Molecular characterization of the *CRa* gene conferring clubroot resistance in *Brassica rapa*

Abstract

Clubroot disease is one of the major diseases affecting Brassicaceae crops. For example, Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) is known to be highly susceptible to clubroot disease. For protection from this disease, genes for resistance to clubroot were introduced from the European turnip. *CRa* is a gene that confers specific resistance to the clubroot pathogen *Plasmodiophora brassicae* isolate M85. Fine mapping of the *CRa* locus using synteny to the *Arabidopsis thaliana* genome and partial genome sequences of *B. rapa* revealed a candidate gene encoding a TIR-NBS-LRR protein. There were several structural differences in this candidate gene between susceptible and resistant lines, and *CRa* expression was observed only in the resistant line. Additionally, four mutant lines lacking clubroot resistance were obtained by the UV irradiation of pollen from a resistant line. All of these mutant lines carried independent mutations in the candidate TIR-NBS-LRR gene. This genetic and molecular evidence strongly suggests that the revealed gene is *CRa*. This is the first

report, on the identification of a clubroot Resistance gene in Brassicaceae, and on the identification of the disease resistance gene in *B. rapa*.

Keywords

Clubroot, resistance gene, *CRa*, Chinese cabbage, Brassicaceae, UV mutation

Introduction

Resistance to pathogens in higher plants has been attributed to two different mechanisms: pathogen race-specific resistance and basal resistance (Kou and Wang 2010; Bernoux et al. 2011). Both types of resistance are important in the protection of plants against pathogens. Molecular aspects of pathogen race-specific resistance have been extensively studied, and this type of resistance is known to be mediated by resistance (R) genes through the recognition of specific pathogen effectors. The R gene-effector interaction is observed between various pathogens and plants, for example for rust disease in flax (Ayliffe et al. 1999; Schmidt et al. 2007), blast disease in rice

(Okuyama et al. 2011), late blight in potato and tomato (Ballvora et al. 2002; Jia et al. 2010), and bacterial speck in tomato and Arabidopsis (Zhang and Gassmann 2007). These effectors initially functioned as virulence/pathogenicity factors for the pathogen, but then became recognized by the hosts as signals of attempted infection (Bernoux et al. 2011).

Most R genes encode proteins carrying nucleotide-binding site (NBS) in the central region and a leucine-rich repeat (NBS-LRR) domain at the C-terminus. These genes form a large family in the plant genome and can be separated into two subclasses, the Toll-Interleukin-1 (TIR) class and the coiled-coil (CC) class (Rafiqi et al. 2009). The NBS-LRR proteins are thought to recognize effectors or effector-specific signals from the pathogen and activate the plant immune system to provide hypersensitive responses (Bernoux et al. 2011).

Clubroot is a critical disease for various Brassicaceae crops and causes significant damage to the production of this crop (Diederichs et al. 2009). Clubroot is caused by the protist *Plasmodiophora brassicae*, which produces spores that are stable for several years in soil (Kuginuki et al. 1999). Without eradication of these spores in the soil, the cultivation of susceptible crops is obstructed (Howard et al. 2010). Therefore, development of clubroot-resistant cultivars has been desired for Brassica crops, and the

identification of R genes for clubroot is necessary.

The Brassicaceae family includes several important vegetable crops, such as *Brassica rapa* (Chinese cabbage, pak choi and turnip), *Brassica oleracea* (cabbage and broccoli) and the oilseed crop *Brassica napus* (rapeseed and field mustard). The genome of the *Brassica* species is thought to be composed of three different genomes: genome A (*B. rapa*), genome B (*B. nigra*) and genome C (*B. oleacea*) (Schranz et al. 2006). Each *Brassica* species is a diploid or amphidiploid composite of these three genomes, and some of the species can be naturally crossed, which indicates genetic relationships among these crops.

Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) is highly susceptible to clubroot disease. In contrast, European turnips (*B. rapa* L. ssp. *rapifera*) are known to carry strong resistance to clubroot (Matsumoto et al. 1998). This resistance is suggested to be controlled by several R genes. Several commercial clubroot-resistant (CR) Chinese cabbage cultivars have been developed in Japan by crossing with resistant turnips (Matsumoto et al. 2005). The early CR cultivars showed partial resistance to clubroot disease, possibly because of the introduction of only one resistance gene (Kuginuki et al. 1999). Several CR loci (CR genes) were found in Chinese cabbage cultivars, although none of the genes responsible for these loci have been isolated.

CRa was the first genetically identified CR locus, and a doubled haploid line T136-8 carrying *CRa* showed strong, dominant resistance to the M85 isolate of *Plasmodiophora brassicae* (Matsumoto et al. 1998). This locus was originally derived from the ECD02 turnip line of the European Clubroot Differential (ECD) series and introduced into Chinese cabbage. For genetic mapping of *CRa*, the restriction fragment length polymorphism (RFLP) marker HC352b was identified as closely linked to the *CRa* locus, and HC352b was converted into a sequence-characterized amplified region (SCAR) marker for practical application to *CRa* breeding (Hayashida et al. 2008).

In the present study, we report the isolation and molecular characterization of the *CRa* gene.

Results

Fine mapping of the *CRa* locus

A previously developed RFLP marker for *CRa* mapping, HC181, was converted to a SCAR marker (HC181-SCAR) by the technique described in the previous report

(Hayashida et al, 2008). The genetic distances of HC352b-SCAR (Hayashida et al., 2008) and HC181-SCAR were estimated to be 2 cM and 4 cM from the *CRa* gene, respectively (Fig. 1a). Because the sequences of both SCAR markers were highly similar to that of the *Arabidopsis thaliana* genes on chromosome 4 (AT4G19210 and AT4G29750), it was expected that the structure in this region would be conserved between *B. rapa* and *A. thaliana*. Thus, additional *CRa*-linked markers were developed based on the genomic sequence of the corresponding region in *A. thaliana*. Additionally, because partial sequence information was available from the Bacterial Artificial Chromosome (BAC) clones of the *B. rapa* line chiifu-401-42 genome, these data were also employed as a resource for mapping. For the fine mapping of *CRa*, 80 F_2 progeny, derived from a cross between resistant T136-8 and susceptible Q5 lines, were selected. These progeny represent all recombinations occurred between HC352b-SCAR and SC2930 in 1,622 F_2 individuals, selected using the fast isolation of recombinant (FIRE) strategy. Based on the genotyping of these materials, 11 SCAR markers were established around *CRa* (Fig. 1a).

The GC2360 marker, which was one of closest markers to *CRa*, was estimated to be positioned 150 kb from *CRa* (assuming that 1 cM is equivalent to 450 kb). Considering the predicted distance between *CRa* and GC2360 and their location in the genome, the

gene responsible for *CRa* was presumed to be located in the end region of BAC clone KBrH102F05 (Fig. 1b). Three predicted genes (ORF1, ORF2 and ORF3), were found in the corresponding region (Fig. 1c). To define the position of *CRa* among these ORFs, four additional DNA markers were designed for the region adjacent to each predicted gene and applied to genotyping in F_2 plants (Fig. 1c, d). Consequently, no recombination between the marker adjacent to the ORF1 and the *CR*a phenotype was found in all of the analyzed F_2 plants (Fig. 1d). The ORF1 in the KBrH102F05 BAC clone encoded a putative TIR-NBS-LRR protein.

Analysis of the *CRa* candidate gene

The KBrH102F05 BAC clone was derived from the genome of chiifu-401-42, which is susceptible to clubroot and may have susceptible gene, cra. Therefore, the structures of the ORF1 in the *CRa*-carrying line T136-8 and clubroot-susceptible line Q5 were determined by analyzing the PCR amplified products obtained using specific primers (Fig. 1c). A single 7.6-kb fragment was amplified from the genomic DNA of T136-8 using PCR with the FW1 and RV2 primer sets (Fig. 2a). From Q5, 7.5 kb and 6.6 kb products were obtained by PCR using primer sets FW1-RV1 and FW2-RV2,

respectively (Fig. 2b). The sequence of the product from T136-8 showed a high similarity to that of the ORF1 in KBrH102F05 (Supplementary Fig. 1), with the exception of a large deletion in the intron region and an extension in the region coding the LRR domain (Supplementary Fig. 2), whereas the PCR products from Q5 showed sequences that were identical to the corresponding regions of the Chiifu-401-42 genome.

