of inoculated areas showing symptoms. CBB resistant (HR45) and susceptible (Dresden) checks were included in each block. Excel Macros programmed by QI Macros (http://www.qimacros.com/) was used to conduct Kolmogorov–Smirnov test of normality.

The $F_{4:5}$ RILs derived from a cross between Sanilac and OAC09-3 were evaluated in field in the summer of 2011 and in growth room using a mix of the four *Xap* isolates mentioned above with two replicates. The area under the

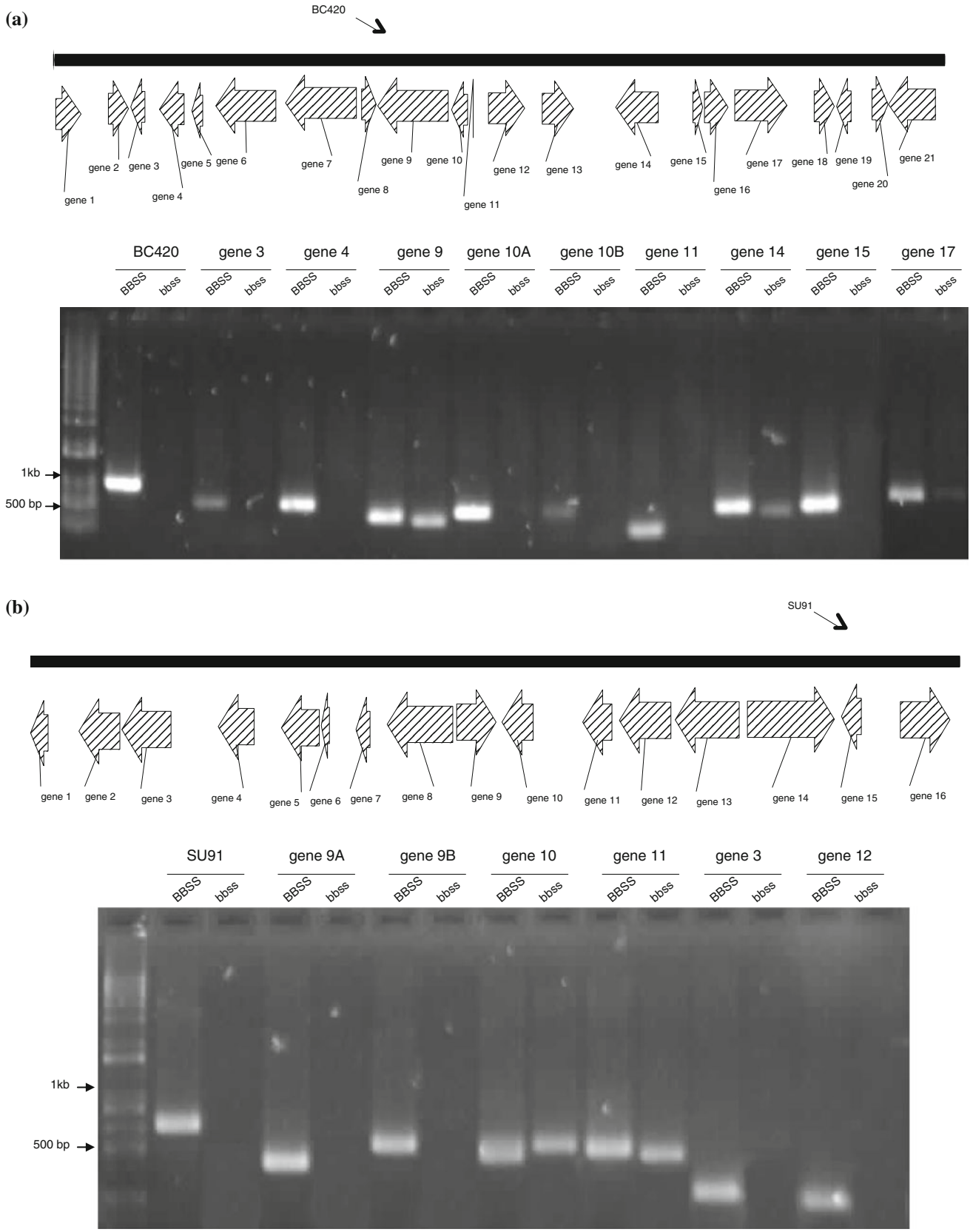

Fig. 1 BAC gene prediction and CGMs (candidate gene markers) screening in contrasting NILs (near isogenic lines). **a** BAC 4K7; **b** BAC 32H6. BBSS and bbss are the resistant and susceptible NILs

disease progress curve (AUDPC) was calculated based on three times rating for each replicate in the field test as:

$$\text{AUDPC} = \sum_{i=1}^{4} \left[\left(\frac{X_{i+1} + X_i}{2} \right) (t_{i+1} - t_i) \right]$$

where, X_i is the severity rating of the host tissue damaged at the ith rating, and t_i is the time in DAI at ith rating.

