Theor Appl Genet (2000) 101:1109–1113

C.E. Vallejos · J.J. Malandro · K. Sheehy
M.J. Zimmermann

Detection and cloning of expressed sequences linked to a target gene

Received: 8 July 1999 / Accepted: 21 March 2000

Abstract DNA markers tightly linked to a target gene are essential starting points for positional cloning. We combined "differential display of mRNA" and "bulked segregant analysis" in order to detect and clone ten expressed sequences as markers linked to a virus resistance gene in *Phaseolus vulgaris*. The combination of these two procedures could be used in lieu of positional cloning, provided polymorphisms detectable by differential display exist in the target gene. Isolation of expressed sequences from specific chromosome regions can also be accomplished by combining these procedures.

Key words Differential display of mRNA · Bulked segregant analysis · Bean common mosaic virus · Linkage analysis

Introduction

The unavailability of a tranposon system in *P. vulgaris* leaves positional cloning as the only viable option for cloning genes that can only be recognized by the phenotype they impart on a plant. The first step in positional cloning is the construction of a high-density linkage map around the target gene. The construction of linkage maps saturated with molecular markers was proposed as a means to provide sufficient markers to initiate a chromosome walk anywhere in a genome. Such maps have been constructed with restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD) and, more recently, with amplified fragment length polymorphism (AFLP) markers (Phillips and Vasil 1999). However, the construction of these maps as a global approach to identify markers linked to a target gene is inefficient. Instead, physical and genetic approaches for the isolation of markers linked to a target gene have been developed. Physical approaches include the construction of a chromosome arm-specific library from microdissected chromosomes (Lüdecke et al. 1989; Schondelmaier et al. 1993) and chromosome sorting for the construction of chromosome-specific libraries (Dolzel 1994). On the other hand, target-oriented genetic approaches include the analysis of near-isogenic lines (NILs) (Young et al. 1988), and bulked segregant analysis (BSA) (Michelmore et al. 1991). These methods are based on the detection of polymorphisms at the DNA level. BSA is a versatile and simple method that requires three steps: (1) determination of the exact genotype of F_2 segregants at a target locus, (2) assembly of DNA pools for each of the two homozygous classes, and (3) screening for markers that distinguish the homozygous pools. Homozygosity of the pools for a given marker is interpreted as linkage between the marker and the target locus. Confirmation of linkage is performed via standard segregation and linkage analysis. This method has been used to identify markers for several disease resistance loci (Michelmore 1995). Polymerase chain reaction (PCR)-based markers, such as RAPDs (Williams et al. 1990) and more recently AFLPs (Vos et al. 1995), have been the primary choice for BSA.

The overall objective of our research program is the positional cloning of the *I* gene of *Phaseolus vulgaris*. A dominant allele of this gene confers resistance to bean common mosaic virus (BCMV) (Drijfhout 1978). Serotype B strains of BCMV induce systemic mosaic in *ii* genotypes but fail to infect *I*- plants, whereas serotype A strains of BCMV induce systemic mosaic in *ii* genotypes and systemic necrosis in *I*- genotypes. A number of dif-

Communicated by M.A. SaghaiMaroof

C.E. Vallejos (✉) · J.J. Malandro · K. Sheehy
Department of Horticultural Sciences,
and Plant Molecular and Cellular Biology Graduate Program,
University of Florida, 1143 Fifield Hall, Gainesville,
FL 32611, USA
e-mail: Vallejos@ufl.edu, Fax: +1 353 392-6479

M.J. Zimmermann
Centro Nacional de Pesquisa de Arroz e Feijao,
Empresa Brasileira de Pesquisa Agropecuaria,
Goiânia, GO, Brazil

Present address:
M.J. Zimmermann, FAO-SDRR, Office: C-692,
Via delle Terme di Caracalla, 00100-Rome, Italy

ferent approaches are being and have been used to identify markers linked to this gene.