To confirm that the amplified fragment from T136-8 was not derived from the paralogous NBS-LRR genes, co-dominant DNA marker CRaim was developed, on the ORF1. In all of the analyzed F_2 plants from T136-8 and Q5, the existence of T136-8 type ORF1 was perfectly correlated with the *CRa* phenotype (Fig. 1d).

Transcript analysis of the *CRa* candidate gene

To confirm the expression of the ORF1, RT-PCR was performed for RNA extracted from leaf tissues using primer sets that amplify the TIR and NBS domains as described (Fig. 2c). For T136-8, an amplified product of the expected size was observed (Fig. 2c), whereas a fragment was not amplified in Q5. Sequencing analysis of the obtained RT-PCR product, as well as PCR fragments from 3’- or 5’- rapid amplification of cDNA

ends (RACE) from T136-8 RNA, demonstrated that the cDNA of ORF1 had a size of 4223 bp and comprised 4 exons (Fig. 3). The size of the encoded protein was predicted to be 1325 amino acids (Fig. 4). The predicted amino acid sequences of the ORF1 alleles were compared between the resistant and susceptible lines. The TIR and NBS domains were partially conserved (79.5% identical), although many In/Dels were present in the LRR domain (Fig. 4).

Additionally, putative variants of the transcripts were also found in the RT-PCR products (Fig. 5). The longest transcript variant involved partial intron sequences between the 2nd and 3rd exons, and consequently, its product protein could be truncated. A shorter transcript, which lacked a large region encoding the NBS and LRR domain, was also observed. These transcripts suggest the possibility of alternative splicing of RNA in ORF1.

Analysis of *CRa* mutants

As an alternative approach to demonstrate that ORF1 is the gene corresponding to *CRa*, several genetic mutants from T136-8 lacking clubroot resistance were produced. For mutagenesis, pollen samples from T136-8 flowers were exposed to UV light and

pollinated to Q5 flowers. The obtained seeds (MF_1) were sown, and these plants were inoculated with the *P. brassicase* M85 isolate. Symptoms in the infected plants were evaluated 1 month after the inoculation. Of the 1078 MF_1 plants that were tested, 5 were found to have galled roots indicating to be deficient in *CRa* resistance (susceptible). Each susceptible MF_1 plant was self-fertilized, and offspring plants (MF_1S_1) were subjected to the genotyping to select T136-8 derived cra homozygotes by DNA markers GC1680 and GC3060, and resistance analysis to confirm susceptibility. Consequently, four lines, designated *cra1*, *cra3*, *cra4* and *cra5*, were established as lines that are susceptible to clubroot but carry *CRa* region derived from T136-8. PCR analysis of the ORF1 resulted in the amplification of 7.6 kb fragments from *cra1*, *cra3*, *cra4* and CR controls (MF_1 plants showing resistance), whereas a 9.5 kbp fragment was amplified from *cra5* (Fig. 6a). Sequencing of these amplified products indicated the presence of independent mutations among these mutant lines, whereas no mutations in the ORF1 were found in the CR controls. In *cra1* and *cra4*, nucleotide substitutions with amino-acid changes were observed in the sequence of the NBS domain (Supplementary Figs. 3, 4). In *cra3*, a displacement of 21 amino-acids occurred in the LRR domain, and an insertion of 12 amino-acids were observed in the LRR domain (Supplementary Fig. 5). A large insertion (approximately 2 kb) was observed in *cra5* (Fig. 3, Supplementary

Fig. 6).

To confirm the genetic linkage of mutations to the susceptible phenotype, the *cra*3 and *cra*5 lines were back-crossed with T136-8, and their BC_1S_1 plants were used in genotyping of the ORF1 and resistance analysis with the M85 isolate. In both BC_1S_1 populations, perfect correlations were observed between the presence of mutations in the ORF1 and disease susceptibility (Table 1). These results strongly suggest that the ORF1 is the gene corresponding to *CRa*.

Distribution of *CRa* in clubroot-resistant *B. rapa*

The distribution of the *CRa* gene in cultivars of *B. rapa* was investigated based on PCR amplification and a pathogen-infection assay. The clubroot-resistant European turnips Debra and Gelria R, as well as two commercially available clubroot-resistant Chinese cabbage cultivars, were assessed. All four of these cultivars showed strong resistance to the *P. brassicae* M85 isolate, and amplification of the 7.6 kb fragment of the *CRa* gene was also observed as shown in T136-8 (Fig. 6b), which indicates that they are carrying the *CRa* gene.

Discussion

CRa encodes a TIR-NBS-LRR protein

We have identified the gene responsible for clubroot resistance, *CRa*, in *B. rapa* based on genetic mapping, sequencing, expression analysis and mutation analysis. The candidate gene for *CRa* was initially identified as the ORF1 with fine mapping based on the synteny between *B. rapa* and *A. thaliana* genomes and the partial genome sequences of *B. rapa*. ORF1 encodes a TIR-NBS-LRR protein, which is a typical product of the plant disease resistance gene (R-gene). Generally, genes encoding TIR-NBS-LRR proteins are known to be responsible for race-specific resistance in higher plants (Bernoux et al. 2011). This candidate *CRa* gene was observed in the genomes of both resistant and susceptible lines, although several structural differences in this gene were observed between these lines. Obvious differences were identified in the DNA sequences of the LRR domain that caused amino acid changes and likely resulted in functional alternations. In the T136-8 plant, an intact transcript was detected using RT-PCR. However, transcripts were not detected in the susceptible line Q5, most likely

because of a large insertion in the allele that was highly similar to a retrotransposon. This apparent presence or absence of transcripts of the putative responsible genes was suggested to determine the resistance or susceptibility to the M85 isolate. In addition, mutant analysis was used to demonstrate that the loss of *CRa* resistance was caused by non-synonymous mutations in the putative *CRa* (ORF1) gene. The complementation experiment was difficult, because of the low efficiency of genetic transformation in *B. rapa* (data not shown).

Structure and putative function of the *CRa* gene

PCR analysis of the *CRa* gene in several *B. rapa* cultivars showed that the *CRa* allele in commercial resistant cultivars originated from European turnips. A large intron was present in the susceptible allele (*cra*) in Q5 and Chiifu-401-42, and a retoroposon-like sequence was observed in this large intron, which suggests that this allele was inactive. Further studies of the distribution and structural variation of the *CRa* locus in *B. rapa* varieties will be helpful for elucidating the evolution of the *CRa* gene.

The genetics of clubroot resistance in Chinese cabbage have been extensively studied among crucifer crops. In addition to *CRa*, seven CR loci are known in *B. rapa*,

including *Crr1*, *Crr2* (Suwabe et al. 2003), *Crr3* (Saito et al. 2006), *Crr4* (Suwabe et al. 2006), *CRb* (Piao et al. 2004), *CRc* and *CRk* (Sakamoto et al. 2008). All of these loci have been shown to be responsible for race-dependent resistance (Piao et al. 2009). Among these loci, *CRa* was the first CR gene to be isolated. Because *CRa* encodes a TIR-NBS-LRR protein, as shown in the present study, other CR genes in Chinese cabbage are also likely to encode TIR-NBS-LRR proteins. Although the evolution or divergence of CR genes in *B. rapa* was previously unknown, resistance genes (R-genes) often show similar structures because they were most likely derived from a common ancestral gene (Leister 2004). It is known that hundreds of NBS-LRR genes are present in the plant genome (Mun et al. 2009), and it is possible that defined CR genes could be identified by employing the strategy used in this study (i.e., mapping and focused candidate gene analysis).

After each CR gene is isolated, multiple CR loci can rapidly accumulate in a cultivar (i.e. pyramiding), thereby conferring strong resistance to a wide range of *P. brassicae* isolates. Furthermore, it is expected that pyramiding of CR genes may be not only additive, but also synergetic (Matsumoto et al. 2012).

Some insights were obtained in the mechanisms of recognition *P. brassicae* conferred by the *CRa* gene product. According to the mutational analysis, each TIR,

NBS and LRR domain was essential for the function of the *CRa* gene product because mutations in each domain caused a loss of *CRa*-mediated resistance. Furthermore, a few transcript variants encoding truncated proteins were found in T136-8 in addition to the full-length encoding transcript. It has been reported that several TIR-NBS-LRR genes also produce alternatively spliced transcripts (Ayliffe et al. 1999; Dinesh-Kumar et al. 2000; Takabatake et al. 2006; Schmidt et al. 2007; Zhang and Gassmann 2007), and products from certain alternative transcripts that carry the TIR and NBS domains or only the TIR domain have been shown to be functional (Marathe et al. 2002; Frost et al. 2004; DeYoung et al. 2006; Gassmann 2008). These observations will be helpful for identifying the molecular and biological functions of *CRa* gene products.

