Plant Molecular Biology **51:** 703–718, 2003.

Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between *Cucumis melo* and *Arabidopsis thaliana*

Hans van Leeuwen[1], Amparo Monfort[1,2], Hong-Bin Zhang[3] and Pere Puigdomènech[1,*]
[1]*Departament de Genètica Molecular, Institut de Biologia Molecular de Barcelona, CID-CSIC, Jordi Girona 18, 08034, Barcelona, Spain;* [2]*Departament de Genètica Vegetal, IRTA, Ctra. de Cabrils s/n, 08348 Cabrils, Barcelona, Spain;* [3]*Department of Soil and Crop Sciences and Institute for Plant Genomics and Biotechnology, 2123 TAMU, Texas A&M University, College Station, Texas 77843-2123, USA; (*author for correspondence; e-mail pprgpm@ibmb.csic.es)*

Received 2 April 2001; accepted in revised form 29 August 2002

Key words: Bacterial Artificial Chromosome, colinearity, Cucurbitaceae, genome organisation, simple sequence repeats and retrotransposons, synteny

Abstract

A bacterial artificial chromosome (BAC) library from the dihaploid melon line 'PIT92' was constructed with a 6 times coverage of the haploid melon genome. A contig of four BACs around the MRGH63 resistance gene homologue fragment was created. The complete sequence of a 117-kb BAC clone allowed to determine two clearly defined regions, the first one containing a cluster of three resistance gene homologues. Separated by a retrotransposon, that contains large long terminal repeats, the second region presents a group of genes with a conserved distribution in two regions of the *Arabidopsis* genome. The detailed analysis of this region provides a description of the gene structure and the presence of repetitive sequences in a defined fragment of the genome of *Cucumis melo*.

Introduction

Cucumis melo L., the cultivated melon, is an important culture world-wide and Cucurbitaceae are only second to Solanaceae in their economic value among horticultural crops. This group of species constitutes a family within the group of Rosidae and their genome is relatively compact. The haploid genome of melon contains 4.5×10^8 bp (Arumuganathan and Earle, 1991), only three times larger than the *Arabidopsis* genome, the plant model whose sequence has recently been elucidated (The *Arabidopsis* Genome Initiative, 2000). Global estimations of the composition of the melon genome indicate that around 30% of it is formed by repeated sequences (Bendich and Anderson, 1974). Up to now, one type of repeated sequences, microsatellites, has been used to construct the *C. melo* genetic map (Danin-Poleg *et al.*, 2000) and to compare it with the maps of other species. From a horticultural point of view, melon is a plant susceptible to many diseases, for that reason the identification and mapping of resistance genes may be an important line of research aiming to obtain melon varieties of increased agronomic interest (Morales *et al.*, 2002).

In plants, many genes have been identified that confer resistance to the infection by pathogens. These genes can be classified into different categories according to the similarity that they present with genes that have previously been identified in different groups of organisms. A number of motifs often present in these proteins have been observed. Common to this type of genes is a leucine-rich repeat (LRR) domain in their proteins. An example of sequences having this structure is the tomato family of *Cf* genes that provide resistance to the fungal pathogen *Cladosporium fulvum*. These genes code for a putative receptor with an extracellular domain (Parniske *et al.*, 1997). The

majority of the cloned resistance genes contain a nucleotide binding site (NBS) in their predicted proteins, followed by a LRR domain in the C-terminal region of the protein (Richter and Ronald, 2000).

Disease resistance genes are often arranged in clusters in the genome (Sudupak *et al.*, 1993). For example, in rice, the *Xa21* multigene family is found in a 230-kb region on chromosome 11. In this locus, different retrotransposable elements (RTPs) and three families of miniature inverted repeat transposable elements (MITEs) are also present. In some cases it has been observed that these elements may produce truncated proteins by inserting themselves within resistance genes (Song *et al.*, 1997). In tomato, two clusters of *Cf9* homologous genes have been found in the short arm of chromosome 1, suggesting a duplication event. Another member of the *Xa21* family, after insertion of the transposable element named *Retrofit*, resembles a gene of the *Cf* class of resistance genes (Wang *et al.*, 1998). These results have suggested that transposable elements contribute to the evolution of resistance genes (Richter and Ronald, 2000). Analysis of the regions containing clusters of resistance genes has been done by direct cloning of the genomic regions in cosmid or BAC vectors. BAC libraries have been one of the most useful tools to clone and to sequence large genomic regions and for map-based cloning of genes (Wang *et al.*, 1995; Xu *et al.*, 2001).

In melon, following a cloning strategy of PCR amplification of the conserved domains, four families of gene fragments homologous to resistance genes have been described (Garcia-Mas *et al.*, 2001). These gene fragments are similar to the resistance genes *L6* and *M* of *Linum usitatissimum* that confer resistance to rust, and *N* of *Nicotiana glutinosa* that confers resistance to tobacco mosaic virus. A linkage map was constructed with an F2 segregating population produced from the resistant Korean accession 'PI161375' and a Spanish line 'Piel de Sapo T111' parentals of *C. melo* (Oliver *et al.*, 2001) and used to map the gene *nsv* that confers resistance to Melon Necrotic Spot Virus (Morales *et al.*, 2002). The melon resistance gene homologous fragments MRGH63 and MRGH4 have been mapped and localised in Linkage Group 4. The structure of the regions containing resistance genes in the genome of species of the Cucurbitaceae family and their relation with the genome of other species is not known. In the present work we present the construction of a library of BACs made from melon and its use to analyse a contig of BACs in the region containing MRGH63. The complete sequence of a BAC from the contig allowed determination of the genes present in this region and the extent of a conserved distribution among dicot species of these genes when compared with the genome sequence of *Arabidopsis*. Colinearity between the genomes of *Arabidopsis* and other dicots has been found (reviewed in Schmidt, 2002). To our knowledge this is the first report of the analysis of a large genomic sequence of a Cucurbitaceae.

Materials and methods

Preparation of high-molecular-weight genomic DNA

Seeds of the dihaploid melon line 'PIT92' ('PI161375' × 'Piel de Sapo T111') were kindly provided by Ramon Dolcet from the 'Institut de Recerca i Tecnología Agroalimentaria' (IRTA), and germinated in a greenhouse of the CID-CSIC. Young leaves of 4–7 weeks old plants were collected after the plants were kept in the dark for 48 h. The leaves were frozen in liquid nitrogen and then stored at −80 °C. Nuclei were isolated from the leaf material as described by Zhang *et al.* (1995) with modifications. 80 g of young leaves were ground in liquid nitrogen for 45 min with a pre-cooled (−80 °C) mortar and pestle. The powder was transferred to a 1000-ml beaker on ice, containing 800 ml ice-cold 1× HB plus 0.15% β-mercaptoethanol and 0.5% Triton X-100. The content was gently stirred with a spoon. After centrifugation, the centrifuge bottles were washed with 1 ml of the washing buffer to obtain the majority of the nuclei. The nuclei were embedded in agarose plugs as described by Zhang *et al.* (1995) with modifications. The agarose plugs containing the nuclei were incubated in the lysis buffer for 36 h. After lysis, the plugs were directly washed three times for 1 h each wash in 40 ml ice-cold TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0) plus 0.1 mM PMSF on ice and three times for 1 h each wash in 40 ml ice-cold TE on ice. The plugs were stored in 45 ml TE at 4 °C.

Digestion and size selection of DNA

Partial digestion and size selection of the high-molecular-weight DNA were performed as described by Hamilton *et al.* (1999) with modifications. The restriction endonuclease used was *Bam*HI (Life Technologies, Inc.). Optimal partial digestion conditions were determined in a pilot experiment using different restriction endonuclease concentrations. Each agarose

plug was cut into 9 slices with a glass cover slip and incubated on ice twice for 30 min each time in the buffer recommended by the manufacturer with the addition of 2 mM spermidine and 1 mM DTT (final concentrations). The washed slices were incubated in Eppendorf tubes on ice for 1 h in the reaction buffer that contained 0.5 mg ml^{-1} BSA and the required amount of restriction endonuclease (2 μl of dilution prepared with the incubation buffer) in addition to the incubation buffer. The partial digestions were performed at 37 °C for 5 min, after which the reactions were stopped by transferring the tubes onto ice and then addition of 0.1 volume of 0.5 M EDTA, pH 8.0.

The pulsed-field gel conditions used for visualisation of partial digests were 1% agarose, 0.5 × TBE, 12.5 °C, pump setting 80, 120 degree angle, 6 V/cm, 50 sec of initial and final pulse time for 20 h with the CHEF-DRIII System (Bio-Rad). The partially digested DNA was stained with ethidium bromide and the size compared with the Lambda Ladder PFG Marker (New England BioLabs, Inc.). The large-scale partial digestion was performed with 0.6, 0.8 and 1.0 units of *Bam*HI. The conditions for the pulsed-field gel only differed from the pilot experiment in the 90 seconds initial and final pulse time.