Differential display of mRNA (DDMR) is based on the selective amplification of a small subset of mRNAs that have in common the last two bases preceding the poly A tail and a common decamer sequence at varying lengths upstream from the inception of the poly A tail. Comparing the same subset of mRNAs from tissues that are at different stages of development, from the same genotype, can lead to the identification of sequences that are differentially expressed (Liang and Pardee 1992). We have used the same principle for the comparison of subsets of mRNAs from tissues that were at the same stage of development but were extracted from different genotypes. Thus, the presence or absence of an amplicon can be interpreted as a polymorphism due to base substitutions in either of the two bases preceding the poly A tail or in the sequence homologous to the random decamer primer upstream from the poly A tail. Insertion/deletions events are also a possibility. Thus, the specific objectives of this project were to determine whether expressed sequences linked to the target gene can be cloned by combining DDMR with BSA.

Materials and methods

Plant material

The Andean BCMV-susceptible genotype Calima (*ii*) and the Mesoamerican BCMV-resistant genotype Jamapa (*II*) were used to generate a recombinant inbred (RI) family (Burr et al. 1988) of 76 lines. This RI family was used for BSA, and segregation and linkage analysis of expressed sequences identified by DDMR. RIs are obtained by taking an F_2 progeny to an advanced generation (F_7 or greater) by a program of single seed descent. This process leads to the fixation of each locus; the extent of fixation "f" at each locus is given by $f=1-(0.5)^{n-1}$, where n is the generation number. Another consequence of this process is the fixation of recombination in the population. BSA was conducted with RI lines from the F_5 generation that were already fixed at the *I* locus, while the linkage relationships between the *I* locus and BSA/DDMR-derived sequences were examined in the F_8 generation.

Typing at the *I* locus

The genotype at the *I* locus of each RI line was deduced from the inoculation responses of 20 seedlings from each line. This screening was performed twice, first at the F_5 level, and later at the F_8 level with those lines that were found to be segregating at the F_5 level. Screening at the F_5 level was conducted to identify lines to be used for BSA.

Seven- to ten-day-old seedlings were grown in Conviron E15 growth chambers (Conviron, Pembina, N.D.) under a 12-h photoperiod, a thermoperiod of 25°/18°C, and a PAR of 450 µmol m^{-2} s^{-1}. One of the primary leaves of each seedlings was sprinkled with carborundum and then gently rubbed with a camel hair brush that had been previously dipped in a suspension of NL3, a necrotic strain of BCMV (A generous gift from Dr M.J. Silbernagel, USDA/ARS IAREC). Five days after inoculation plants were screened for presence or absence of systemic necrosis. When 100% of the plants in a family displayed systemic necrosis, the genotype of the line was assumed to be fixed for the dominant allele *I*. In contrast, an RI line was assumed to be fixed for the recessive allele *i* when all of its 20 seedlings failed to develop systemic necrosis and developed systemic mosaic instead.

Preparation of RNA samples

In addition to the parental genotypes, 4 plants of each of 12 lines of each genotypic class (*II* and *ii*) were grown in Conviron E15 growth chambers under the conditions described above. Plants were grown in half-gallon pots in Metromix 220 media with a fertilizer supplement. Total RNA was extracted from fully expanded healthy leaves. Approximately 8–12 g of leaf tissue was ground in liquid nitrogen and extracted with 4 vol. of extraction buffer (125 m*M* Tris HCl pH 7.6, 6.25 mM NaEDTA, 2.5% Sarkosyl, 500 mM NaCl, and 0.5% $NaSO_3$). The homogenate was subsequently extracted with an equal volume of phenol/chloroform (4:1) and then with chloroform/octanol (24:1); in each extraction the aqueous and organic phases were separated by centrifugation in a JA14 Beckman rotor at 8,000 rpms for 20 min at 4°C. The RNA from the aqueous phase was precipitated by the addition 12 *M* LiCl to a final concentration of 2 *M*, incubation on ice for at least 8 h, and a 20-min centrifugation at 8,000 RPMs at 4°C. The RNA precipitate was re-suspended in 25 ml of 2 M LiCl and spun over a 5-ml pad of 4 *M* LiCl at 12,000 g_{max} in a JA20 Beckman rotor. The RNA pellet was dissolved in TE buffer and precipitated by the addition of 1/9 vol. of 3 *M* Na acetate and 2.5 vol of ethanol. The RNA pellet was then dissolved in TE buffer, and the RNA concentrations were determined spectrophotometrically (A_{260}). RNA samples for BSA were prepared by pooling equal amounts of total RNA from each of the 12 lines with the same genotype, *ii* or *II*. Poly(A)$^+$ RNA was isolated by oligo dT chromatography (Aviv and Leder 1972) from each of the two bulks and from the parental genotypes.