An advanced mutation method using UV-irradiated pollen

In the present study, we succeeded in isolating four independent mutant lines lacking *CRa* resistance from among 1,078 MF_1 plants derived from the pollination of mutagenized pollen from resistant lines to flowers of a susceptible line. The calculated frequency of independent mutations per gene in this method was 0.5%, which is higher than the typical rate obtained using the mutagenizing reagent EMS (approximately

0.06%) (Nakanishi et al. 2004). This high mutation rate probably resulted from the mutagenic treatment of the haploid cells. Various types of mutations in the genomic sequences of the analyzed lines were observed as single base substitutions, small-size In/Dels and large-fragment insertions such as transposable elements. Therefore, it is expected that mutants showing various phenotypes, including a loss-of-function or leaky phenotype, could be obtained. This method could be applicable to any plant, and has several other advantages, including the ease of use and a lack of exposure to hazardous reagents (Nakanishi et al. 2004) or radiation (Naito et al. 2005). Additionally, the mutagenized population can also be easily scaled up in this approach. This method is useful in the functional analysis of defined genes in higher plants with less transformation efficiency.

Materials and methods

Plant materials

Two *B. rapa* doubled haploid lines (T136-8 and Q5; Matsumoto et al. 1998), two

Chinese cabbage cultivars, (Ryutoku and CR Shinki), and two fodder turnips (Debra and Gelria R), were employed as plant materials in this study. T136-8, Ryutoku, CR Shinki and fodder turnips are resistant to the *P. brassicae* M85 clubroot isolate, and Q5 is susceptible. The plants were grown in glass house at Nagano prefecture (Japan). Fine mapping was conducted using 80 F_2 populations derived from T136-8 and Q5, which were selected from 1,622 F_2 individuals using the FIRE strategy (Horiguchi et al. 2001).

FIRE strategy

The FIRE strategy enables the selection of F_2 plants carrying recombinations in the objective genomic region based on genotypes of two closely linked co-dominant DNA markers. For selection, 1,622 F_2 plants derived from the T136-8 × Q5 cross were applied for genotyping using the HC352b and SC2930 DNA markers. Then, 80 F_2 individuals were defined as recombinants. The *CRa* genotypes of the 80 recombinants were determined as described for the clubroot resistance analysis.

Mutagenesis of *B. rapa* pollen

Flowers of T136-8 plants were exposed to UV light for 20 minutes at an intensity of

1,275 μW/cm^2 (total quantity, 1,530 mW/cm^2). After exposure, pollen samples from treated flowers were immediately pollinated to pistils of Q5 flowers. After the fertilization, seeds were harvested and germinated. The seedlings (MF_1 generation) were inoculated with the *P. brassicae* M85 isolate. Individuals showing the susceptible phenotype were screened as mutant candidates and these plants were self-fertilized. To select homozygotes of the mutated *CRa* region derived from T136-8, MF_1S_1 plants were analyzed using the DNA markers GC1680 and GC3060. The selected offspring plants were subjected to resistance analysis, and candidate lines showing the susceptible phenotype were defined as *CRa* mutants. The *cra3* and *cra5* were crossed with T136-8 to confirm the segrigation of their genotype and phenotype in their offspring. Additionally MF_1 plants which were resistant to clubroot were analyzed together defined as CR controls.

Clubroot resistance analysis

Resistance of *B. rapa* plants to *P. brassicae* isolate M85 was evaluated using the method described by Matsumoto et al. (2005). The M85 isolate is a field isolate identified as race 2 according to Williams classification (Williams 1966). Spore

suspensions of the M85 isolate, collected from galled roots of Chinese cabbage ‘Sakumidori No. 3’, were mixed in pasteurized organic soil. Seeds were sown in the contaminated soil, and the extent of clubroot resistance was evaluated 30 days after inoculation. The *CRa* genotype in the selected 80 F_2 recombinants was determined by the segregation of resistance in the F_3 individuals (11-24 plants each).

DNA and RNA extraction

DNA extraction from the parental lines was performed using a CTAB method as previously described (Matsumoto et al. 1998). Genomic DNA of F_2 individuals was extracted using the method described by Edwards et al. (1991) with an additional phenol-chloroform treatment. RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN, http://www.qiagen.com/) from young plant leaves according to the manufacturer’s instructions.

DNA marker development

All of the DNA markers in the present study were developed as SCAR markers. Primers for amplifying genomic regions described as “informative sequences” in Supplementary Table 1 were designed in this study (*Arabidopsis* Genome Initiative 2000; Choi et al. 2007; Mun et al. 2008). PCR products from T136-8 and Q5 were sequenced to confirm polymorphisms and were designated as PCR-based DNA markers.

PCR Amplification

PCR amplification of DNA markers was performed using ExTaq (TAKARA, http://www.takara-bio.com/index.htm) according to the manufacturer’s instructions (Nakanishi et al. 2005). The reaction conditions were as follows: 35 cycles of 94°C for 30 sec, annealing for 30 sec and 68°C for 2 min with an initial step of 94°C for 2 min. The annealing temperature was customized for each DNA marker as indicated in Supplementary Table 1.

The amplification of the *CRa* region was performed using KOD FX polymerase (TOYOBO, http://www.toyobo-global.com/) according to the manufacturer’s instructions. The reaction conditions were as follows: 35 cycles of 98°C for 10 sec and 68°C 10 min with an initial step of 94°C for 2 min.

Calculation of genetic distance

The genetic distances between loci calculated as recombination values were converted into centi-Morgans (cM) using the Kosambi function (Kosambi 1944). The physical distance between loci was roughly estimated using the relationship 1 cM = 450 kb (Johnston et al. 2005; Choi et al. 2007).

Sequencing of PCR products

The amplified PCR products were extracted from gels using the High Pure PCR Product Purification Kit (ROCHE Applied Science, https://www.roche-applied-science.com/). Gel-extracted PCR products were used for sequencing reactions, which were conducted using the BigDye Terminator v.3.1 kit (Life Technologies, http://www.lifetechnologies.com/) with the primers indicated in Supplementary Table 1. The products were analyzed using an ABI PRISM 3100 Genetic Analyzer (Life Technologies) as previously described (Taguchi et al. 2010).

Transcript analysis

For RT-PCR, 2 μg of total RNA was treated with DNaseI and reverse transcribed using an oligo-dT primer. The synthesized single-stranded cDNA was diluted five-fold with TE buffer and used as the template for PCR amplifications. The following primer sets were used in this study: FW5-RV3 for the TIR region, FW5-RV4 for the TIR-NBS region of T136-8 and FW5-RV6 for the TIR-NBS region of Q5. Sequences of the RT-PCR amplified products were analyzed using the same procedure as described above.

Full-length cDNA was isolated using the GeneRacer KIT (Invitrogen) according to the manufacturer's instructions (Matsumura et al. 2006).

Figure Legends

Fig. 1 Genetic and physical maps of the regions surrounding the *CRa* locus in *Brassica rapa.*

a: Genetic map of the *CRa* region. DNA markers and their relative positions around

CRa are illustrated. DNA markers designated with the letters 'SC' were designed based on synteny between the *Arabidopsis thaliana* and *B. rapa* genomes, and the markers starting with 'GC' were designed based on BAC sequence information.

b: The relationship of *B. rapa* BAC clones KBrH102F05 and KBrB078F08, whose DNA sequence was used to develop GC2360.

c: Predicted genes located at the last 40 kb of KBrH102F05. Predicted exons and introns are shown by squares and lines, respectively. Predicted genes are denoted tentatively as ORF1, ORF2 and ORF3.

d: *CRa* genotypes predicted from resistance analysis of the selected recombinants and the DNA marker genotypes are shown. DNA markers in the table were developed based on the sequence information of KBrH102F05.

For *CRa* genotype A: resistant genotype, B: susceptible genotype and H: heterotype.

For most DNA marker A: T136-8 genotype, B: Q5 genotype and H: heterotype.

For DNA marker GC30 + : resistant genotype (homo or hetero), and – : susceptible genotype.