Three gel pieces with different size classes of digested DNA (100–200 kb, 200–300 kb, 300–400 kb) were prepared for electroelution as described by Strong *et al.* (1997) with 1/4 inch diameter dialysis tubing (Life Technologies, Inc.). The dialysis tube was rinsed with ice-cold distilled H_2O three times, then with ice-cold 0.5 × TBE three times. The pulsed-field gel conditions for electroelution of the DNA from the gel pieces in the dialysis tubing were 0.5 × TBE, 12.5 °C, pump setting 80, 120 degree angle, 6 V/cm, 30 sec of initial and final pulse time for 4 h.

The eluted DNA was collected with cut-off tips and directly loaded in a new gel for the second size-selection with the conditions 1% agarose, 0.5 × TBE, 12.5 °C, pump setting 80, 120 degree angle, 4 V/cm, 5 sec of initial and final pulse time for 9 h. The compressed DNA zones (>100 kb) were excised from the gel and prepared for electroelution under the same conditions as the first electroelution. After electroelution the gel piece was removed and the DNA was dialysed in the same tubing in ice-cold 0.5 × TE on ice three times for 2 h.

Vector preparation and ligation

The pECBAC1 vector (Frijters *et al.*, 1997) was purified by the alkaline lysis method, followed by two rounds of cesium chloride/ethidium bromide ultracentrifugation as described by Sambrook *et al.* (1989). The vector DNA was completely digested and dephosphorylated as described by Zhang *et al.* (1996). The concentration of the genomic DNA was estimated on a 1% agarose gel containing known concentrations of λ DNA. The three classes of size selected DNA were ligated into the pECBAC1 vector separately. Prior to ligation, the vector DNA was heated at 60 °C for 6 min. Ligation reaction mixtures were prepared on ice and contained vector DNA and genomic DNA in a molecular weight ratio of 1:5 (genomic DNA classes 100–200 kb and 200–300 kb) or 1:6 (genomic DNA class 300–400 kb). Ligations and test ligations were performed as described by Hamilton *et al.* (1999) with ligation buffer (5×) and T4 DNA ligase (Life Technologies, Inc.).

Transformation, selection, analysis of clones and library construction

Transformation of the ligated DNA into *E. coli* and selection for recombinant transformants were performed as described by Zhang *et al.* (1996). Plasmid DNA of recombinant clones was isolated by BAC miniprep (Zhang *et al.*, 1996), and genomic DNA inserts liberated by digestion with the *Not1* restriction endonuclease (Bio Labs, Inc.) and analysed by PFGE with the conditions of 1% agarose, 0.5 × TBE, 12.5 °C, pump setting 80, 120 degree angle, 6 V/cm, 5 sec initial, and 15 sec final pulse time for 16 h. Recombinant BAC clones were ordered in 60 384-well micro titer dishes as described by Zhang *et al.* (1996).

Filter preparation and hybridisation

The Biomek 2000 High Density Replicating System (Beckman) and Hybond N+ membranes (Amersham) were used to prepare filters for colony hybridisation. The filters were processed as described by Woo *et al.* (1994) and then baked at 80 °C to link the DNA to the membrane. Each filter (8 × 12 cm) contains 1,536 double-spotted individual clones from four microtiter dishes of the library. The probes were liberated from the cloning vector by restriction endonuclease digestion and purified from agarose gels. The probes were labelled with α-^{32}P dCTP using the random hexamer

primer method (Feinberg and Vogelstein, 1983). Hybridisation and washing conditions were as described by Church and Gilbert (1984).

Construction of a contig containing a Resistance Gene Homologue

Positive clones from the radioactive screening of the library were analysed for the sizes of the inserts as described earlier and tested for the presence of the probed sequence by PCR. The BAC ends of the positive clones from the PCR screening were sequenced using the Terminator cycle sequencing ready reaction kit (Perkin Elmer) according to the manufacturers recommendations. Prior to PCR amplification, the BAC plasmid DNA (10 μl) was mixed with H_2O (10 μl) and sheared with a cut-off 0.8 mm syringe, followed by a 30 min incubation at 65 °C. The sequencing reactions were visualised and processed with ABI PRISM 377. Primers based on the BAC end sequences were used to create a contig of the region.

Subcloning, sequencing and assembly

A BAC was sequenced using the shotgun subcloning strategy (GATC Biotech AG, Germany). The shotgun subclone library consisted of 1344 subclones with an insert size varying from 1100–1400 bp. The cloning vector used for the subcloning was pCR® 4Blunt-TOPO® (GATC Biotech AG, Germany). BAC vector DNA of the subclones was isolated from 3 ml overnight cultures (Terrific Broth; 50 μg ml^{-1} kanamycine), by the alkaline lysis method as described by Sambrook *et al.* (1989) with modifications. The final DNA pellet was resuspended in 50 μl of H_2O and 1 μl was evaluated on a 1% agarose gel for estimation of DNA concentration and for the presence of insert DNA. Sequencing was performed with the Terminator cycle sequencing ready reaction kit (Perkin Elmer), using the T3 or M13 forward primers. The sequencing reactions were visualised and processed with ABI PRISM 377. The ordination of the sequences was performed with the pregap and gap4 software of the STADEN package (Staden, 1996). Remaining gaps in the contig were resolved by PCR amplification with flanking primers, and purification and subsequent sequencing of the amplified regions.

Sequence analysis, characterisation and annotation

Analysis of DNA and protein sequences was done with the Genetics Computer Group software (Madison, WI, USA), the BCM Search Launcher Interface (Smith *et al.*, 1996; http://searchlauncher.bcm.tmc.edu/) and the GeneDoc software (Nicholas *et al.*, 1997, http://www.hgmp.mrc.ac.uk/embnet.news/vol4_2/genedoc.html). Characterisation of the 60K17 BAC DNA sequence was done with the BLASTX program (Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/BLAST/) and several gene prediction programs: NetPlantGene (Hebsgaard *et al.*, 1996; http://www.cbs.dtu.dk/services/NetPGene/), Glimmer M (Salzberg *et al.*, 1999; http://www.tigr.org/tdb/glimmerm/glmr_form.html), GeneMark (Borodovsky and McIninch, 1993; http://dixie.biology.gatech.edu/GeneMark/eukhmm.cgi), GeneMark.hmm (Lukashin and Borodovsky, 1998; http://dixie.biology.gatech.edu/GeneMark/eukhmm.cgi), GENSCAN (Burge and Karlin, 1997; http://genes.mit.edu/GENSCAN.html), Genie (Reese *et al.*, 1997; http://www.fruitfly.org/seq_tools/genie.html) and SpliceView-WebGene (Rogozin and Milanesi, 1997; http://l25.itba.mi.cnr.it/~webgene/wwwspliceview.html). Repetitive sequences were detected and analysed with STADEN (Staden, 1996), Sputnik (Abajian, unpublished; http://rast.abajian.com/sputnik/) and Palindrome (Rice *et al.*, 2000; http://bioweb.pasteur.fr/seqanal/interfaces/palindrome.html). Further analysis of the putative proteins was done with BLASTP and TBLASTN (Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/BLAST/), TMHMM (Sonnhammer *et al.*, 1998; http://www.cbs.dtu.dk/services/TMHMM/), TargetP (Emanuelsson *et al.*, 2000; http://www.cbs.dtu.dk/services/TargetP/), Blast 2 Sequences (Tatusova and Madden, 1999; http://www.ncbi.nlm.nih.gov/BLAST/), InterProScan (Apweiler *et al.*, 2001; http://www.ebi.ac.uk/interpro/scan.html) and SMART (Letunic *et al.*, 2002; http://smart.embl-heidelberg.de/).

Analysis of micro-synteny between Arabidopsis *and melon*

Exon sequences from 12 putative genes were analysed with the TBLASTX program (Altschul *et al.*, 1997) at the TAIR BLAST web site (http://www.arabidopsis.org/Blast/) with the parameters: Genes from AGI - Total Genome (DNA), no filter and expectation 0.0001. To detect micro-synteny, we searched for small-sized regions in the *Arabidopsis* genome that contained a substantial number of the genes from the melon BAC.

