Theor Appl Genet (2003) 107:191–202
DOI 10.1007/s00122-003-1239-7

D. N. Kuhn · M. Heath · R. J. Wisser · A. Meerow · J. S. Brown · U. Lopes · R. J. Schnell

Resistance gene homologues in *Theobroma cacao* as useful genetic markers

Received: 1 May 2002 / Accepted: 17 October 2002 / Published online: 21 March 2003

Abstract Resistance gene homologue (RGH) sequences have been developed into useful genetic markers for marker-assisted selection (MAS) of disease resistant *Theobroma cacao*. A plasmid library of amplified fragments was created from seven different cultivars of cacao. Over 600 cloned recombinant amplicons were evaluated. From these, 74 unique RGHs were identified that could be placed into 11 categories based on sequence analysis. Primers specific to each category were designed. The primers specific for a single RGH category amplified fragments of equal length from the seven different cultivars used to create the library. However, these fragments exhibited single-strand conformational polymorphism (SSCP), which allowed us to map six of the RGH categories in an F_2 population of *T. cacao*. RGHs 1, 4 and 5 were in the same linkage group, with RGH 4 and 5 separated by less than 4 cM. As SSCP can be efficiently performed on our automated sequencer, we have developed a convenient and rapid high throughput assay for RGH alleles.

Keywords Resistance gene homologue · *Theobroma cacao* · Marker-assisted selection · *Crinipellis perniciosa* · SSCP

Communicated by D.B. Neale

M. Heath · R. J. Wisser · A. Meerow · J. S. Brown · R. J. Schnell (✉)
USDA-ARS, Subtropical Horticulture Research Station, 13601 Old Cutler Road, Miami, FL 33158, USA
e-mail: miars@ars-grin.gov
Tel.: +1-305-2543611
Fax: +1-305-9696410

U. Lopes
Centro de Pesquisas do Cacau (CEPEC)/CEPLAC, Caixa Postal 7.45600-000, Itabuna, Bahia, Brazil

D. N. Kuhn
Department of Biological Sciences, Florida International University, University Park, Miami, FL 33199, USA

Introduction

Theobroma cacao L. is an important tropical tree crop whose seeds (beans) are used to produce cocoa butter and chocolate. Most cacao cultivation occurs on small subsistence-level farms where harvested beans serve as the primary cash crop. During the past two decades the cacao industry in Central and South America has undergone a severe crisis. In Brazil, production has been devastated by an epiphytotic caused by witches' broom [*Crinipellis perniciosa* (Stahel)]. This aggressive basidiomycete is responsible for the decrease in production from 405,000 metric tons in 1986 to less than 130,000 metric tons in 1998. Another basidiomycete, *Moniliophthora roreri* (Cif. and Par.), the causal agent of frosty pod rot, is an extremely destructive pathogen which has caused serious losses in Ecuador, Colombia, Peru and Costa Rica. In 1998, the USDA-ARS revived a cacao genetics program based at the Subtropical Horticulture Research Station in Miami, Florida. The primary goal of the program is to assist with the development of a biotechnology based approach to solve these two destructive disease problems.

Cacao breeding through traditional methods is a slow process and, although much variability exists, most breeding work has utilized only a narrow genetic base (Motamayor et al. 2002). Breeding programs have been supported at many national institutes in the Western Hemisphere. However, except for a few selected clones, very little superior planting material is available for farmers (Hunter 1990). Cacao poses a challenge in regard to improvement through traditional breeding methods. Yield trials are costly and take from 5 to 8 years to complete (Irizarry and Rivera 2002). Many cultivars are self-incompatible, which makes it difficult to produce F_2 populations. Because of the length of time and the large areas needed for such breeding experiments, maintenance of families has been disrupted by political and programmatic changes. Most planted cultivars are mixed hybrid varieties that have been reproduced in seed gardens. Their pedigree and genotype may be uncertain or unknown

(Hunter 1990). Finally, disease resistance in *T. cacao* is multigenic, making the discovery of individual genes responsible for resistance unlikely (Pires et al. 1996).

Marker-assisted selection (MAS) can overcome some of these concerns by developing molecular markers associated with the desired phenotype, in this case disease resistance (Lande 1992). These markers can be used to pre-select material for the assay of disease resistance or for breeding before the phenotype is expressed. For example, by identifying a molecular marker associated with pod rot, one could identify material for breeding without having to wait until it had actually produced fruit to assay for disease resistance. Such measures should greatly reduce the cycle time for selection of cacao genotypes.