Genotyping

Young leaf samples (100 mg) were frozen in liquid nitrogen and ground using an AutoGrinder 48 (AutoGen Inc., Holliston, MA, USA). After incubation with plant lysis buffer (AutoGen AG00121) at 65 °C for 30 min, the DNA was automatically extracted using an AutoGen 850 Alpha system following the manufacturer's directions (AutoGen Inc.).

One or two primer pairs per each putative gene were designed using Primer3 (http://frodo.wi.mit.edu/). PCR was performed in 25 µl volumes containing 1 µl genomic DNA (25 ng/µl), 0.5 µl dNTP mixtures (10 mM each), 5 µl 5 × Green GoTaq PCR buffer (Promega, USA), 2 µl primers (1.5 mM each), 0.2 µl GoTaq polymerase (5 Units/µl) (Promega, USA), and 16.5 µl double-distilled water. The PCR program was optimized and consisted of initial denaturation at 94 °C for 3 min, 20 cycles of 30 s at 94 °C, 45 s from 56 °C to a "touchdown" (the annealing temperature was lowered 1 °C after every two subsequent cycles) at 47 °C for primer annealing, 1 min at 72 °C for primer extension, followed by 21 cycles with a 47 °C annealing temperature, and ended with 1 cycle of 10 min at 72 °C. The PCR products were analysed on 1.5 % agarose gel and visualized by SYBR® Safe staining (Invitrogen, USA). The $F_{4:5}$ RILs, derived from a cross between Sanilac and OAC09-3, were genotyped using SU91, SU91-CG10, and SU91-CG11. The PCR protocol includes: 34 cycles of 30 s at 94 °C, 45 s at 60 °C, and 60 s at 72 °C, followed by one cycle of 10 min at 72 °C. The PCR products were electrophoresed through 2 % high resolution agarose gel containing ethidium bromide for 3 h at 100 V.

Statistical analysis

Association study analyses were carried out with TASSEL 2.1 software, available at http://www.maizegenetics.net/index.php?option=com_content&task=view&id=89&Itemid=119. The MLM analyses were performed using a kinship K matrix and population structure Q matrix. The K matrix was generated based on 75 SNPs using kinship matrix function in TASSEL (Shi et al. 2011). Population structure consisted of a Q matrix that describes the percent subpopulation parentage for each line in the analysis (Shi et al. 2011). These percentages were calculated by STRUCTURE 2.3.3 software, available at http://pritch.bsd.uchicago.edu/structure.html. We set k (the number of subpopulations) from 1 to 10 and performed 10 runs for each k value. For each run, a burn in of 5,000 iterations was followed by an additional 5,000 iterations. Since the likelihood for model parameter $k = 2$ was much higher than $k = 1$ and comparable with $k = 3$ or higher, we chose $k = 2$ and generated a Q matrix from 75 SNPs (Shi et al. 2011).

Results

Screening HR45 BAC library using SU91 marker

As the HR45 germplasm line contains SU91-QTL, a BAC library was constructed from high molecular weight DNA of HR45 (Liu et al. 2010). The library is composed of 33,024 clones with insert sizes of ranging from 30 to 280 kb, with an average size of 107 kb (Liu et al. 2010). Since the library is estimated to represent approximately a sixfold genome coverage (Liu et al. 2010), only one-third of library was screened with twofold genome coverage. The marker SU91 was used to screen plate pools, column pools, and row pools from the library to identify positive clones through PCR screening. Only one clone 32H6, with an insert size of ~90 kb, was positive for the SU91 marker.

BAC sequencing and de novo assembly

Based on the physical map of BC420-QTL locus, four BACs, 4K7, 52C11, 59I10, and 79B3, were selected in a previous study for sequencing (Liu et al. 2010). Adding BAC 32H6 for SU91-QTL, five BACs were sequenced using the Roche 454 pyrosequencing platform. Because of the short 454 reads generated, no BAC was assembled into one contig by Roche 454 assembly software Newbler. Newbler uses the overlap–layout–consensus (OLC) algorithm, whereas CLC Genomics Workbench and Geneious use de Bruijun and Greedy Algorithms (Miller et al. 2010). As the programs used different underlying algorithms to generate contigs, we explored merging of assemblies and found that the merged data sets not only aligned better to reference sequences than individual assemblies, but were also more consistent in the number and size of contigs. BAC 32H6, 4K7, 52C11, and 59I10 were assembled into one contig totalling 96, 101, 75, and 78 kb, respectively, whereas BAC 79B3 was assembled into two contigs totalling 92 kb (Table 1). The contig sizes of BAC 32H6, 4K7, and 52C11 are close to the size estimates made through pulse field gel electrophoresis, whereas BAC

Table 1 The summary of BAC assembly

BAC	Estimated size[a] (kb)	454 reads	Contig[b] (kb)	Consensus sequence[c] (kb)	Identical to reference sequence
32H6	90	115,197	96	65	pIndigoBAC-5 (97.5 %)[d], SU91 (99.7 %)
4K7	90	47,209	101	88	pIndigoBAC-5 (99.9 %), BC420 (99.8 %)
52C11	70	38,656	75	75	pIndigoBAC-5 (90.2 %), BC420 (54.9 %)
59I10	100	20,652	78	76	pIndigoBAC-5 (99.7 %), STS183.142 (92.7 %)
79B3	148	20,848	55	51	pIndigoBAC-5 (50.0 %), STS333.274 (55.3 %)
			37	34	pIndigoBAC-5 (12.3 %), STS333.274 (53.5 %)