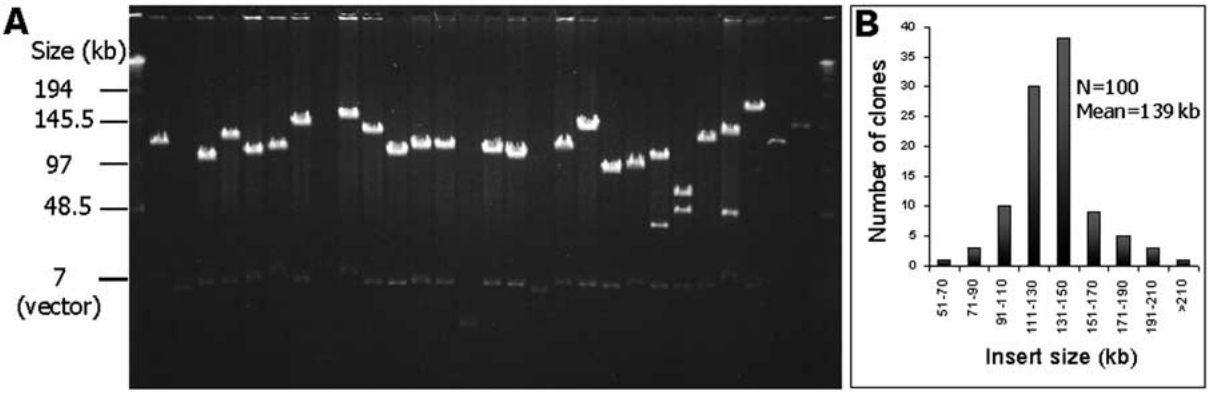

Figure 1. Insert DNA analysis of random BAC clones of the melon 'PIT92' BAC library. **A.** BAC clones analysed by pulsed-field gel electrophoresis. BAC DNA was isolated, digested with *Not* I to release the insert DNA, separated on a 1% agarose gel and stained with ethidium bromide. The marker used in the outer lanes is the Lambda Ladder PFG Marker (New England BioLabs). The common ~7.4 kb band is the pECBAC1 cloning vector. **B.** Distribution of size classes for 100 BAC clones with inserts. BAC DNA was analysed as shown in Figure 1A and insert sizes grouped in nine size classes.

Results

Melon BAC library construction and characterisation

A melon BAC library was constructed from the dihaploid 'PIT92' melon line, which was obtained from a F1 plant from 'PI161375' × 'Piel de Sapo T111'. The 'PIT92' BAC library contains 23040 clones in 60 384-well microtiter plates. The genomic DNA was obtained from nuclei embedded in agarose plugs, and partially digested with *Bam*HI. High-molecular-weight DNA fragments were size-selected twice with Pulsed Field Gel Electrophoresis (PFGE), and after electroelution they were ligated into the pECBAC1 vector. BAC DNA was isolated from 100 randomly picked clones, digested with *Not*I to release the melon DNA insert from the cloning vector and analysed by PFGE (e.g., Figure 1A). The insert size of each clone was determined by adding up the sizes of all fragments in each lane, except the 7.4-kb pECBAC1 vector fragment. The clones were grouped in nine different size classes (Figure 1B). The insert sizes of clones in the library range from 60 to 300 kb, with 94% of the clones having an insert size greater than 100 kb. According to these data, the 'PIT92' BAC library has an average insert size of 139 kb and 86% of the clones contain melon insert DNA. With the size of the melon genome being 445 Mb, the genome coverage of the library is approximately 6.2 times.

Screening the BAC library with the MRGH63 Resistance Gene Homologue

Previously, the isolation of four families of resistance gene homologue fragments from the melon genome was reported (Garcia-Mas *et al.*, 2001). MRGH63, a homologue mapping near several disease resistance genes in the melon genetic map, was used to screen the high-density colony filters of the 'PIT92' library in order to obtain a BAC clone that contains this sequence. Southern blot hybridisation showed the presence of one copy of MRGH63 in the 'PI161375' and 'PIT92' genomes (result not shown). With MRGH63-specific PCR primers, the presence of MRGH63 was confirmed for six of the BAC clones positive in the radioactive screening. Insert sizes were determined with PFGE. Four clones were used to create a 150-kb contig of the MRGH63 region in the melon genome. Primers were designed based on the sequenced borders of the BAC clones and used to orientate the clones in the contig (Figure 2).

Sequencing and characterisation of a melon BAC clone containing MRGH63

Analysis of predicted gene sequences.
Fragments of a BAC clone (60K17) containing MRGH63 were subcloned with the shotgun method and subsequently sequenced. The complete sequence of the BAC clone is 116,877 bp and it was obtained with a sequencing redundancy of 4.2 times. Seven gene prediction programs were used to analyse the exon structure of the putative genes: Net-

Table 1. Analysis of predicted genes from the melon 60K17 BAC.

Gene	Protein Size (aa)	BLAST				ESTs					
		No. Hits (*E-value* ≤ *1e–10*)	*Acc. No.*	*Best Hit Organism*[a]	*E-value*	*No Hits* (*E-value* ≤ *1e–10*)	*Top 5* Organisms[a]				
MRGH6[b]	84	128	AAG01052	*G.m.*	6e-17	58	*L.j.*	*S.t.*	*M.t.*	*G.m.*	*C.p.*
MRGH5	1092	489	At5g17680	*A.t.*	1e-152	515	*M.t.*	*G.m.*	*S.t.*	*A.t.*	*L.e.*
MRGH63	943	441	AAG48132	*G.m.*	1e-130	382	*M.t.*	*S.t.*	*G.m.*	*L.e.*	*L.j.*
Hlh	299	16	At1g72210	*A.t.*	1e-49	53	*R.p.*	*G.m.*	*P.tr.*	*M.t.*	*T.a.*
Drzf	379	101	At3g05200	*A.t.*	2e-69	365	*M.t.*	*G.h.*	*L.e.*	*S.t.*	*A.t.*
Spp	167	12	At5g27430	*A.t.*	8e-72	252	*G.m.*	*M.t.*	*S.t.*	*S.c.*	*L.e.*
Hsp1	391	4	At1g54840	*A.t.*	5e-67	16	*G.m*	*A.t.*	*S.t.*	*S.r.*	*P.ta.*
Hsp2	171	4	At1g54850	*A.t.*	1e-52	14	*G.m.*	*S.t.*	*A.t.*	*S.r.*	*G.a.*
Sbp2[c]	466	336	At3g05240	*A.t.*	1e-113	206	*S.t.*	*G.m.*	*M.t.*	*A.t.*	*L.e.*
Mki2[d]	383	24	AAL18925	*H.b.*	1e-137	127	*S.t.*	*G.m.*	*M.t.*	*L.e.*	*G.h.*
Sbp1	565	438	At3g05240	*A.t.*	1e-170	276	*S.t.*	*L.e.*	*M.t.*	*A.t.*	*G.m.*
Ctp	225	0	At5g27440	*A.t.*	2e-09	0	*C.r.*	*H.s.*	*M.c.*	*L.j.*	*H.v.*
Rzf	209	1	At3g05250	*A.t.*	4e-34	11	*G.m.*	*A.t.*	*H.v.*	*P.tr.*	*L.e.*
Mki1[b]	135	5	AAL18925	*H.b.*	3e-37	28	*L.e.*	*G.m.*	*L.j.*	*L.p.*	*M.t.*

[a]*A.t.: Arabidopsis thaliana; C.p.: Citrus x paradisi; C.r.: Ceratopteris richardii; G.a.: Gossypium arboreum; G.h.: Gossypium hirsutum; G.m.: Glycine max; H.b.: Hevea brasiliensis; H.v.: Hordeum vulgare; H.s.: Homo sapiens; L.e.: Lycopersicon esculentum; L.j.: Lotus japonicus; L.p.: Lycopersicon pennellii; M.c.: Mesembryanthemum crystallinum; M.t.: Medicago truncatula; P.ta.: Pinus taeda; P.tr.: Populus tremula x Populus tremuloides; R.p.: Robinia pseudoacacia; S.c.: Secale cereale; S.r.: Stevia rebaudiana; S.t.: Solanum tuberosum; T.a.: Triticum aestivum.*
[b]Partial genes from BAC borders.
[c]Pseudogene with stop codons and deletions in reading frame.
[d]Duplicated exon only considered once.