Fig. 2 Genomic and expression analysis of ORF1 in *Brassica rapa*

a: PCR product from the clubroot-resistant Chinese cabbage line T136-8 obtained using the primer set 1: FW1- RV2.

b: PCR product from the clubroot-susceptible Chinese cabbage line Q5 obtained using two primer sets, 1: FW1-RV1 and 2: FW2-RV2.

c: Expression of the ORF1 demonstrated by RT-PCR using the primer set targeting the TIR region（T136-8: FW5-RV3, Q5: FW5-RV3 and TIR-NBS region (T136-8: FW5-RV4, Q5: FW5-RV6). Total RNA from each plant tissue was treated with DNaseI prior to reverse transcription; actin cDNA was also amplified as a control using an *A. thaliana* ACT2 (β-actin) primer set.

Fig. 3 Gene structure of ORF1 in *Brassica rapa*.

The upper panel indicates the ORF1 structure in the resistant line T136-8 based on the full-length cDNA. The lower panel indicates the predicted gene structure of the ORF1 in the susceptible lines Q5 and Chiifu-401-42.

The gray boxes and lines indicate UTR and introns, respectively, and the white boxed regions indicate the predicted coding sequences (CDS). The positions of primers used for full-length cDNA amplification are indicated in both illustrations.

Mutational positions in *cra* mutants are also indicated in the upper panel (T136-8). TIR: Toll-Interleukin-1 domain; NBS: nucleotide-binding site; LRR: leucine-rich repeat.

Fig. 4 Comparison of the amino-acid sequences of ORF1 products of T136-8 and Q5.

Amino acid sequences were predicted from the ORF1 sequences of T136-8 and Q5 (Fig. 3) and compared using ClustalW. Identical positions are shown in black boxes.

Single solid line: Toll-Interleukin-1 domain (TIR); dotted line: nucleotide-binding site (NBS); double solid line: leucine-rich repeat (LRR).

Fig. 5 Variety of ORF1 transcripts in *Brassica rapa*.

Three transcript variants are illustrated by boxes indicating exons. Each domain structure is indicated. The arrow in the second variant shows the position of the termination codon in the transcript. TIR: Toll-Interleukin-1 domain; NBS: nucleotide-binding site; LRR: leucine-rich repeat.

Fig. 6 PCR products of the ORF1 in mutants and *Brassica rapa* cultivars.

a: Clubroot-infected roots of the susceptible mutants and CR controls (resistant line). PCR (FW1-RV2) products of the ORF1 are shown for each line and T136-8.

b: PCR (FW1-RV2) products amplified from Debra and Gelria R (CR European turnip cultivars) as well as CR Shinki, Ryutoku and T136-8 (CR Chinese cabbage cultivars).

Table 1. Correlation of the clubroot phenotype and the ORF1genotype in the progeny of mutants

F_2 parent	Clubroot phenotype		ORF1 genotype*	
	Resistant	Susceptible	wild type homo/hetero	Mutant homo
cra3×T136-8	15		15	0
		11	0	11
cra5×T136-8	7		7	0
		3	0	3

* ORF1 was detected by Nested PCR using FW1-RV2 as first and FW1-RV5 as second primer set.

In both cases, only the wild type ORF1 genotype was detected but not mutated allele.

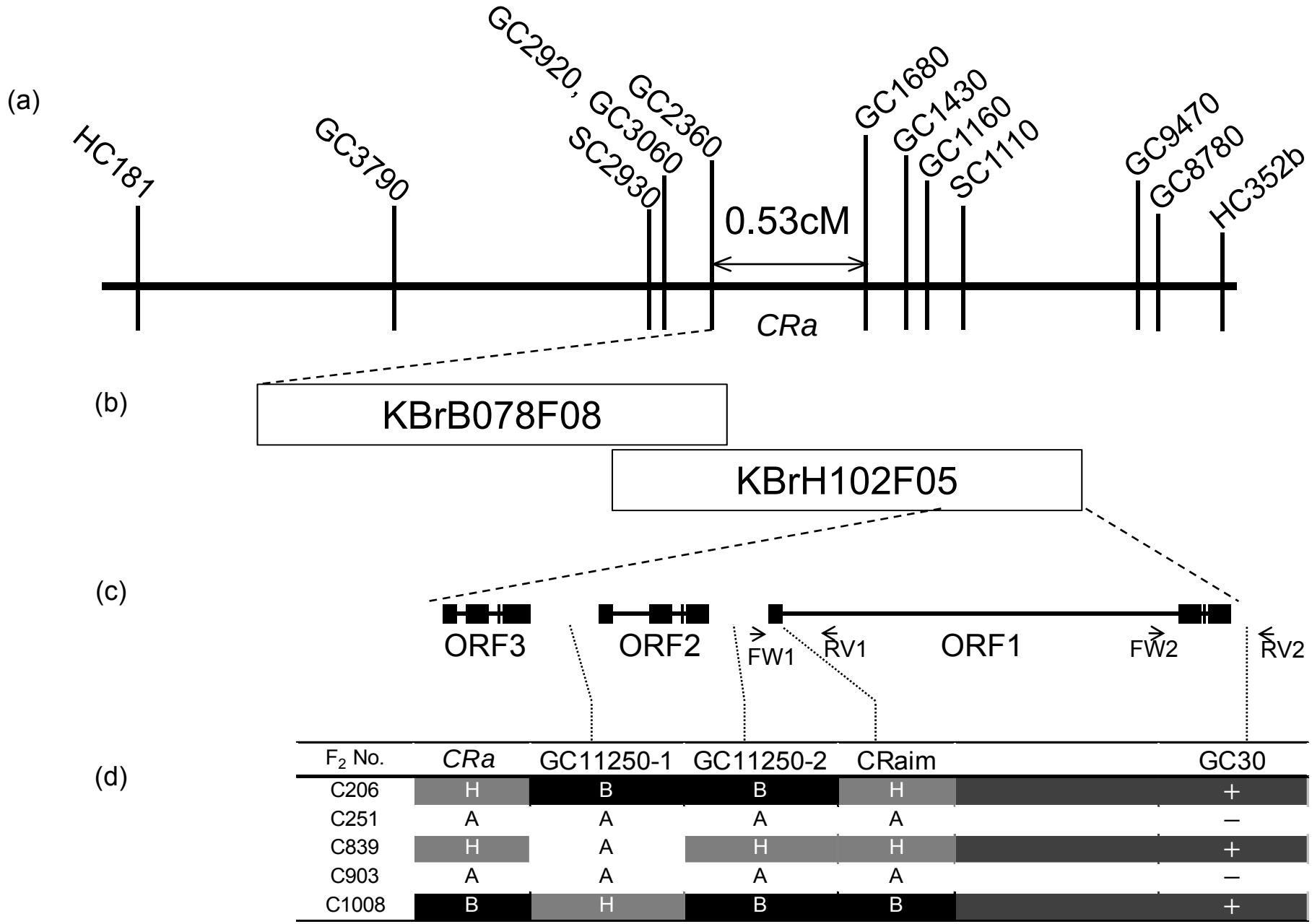

(a)
HC181
GC3790
GC2920, GC3060
SC2930
GC2360
0.53cM
GC1680
GC1430
GC1160
SC1110
GC9470
GC8780
HC352b
CRa
(b)
KBrB078F08
KBrH102F05
(c)
ORF3
ORF2
FW1
RV1
ORF1
FW2
RV2
(d)
F2 No.
CRa
GC11250-1
GC11250-2
CRaim
GC30
C206 H B B H +
C251 A A A A –
C839 H A H H +
C903 A A A A –
C1008 B H B B +