[a] The sizes were estimated by pulse field gel electrophoresis

[b] Largest contig in final assemble using Geneious software

[c] The contig essentially appears as a multiple sequence alignment of the fragments. After some manual editing of the contig to resolve disagreements between fragments which result from read errors, the consensus sequence of the contig is extracted as the sequence being reconstructed

[d] The percent identity, i.e. the percent that the BAC consensus sequences match up respective reference sequences within the limits of the full match

59I10 and 79B3 were much less than expected. Consensus sequence was constructed to remove the gaps and keep the most frequent residues at each site for the contigs. In comparison with pIndigoBAC-5 vector sequence, more than 97 % of the sequences were recovered from the contigs of BAC 32H6, 4K7, and 59I10, but <50 % from BAC 79B3. Moreover, the consensus sequences of BAC 4K7 and 32H6 were found to have high homology to the original BC420 (99.8 %) and SU91 (99.7 %) marker sequences, respectively. Thus, accounting for assembly size and relative to known target sequence, the sequences of BAC 32H6 and 4K7 had the best value for CGM development.

BAC sequence annotation

Both ab initio and homology-based methods were used to search the genes in the BAC sequences (Fig. 1; Table 2). For BAC 4K7, 21 novel genes were ab initio predicted by FGENESH using *Medicago* gene model, including 10 from sense chain and 11 from anti-sense chain (Table 2). Due to limited resources, not all of them could identify similar nucleotide sequences that exist in current EST databases by BLASTN. Of the 21 putative genes, four and three genes were supported by common bean and tepary bean ESTs, respectively, there were 16 genes in BAC 32H6 including 3 from sense chain and 13 from anti-sense chain. Of them, three, five, and one putative genes were supported by common bean, tepary bean, and soybean ESTs, respectively (Table 2). In addition, according to current protein database, genes 3, 19, and 21 identified from BAC 4K7 encoding the proteins similar to common bean gag-pol protein (AAR13313) and potato integrase core domain containing protein (ABI34329), whereas genes 3, 10, 11, and 14 from BAC 32H6 encoding the protein homologues of soybean UDP-glycosyltransferase 82A1-like protein (XP_003544918), anthocyanidin 3-*O*-glucosyltransferase 1-like protein (XP_003518582), and choline-phosphate cytidylyltransferase B-like protein (XP_003545206) (Table 2).

Screening CGMs using contrasting NILs

For each putative gene, one or two primer pairs were designed and tested in the contrasting NILs. NIL BBSS and bbss represent resistant and susceptible genotypes. Nine CGMs, derived from BAC 4K7, showed polymorphisms, including six dominant markers only present in BBSS and three co-dominant markers with size polymorphism between BBSS and bbss (Fig. 1a). These CGMs come from eight putative genes (Fig. 1a, b). BC420 was located between genes 8 and 9 (Fig. 1a). Meanwhile, six CGMs derived from BAC 32H6 showed polymorphisms. They include four dominant markers only present in BBSS and two co-dominant markers with size polymorphism between BBSS and bbss (Fig. 1b). Five candidate genes contribute these CGMs. SU91 was located between genes 14 and 15 (Fig. 1b).

The evaluation of CGMs by association study

Polymorphic CGMs were further evaluated in a bean breeding population, including 392 advanced breeding lines (Fig. 2a). Each line was rated twice for CBB resistance at 14 and 21 DAI. The resistant check, HR45, was scored 0 at both disease observation dates, whereas the susceptible check, Dresden, was scored 5 (Fig. 2b). The frequency distribution of CBB severity scores showed a continuous variation with population mean shifted towards susceptibility (Fig. 2b). The Kolmogorov–Smirnov test of normality for the whole population was significant ($P \leq 0.05$) for both 14 and 21 DAI. Since a complex familial relationships and population structure was found in this breeding population, associations between the markers

Table 2 List of genes predicted by FGENESH program and its EST and protein homologues from BlastN and BlastX output