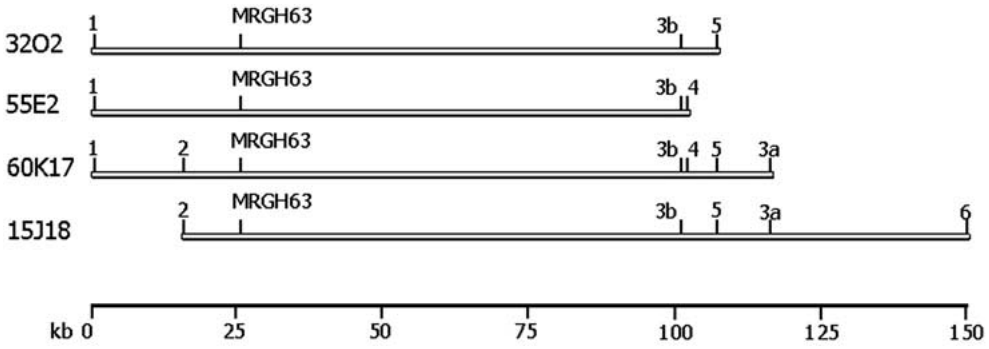

Figure 2. Schematic representation of the MRGH63 contig. The 'PIT92' BAC library was screened with a radioactive MRGH63 probe, and positive clones were confirmed with MRGH63 specific PCR primers. Common markers in the BAC clones are indicated with equal numbers and were obtained by BAC-end sequencing followed by PCR with all clones. 1: BAC-end (SP6) 32O2 clone (identical to 55E2 and 60K17 clones), 2: BAC-end (SP6) 15J18 clone (not determined in 32O2 and 55E2 clones), 3a: BAC-end (SP6) 60K17 clone, 3b: second amplification product with 3a specific primers (identical regions were later confirmed in sequence of 60K17), 4: BAC-end (T7) 55E2 clone (not determined in 32O2 and 15J18 clones), 5: BAC-end (T7) 32O2, 6: BAC-end (T7) 15J18. The size of the 60K17 clone was determined by sequencing and measured 117 kb. The sizes of the 32O2 and 55E2 clones (108 and 103 kb, respectively) were determined with the sequence information of the 60K17 clone. The size of the 15J18 clone, 135 kb, was determined as shown in Figure 1A.

PlantGene, GlimmerM, GeneMark, GeneMark.hmm, GENSCAN, Genie and SpliceView-WebGene (see *Materials and methods*). The BLASTX program was used to identify potential gene encoding regions and the BLASTX, BLASTP and TBLASTN (EST database) programs were used to determine the most likely start codons, acceptor splice sites, donor splice sites and stop codons. The sequence annotation (Figure 3) shows 14 putative genes, one retrotransposon (RTP) of the Ty3/Gypsy type and two fragments with similarity to this RTP, and 335 microsatellites (Figure 3). The similarities of the predicted genes with proteins and ESTs from other species are presented in Table 1. The sequence of the 60K17 BAC clone appears in the GenBank database under the accession number AF499727.

At one border of the BAC clone analysed (Figure 3), a resistance gene homologue cluster was found that contains the complete gene MRGH63 (*M*elon *R*esistance *G*ene *H*omologue), a new resistance gene homologue MRGH5 and one partial resistance gene homologue MRGH6. MRGH5 and MRGH63 both have four exons and they contain the characteristic domains TIR (*T*oll-*I*nterleukin *R*esistance, accession number PF01582 from the Pfam Database, http://www.sanger.ac.uk/Software/Pfam/index.shtml), NBS (*N*ucleotide *B*inding *S*ite, PF00931) and LRR (*L*eucine *R*ich *R*epeat, PF00560) of disease resistance genes (see Figure 4). The main difference between the proteins is found in the LRR domain. MRGH5 contains at least 9 LRRs compared to MRGH63 that only shows 4 LRRs as detected by InterProScan. The MRGH5 and MRGH63 genes are in opposite orientation with two RTP fragments in between (Figure 3), and two fragments with homology to the TIR and to LRR domains of resistance genes in between (result not shown). MRGH5 has the strongest similarity with the disease resistance RPP1-WsB-like protein from *Arabidopsis* (Table 1). The MRGH63 protein is most similar to the functional candidate resistance protein KR1 from *Glycine max*. Northern blot analysis of different tissues from melon varieties 'Piel de Sapo T111' and 'PI161375' with a MRGH63 probe showed a positive signal of an mRNA with the expected size (result not shown).

Other genes found in the BAC sequence may have a function in the regulation of transcription. The predicted protein of the Hlh (*H*elix *l*oop *h*elix) putative gene contains a helix-loop-helix DNA binding domain (PF00010). It has the strongest similarity with an expressed protein from *Arabidopsis* and similar homology with several proteins involved in transcriptional regulation in other systems. Two genes that encode zinc-finger proteins are present in the same orientation separated by approximately 35 kb. The Drzf (*D*efence *R*ING *z*inc *f*inger) putative gene contains a RING zinc finger domain (PF00097) and a putative transmembrane domain as predicted by TMHMM, and has highest homology with the ATL6 protein from *Arabidopsis* (Table 1). It also has homology with the EL5 protein from *O. sativa* and the protein 132 from *N. tabacum*. The genes encoding ATL6, EL5 and protein 132 are rapidly induced in *A. thaliana* (Salinas-Mondragón *et al.*, 1999), *O. sativa* (Takai *et al.*, 2001) and *N. tabacum* (Durrant *et al.*, 2000), respectively, in response to fungal elicitors and might play a role in the early stages of the defence response induced in plants upon pathogen attack. The Rzf (*R*ING *z*inc *f*inger) putative gene has the strongest homology with an unknown protein from *Arabidopsis*. Furthermore, it has homology with several transcriptional activator proteins. The RING domain identity between the Drzf and Rzf genes is 32%. In contrast to Drzf, which is a single exon gene, Rzf is composed of 5 exons. Another difference between Drzf and Rzf is that Rzf does not have a putative transmembrane domain. Alignment of the two melon proteins shows a common structure (result not shown), mainly in the RING domain. The Drzf protein contains an additional sequence of 40 aa at the N-terminal end and several extra internal sequences when compared to Rzf.

Other genes present in the 60K17 BAC include a putative signal peptidase gene that contains 5 introns and encodes a 167 aa protein. The highest similarity of this Spp (*S*ignal *pep*tidase) gene was found with a signal peptidase gene from *Arabidopsis* (Table 1). Two putative heat shock protein genes are present in the melon BAC. The carboxy-terminal end of the Hsp2 (*H*eat *s*hock *p*rotein) protein contains the HSP20 heat shock protein domain (PF00011) as detected with InterProScan. The heat shock protein domains of the two sequences are 60% identical, but the prediction programs do not detect the HSP20 domain in Hsp1. Two genes were found with homology to, amongst others, selenium-binding protein genes. Sbp1 (*S*elenium *b*inding *p*rotein) is a single exon gene that encodes a protein containing five PPR repeats (PF01535). Sbp2 is most likely a pseudogene with several stop codons and deletions in its ORF. The two genes are 64% identical at the amino acid level. There are also two genes that encode putative mevalonate kinases. Mki2 (*M*evalonate *ki*nase) contains the GHMP kinase do-

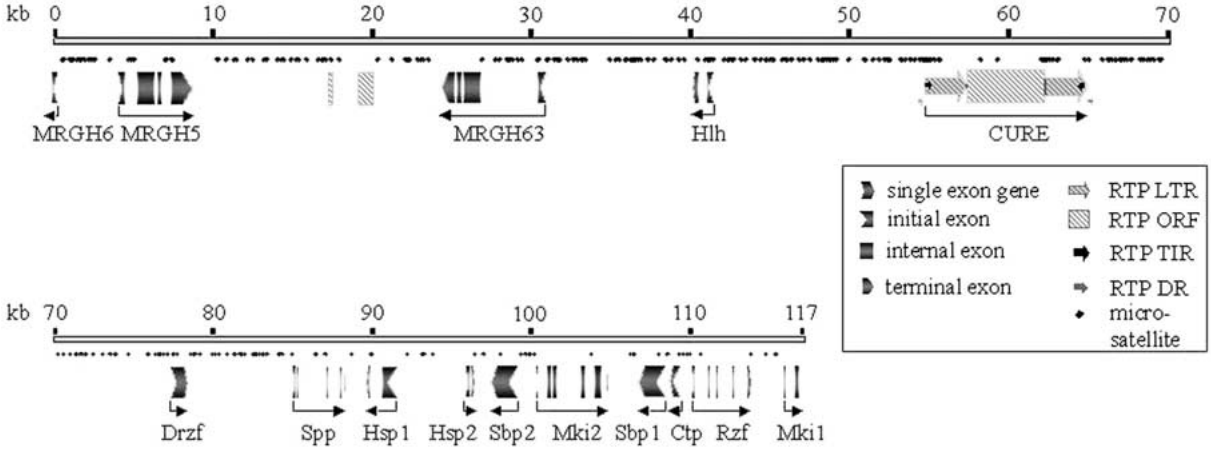

Figure 3. Schematic representation of the annotation of the 60K17 BAC clone. The clone was sequenced and analysed as described in Materials and Methods. The sizes of the exons, introns and other elements are at scale, except for the microsatellites which are drawn bigger for better visualisation.

main (PF00288). The third and fourth exons of this gene (Figure 3) are highly identical, and further analysis showed that the sequence upstream and downstream of these exons were also identical, suggesting a duplication event of 300 nucleotides within the Mki2 gene. Furthermore, BLASTX analysis of the region upstream of exon 4 revealed a low homology with proteins from transposable elements. The available Mki1 sequence (exons 1, 2 and 3 partially) is 82% identical at the amino acid level to Mki2. Ctp (*C*hloroplast *t*ransmembrane *p*rotein) is a single exon gene that encodes a 225 aa protein containing three putative transmembrane domains as predicted by TMHMM, and a putative chloroplast transit peptide as predicted by TargetP. As can be seen in Table 1, homology of Ctp with ESTs from several organisms was detected. The homology was low but significant, and in a specific region of the protein. BLASTP analysis revealed that the only protein that shows significant homology to Ctp is a hypothetical protein from *Arabidopsis* (Table 1).