The development of polymorphic molecular markers linked to loci controlling disease resistance is critical. In addition to standard molecular markers (e.g., SSR, RAPD and AFLP), we have developed molecular markers using the candidate-gene approach (Gentzbittel et al. 1998). In such an approach, known genes that affect similar processes in other plants are examined. For example, one might look for genes that code for enzymes in anthocyanin biosynthesis to find genetic markers for flower or fruit color (Deng and Davis 2001). Resistance gene homologues are appealing candidate genes for several reasons. One can amplify resistance gene homologues from any plant with degenerate primers (Aarts et al. 1998; Shen et al. 1998). Importantly, in plants where the genome has been extensively mapped, resistance genes and resistance gene homologues are observed to be clustered (Kanazin et al. 1996; Aarts et al. 1998; Meyers et al. 1998a). In addition, identifying a marker for a single cluster could potentially generate a marker for resistance to several different diseases because adjacent genes in a cluster may condition resistance to radically different pathogens (e.g., virus and nematodes) (van der Vossen et al. 2000).

A resistance gene homologue (RGH) is a DNA sequence that contains a nucleotide binding site (NBS) motif and a leucine-rich repeat (LRR) that shares significant identity of the amino-acid sequence with known and mapped resistance genes from other plants (Hammond-Kosack and Jones 1997; Hulbert et al. 2001). The NBS region has some highly conserved portions from which degenerate primers have been designed (Aarts et al. 1998; Shen et al. 1998). These primers allow the amplification of sequences (RGHs) from both monocots and dicots. Thus, these primers have proved useful in amplifying resistance gene homologues from a variety of plants whose genetic structure is unknown or only poorly known, such as cacao, reported here, and coffee (Noir et al. 2001).

Once RGHs have been amplified using these primers, the challenge is to turn these fragments into useful genetic markers, i.e., molecular markers that represent a single locus on the chromosome and are polymorphic within the target population. We report here on the isolation and characterization of 74 unique RGHs from cacao and on the use of single-strand conformational polymorphism (SSCP) to establish the segregation of six classes of these markers in an F_2 population segregating for resistance to witches' broom disease. SSCP is quite frequently used in medical diagnosis and animal studies but has been neglected as an analysis tool in plants and ecological studies (Sunnucks et al. 2000). SSCP has typically been analyzed on gel-based systems but, here, we report on the use of the ABI3100 16 capillary automated genetic analyzer to generate the segregation data for the RGH.

Materials and methods

T. cacao cultivars used in this study

Sca6 (MIA1), EET400, Pound 12, GS46, Amelonado, 106R and IMC67 × Sca12 were used to prepare the RGH library. Subsequent genotyping has revealed that the Sca6 cultivar representative in Miami is not identical to the Sca6 cultivars from Brazil or the Reading Germplasm Collection. Therefore, the Miami Sca6 cultivar will be described as MIA1. These cultivars are a part of the National Germplasm Repository at the USDA-ARS station in Miami.

Leaf material from Sca6, ICS1, F_1 and P1-P192 was obtained from CEPEC/CEPLAC, Itabuna, Brazil. These plants are the parents (Sca6, ICS1), a single F_1 tree (TSH516) from the cross and 146 F_2 progeny (Ahnert 2000) from the selfing of the single F_1 tree, respectively.

DNA isolation from leaves

DNA was isolated from fully expanded, green leaves of *T. cacao*. Fresh leaf tissue (200 mg) was homogenized by disruption with beads in a FastPrep FP120 (BIO101/Savant) and DNA was isolated following the FastPrep Kit procedure (BIO101, Qbiogene, Inc., Carlsbad, Calif.).

PCR with RGH degenerate primers

DNA was amplified with either the RG1 and RG2 primers from *Arabidopsis* (Aarts et al. 1998), the PLOOPGA and GLPL6 primers from lettuce (Shen et al. 1998), or in one experiment with Pound 12, the RG1 and GLPL6 primers. Amplification conditions were: 1 × PCR buffer with 1.5 mM of $MgCl_2$, 1 μM of primers, 200 μM of dNTPs, 1 U of Ampli*Taq* DNA Polymerase (Perkin Elmer) and approximately 10 ng of template DNA. Thermocycler (MJ Research) parameters were: 95 °C for 5 min, followed by 30 cycles of 95 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min.