FGENESH (Medicago)			BlastN		BlastX	
Gene	Chain	Start base–end base	Supported EST	*E* value (max identity)	Protein homologue	*E* value (max identity)
BAC 4K7 annotation						
1	Plus	62–2,372	HO780792, tepary bean	1*E*−89 (92 %)		
2	Plus	4,873–6,724				
3	Minus	6,931–8,231	CB539386, common bean seedling	0 (95 %)	AAR13313, common bean gag-pol polyprotein	2*E*−20 (61 %)
4	Minus	9,541–11,816				
5	Minus	12,114–13,118				
6	Minus	14,301–19,789				
7	Minus	20,683–27,177	FE703648, common bean	8*E*−161 (82 %)		
8	Plus	27,647–28,981	FE703648, common bean	8*E*−166 (82 %)		
9	Minus	29,105–35,574				
10	Minus	35,873–37,331				
11	Minus	37,853–38,323				
12	Plus	39,255–42,465				
13	Plus	44,163–47,036				
14	Minus	50,877–54,764				
15	Plus	57,924–58,807				
16	Plus	58,990–61,126				
17	Plus	61,762–66,526	HO780792, tepary bean	1*E*−89 (99 %)		
18	Plus	69,025–70,874				
19	Minus	71,081–72,381	CB539386, common bean seedling	0 (95 %)	AAR13313, common bean gag-pol polyprotein	2*E*−20 (61 %)
20	Plus	74,304–75,762				
21	Minus	75,812–80,096	HO777542, tepary bean	0 (99 %)	ABI34329, potato integrase core domain containing protein	0 (59 %)
BAC 32H6 annotation						
1	Minus	867–1,114				
2	Minus	3,047–5,616				
3	Minus	5,733–8,815	CX546984, water stressed soybean	3*E*−164 (85 %)	XP_003544918, soybean UDP-glycosyltransferase 82A1-like	0 (72 %)
4	Minus	11,784–14,035	FE676335, common bean	0 (97 %)		
5	Minus	15,745–18,101	HO780614, tepary bean	0 (98 %)		
6	Minus	18,228–18,751				
7	Minus	20,404–21,286				
8	Minus	22,376–26,033	HO796722, tepary bean	0 (99 %)		
9	Plus	26,066–28,815	HO787592, tepary bean	0 (99 %)		
10	Minus	29,180–31,139	HO797100, tepary bean	0 (98 %)	XP_003518582, soybean anthocyanidin 3-*O*-glucosyltransferase 1-like	0 (66 %)
11	Minus	34,246–36,076	HO797100, tepary bean	0 (96 %)	XP_003518582, soybean anthocyanidin 3-*O*-glucosyltransferase 1-like	0 (65 %)
12	Minus	36,542–39,769				

Table 2 continued

FGENESH (Medicago)			BlastN		BlastX	
Gene	Chain	Start base–end base	Supported EST	*E* value (max identity)	Protein homologue	*E* value (max identity)
13	Minus	40,050–44,039				
14	Plus	44,568–50,027	FE684024, common bean	0 (99 %)	XP_003545206, soybean choline-phosphate cytidylyltransferase B-like	2*E*−135 (94 %)
15	Minus	50,458–51,720	FE694947, common bean	0 (94 %)		
16	Plus	54,167–57,260				

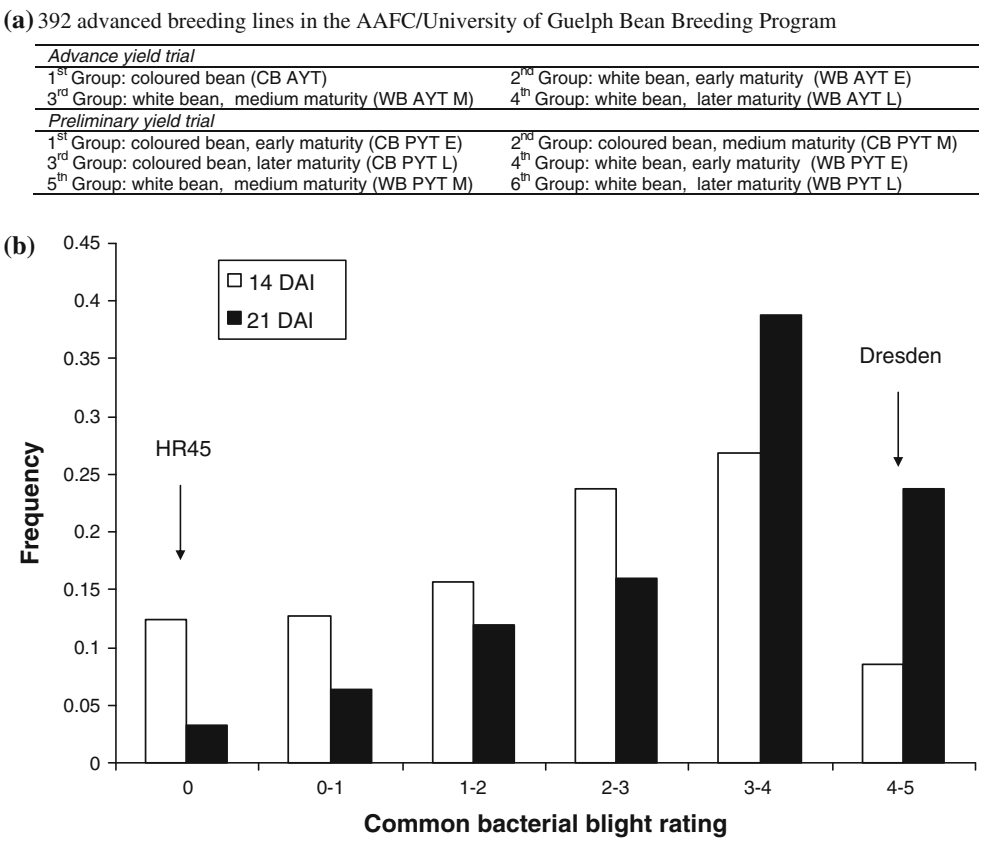

Fig. 2 The frequency distribution of CBB ratings of breeding population in 2009 CBB nursery. Rating scales: 0 = no symptoms and 5 = more than 80 % of inoculated areas showing symptoms