DDMR and cloning of selected sequences

DDMR of the four samples was conducted exactly as described by Liang and Pardee (1992). Briefly, the first cDNA strand was prepared in a 50-µl reaction containing 3.75 µg of poly (A)$^+$ RNA, 750 U M-MLV reverse transcriptase (Life Technologies, Rockville, Md.), 1×RT buffer, 20 µ*M* dNTPs, 10 m*M* DTT, and 2.5 µ*M* of a single $(dT)_{11}NN$ primer. The RT reaction was run at 35°C for 1 h. A 2-µl aliquot of the RT reaction was used for a 20-µl PCR reaction that also contained 1.5 m*M* $MgCl_2$, 20 µ*M* dNTPs, 4.62 Mbq ^{35}S–α thio-dATP, 2.5 µ*M* $(dT)_{11}NN$ primer, 0.5 µ*M* random primer (kit E, Operon Technologies, Alameda, Calif.), 1 U *Taq* DNA polymerase (Life Technologies, Rockville Md.), and 1×reaction buffer. PCR was carried out in a Perkin-Elmer GeneAmp96 thermocycler under the following protocol: 2 min at 94°C, 35 cycles of 94°C for 30 s, 42°C for 30 s, 72°C for 30 s, and 7 min at 72°C. The amplification products were denatured by the addition of 12 µl of denaturation solution (98% formamide, 5 m*M* EDTA, 0.025% bromophenol, and 0.025% xylene cyanole), heating at 95°C for 5 min, and immediate quenching on ice. An aliquot of 3 µl was resolved by urea-PAGE. Gels were prepared with 6% Long Ranger (AT Biochemical, Malvern Pa.), 7 *M* urea, and 1×TBE. Samples were electrophoresed for 2.5 h at a constant power of 50 W. Gels were dried onto filter paper, and labeled DNA sequences were visualized by autoradiography in Kodak X-Omat film.

DNA fragments that appeared to be unique to one bulked sample and its corresponding parental genotype were isolated and cloned according to the following protocol. The X-film was first aligned with the dried gel, the selected band was excised with a scalpel, and the dried gel segment was transferred to a microcentrifuge tube. The dried gel and backing filter paper were allowed to hydrate in 100 µl of water for 10 min. After hydration, the sample was boiled for 15 min and then spun at 16,000 g_{max} for 2 min at 4°C. The supernatant was transferred to a new tube, and the DNA was precipitated after the addition of 10 µl 3 *M* Na acetate and 5 µl 5 mg/ml glycogen (Boehringer Mannheim Biochemicals, Indianapolis, Ind.), 450 µl of ethanol and incubation at –80°C for 30 min. The DNA was pelleted at 16,000 g_{max} for 15 min at 4°C and the pellet washed in 70% ice-cold etahnol, dried, and resuspended in 10 µl H_2O. A 5-µl aliquot was used for a 40-µl PCR reaction under the conditions described above but without the radio-

labeled nucleotide. PCR products were checked by standard agarose gel electrophoresis. When low yield was obtained, a 10% volume of the first reaction was used for a second PCR reaction.

Cloning of the PCR products was carried out by either a T/A or a blunt end cloning procedure. In the T/A cloning procedure, the cloning vector pBlueScript was prepared by digestion with *Eco*RV, followed by phenol and chloroform extraction, and then by incubation of the vector with *Taq* polymerase and 5 mM of dTTP for 2 h at 75°C. Ligations of PCR-derived inserts to T-vector were performed in a 10-µl reaction volume according to the following conditions: 200–300 ng T-vector, 100 ng insert DNA, 13% polyethylene glycol (8,000 MW), 1 µ*M* hexammine cobalt chloride, 5 U ligase (Promega, Madison, Wis.), 1×ligase buffer, 0.5 m*M* ATP. The mixture was incubated overnight at 14°C. Ligation products were transformed into XL-1 Blue supercompetent cells (Stratagene, La Jolla, Calif.), and transformants were selected on X-Gal SOB plates containing 20 µg/ml ampicillin and 80 µg/ml methicillin . PCR products were also cloned using the PCR cloning kit from Stratagene according to the kit's manual.