Library preparation and screening

Amplification products from all seven cultivars amplified with either RG1-RG2 or PLOOPGA-GLPL6 and those from the single amplification with RG1-GLPL6 were pooled and ethanol precipitated. DNA was cloned into the pCR4-TOPO plasmid as directed in the kit protocol (Invitrogen). Individual colonies were grown in modified L broth for plasmid isolation and for storage in 50% glycerol at –80 °C. The 605 clones of the library were grown in grids on agar plates, transferred to 9-cm nylon disks and prepared for hybridization as described in Molecular Cloning (Sambrook and Russell 2001). The amplified chloroplast DNA was labeled with α^{32}P-dATP (3,000 Ci/mmol, NEN) using a nick translation kit (GibcoBRL-Life Technologies). Hybridization was at 65 °C in a hybridization oven. After washing at 65 °C, disks were exposed to Kodak X-Omat film for 48 h. Colonies that hybridized to the probe

were removed from the library. Remaining colonies were screened for insert size by amplifying the plasmid DNA with the M13 forward and reverse primers (New England Biolabs) and determining the size by electrophoresis on 2.0% agarose gels in 0.5 × TBE at 120 V. Plasmids containing the putative RGH DNA were amplified on an ABI877 Integrated Thermal Cycler. Amplification conditions were: 1 × PCR buffer with 1.5 mM of $MgCl_2$, 0.4 μM M13 primers, 200 μM of dNTPs, 0.2 U of Ampli*Taq* DNA Polymerase (Perkin Elmer) and approximately 40 ng of plasmid DNA. Thermocycler parameters were: 95 °C for 5 min, followed by 30 cycles of 95 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min and one cycle of 72 °C for 10 min.

DNA sequencing and analysis

Plasmid DNA was isolated with the Wizard Plus SV Minipreps kit (Promega). Sequencing reactions consisted of approximately 300 ng of plasmid DNA in a 10 μl reaction with 2 μl of Big Dye reaction mix (ABI), 1 μl of 5 × Sequencing Buffer (ABI) and 3 pg of either the T3 or T7 sequencing primers (New England Biolabs). Sequencing reactions were prepared on an ABI877 Integrated Thermal Cycler in 96-well racks. Reaction products were ethanol precipitated as described in the ABI manual, centrifuged in an IEC CentraMP4 centrifuge with Rotor 244 at 1,400 × g for 45 min. Products were dried by inverting the plates and centrifuging at 700 × g for 1 min. Samples were sequenced on an ABI3700 either at USDA, Ft. Pierce, Fla. or USDA, Beltsville, Md. Output from the ABI3700 was assembled using Vector NTi Suite 6 (Informax). Assembled sequences were trimmed for vector contamination and compared against the GenBank database using the BLAST program (Altschul et al. 1997). Nucleotide sequences were translated into amino-acid sequences using either Vector NTi Suite 6 or GCG (Genetics Computer Group, Madison, Wis.). The amino-acid sequences were compared against the GenBank database using the BLASTp program.

Amino-acid sequences with E scores (expectation scores) less than e^{-15} for matches with known NBS/LRR homologues were designated as resistance gene homologues. Sequences were submitted to GenBank and are accession numbers AF402695–AF402768. Additional allele sequences were obtained by amplifying the DNA of Sca6, ICS1, TSH516 and individuals from the F_2 population, cloning the amplicons in pCR-4 TOPO and sequencing as described above. These allele sequences were submitted to GenBank and are accession numbers AY050259–AY050268. RGH nucleotide sequences were compared and aligned using PileUp in GCG. Multiple sequence files (MSF) from the aligned nucleotide sequences were used to generate an uncorrected pairwise distance matrix with Distance in GCG. Sequences that had fewer than 20 substitutions per 100 were grouped into a category. All sequences within a category were realigned using PileUp and translated. Representative amino-acid sequences from all categories with uninterrupted reading frames and amino-acid sequences of known resistance genes retrieved from GenBank (BAA25068, XA1; AAK00132, Pi-ta; AAC49408, Prf; AAD47197, Rp1; CAA61131, Rpm1; AAC78631, Rpp8; AAC72979, Rpp1; AAF08790, Rpp5; AAF42832, Rpp13; AAA21874, Rps2; T30558, Dm3; AAF04603, Gpa2; CAB56299, Rx; AAD27815, I2; T06404, I2-C2; AAC67238, Mi; AAG37354, Mla; BAA7628, PiB; T00020, Xa-1; AAC26126, RPS5; AAA50763, N; AAA91022, L6; AAB47618, M; CAB50708, Rps4) were trimmed and aligned using PileUp with the PAM250 weighting matrix. The MSF file generated by PileUp was converted into a NEXUS format using the Paupsearch function in GCG. The NEXUS file was exported to PAUP (Swofford 2002) where it was edited to remove characters incompatible with PHYLIP (Felsenstein 1989). The edited file was analyzed sequentially in the following PHYLIP programs: SEQBOOT (100 data sets generated), PROTDIST with the PAM100 weighting matrix, NEIGHBOR and CONSENSE. This generated a neighbor joining analysis with 100 bootstrap replicates. The resulting consensus tree was printed with TreeView (Page 1996).