Analysis of repetitive sequences

Retrotransposons A retrotransposable element (RTP) of 10 kb was found in a gene-poor region of the melon 60K17 BAC between the Hlh and Drzf genes, and has been named CURE (*Cu*cumis *R*etrotransposable *E*lement). The RTP contains two long terminal repeats (LTRs) of 2.6 kb, similar in length to the LTRs of the RTP *delI* from *Lilium henryi* (Smyth *et al.*, 1989). The coding region of the RTP is 4.8 kb, relatively short when it is compared to the typical size of coding regions of RTPs in plants. The gene and protein domain distribution of this mobile element is the typical of the Ty3/Gypsy type of RTPs. The LTRs contain the conserved LTR feature consisting of the dinucleotide end-sequences (5′-TG-CA-3′) that are part of 19-nucleotide terminal inverted repeats. The two LTRs are highly similar with only two mismatches and a $(TA)^n$ microsatellite that differs in length. The RTP is flanked by a direct repeat of 5 nucleotides that may be the result of a duplicated target site. The 5′ end of the coding region of the RTP contains a primer binding site (PBS) with similarity to the $tRNA_{Met}$, including the conserved TATCA motif that is recognised by the integrase (Langdon *et al.* 2000). At the 3′ end of the coding region a polypurine track (PPT) is present. The coding region contains two ORFs of 429 (ORF1) and 923 (ORF2) amino acids, respectively. ORF1 contains the $CX_2CX_4HX_xC$ conserved domain of RTP gag proteins (Gorelick *et al.*, 1988; Jentoft *et al.*, 1988) and the aspartyl protease domain as detected with InterProScan, and has highest homology with a Ty3-Gypsy type RTP from *O. sativa* (29% aa identity). ORF2 contains the three conserved domains reverse transcriptase, RNaseH and integrase, in the typical order of Ty3/Gypsy type of RTPs, and has highest homology with a sequence from *O. sativa* (60% aa identity).

Microsatellites The STADEN and Sputnik programs were used to analyse the 60K17 BAC clone sequence for the presence of microsatellites. Minimum lengths of 8 (mononucleotide and dinucleotide), 9 (trinucleotide), 12 (tetranucleotide) or 15 (pentanucleotide) nucleotides were used as criteria for the

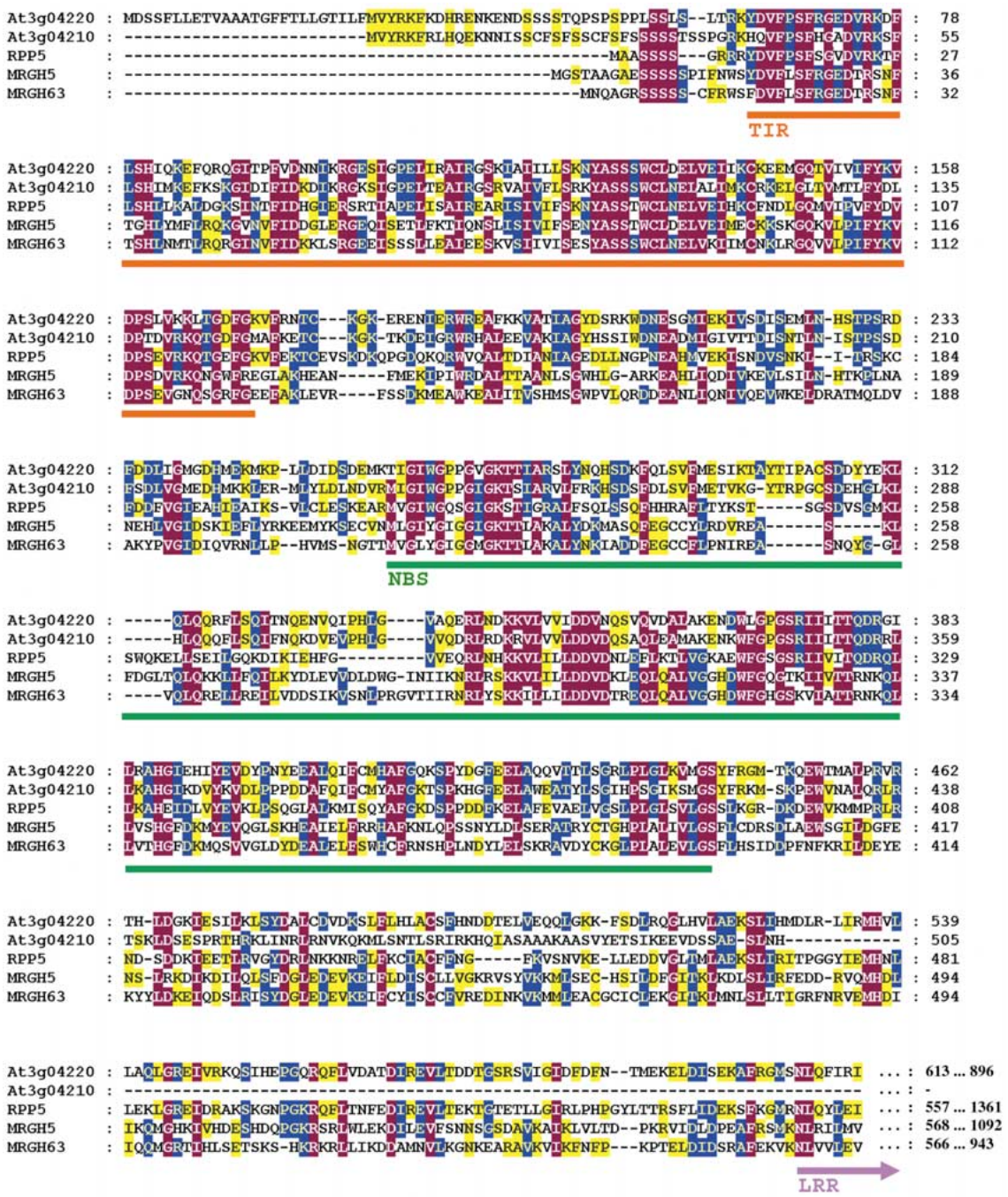

Figure 4. Alignment of partial amino acid sequences of the melon MRGH5 and MRGH63 genes from the melon resistance gene cluster with two *Arabidopsis* resistance gene homologous proteins (see also Figure 5) and the RPP5 resistance protein from *Arabidopsis*. Identical amino acids in 80–100% of the sequences are illustrated in purple, identical amino acids in 60% or 40% of the sequences are illustrated in blue and yellow, respectively. The C-terminal part of the proteins, containing the LRRs, is not shown.

Table 2. Analysis of microsatellites in the melon 60K17 BAC.

MS[a] type	No. MSs	kb/ MS	MS type	No. MSs	kb/ MS
Mononucleotide	*167*	*0.7*	*Tetranucleotide*	*19*	*6.2*
T/A	164	0.7	TAAA/TTTA	9	13.0
G/C	3	39.0	CCTT/AAGG	3	39.0
			CTTT/AAAG	2	58.4
Dinucleotide	*84*	*1.4*	GAGT/ACTC	1	116.9
TA/AT	64	1.8	TTAA/TTAA	1	116.9
TC/GA	16	7.3	TCCC/GGGA	1	116.9
TG/CA	4	29.2	TGAA/TTCA	1	116.9
			TTTG/CAAA	1	116.9
Trinucleotide	*57*	*2.1*			
TTA/TAA	25	4.7	*Pentanucleotide*	*8*	14.6
AAG/CTT	10	11.7	TTTTA/TAAAA	5	23.4
CCG/CGG	8	14.6	AGCCG/CGGCT	2	58.4
AAC/GTT	8	14.6	CAATA/TATTG	1	116.9
TCC/GGA	2	58.4			
ACC/GGT	1	116.9	MS Length (bp)	No. MSs	kb/ MS
AGC/GCT	1	116.9	8–11	241	0.5
TAC/GTA	1	116.9	12–19	65	1.8
TGA/TCA	1	116.9	≥20	29	4.0

[a]MS: microsatellite

analysis. Nearly half of the microsatellites are of the T/A mononucleotide type (Table 2). The majority of the microsatellites are located in intergenic regions and introns. In the gene-rich region (between 85 and 117 kb), the number of microsatellites in the intergenic regions and introns is lower than in the rest of the BAC clone. Comparison of the microsatellite frequency in the *Arabidopsis* genome (Casacuberta *et al.*, 2000) with that of the melon BAC revealed that the T/A and AT/TA microsatellites are approximately two times more frequent in melon (Table 2). When the location of microsatellites can be related to coding sequences (G-C rich or putative codons) the frequency is lower in melon than in *Arabidopsis*. In rice, microsatellites with a length of 12–19 occur every 1.9 kb, a similar value as the 1.8 for melon, while the ones with a length equal or greater than 20 occur every 16 kb in rice (Temnykh *et al.*, 2001) and in melon every 4.0 kb (Table 2).