and CBB rating were determined by unified MLM method, taking population structure and kinship into account. Because CBB ratings varied between disease observation dates (Fig. 2b), these associations were determined for respective DAI. Table 3 presents the markers significantly associated with CBB ratings for each DAI analyses. The *P* value determines whether a QTL is associated with the marker. The R^2 statistic is commonly used in QTL mapping studies to measure the proportion of phenotypic variation explained by molecular markers. However, unlike fixed linear regression models, linear mixed models have no well-established R^2 statistic for assessing goodness-of-fit and prediction power. The R^2 marker only measures the contribution of the marker to sum square after accounting for all other effects in the model. Eighty-eight percent (15 of 17) of markers were significant in at least one date. Of them, 7 and 15 markers were significantly associated with the 14 and 21 DAI CBB rating, respectively. Seven markers were significant for both dates, with high associations in markers SU91, SU91-CG9A, SU91-CG9B, SU91-CG10, and SU91-CG11 ($P \leq 0.001$) located at the SU91_QTL locus.

Table 3 Testing of association between marker loci and common bacterial blight severity using unified MLM (mixed linear model) method

Marker	Target gene	Forward (5′–3′)	Reverse (5′–3′)	Size (bp)	14 DAI		21 DAI	
					P	R^2_marker[a]	*P*	R^2_marker
BC420-QTL locus								
BC420		GCAGGGTTCGAAGACACACTGG	GCAGGGTTCGCCCAATAACG	896	NS	0.0051	NS	0.0014
BC420-CG3	3	GGACTTAGCGTACGGTTGGA	TGTGGTCGATGAGAACAAGG	558/540	NS	0.0081	*	0.0200
BC420-CG4	4	ACCATCCCTCTGCCTTTTCT	TCATCTTCTGATCGGCCTTT	590	NS	0.0077	**	0.0196
BC420-CG9	9	AAGCAAACCCTTCCATTCC	TCCCAAACACCAATGGAAAT	415/375	NS	0.0104	*	0.0184
BC420-CG10A	10	AAGGCTGCAAAGATTGGAGA	TTGATGAAGCCTTTGGAACC	486	NS	0.0049	*	0.0124
BC420-CG10B	10	CCACCTGCCACATAGACCTT	TCTCGAGAAGGGCAGAGGTA	459	*	0.0136	**	0.0256
BC420-CG11	11	GTGTCCATCTCTGGGTGCTT	GGATGCAAAGAAGAGGCAAA	227	NS	0.0077	**	0.0196
BC420-CG14	14	CGAGACTCGTGTGCTCTCTG	ACGAAGGTTGATTCCCAGTG	519/425	*	0.0180	**	0.0197
BC420-CG15	15	GATCCCAAGAAAATGGCAGA	CAAGTCGTGGGATTCTGTGA	486	NS	0.0046	*	0.0132
BC420-CG17	17	AGCCAGAATGTATCGAATTG	TATGCAACCAAAACCAAAGG	500/510	NS	0.0070	*	0.0177
SU91-QTL locus								
SU91		CCACATCGGTTAACATGAGT	CCACATCGGTGTCAACGTGA	628	***	0.1253	***	0.0953
SU91-CG3	3	GCAGAAGATGCCAAGAGGTC	CTCTATTCACCGCCAGCTTC	215	NS	0.0001	NS	0.0024
SU91-CG9A	9	AGCTGTTATTGGTCATTCATTTG	GATCTCCCCTTATCGTCTTCG	383	***	0.0532	***	0.0469
SU91-CG9B	9	CCCGAGTTAGAAGTAGGTGGAG	TGTTGAAAACAAACTATCGTGAG	501	***	0.0557	***	0.0475
SU91-CG10	10	ATGGTGGAGACGAGATGACC	TCCGACATTGAAACCAGTTG	425/350	***	0.2735	***	0.2554
SU91-CG11	11	GGCGACGGCTTCTTTGAC	TCCAAAGACCAAAGGGTGAG	464/425	***	0.2257	***	0.2193
SU91-CG12	12	ACGAAACACCATACCCCAAA	CGGTCAGCAGTTTCTTCCTC	179	NS	0.0011	NS	0.0000

NS not statistically significant

[a] R^2_marker was calculated as the proportion of sum square due to marker after accounting for all other effects in model

* $P \leq 0.05$

** $P \leq 0.01$

*** $P \leq 0.001$

In contrast, BC420 was not significant for both dates. This finding is consistent with a recessive epistatic model of inheritance between the two loci, i.e. the SU91_QTL locus is essential for CBB resistance and the greatest gain in resistance to CBB is to select breeding materials that are fixed for both QTL. Based on R^2_marker value, two CGMs per each locus show better association to CBB resistance than the BC420 and SU91, including BC420-CG10B and BC420-CG14 for BC420_QTL locus, and SU91-CG10 and SU91-CG11 for SU91_QTL locus.