Primer design

Primers were designed to each representative of the 11 categories of RGHs using the PRIME(+) program (GCG). To identify primers specific to a single category, we created a local BLAST database consisting of each primer sequence and the representative sequences of the 11 categories with GCGTOBLAST (GCG). Primers that only matched the sequences from which they were designed were chosen.

Cloning of RGH1 alleles

DNA from Sca6, ICS1, TSH516, P4 and P20 was amplified with TcRGH1for405 (5'GCTGTTGTCTCTCAGACTCC3') and TcRGH1rev405 (5'TTTGCATTCTCCAGCAACC3') which generated a single 405-bp amplicon. Amplicons from each PCR reaction were cloned in pCR4-TOPO (Invitrogen). At least ten clones from each cloning reaction were sequenced as described above. RGH 1 alleles were submitted to GenBank (AY050259–AY050262).

Single-strand conformational polymorphism (SSCP) analysis

SSCP analysis was performed on an ABI 3100 automated sequencer according to the recommended protocol in the ABI310 GeneScan Reference Guide with the following modifications. PCR was performed in 10-μl volumes containing 1 × PCR buffer with 1.5 mM of $MgCl_2$, 200 μM of each dNTP, 200 nM of upper and lower primers specific for each RGH category (Table 1), 1 mg/ml of BSA, 0.4 U of Ampli*Taq* DNA polymerase (Perkin Elmer) and 2.5 ng of template DNA. After an initial denaturation at 94 °C for 2 min, 39 cycles of PCR consisting of 94 °C for 30 s, 49 °C for 40 s and 72 °C for 1 min were performed with a final 72 °C extension step of 5 min. Prior to electrophoresis, 1 μl of PCR product was added to 10.5 μl Hi Di Formamide (Applied Biosystems), 0.5 μl of GeneScan Rox 1000 internal size standard (Applied Biosystems) and 0.5 μl of freshly prepared 0.3 N NaOH. Samples were denatured in a heat block at 95 °C for 5 min and thereafter were immediately placed into an ice-water bath for at least 2 min. PCR products were separated on a 36-cm capillary containing a 5% GeneScan Polymer in 1 × TBE, 10% glycerol. Injection was for 10 s at 3 kV and products were separated at 20 or 30 °C depending on the fragment size and pattern complexity.

Identification and characterization of microsatellite sequences

Total genomic DNA was extracted from 200 mg of *T. cacao* leaf tissue using the BIO 101 FastDNA Kit (BIO 101, Carlsbad, Calif., USA), and stored at –80 °C until needed. Genomic DNA was prepared for SSR enrichment and enriched for SSR sequences following the protocol of Edwards et al. (1996) modified for the use of custom made biotinylated probes (Invitrogen, Carlsbad, Calif.) and Dynabeads M-280 Strepavidin (DYNAL Inc., Lake Success, N.Y., USA) to replace the Hybond N^+ membranes (Amersham, Arlington Heights, Ill, USA) with bound probes. In addition, a second round of SSR enrichment was performed. Prior to cloning, the enriched DNA was size-fractionated using the Equilibrate SizeSep 400 Spun Columns Sepharose CL-4B (Amersham Pharmacia Biotech, Inc., PiScataway, N.J., USA) and 4 μl of fractionated DNA was cloned using the TOPO cloning kit for sequencing – version E (Invitrogen, Carlsbad, Calif., USA) following the manufacturers instructions. Plasmid recovery, M13 insert amplification, Exonuclease I treatment, and cycle sequencing were done using the protocol of Smith et al. (2000), with modifications; a 96-well plate format was used and colonies were grown in 100-μl of SOC broth containing 100-mg/l of ampicillin instead of 50 μl of LB broth containing 50 mg/l of ampicillin. Cycle sequencing was initiated by using the T3 or T7 primer [1.6 pmol/μl] in a 10-μl sequencing reaction mixture. Sequencing was accomplished by use of the ABI Prism Dye terminator cycle sequencing Ready Reaction