Other types of repetitive elements A repeated DNA sequence was found with one copy (DR1b) just upstream of the Hsp2 gene and another copy (DR1a) just downstream of the Mki2 gene. Putative terminal inverted repeats (TIRs) were detected with the Palindrome program at the 5′ and 3′ ends of the repetitive element; 16 nucleotides with 4 mismatches for DR1a and 17 nucleotides with 4 mismatches for DR1b. The TIRs of DR1a and DR1b are not similar to each other and are followed by 10–20 nucleotides that are neither the same in the two copies. The size of DR1a and DR1b, including the putative TIRs, is 933 and 930 nucleotides, respectively. DR1a and DR1b, excluding the non-conserved TIRs, are 92% identical. The sequence contains several short ORFs, some of these having low similarity with retroelements from *Arabidopsis* (TBLASTX E-values ranging from 0.11 to 1.0). A low degree of similarity was also detected using BLASTX with a cassette chromosome recombinase A protein from *Staphylococcus aureus* (E-value 7.4) and with a hypothetical protein from *Chlamydophila pneumonia* (E-value 9.7). The DR1a copy is located 41 nucleotides downstream of the stop codon of the terminal exon of the Mki2 gene. The DR1b copy is located 735 nucleotides upstream of the putative start codon of the Hsp2 gene.

Approximately 40 sequences with the features expected for MITEs (Miniature Interspersed Transposable Elements) with the corresponding TIRs (Terminal

Inverted Repeats) were found. The sizes of the TIRs are equal or greater than 16 bp and are situated at approximately 600 bp distance. The internal sequences are AT-rich, do not contain ORFs, and no similarity was found between the putative MITEs. Future analysis consisting of searching for other copies of the elements and determination of the copy number in the melon genome can help to elucidate the features of MITEs in melon.

Analysis of micro-synteny between melon and Arabidopsis

Exon sequences of the predicted melon genes were compared with the *Arabidopsis* genome using the TAIR TBLASTX program. Two regions in the *Arabidopsis* genome were detected, which contained several sequences significantly similar to several genes from the melon BAC. In both regions, located in chromosomes 3 (At3) and 5 (At5) of *Arabidopsis*, five genes homologous to six genes from the melon BAC were detected. Three of the five genes from At3 and At5 were homologous to each other and to the same melon genes. The majority of the genes are present only one to three times in *Arabidopsis*, except for the MRGHs, Drzf, Sbp1 and Sbp2 genes (see Table 3). A detailed view of the syntenic regions is presented in Figure 5. The synteny starts with the Drzf melon gene and ends with the Rac melon gene located at the right border of the melon contig. The synteny between At3 and At5 and the melon BAC shows not only a conserved gene order, but also a conserved gene orientation. Moreover, the exon structure of the *Arabidopsis* and melon genes is highly similar.

In Table 3 are presented the BLAST homologies between the genes from *Arabidopsis* and melon in the syntenic region. The melon Rac gene is present in the contig and was detected by end sequencing of the 15J18 BAC clone (see Figure 2). The Rac melon gene fragment is homologous to plant rac proteins that induce superoxide production in mammalian cells and might play a role in the oxidative burst as reaction to pathogen attack (Hassanain *et al.*, 2000). In the syntenic region on At3, the At3g05250 gene is homologous to Sbp1 and Sbp2 from melon. Sbp1 and Sbp2 are 64% identical at DNA level. A similar situation was detected with At5 where the At5g27450 gene is homologous to Mki2 and Mki1 from melon. Mki1 is a 5′ gene fragment because it is situated at the BAC end, but it is likely that sequencing of the complete gene would show a gene similar to Mki2.

The genes in these regions of At3 and At5 that did not have homologous sequences in the melon BAC were compared to each other and two different genes with homology between the two *Arabidopsis* regions were detected. One encodes a putative amino acid transferase (At3g05190 and At5g27420) and the other one encodes a putative protein with unknown function (At3g05280 and At5g27490). The Hsp1 and Hsp2 melon genes do not have homologous sequences in these regions of At3 and At5, but we did detect two homologous genes on chromosome 1 of *Arabidopsis* (see Table 1). These genes, At1g54840 and At1g54850, are situated next to each other as in melon, and in the same orientation as in melon. TBLASTX analysis of the *Arabidopsis* genome with the melon Hsp genes showed four and three hits for hsp1 and hsp2, respectively.

In Figure 5, the nearest disease resistance genes on At3 and At5 are represented in green. Disease resistance genes from the same class as the MRGHs are present on At3 at approximately 350 kb from the syntenic region. Figure 4 shows the protein alignment between MRGH5, MRGH63, At3g04210, At3g04220, and the disease resistance gene RPP5 from *Arabidopsis*. The TIR and NBS domains are highly conserved while the LRR domains show a lower degree of similarity and different number of LRRs. On the other side of the syntenic region two *Cf*-2 like disease resistance genes (Dixon *et al.*, 1996) are present on At3. On At5, the nearest disease resistance gene is located at approximately 177 kb from the syntenic region, but it does not belong to the same class as the MRGHs.

Discussion

A BAC library from a dihaploid melon line coming from a cross between two distant varieties has been constructed. One of the advantages of this material is the uniformity of the allelic composition of the genome that may be useful for physical mapping using the library. The library has a 6 times coverage of the melon haploid genome with an average insert size of 139 kb. These features make it a useful tool for the molecular genetic analysis of the melon genome.

A contig of melon BACs has been constructed around the resistance gene homologue fragment MRGH63 (Garcia-Mas *et al.*, 2001). One of the BACs of the contig (60K17) has been fully sequenced in order to get a first insight on the physical structure of the melon genome. It appears that in this BAC, 14

Table 3. Identities between melon and *Arabidopsis* in syntenic regions.

Melon gene	No. hits in A.t. genome[a]	At3		At5	
		Acc. no.	*Identity*[b]	*Acc. no.*	*Identity*[b]
MRGH6[c]	>50	At3g04210	41%[d]	–	–
		At3g04220	48%[d]		
MRGH5	>50	At3g04210	35%	–	–
		At3g04220	31%		
MRGH63	>50	At3g04210	33%	–	–
		At3g04220	26%[e]		
Drzf	>50	At3g05200[g]	44%	At5g27420	44%
Spp	2	At3g05230	74%	At5g27430	77%
Sbp2[f]	>50	At3g05240[g]	41%[e]	–	–
Mki2	1	–	–	At5g27450	62%[e]
Sbp1	>50	At3g05240[g]	51%	–	–
Ctp	1	–	–	At5g27440	35%
Rzf	1	At3g05250	40%	–	–
Mki1[c]	1	–	–	At5g27450	56%[d,e]
Rac[c]	3	At3g05310	49%[d,e]	At5g27540	52%[d,e]

[a]TBLASTX E-value < 1e-10, [b]BLASTP identities, [c]Partial gene at BAC border, [d]Identity between partial proteins melon and A.t., [e]ClustalW identity, [f]Pseudogene with stop codons and deletions, [g]Best hit from A.t.