Fig. 1

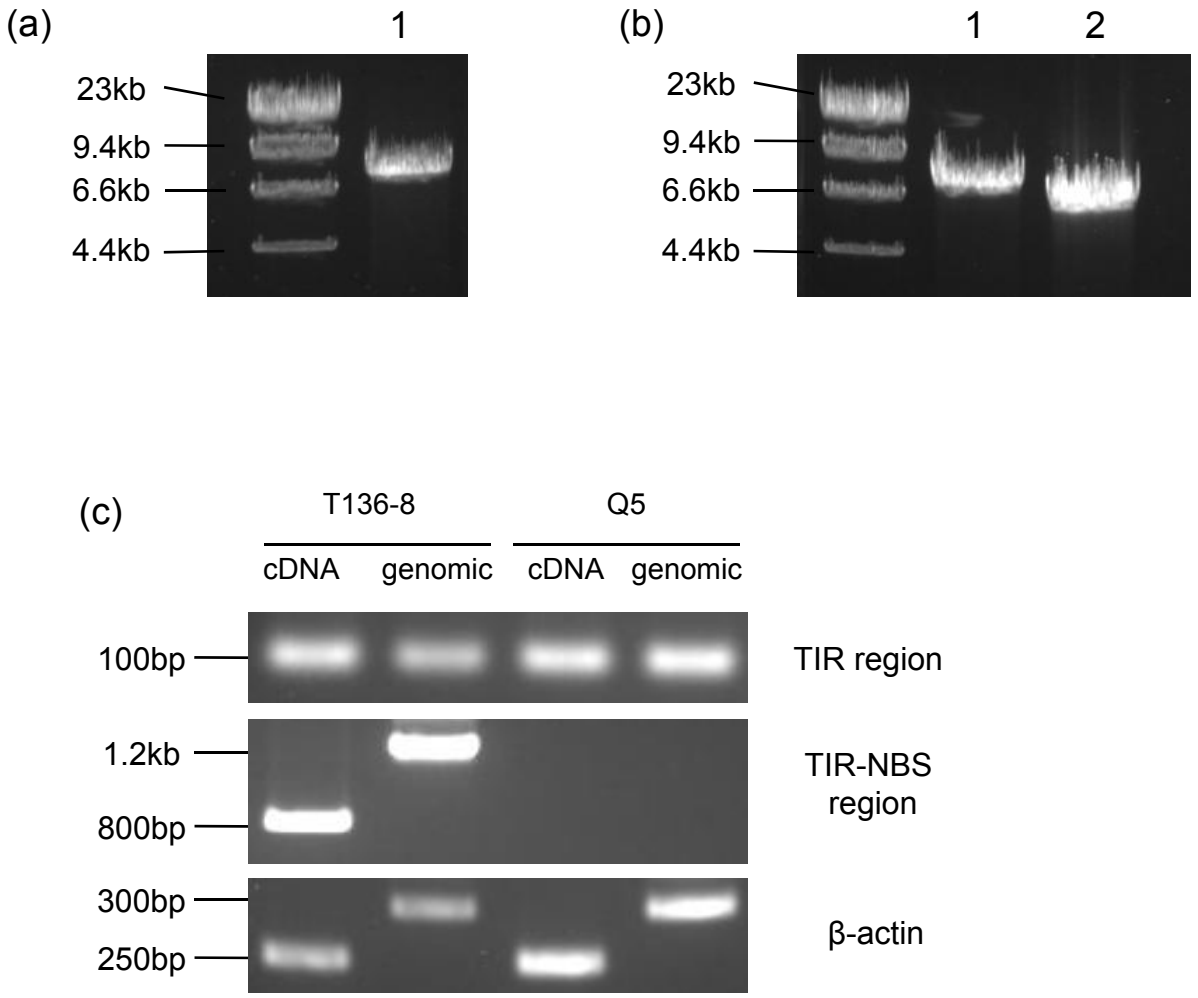

(a)
1
23kb
9.4kb
6.6kb
4.4kb
(b)
1
2
23kb
9.4kb
6.6kb
4.4kb
(c)
T136-8
Q5
cDNA
genomic
cDNA
genomic
100bp
TIR region
1.2kb
800bp
TIR-NBS region
300bp
250bp
β-actin

Fig. 2

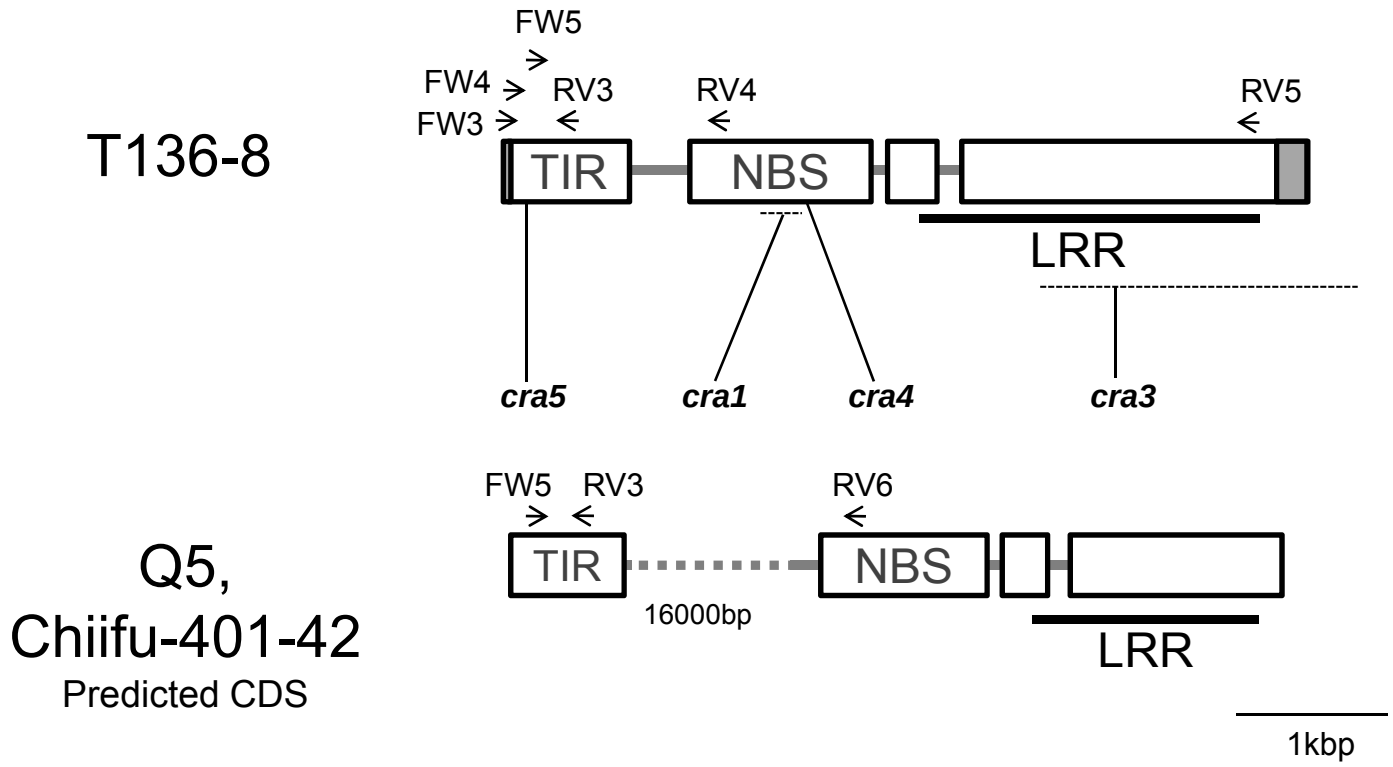

T136-8
FW5
FW4
FW3
RV3
RV4
RV5
TIR
NBS
LRR
cra5
cra1
cra4
cra3
Q5,
Chiifu-401-42
Predicted CDS
FW5
RV3
RV6
TIR
16000bp
NBS
LRR
1kbp