Segregation and linkage analysis

Extraction of genomic DNA and Southern hybridization and analysis were conducted as described previously (Vallejos et al. 1992). The DDMR-derived cDNA clones were used as probes to detect RFLPs between the parental genotypes (Fig. 1C) with the aid of 16 restriction enzymes (Set 1: *Bam*HI, *Dra*I, *Eco*RI, *Eco*RV, *Hin*dIII, *Xba*I; Set 2: *Alu*I, *Bgl*II, *Bst*EII, *Hae*III, *Hpa*II, *Msp*I, *Pst*I, *Rsa*I, *Sst*I, *Taq*I). Informative probe-restriction enzyme combinations were used to follow the segregation of loci encoding the cDNAs (Fig. 1D). The linkage relationships between the *I* locus and the cDNAs were examined in the F_8 generation of the RI family. Linkage analysis was conducted with MAPMAKER/EXP 3.0 (Lander et al. 1987; Lincoln et al. 1992). Map distances were calculated with the Kosambi function.

Results

One hundred and two primer pair combinations were used to compare the DDMR profiles of Calima (*ii*), the susceptible bulk (*ii*), the resistant bulk (*II*), and Jamapa (*II*). On average, 40 amplification products between 140 and 500 nt long were detected with each primer combination. A total of approximately 4,000 amplification products were detected in this project. In general, all four lanes displayed similar amplicon profiles. Nonetheless, numerous differences in the profiles of the parental genotypes were detected. These were noted by the presence of an amplicon in one genotype, while the same amplicon was clearly absent in the other parental genotype. These differences could be attributed to differences in the expression profiles of the parental genotypes or to polymorphisms detected at the RNA level. In most cases, the genotype-specific amplicon was also present in both the susceptible and the resistant bulks. This pattern indicted that whatever the basis for the difference was, it was not associated in any way with the target gene *I*, the gene used as the only criterion for the assembly of the bulks. On average, 5–6 genotype- specific amplicons not associated with the *I* gene were detected with each primer pair. Among all the genotype-specific amplicons, 48 showed a pattern that was indicative of linkage to the *I* locus. Of these, 31 were detected only in the Calima and susceptible lanes, and 17 only in the Jamapa and resistant lanes. These sequences were excised, eluted from the dry gel, PCR-amplified, and cloned (Fig. 1A, B).

The BSA/DDMR-derived cDNA clones were used as probes for segregation and linkage analyses to determine whether their respective encoding loci were linked to the *I* locus. This objective was accomplished in three steps. First, we identified the restriction enzyme that detected polymorphisms between the parental genotypes for a particular cDNA clone (Fig. 1C). Second, we digested genomic DNAs of the RI family with a suitable restriction enzyme and prepared Southern blots that were hybridized with the appropriate cDNA (Fig. 1D). Finally, the linkage relationships between the *I* locus and the loci encoding the BSA/DDMR-derived cDNA clones were investigated using MAPMAKER 3.0. The segregating loci were grouped using a minimum linkage criteria of 40 cM and an LOD of 5.0. Grouped loci were ordered with an LOD of 3.0.

Only 40 of the 48 cDNA clones detected RFLPs between the parental genotypes with at least 1 of 16 different restriction enzymes. Of the 40 clones, 2 yielded identical restriction and segregation patterns and were assumed to be alleles from the same locus. Of the 39 remaining clones, 4 detected duplicate loci, and 35 clones detected single-copy loci. Thus, a total of 43 segregating loci were analyzed. Ten (25%) cDNAs for which polymorphisms were detected were linked to the *I* locus in linkage group *D* (Fig. 1 E). We had previously assigned the *I* locus to this linkage group by low-resolution mapping with previously mapped genomic clones in an F_2 progeny (A301×Calima) (Zimmermann and Vallejos 1992). All of the cDNA clones linked to the *I* gene were located on the proximal side suggesting that this locus is near the telomere. The linked expressed sequences were found in a span of 32 cM. Expressed sequences not linked to the *I* locus were detected in each 1 of the 11 linkage groups of the common bean. However, five clusters of 3 loci each were detected.