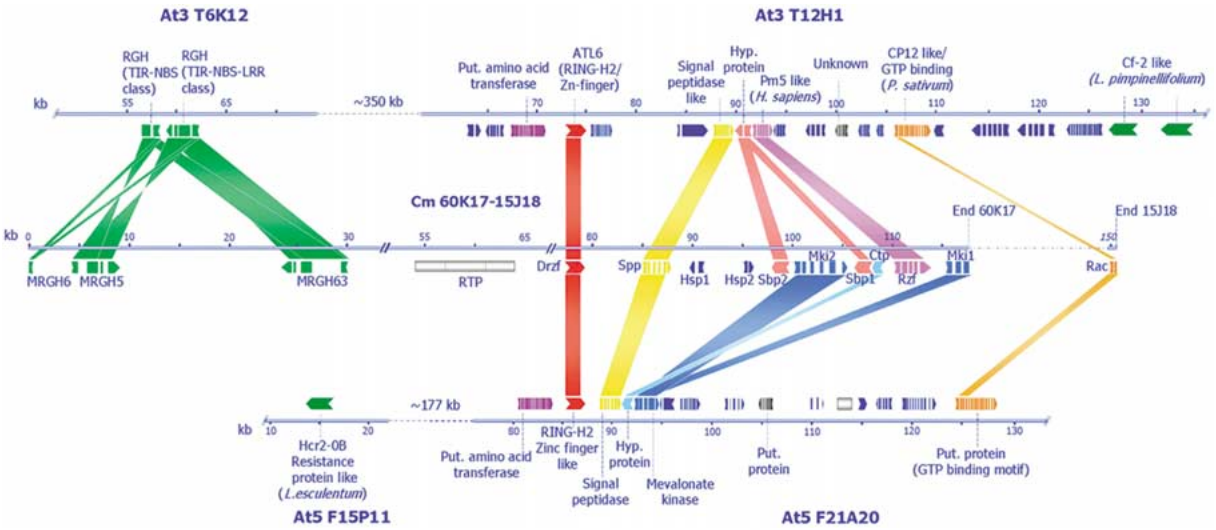

Figure 5. Microcolinearity between the melon 60K17 BAC and two regions of *Arabidopsis* on chromosomes 3 (At3) and 5 (At5). Gene elements are illustrated as explained in the legend of Figure 3. The proportion of exon and intron size is modified in some cases for better visualisation. Homologous genes in melon, At3 and At5, are illustrated with identical colours and connecting bars of the same colour. Genes that were homologous in the two regions of *Arabidopsis*, but that did not have a homologue in melon are only indicated with equal colours. The names of the BAC clones from *Arabidopsis* containing these regions is indicated behind the chromosome numbers. The nearest (putative) resistance genes in *Arabidopsis*, as well as the putative melon resistance genes, are coloured in green.

Table 4. Melon 60K17 BAC sequence characteristics.

Sequence length	116.877 bp
Number of genes	14
Gene density[a]	1 gene/9 kbp
Average exon length	455 bp
Average intron length	638 bp
Exons per gene	3.0 (range 1 to 6)
Average gene length[b]	2.6 kbp
GC content	32.4%
Microsatellites number	335 (3.4%)
Retrotransposons number	1 (8.6%)

[a]Partial genes at BAC borders counted as 0.5 gene.
[b]From start to stop codon.

genes are present, including at least four genes that may be related to a defence function and another fifth gene possibly related to defence that is included in the contig but in another BAC. These genes include three MRGH genes but also the Drzf gene and the partial Rac gene present in a border of the contig (see Figures 3 and 5). It is even possible that in bordering regions other genes related to resistance may be found as the results available in other species confirm that these genes tend to accumulate in clusters in plant genomes (Sudupak *et al.*, 1993).

The gene density in the 60K17 BAC is 1 gene per 9 kb, more than two times lower than in *Arabidopsis* (The *Arabidopsis* Genome Initiative, 2000), and similar to the values found in a first draft of the rice genome (1 gene per 6.1–10.3 kb) (Mayer *et al.*, 2001; Tarchini *et al.*, 2000). The intergenic length is also higher than in *Arabidopsis*. In average, the size of exons is larger, but at the moment the number of exons per gene is lower (3 in melon respect to 5 in *Arabidopsis*) and the average length of introns is around three times longer than in *Arabidopsis* (see Table 4). In general, it may be concluded that the longer length of the melon genome lies in its longer non-coding sequences, but it is difficult to generalise with having only the sequence of one BAC.

The analysis of coding sequences and repetitive elements of the melon 60K17 BAC can help us to deduce some characteristics of the melon genome. In this respect, CURE, the only retrotransposon found in the BAC, is a member of the Ty3/Gypsy family having large long terminal repeats and a relatively short internal fragment that includes the coding sequences. The most similar retrotransposon found in plants is *dell*, a retroelement from *Lilium* that is present in 13,000 copies in this genome (Smyth *et al.*, 1989). Microsatellites are also abundantly present in the melon genomic fragment analysed. It appears that they are not evenly distributed and more abundant in the regions that are less rich in genes than in the proximity of coding sequences (see Figure 2). It has been reported that the melon genome has a very repetitive genome (approximately 30%) (Bendich and Anderson, 1974). In the sequenced BAC only 13% of the sequence belongs to known families of repeated sequences as retroelements, direct repeats and microsatellites. However, other families specific of melon may appear and the MITE sequences have not been considered, in particular because, although they have the features of MITEs, no evidence was found that they are repeated in the melon genome. Some possible MITEs have been found and if they belong to repeated sequences they would account for approximately 30%. In this respect, MITEs would be essential to account for the repeats in the melon genome. In any case, a comparative analysis with other regions of the melon genome will be needed to confirm if they do belong to this type of repeated sequences.

When comparing the coding sequences found in the 60K17 BAC, it can be seen that most of the genes (12 out of 14) have a duplicate in the sequence with between 60% and 80% of identity at the amino acid level. However, generally they are not in tandem and a simple mechanism for the generation of the sequence is not easily found. This is especially clear when comparing the melon sequence with the sequenced *Arabidopsis* genome. Two regions of *Arabidopsis* with similar genes in the same order and orientation as in melon were found. In fact, the sequences similar to the MRGH resistance genes have more than 50 similar sequences in the *Arabidopsis* genome. However, a group of three genes allows to define the two regions in chromosomes 3 and 5 of *Arabidopsis* that contain a significant degree of microsynteny. The comparison of the genomic sequences reveals many different types of variations such as: genes that are only present in one of the *Arabidopsis* chromosomes but not in the other (Sbp and Rzf homologues in chromosome 3, and the Mki and Ctp homologues in chromosome 5 of *Arabidopsis*; Figure 5), genes that are present in the two *Arabidopsis* chromosomes but not in melon (amino acid transferase and unknown/putative protein genes; Figure 5), genes that are duplicated in melon (Sbp and Mki genes; Figure 5), and genes that have homologous sequences in other parts of the genome (e.g., melon heat shock protein genes with homologues in chro-

mosome 1 of *Arabidopsis*). For this reason, although some degree of localised synteny and colinearity is found between *Arabidopsis* and melon, it appears that many diverse mechanisms of gene reshuffling have acted in the evolutionary period separating the two species.

Similar analysis was made between *Arabidopsis* and rice, and microcolinearity was found (Mayer *et al.*, 2001). Microcolinearity was also found between a tomato region containing the *ovate*-gene and the *Arabidopsis* genome (Ku *et al.*, 2000), and the *Lateral Suppressor* regions of tomato and two Brassicaceae (Rossberg *et al.*, 2001). Tomato is a member of the Solanaceae family that phylogenetically is more distant to *Arabidopsis* than the Cucurbitaceae family (Soltis *et al.*, 1999). It may be concluded that the similarity of genes between *Arabidopsis* and melon is relatively high and allows to screen for genes similar in these species. The microcolinearity between melon and *Arabidopsis*, and between other dicots and monocots and *Arabidopsis*, suggests that mapping data from one species can be used for localisation of orthologous genes in another species. However, local genome rearrangements, as shown in this work, can prevent this strategy from being successful.

Acknowledgements

The authors would like to thank Ramon Dolcet from the 'Institut de Recerca i Tecnología Agroalimentaria' (IRTA) for providing seeds for 'PIT92', Chantel Scheuring, Scott McCuine and the rest of the staff at the TAMU BAC Center for their technical assistance for the BAC library construction, replication and filter production, Anna Vidal and Merçé Miquel for the DNA sequencing work, and Jordi Garcia-Mas and Pere Arús for their valuable comments about the manuscript. Funds for this project were provided by grant 2FD97-0286-(02-01) from the Spanish 'Comisión Interdepartamental de Ciencia y Tecnología' (CICYT) and grant 2000FI-00316 to Hans van Leeuwen from the Catalan 'Comisión Interdepartamental de Recerca i Tecnología' (CIRIT).

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D.R., *et al.* 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. 29: 37–40.

Arumuganathan, K. and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9: 208–218.

Bendich, A.J. and Anderson, R.S. 1974. Novel properties of satellite DNA from muskmelon. Proc. Natl. Acad. Sci. USA 71: 1511–1515.

Borodovsky, M. and McIninch, J. 1993. GeneMark: parallel gene recognition for both DNA strands. Computers & Chemistry 17: 123–133.

Burge, C. and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268: 78–94.

Casacuberta, E., Puigdomènech, P. and Monfort, A. 2000. Distribution of microsatellites in relation to coding sequences within the Arabidopsis thaliana genome. Plant Sci. 157: 97–104.

Church, G.M. and Gilbert, W. 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA 81: 1991–1995.

Danin-Poleg, Y., Reis, N., Baudracco-Arnas, S., Pitrat, M., Staub, J.E., Oliver, M., Arús, P., deVicente, C.M. and Katzir, N. 2000. Simple sequence repeats in Cucumis mapping and map merging. Genome 43: 963–974.

Dixon, M.S., Jones, D.A., Keddie, J.S., Thomas, C.M., Harrison, K. and Jones, J.D. 1996. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84: 451–459.