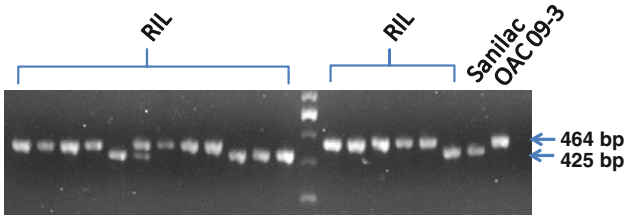

Fig. 3 Co-dominant marker SU91-CG11 was used to genotype parental lines Sanilac and OAC 09-3, and the recombinant inbred lines derived from the cross between OAC 09-3 and sanilac

The evaluation of SU91 CGMs by genetic mapping

The $F_{4:5}$ RIL population derived from the cross between Sanilac and OAC 09-3 showed continuous variations of CBB scores with population mean shifted towards susceptibility in both field and growth room phenotying (Table 4). Both Sanilac and OAC 09-3 do not have BC420 marker and BC420 CGMs. SU91 is a dominant marker present in OAC 09-3. SU91-CG10 is also a dominant marker, but present in Sanilac. SU91-CG11 is a co-dominant marker amplifying a 425 bp band for Sanilac and a 464 bp band for OAC 09-3 (Fig. 3). All three markers were used to genotype 90 $F_{4:5}$ RILs derived from the cross between Sanilac and OAC 09-3. As shown in Table 4, SU91 is significantly skewed towards resistant allele, SU91-CG11 is less deviated than SU91, while SU91-CG10 is normally segregating (Table 4). In both field and growth room phenotyping conditions, SU91-CG11 accounted for greater portion of phenotypic variation than SU91 (Table 4). Since SU91-CG11 is a co-dominant marker, it is able to distinguish homozygous QTL from heterozygous QTL in plants (Fig. 3). The phenotypic value of the heterozygous genotypes were between the two parental type homozygous genotypes in both field and growth room test (Supp. Table 1).

Discussion

Traditionally, BAC clones were sequenced by the Sanger-based, shotgun approach. The Roche 454 platform, with its longer reads (400–500 bp) provided a much higher throughput than Sanger sequencing at a lower cost per base (Varshney et al. 2009). Since reference bean genome was not available, our BAC assembly could not be completed solely using Newbler software. Analysis of 89,017 BAC-end sequences from G19833 in combination with 1,404 shotgun sequences from Bat7 revealed that ~49.2 % of the bean genome contains repetitive sequence (Schlueter et al. 2008). When compared to the overall sequence composition of other legume BAC-end sequencing analyses, common bean appears to have a much higher percentage of repetitive sequence than the 8.5 % identified for white clover or the estimated 33.5 % for soybean (Schlueter et al.

Table 4 Genotypic frequencies and single marker QTL analysis of SU91 and SU91 candidate gene markers in a recombinant inbred line population derived from the cross between OAC 09-3 and Sanilac

Marker	Allele			X^2	Note	Field AUDPC							Growth room						
	A	H	B			0[a]	1	2	3	4	5	R^{2c}	0	1	2	3	4	5	R^2
SU91	28	0	61	14.55^{***}	Dominant	6[b]	10	29	32	13	0	0.4^{***}	1	4	21	33	27	2	0.27^{***}
SU91-CG10	49	0	40	1.36^{NS}	Dominant, Repulsion to SU91	(2)	(2)	(23)	(50)	(13)	(0)	0.28^{***}							0.23^{***}
SU91-CG11	27	20	41	2.88^{*}	Co-dominant							0.42^{***}							0.36^{***}

A Sanilac allele, *H* heterozygous, *B* OAC 09-3 allele, *NS* not statistically significant

[a] CBB rating: 0 = no symptoms, 1 = less than 10 %, 2 = 11–30 %, 3 = 31–50 %, 4 = 51–80 %, and 5 = more than 80 % of inoculated areas showing symptoms

[b] CBB rating at 14 days after inoculation (DAI), whereas 21 DAI in parentheses

[c] R^2 is the proportion of phenotypic variance accounted for by the marker

* $P \leq 0.05$

** $P \leq 0.01$

*** $P \leq 0.001$

2008). As a result of this, the assembly of the HR45 BAC clones was challenged by the amount of repetitive sequences present at the bean genome. Kumar and Blaxter (2010) have carried out a systematic comparison of five assemblers (CAP3, MIRA, Newbler, SeqMan, and CLC) and concluded combining optimal assemblies from different programs would give a more credible final product. We explored combining three software, Newbler, CLC, and Geneious, and found 97.5 % and 99.9 % vector sequences recovered from BAC 32H6 and 4K7. This demonstrated that merged assemblies aligned better to reference sequences than individual assemblies. However, this approach could not be used to assemble all the BAC clones in this study, and additional analysis will be required to fully assemble those fragments.