Discussion

In contrast to results obtained with low multiplexing approaches such as RAPD markers screening, the combination of BSA and DDMR appeared to be a very efficient approach for the isolation of expressed sequences linked to a target gene. The number of linked expressed sequences could be increased by increasing the number of primer combinations for DDMR and by using as many tissues at different stages of development as possible. The locus of the expressed sequence closest to the target gene was 6.8 cM, or 3.6 Mbp, according to a previous calculations of the average relationship between physical and map distances in the bean genome (Vallejos et al. 1992). However, the distal position of the *I* locus suggests that it may be near a subtelomeric region where high rates of recombination and a concomitant decrease in the Kb/cM ratio are expected. This case is strength-

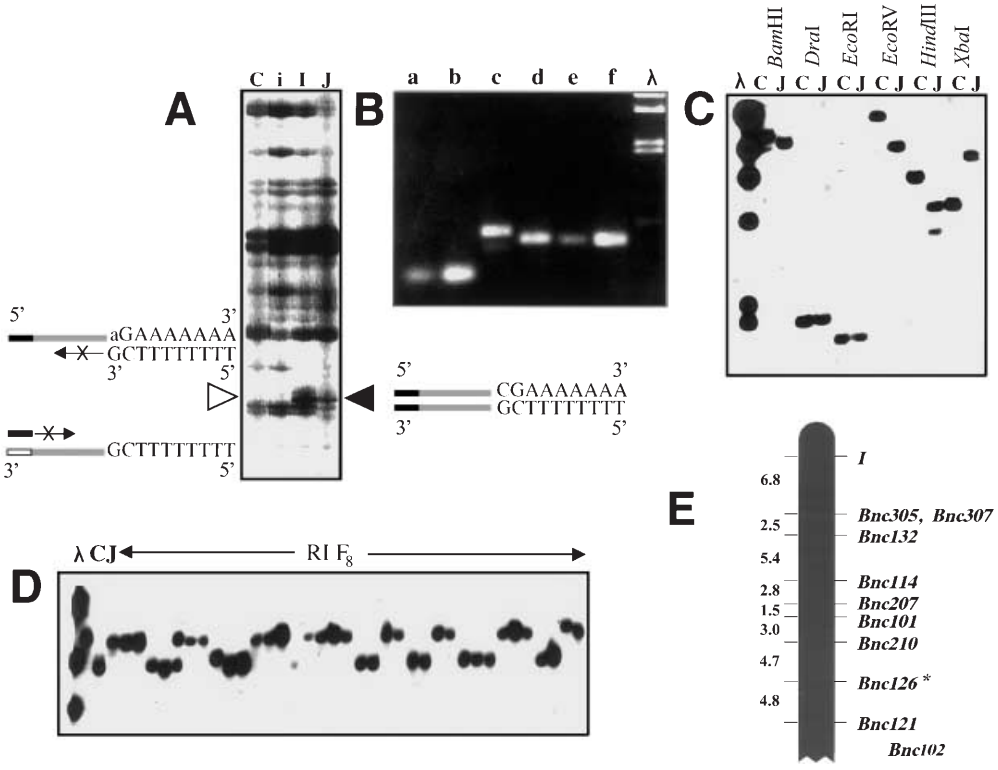

Fig. 1A–E Isolation of expressed sequences linked to the I locus by a combination of BSA and DDMR. **A** Sector of urea-PAGE gel showing an example of amplification products of Calima (*C*), susceptible pool (*i*), resistant pool (*I*), and Jamapa (*J*). An amplification product unique to the resistant pool and the parental Jamapa genotype is marked by a *black arrow head*; absence of the amplicon band in both Calima and the susceptible pool is assumed to be due to polymorphisms at either the nucleotides preceding the poy A tail or in the sequence that corresponds to the random decamer upstream. **B** Differential amplification products (*a–f*) were eluted out of the urea-PAGE gel, amplified by PCR, and displayed on an agarose gel; these fragments were later cloned in BlueScriptII. **C** Identification of RFLPs via Southern hybridization of genomic DNA from the parental genotypes Calima (*C*) and Jamapa (*J*) with the clones obtained in **B**. **D** A RI family of the cross (Jamapa×Calima) was used for segregation and linkage analysis. **E** Linkage map in the region of the *I* locus. Markers were assigned to this linkage groups with an LOD >5.0, and were ordered with an LOD >3. *The two alleles of this locus were cloned independently. The Kosambi function was used to calculate map distances

ened by the fact that all the expressed sequences linked to *I* were found on the proximal side.