Durrant, W.E., Rowland, O., Piedras, P., Hammond-Kosack, K.E. and Jones, J.D.G. 2000. cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles. Plant Cell 12: 963–977.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. 2000. Predicting subcellular localisation of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300: 1005–1016.

Feinberg, A.P. and Vogelstein, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6–13.

Frijters, A.C.J., Zhang, Z., van Damme, M., Wang, G.L., Ronald, P.C. and Michelmore, R.W. 1997. Construction of a bacterial artificial chromosome library containing large EcoRI and HindIII genomic fragments of lettuce. Theor. Appl. Genet. 94: 390–399.

Garcia-Mas, J., van Leeuwen, H., Monfort, A., de Vicente, M.C., Puigdomènech, P. and Arús, P. 2001. Cloning and mapping of resistance gene homologues in melon. Plant Sci. 161: 165–172.

Gorelick, R.J., Henderson, L.E., Hanser, J.P. and Rein, A. 1988. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a 'zinc finger-like' protein sequence. Proc. Natl. Acad. Sci. USA 85: 8420–8424.

Hamilton, C.M., Frary, A., Xu, Y., Tanksley, S.D. and Zhang, H.B. 1999. Construction of tomato genomic DNA libraries in a binary-BAC (BIBAC) vector. Plant J. 18: 223–229.

Hassanain, H.H., Sharma, Y.K., Moldovan, L., Khramtsov, V., Berliner, L.J., Duvick, J.P. and Goldschmidt-Clermont, P.J. 2000. Plant rac proteins induce superoxide production in mammalian cells. Biochem. Biophys. Res. Commun. 272: 783–788.

Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouze, P. and Brunak, S. 1996. Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. Nucleic Acids Res. 24: 3439–3452.

Jentoft, J.E., Smith, L.M., Fu, X.D., Johnson, M. and Leis, J. 1988. Conserved cysteine and histidine residues of the avian myeloblastosis virus nucleocapsid protein are essential for viral

replication but are not 'zinc-binding fingers'. Proc. Natl. Acad. Sci. USA 85: 7094–7098.

Ku, H.M., Vision, T., Liu, J. and Tanksley, S.D. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. Proc. Natl. Acad. Sci. USA 97: 9121–9126.

Langdon, T., Seago, C., Mende, M., Leggett, M., Thomas, H., Forster, J.W., Thomas, H., Jones, R.N. and Jenkins, G. 2000. Retrotransposition evolution in diverse plant genomes. Genetics 156: 313–325.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P. and Bork, P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res. 30: 242–244.

Lukashin A. and Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26: 1107–1115.

Mayer, K., Murphy, G., Tarchini, R., Wambutt, R., Volckaert, G., Pohl, T., Düsterhöft, A., Stiekema, W., Entian, K., Terryn, N. *et al.* 2001. Conservation of Microstructure between a Sequenced Region of the Genome of Rice and Multiple Segments of the Genome of Arabidopsis thaliana. Genome Res. 11: 1167–1174.

Morales, M., Luís-Arteaga, M., Alvarez, J.M., Dolcet-Sanjuan, R., Monfort, A., Arús, P. and Garcia-Mas, J. 2002. Marker saturation of the region flanking the gene NSV conferring resistance to the Melon Necrotic Spot Carmovirus (MNSV) in melon. J. Amer. Soc. Hort. Sci. 127: 540–544.

Nicholas, K.B., Nicholas H.B. Jr. and Deerfield, D.W. II. 1997. GeneDoc: Analysis and Visualization of Genetic Variation, EMBNET.NEWS 4: 14, http://www.hgmp.mrc.ac.uk/embnet.news/vol4_ 2/genedoc.html.

Oliver, M., Garcia-Mas, J., Cardus, M., Pueyo, N., Lopez-Sese, A.L., Arroyo, M., Gomez-Paniagua, H., Arus, P. and de Vicente, M.C. 2001. Construction of a reference linkage map for melon. Genome 44: 836–845.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B. and Jones, J.D. 1997. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell 12: 821–832.

Reese, M.G., Eeckman, F.H., Kulp, D. and Haussler, D. 1997. Improved splice site detection in Genie. Proceedings of the First Annual International Conference on Computational Molecular Biology (RECOMB), Santa Fe, NM, ACM Press, New York.

Rice, P., Longden, I. and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16: 276–277.

Richter, T.E. and Ronald, P.C. 2000. The evolution of disease resistance genes. Plant Mol. Biol. 42: 195–204.

Rogozin, I.B. and Milanesi, L. 1997. Analysis of donor splice signals in different organisms. J. Mol. Evol. 45: 50–59.

Rossberg, M., Theres, K., Acarkan, A., Herrero, R., Schmitt, T., Schumacher, K., Schmitz, G. and Schmidt, R. 2001. Comparative sequence analysis reveals extensive microcolinearity in the *Lateral Suppressor* regions of the tomato, Arabidopsis, and Capsella genomes. Plant Cell 13: 979–988.

Salinas-Mondragón, R.E., Garcidueñas-Piña, C. and Guzmán, P. 1999. Early elicitor induction in members of a novel multigene family coding for highly related RING-H2 proteins in Arabidopsis thaliana. Plant Mol. Biol. 40: 579–590.

Salzberg, S.L., Pertea, M., Delcher, A.L., Gardner, M.J. and Tettelin, H. 1999. Interpolated Markov models for eukaryotic gene finding. Genomics 59: 24–31.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schmidt, R. 2002. Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Mol. Biol. 48: 21–37.

Smith, R.F., Wiese, B.A., Wojzynski, M.K., Davison, D.B. and Worley, K.C. 1996. BCM Search Launcher–An Integrated Interface to Molecular Biology Data Base Search and Analysis Services Available on the World Wide Web. Genome Res. 6: 454–462.

Smyth, D.R., Kalitsis, P., Joseph, J.L. and Sentry, J.W. 1989. Plant retrotransposon from Lilium henryi is related to Ty3 of yeast and the gypsy group of Drosophila. Proc. Natl. Acad. Sci. USA 86: 5015–5019.

Soltis, P.S., Soltis, D.E. and Chase, M.W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402: 402–404.

Song, W.Y., Pi, L.Y., Wang, G.L., Gardner, J., Holsten, T. and Ronald, P.C. 1997. Evolution of the rice Xa21 disease resistance gene family. Plant Cell 9: 1279–1287.

Sonnhammer, E.L.L., von Heijne, G. and Krogh, A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proceedings of Sixth Conference on Intelligent Systems for Molecular Biology, Menlo Park, CA, USA, AAAI Press, 175–182.

Staden, R. 1996. The Staden Sequence Analysis Package. Mol. Biotechnol. 5: 233–241.

Strong, S.J., Ohta, Y., Litman, G.W. and Amemiya, C.T. 1997. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-seperated partial digests of genomic DNA. Nucleic Acids Res. 25: 3959–3961.

Sudupak, M.A., Bennetzen, J.L. and Hulbert, S.H. 1993. Unequal exchange and meiotic instability of disease-resistance genes in the Rp1 region of maize. Genetics 133: 119–125.

Takai, R., Hasegawa, K., Kaku, H., Shibuya, N. and Minami, E. 2001. Isolation and analysis of expression mechanisms of a rice gene, EL5, which shows structural similarity to ATL family from Arabidopsis, in response to N-acetylchitooligosaccharide elicitor. Plant Sci. 160: 577–583.

Tarchini, R., Biddle, P., Wineland, R., Tingey, S. and Rafalski, A. 2000. The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12: 381–391.

Tatusova, T.A. and Madden, T.L. 1999. Blast 2 sequences – a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. 174: 247–250.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11: 1441–1452.

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796–815.

Wang, G.L., Holsten, T.E., Song, W.Y., Wang, H.P. and Ronald, P.C. 1995. Construction of a rice bacterial artificial chromosome library and identification of clones linked to the Xa-21 disease resistance locus. Plant J. 7: 525–533.

Wang, G.L., Ruan, D.L., Song, W.Y., Sideris, S., Chen, L., Pi, L.Y., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B.S., Whalen, M.C. and Ronald, P.C. 1998. Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10: 765–779.

Woo, S.S., Jiang, J., Gill, B., Paterson, A.H. and Wing, R.A. 1994. Construction and characterisation of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Res. 22: 4922–4931.

Xu, M., Song, J., Cheng, Z., Jiang, J. and Korban, S.S. 2001. A bacterial artificial chromosome (BAC) library of Malus floribunda 821 and contig construction for positional cloning of the apple scab resistance gene Vf. Genome 44: 1104–1113.

Zhang, H.B., Choi, S., Woo, S.S., Li, Z., and Wing, R.A. 1996. Construction and characterisation of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol. Breed. 2: 11–24.

Zhang, H.B., Zhao, X., Ding, X, Paterson, H. and Wing, R. 1995. Preparation of megabase-size DNA from plant nuclei. Plant J. 7: 175–184.