Fig. 3

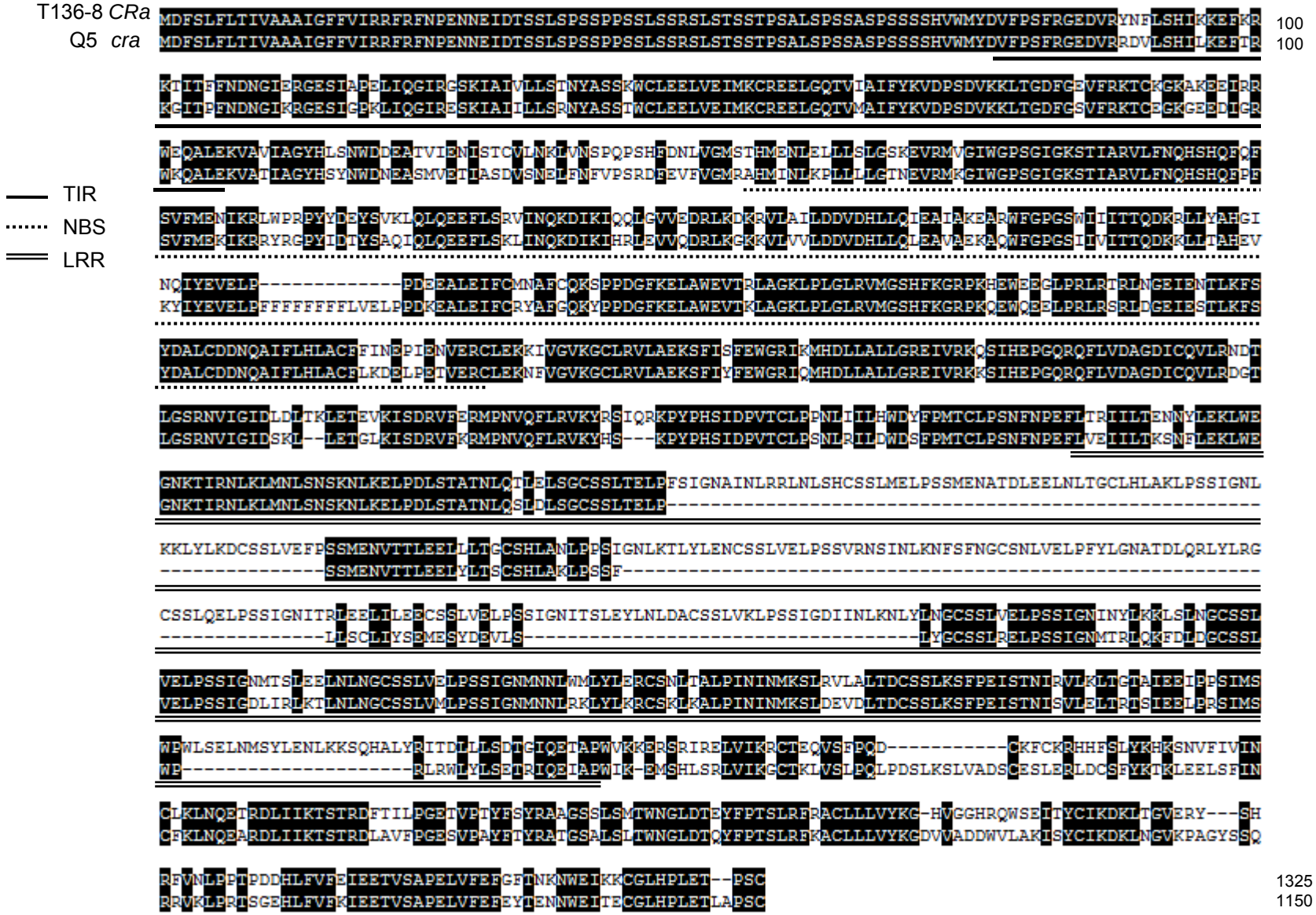


Fig. 4

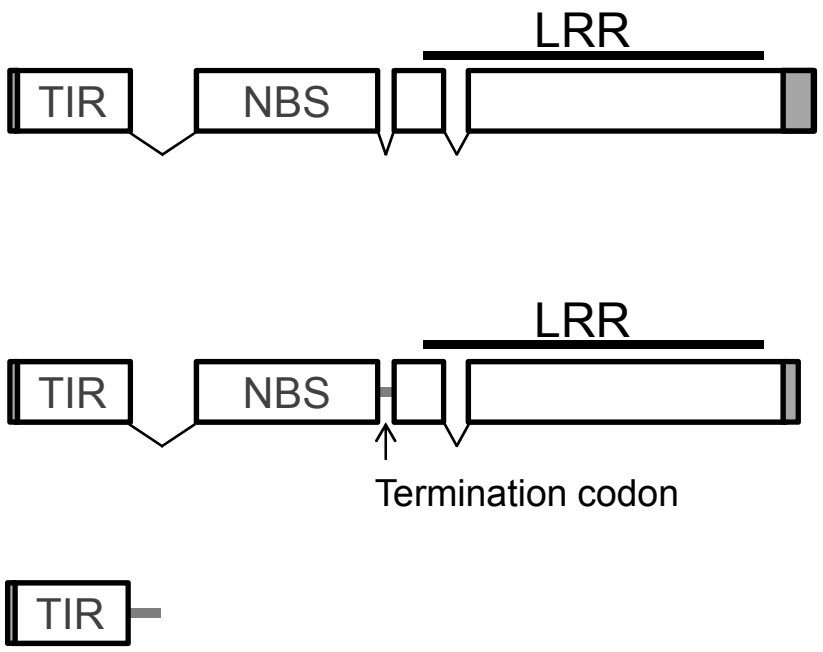


Fig. 5

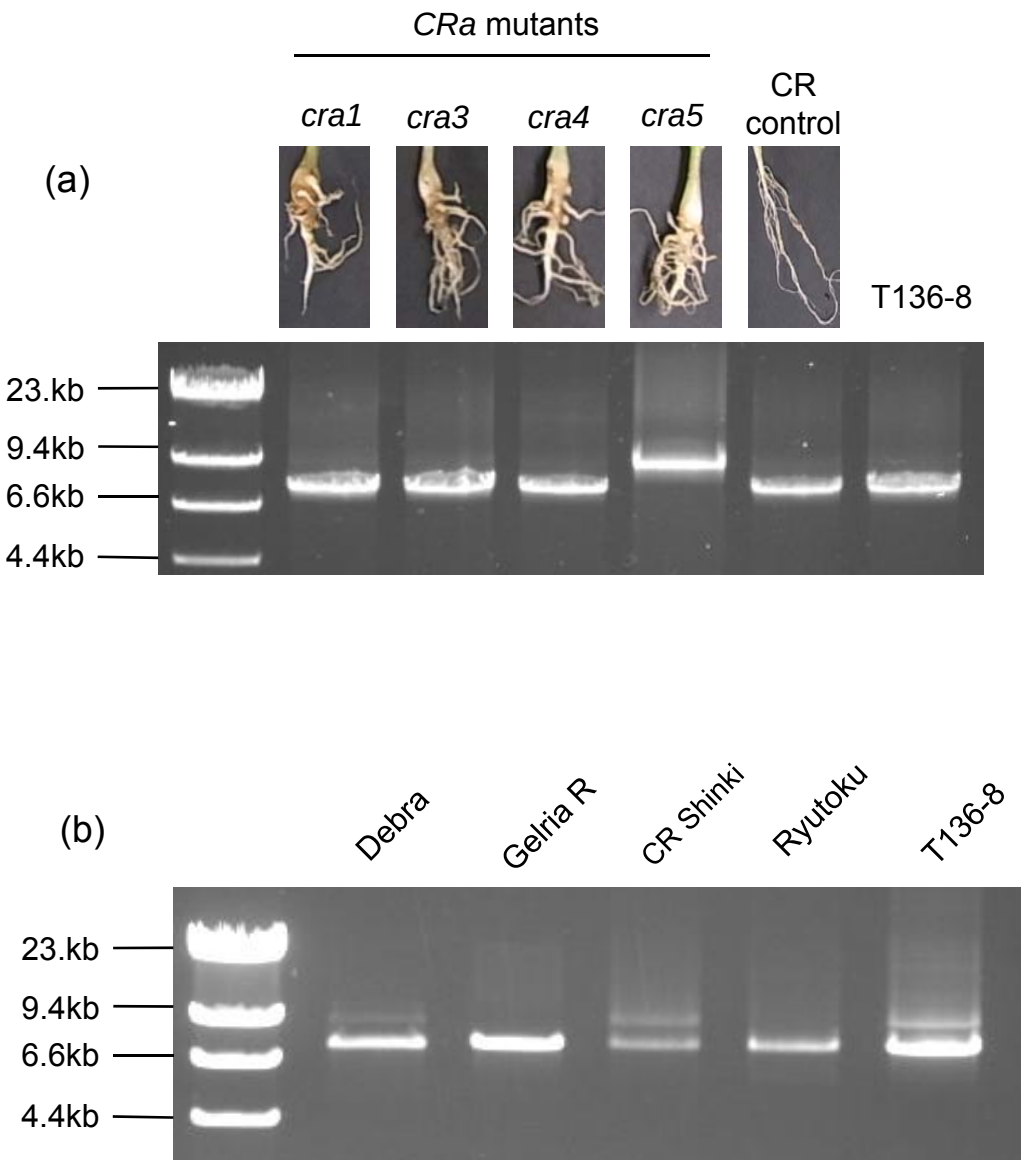

(a)
CRa mutants
cra1
cra3
cra4
cra5
CR control
T136-8
23.kb
9.4kb
6.6kb
4.4kb
(b)
Debra
Gelria R
CR Shinki
Ryutoku
T136-8
23.kb
9.4kb
6.6kb
4.4kb