The DDMR/BSA combined approach could be used to isolate expressed sequences in selected regions of the genome in organisms that already have detailed RFLP maps. Furthermore, DDMR/BSA could be seen potentially as an alternative to standard positional cloning, provided that allelic differences between parental genotypes at the target locus could be detected via DDMR.

The detection and cloning of sequences not linked to the *I* locus could be due to either artifacts inherent to DDMR (Sun et al. 1994) or to differential gene expression controlled by the *I* locus. It is possible that many of the cloned sequences not linked to the *I* locus were indeed artifacts. However, an argument against this possibility is the fact that the presence or absence of an amplification product in a DDMR gel was detected in both the parental genotype lane and that of the bulked samples with the same phenotype. On the other hand, in support of the possibility that the *I* locus may be controlling the expression of at least some of the sequences in question is the fact that some of them were found in clusters. There are several examples in the plant literature where disease resistance-related sequences can be found in clusters (Bent 1996; Kanazin et al. 1996: Yu et al. 1996; Rivkin et al. 1998). Alternatively, some of the clusters could be explained as the result of systematic experimental errors. We are currently testing the possibility of differential gene expression that may be controlled by the *I* gene.

Acknowledgments We want to thank Dr. L.C. Hannah, Dr. R.J. Ferl, and Dr. S.M. Schuster for reviewing this manuscript and offering many valuable suggestions. This work was supported in part by USDA grant 9201766. M. J. Zimmermann was supported by the Brazilian National Research Council and EMBRAPA. Florida Agricultural Experiment Station Journal Series No. R-06694. All experiments were conducted in compliance with current laws and regulations of the United States of America.

References

Aviv H, Leder P (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid-cellulose. Proc Natl Acad Sci USA 69:1408–1412

Bent AF (1996) Plant disease resistance genes: function meets structure Plant Cell 8:1757–1771

Burr B, Burr FA, Thompson KH, Albertsen MC, Tuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Dolzel J, Lucretti S, Schubert I (1994) Plant chromosome analysis and sorting by flow cytometry. Crit Rev Plant Sci 13:275–309

Drijfhout E (1978) Genetic interaction between *Phaseolus vulgaris* and bean common mosaic virus with implications for strain identification and breeding for resistance. Agricultural Research Reports 872. Institute for Horticultural Plant Breeding, Wageningen, The Netherlands

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93:11746–11750

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Liang P, Pardee AB (1992) Differential display of eukaryotic mRNA by means of the polymerase chain reaction. Science 257:967–971

Lincoln SE, Daly M, Lander ES (1992) Constructing genetic maps with MAPMAKER/EXP 3.0. Whitehead Institute Technical Report. 3rd edn. Whitehead Technical Institute, Cambridge, Mass.

Lüdecke H-J, Senger G, Claussen U, Hortsthempke B (1989) Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. Nature 338:348–350

Michelmore RW (1995) Molecular approaches to manipulation of disease resistance genes. Annu Rev Phytopathol 15:393–427

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Phillips RL, Vasil IK (eds) (1999) DNA-based markers in plants, 2nd edn. Kluwer Academic Publ, Dordrecht, The Netherlands

Rivkin, MI, Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. Genome 42:41–47

Schondelmaier J, Martin R, Jahoor A, Graner A, Koop H-U, Herrmann RG, Jung C (1993) Microdissection and microcloning of the barley (*Hordeum vulgare* L.) chromosome 1HS. Theor Appl Genet 86:629–636

Sun Y, Hegamyer G, Colburn NH (1994) Molecular cloning of five messenger RNAs differentially expressed in pre-neoplastic and neoplastic JB6 mouse epidermal cells: one is homologous to human tissue inhibitor of metalloproteinase-3. Cancer Res 54:1139–1144

Vallejos CE, Sakiyama NS, Chase CD (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131: 733–740

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

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Young ND, Zamir D, Ganal MW, Tanksley SD (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm–2a* gene in tomato. Genetics 120:579–585

Yu YG, Buss GR, Maroof MA (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA 93:11751–11756

Zimmermann MJ, Vallejos CE (1992) Genomic localization of monogenic dominant resistance to bean common mosaic virus in *Phaseolus vulgaris.* In: Plant Genome I. (Int Conf Plant Genome). Abstr. 164, pp 56. http://www.intl-pag.org/pag/1/abstracts/22pg1.html.