Although the marker BC420 has been used extensively in the bean breeding programs for CBB resistance, its usage is limited by its additive effect on resistance. Most of the new CBB resistance materials in the AAFC/University of Guelph Bean Breeding Program have SU91 marker band but not BC420, for example, Rexeter (OAC 07-2), Apex (HR199), and ACUG 10-6. All of them have high levels of resistance. Moreover, its tight linkage with the *V* gene limits the usefulness of BC420 in coloured bean market classes (Liu et al. 2008). Previous studies have indicated that some CBB-resistant lines derived from XAN159 in Mexico had the SU91 marker band but not the BC420 marker (Liu et al. 2008). This may be because when they selected the favourable seed coat colour, the BC420 linked QTL was removed. These two major QTL from XAN159 have been transferred into different genetic backgrounds through traditional breeding. With the development of more markers tightly linked to the major QTL on B6, breeders will have more options to pyramid these two QTL in marker-assisted variety development. CGMs are superior to the random DNA markers SU91 and BC420 due to their potential complete linkage with target loci (Andersen and Lubberstedt 2003). The recruitment of polymorphic markers required for CGM development is fairly simple for *Arabidopsis* and rice because the whole genome has been sequenced. Since no detailed sequence information is available for common bean, the sequences from BACs containing QTL markers were explored to develop CGMs. Two CGMs per each locus showed better association to CBB resistance than BC420 and SU91, including BC420-CG10B and BC420-CG14 for the BC420_QTL locus, and SU91-CG10 and SU91-CG11 for the SU91_QTL locus (Table 3). Somehow, SU91-GC10 marker appears to be a co-dominant marker in the NILs shown in Fig. 1, while it is a dominant marker in the RIL population, thus its usefulness may depend on the population being assessed. Overall, BC420-CG14 and SU91-CG11 are co-dominant markers (Figs. 1, 3), and as a result, are more informative than BC420 and SU91 for MAS. However, genetic ratios for SU91 and SU91-CG11 are skewed to the resistant parent (Table 4). Skewed segregation ratios of genetic loci are often observed in the progeny of inter- and intra-specific hybrids and the phenomenon are resulted from distorted inheritance of alleles caused by competition among gametes or from abortion of gamete or zygotes for preferential fertilization (Lyttle, 1991). Since the origin of OAC 09-2 resistance was also from tepary bean via inter-specific hybridization, segregation distortion of SU91 and SU91-CG11 could be largely explained by the result of inter-specific hybridization. Furthermore, the high proportion (23 %, 22 of 88) of heterozygotes was detected by the co-dominant SU91-CG11 marker in the RIL population (Table 4), thus SU91-linked locus appears to be a highly heterozygous locus. Recent genomic research showed the residual heterozygosity occurred in segmental "blocks", thus the heterozygous locus under selection is assumed to be embedded in a larger structural variant that prohibits homologous recombination (http://cphg.virginia.edu/mackey/2010/05/14/residual-heterozygosity-vs-segmental-duplications/). Since linkage drag was detected at BC420-linked locus (Liu et al., 2008), we could assume the size of the introgression segment containing SU91-linked locus may also be much larger than the target region. Thus, excess heterozygosity at SU91-linked locus could result from the selection of a larger segment that prohibited homologous recombination.

Previously, six amplified fragment length polymorphism (AFLP) markers tightly linked to BC420-QTL were screened using bulk segregation analysis (BSA) (Liu et al. 2008). For BSA, the individuals that are pooled are generally derived directly from a population that has already been used to map (or monitor) segregation of a target trait. BSA can be conducted as soon as that information is available and BSA has a high degree of versatility in making genotypically contrasting pools for disparate regions across the genome by using different individual lines from the same population. The resolution of BSA can be varied by increasing or decreasing the number of individuals included in each pool of samples, or by narrowing the genetic interval when marker data is available. However, BSA can also be inherently biased if the amounts of DNA are highly skewed towards one or a few individuals in each pool. One result of an imbalance in pool membership could be elucidated as either high background outside of the target region (which may lead to the suggestion that the target region contains trans-acting factors that influence gene expression at regions across the genome), or result in genes from outside the target interval being falsely included in it. Thus, after converting six AFLP markers into seven sequence tagged site (STS) markers, only three STS markers were mapped to BC420-QTL (Liu et al. 2008).

Alternatively, we used a pair of QTL-NIL to screen CGMs. Since NILs are genetically cleaner than BSA, it avoids some of the BSA's issues. The greatest concern is that the analyses are limited by availability, taking many generations of backcrossing and selection followed by selfing to generate the populations. Moreover, if flanking genetic markers have not been used in their development, then the size of the introgression containing the target region is often large, and as a result, resolution suffers. Because CBB-NILs were developed solely guided by single QTL marker per locus, polymorphic CGMs were further evaluated in large population with 392 lines.

Since common bean is recalcitrant to transformation, association studies have been used to indirectly approve gene function (Andersen and Lubberstedt 2003). The aim of association study is to identify genes and even functional motifs within genes that affect phenotypic characteristics. CGMs developed from genes 10 and 14 in BAC 4K7 are better associated with CBB resistance than BC420 (Table 3), and CGMs developed from genes 10 and 11 in BAC SU91 are better than SU91. Thus, genes 10 and 14 at the BC420 locus are CGs for the BC420_QTL locus, whereas genes 10 and 11 in BAC 32H6 are CGs for the SU91_QTL locus. Moreover, both candidate genes for the SU91_QTL locus were supported by the presence of tepary bean ESTs (Table 2).