Fig. 6

Supplementary Table 1. PCR primers used in this study

Marker - Primer		Distance (cM)	Sequence (5'-3')	TA* (°C)	Informative sequences
HC181	FW	2.84	GTGTCACTTCCCAGTAAATCTCCACACTG	58	HC181, KBrH070D10
	Q-RV		TTGGAATCGGTTATCTTACCGGTTAACTCT		
	T-RV		GACAGGTTTACTTCTGATTCTGATTCTGAT		
SC1110	FW	0.65	CGAGAGCTTGATGCCAAGATGAG	58	AT4g21110
	Q-RV		TACTACAAGGAGGAGCACAAAGTCAAGAA		
	T-RV		GTGTGGCCTCAAACAAGTACCTAGTAGAGT		
GC9460	FW	1.42	ATCGTTACATAGAGCCAGCAAAGCC	65	AT4G19460, AT4G19470, KBrB085J21
	RV		TGGATTTTGAGTTTTTATTAGTCTCGAGGT		
GC8780	Q-FW	1.48	TATCCATAACATTATCAAAAACTTAACTCA	63	AT4G18780, KBrB030D12R
	T-FW		TATCCATAACATTATCAAAAACTTAACTTC		
	RV		GAATGAATGGTGCATTCGTTAGAAC		
GC1160	FW	0.46	TTATGACACGTTTTCTGCTGATGACATAAT	**	AT4G21160, AT4G21190, KBrH091P11
	RV		GGAGTGGGAGTTTAGGTATATCAAAGGAAG		
GC1430	Q-FW	0.37	AGAACAATCTTACCTTGTGCCTCTTCATAC	**	AT4G21430, AT4G21410, KBrH069E01
	T-FW		AAGATTGAGGTAAGTTCCCCTTGGTTTT		
	RV		TGATAGGTATCTACCTGAGCTAAGGCGTGA		
GC1680	FW	0.18	ATTCTACTTCCTTTTAGCTGCTCTAACCGC	61	AT4G21680, AT4G21670, KBrH059N21
	Q-RV		TTCGTAATAATTGATTTGCATATCAATTGG		
	T-RV		ATAATTGATTGCATATATATTATTAATCGG		

Supplementary Table 1. Continued

Marker - Primer		Distance (cM)	Sequence (5'-3')	TA* (°C)	Informative sequences
GC2920	1st-FW	0.52	ATTTTCGAATCATCCAAGCTGAAAGT	50	AT4G22920, KBrH129J18R
	2nd-FW		CAAAGAACTGCCTGTTGTAAGTAAA	61	
	RV		TGTTCAACAAGTTCCCATCTCCAT		
GC3060	FW	0.52	TTGCGGTGATTAAAATACAATCTATATTTC	59	AT4G23060, KBrB080F13F
	RV		TTTGGGTTTCCACAAAACAGATTACTTTA		
GC2360	1st-FW	0.31	CAGCACCAGCATAACCAGCTACAGTC	61	AT4G22360, KBrB78F08R
	1st-RV		AGAACTTTGCAAGTGGCTCAGATAAT		
	2nd-FW		AGTTTTGTAATTTTCACCCAAAGTATCA	63	
	2nd-RV		CAGCTGGAGGAGCACTGCAACGGAGAGA		
GC3790	FW	1.71	AAATGCCATCTTTTCGATTTGTACTCC	**	AT4G23790, KBrH053E08
	RV		ATTTTCCACTCCAGCGACTTTTCTCC		
GC1250-1	1st-FW		GATCTAGTTTATTATTATTATTGTATCAGG	51	KBrH102F05
	RV		GAAGTTTAGGTGCACTATCCACTAAA		
GC1250-2	FW		AATGATCAGAGAGGCAAAACAGAA	63	KBrH102F05
	RV		CAATGATAATCTTACACTAATTAATACAGA		
GC30	FW		TGATGCATTCCTTTTATATCGTAAA	63	KBrH102F05
	RV		AGTATCTGAACCGAAGCCCAACTAA		

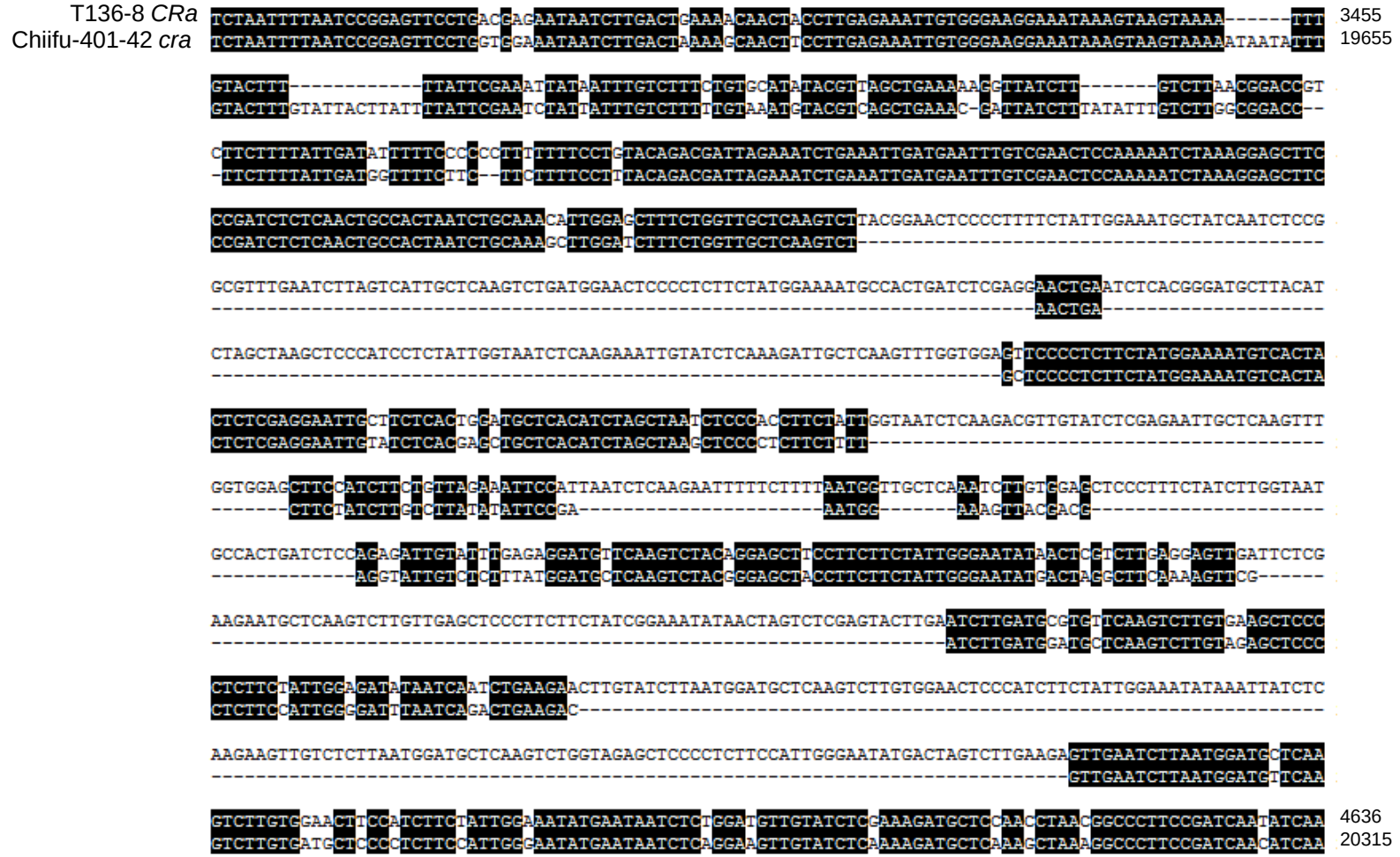
Supplementary Table 1. Continued

Marker - Primer		Distance (cM)	Sequence (5'-3')	TA* (°C)	Informative sequences
CRaim-T	T-FW		TATATTAATGATAAAGCAGAAGAAGAAA	63	This work
	T-RV		AATGCGACTGAGAAAGTTGTAG		
craim-Q	FW1		TGAAGAATGCGGGCTACGTCCTCTGAAATC	59	This work
	Q-RV		GAAGTAGATGAACGTGTTTATTTTAGAAA		
FW1			TGAAGAATGCGGGCTACGTCCTCTGAAATC	**	
FW2			GTAAGGTATTCGTGTATAAGTATATCAAA	**	
FW3			ACAATCAGTAGCAAGCTCATTGATCC	59	
FW4			AGCTCATTGATCCTTTTCCCCACCTCACAA	59	
FW5			AAGATTCAGATTCAATCCAGAAA	59	
RV1			GCTCCCAAAGCCACGAATCTGGAA	**	This work
RV2			AGTATCTGAACCGAAGCCCAACTAA	**	
RV3			GAAGTAGACAAAGACCTAGAAGATAAA	59	
RV4			TTCCTTGGAGCCCAGGCTTAACAA	59	
RV5			TGAGGGAGTTTCCAGAGGATGTAGCCCGCA	**	
RV6			AGGAAGCTCGACCTCGTAAA	59	