The NILs and resistant controls used in this study have obtained their resistance through crosses to XAN159, which was developed through inter-specific crosses to tepary bean introduction line PI319443 (Miklas et al. 2006). Together with another tepary bean introduction line PI440795, they provide the major sources of CBB resistance to the Harrow, Ontario breeding population (Shi et al. 2011). Because genetic sequence of genes 10 and 11 at the SU91 locus are more similar to tepary bean ESTs than common bean, which indicate these regions as possible sites of interspecific introgression of tepary DNA. In addition, both genes 10 and 11 at the SU91 locus are the homologue of the gene encoding UDP-glycosyltransferase 82A1-like protein in soybean (Table 2). UDP-glycosyltransferase (UGT) was found to be a key protein in the plant that creates drug resistance (Zhao et al. 2007). Alternatively, bean gene function could be directly tested by VIGS (virus-induced gene silencing) (Waterhouse and Helliwell 2003). BPMV (bean pod mottle virus) was used as a vector to over-express or silence genes in soybean (Zhang et al. 2010). Because BPMV is also known to infect common bean, it has been tested as a vector in black bean cultivar Black Valentine. GFP genes carried by the BPMV vector were expressed in shoots and roots of Black Valentine when the plants were infected with the single gene expression construct pBPMV-GFP2 or the double gene expression construct pBPMV-GFP-BAR (Zhang et al. 2010). Validating the functions of the CBB candidate genes with VIGS is underway.

Acknowledgments The authors acknowledge the financial supports from the Ontario Bean Produces' Marketing Board, Ontario Coloured Bean Growers' Association, Agriculture and Agri-Food Canada, Ontario Research Fund, and the technical assistance provided by Terry Rupert, Barbara Harwood, Bailing Zhang, and Samantha Krueger.

References

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Kumar S, Blaxter ML (2010) Comparing de novo assemblers for 454 transcriptome data. BMC Genomics 11:571

Liu S, Yu K, Park SJ (2008) Development of STS markers and QTL validation for common bacterial blight resistance in common bean. Plant Breed 127:62–68, 108

Liu SY, Yu K, Haffner M, Park SJ, Banik M, Pauls KP, Crosby W (2010) Construction of a BAC library and a physical map of a major QTL for CBB resistance of common bean (*Phaseolus vulgaris* L.). Genetica 138:709–716

Lyttle TW (1991) Segregation distortion. Annu Rev Genet 25:511–557

McClean PE, Gepts P, Scott JA (2008) *Phaseolus vulgaris*: a diploid model for soybean. In: Stacey G (ed) Genetics and genomics of soybean. Springer, Berlin, pp 55–76

Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica 147:105–131, 279

Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics 95:315–327

Ouyang S, Thibaud-Nissen F, Childs KL, Zhu W, Buell CR (2009) Plant genome annotation methods. In: Somers DJ et al (eds) Methods in molecular biology, plant genomics, vol 513. Humana, Berlin, pp 263–280

Park SJ, Dhanvantari BN (1987) Transfer of common blight resistance from *Phaseolus coccineus* Lam. to *P. vulgaris* L. through interspecific hybridization. Can J Plant Sci 67:685–695

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304, 256

Schlueter JA, Goicoechea JL, Collura K, Gill N, Lin JY, Yu Y, Kudrna D, Zuccolo A, Vallejos CE, Muñoz-Torres M, Blair MW, Tohme J, Tomkins J, McClean P, Wing RA, Jackson SA (2008) BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. Trop Plant Biol 1:40–48

Seo YS, Rojas MR, Lee JY, Lee SW, Jeon JS, Ronald P, Lucas WJ, Gilbertson RL (2006) A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus-specific manner. Proc Natl Acad Sci USA 103:11856–11861

Shi C, Navabi A, Yu K (2011) Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. BMC Plant Biol 11:52

Vandemark GJ, Fourie D, Miklas PN (2008) Genotyping with real-time PCR reveals recessive epistasis between independent QTL conferring resistance to common bacterial blight in dry bean. Theor Appl Genet 117:513–522, 178

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Waterhouse PM, Helliwell CA (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4:29–38

Yu K, Park SJ, Poysa V (2000) Marker-assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. Plant Breed 119:411–415, 484

Yu KF, Park SJ, Zhang BL, Haffner M, Poysa V (2004) An SSR marker in the nitrate reductase gene of common bean is tightly linked to a major gene conferring resistance to common bacterial blight. Euphytica 138:89–95, 104

Zhang C, Bradshaw JD, Whitham SA, Hill JH (2010) The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. Plant Physiol 153:52–65

Zhao Y, Chow TF, Puckrin RS, Alfred SE, Korir AK, Larive CK, Cutler SR (2007) Chemical genetic interrogation of natural variation uncovers a molecule that is glycoactivated. Nat Chem Biol 3:716–721