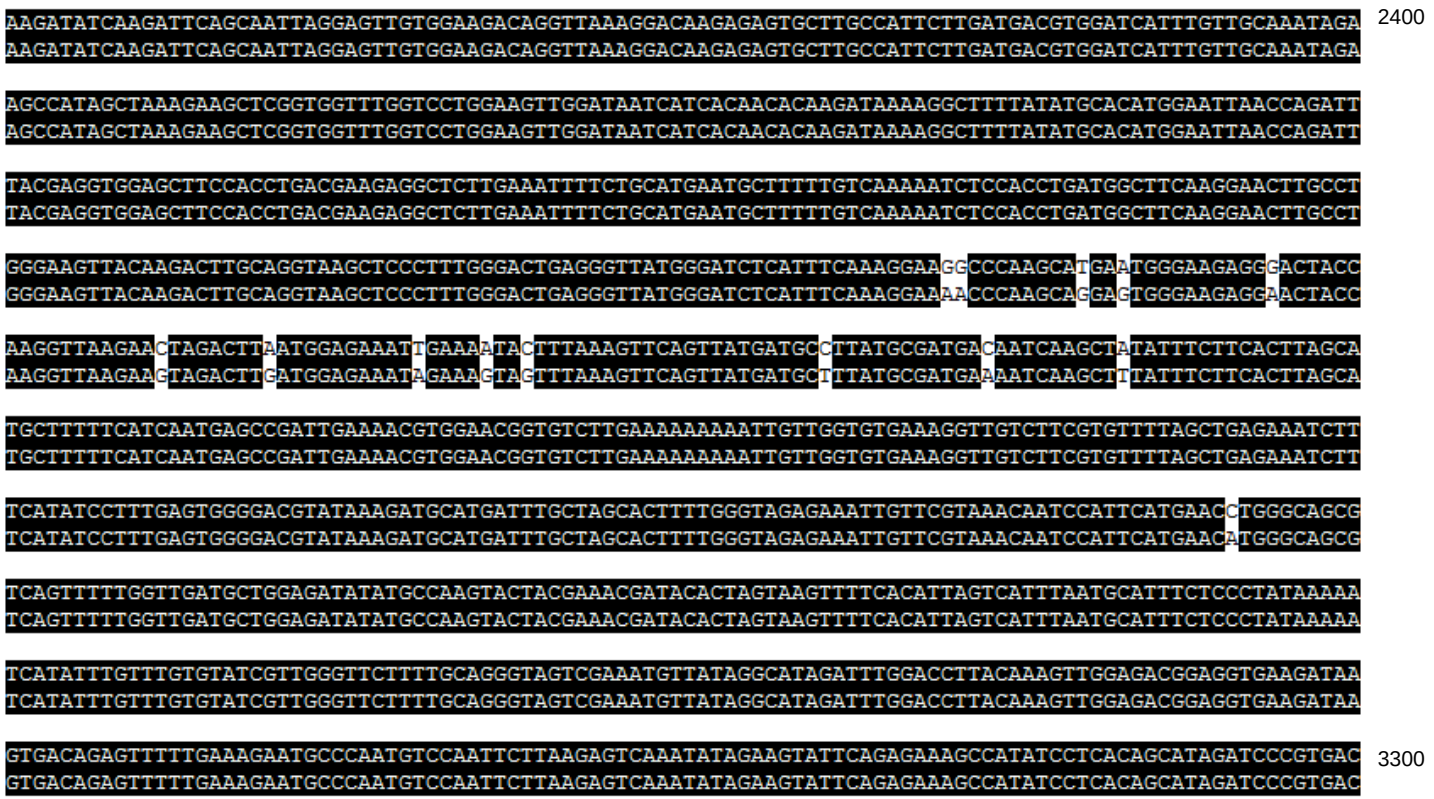
*Anneal temperature used for PCR in genotype detection.

**Not used in 3 step PCR

Supplementary Fig. 1



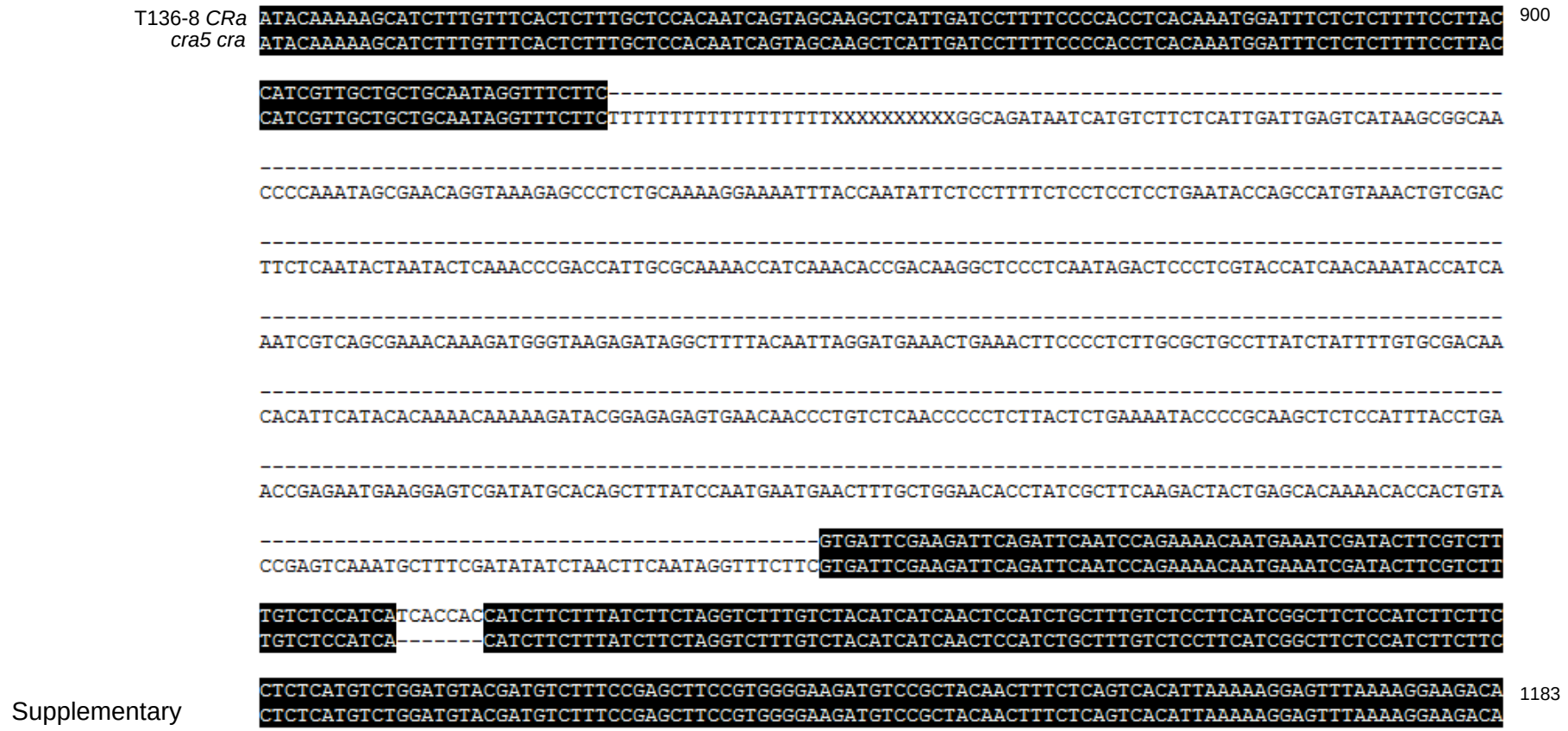
Supplementary
Fig. 2



Supplementary Fig. 3

Supplementary
Fig. 4

Supplementary Fig. 5



Supplementary Fig. 6