Table 1 Primers specific for *T. cacao* RGH categories

Primer name	Primer sequence (5'-3')	Position in nuc. sequence	Amino-acid translation
TcRGH1for437FAM	6CATGGCAAAAGAAGTTGGAAAG	24–45	MAKEVGK
TcRGH1rev437HEX	8CATCAATCAATTCACTGTGGC	490–472	VAGECK
TcRGH2for440FAM	6CGAAGCAAAGGCAGACAAG	45–63	EAKADK
TcRGH2rev440HEX	8CGCATCTCCTACATGCATC	484–466	DACRRCA
TcRGH3afor349FAM	6TTGTAATTTGGGTGGTGACTTC	80–101	VIWVVTS
TcRGH3arev349HEX	8CAAGATGGCTGKCAAGGG	428–411	PCQPSC
TcRGH3bcfor359FAM	6TGGGTGGTKGTGTCTAAAG	88–106	WVVVSK
TcRGH3bcrev359HEX	8GGGTTTYTTCTCCAACCTTC	446–427	KVGEET
TcRGH4for411FAM	6TTGATTTGGGCAACTGTATCC	79–99	LIWATVS
TcRGH4rev411HEX	8ACCACATTCGCAGACAAC	489–472	VVCECG
TcRGH5for581FAM	6AGATYMTCAACAACCAAC	31–48	IINNQ
TcRGH5rev581HEX	8TACCNTKTCATCTAACCC	611–594	GLDEVK
TcRGH6for487FAM	6CAAACATTCACTGAGGAAC	97–115	QTFTEE
TcRGH6rev487HEX	8CAATACGCCTCCATTCAC	583–566	EWRRI
TcRGH7for386FAM	6CTTGCTCAGCTTGTCTAC	25–42	LAQLVY
TcRGH7rev386HEX	8AATACAGAAAGGCAATCTTC	409–390	EDCLSV
TcRGH8for472FAM	6CTTGCTCARCTTGTCTAC	28–45	LAQLVY
TcRGH8rev472HEX	8CACAYTTCTTCACAATCTCC	499–480	EIVKKC
TcRGH9for447FAM	6ATACTTGTGTTCAAAATC	48–65	YLCSKS
TcRGH9rev447HEX	8CACTTTCTCACTATTTTC	494–477	KIVRKC
TcRGH10for468FAM	6GCTCAGCTCGTTTATAAC	28–45	AQLVYN
TcRGH10rev468HEX	8ACACCTTTTCACTATTTCC	495–477	EIVKRC
TcRGH11for300FAM	6GTTCCTGATACCAACGACC	214–232	VPDTND
TcRGH11rev300HEX	8ATGCCCGTCAAAGTTTTC	513–496	ENFDGH

Kit with Ampli*Taq* DNA Polymerase FS version 2.0 (Applied Biosystems, Foster City, Calif., USA). Dye Terminator Cycle Sequencing products were analyzed on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). Electrophoresis was performed using POP6 and capillaries 50-cm in length and 50-μm internal diameter. Analysis was performed using the ABI Prism DNA sequencing analysis software (Version 3.7) with the CE-2 base-caller. A repeat was identified as an SSR if the number of repeats (n) was more than nine for a dinucleotide repeat, more than six for a trinucleotride repeat and more than four for a tetranucleotide repeat. Specific primers were designed in the flanking regions of the SSR using the GCG software package, Prime(+) program (Accelrys, Madison, Wis., USA). Both SHRS and CIRAD microsatellites were amplified following the protocol of Lanaud (Lanaud et al. 1999) with modification of the annealing temperature as required by the individual microsatellite sequence. SHRS microsatellite sequences and the primers designed from them have been submitted to GenBank.

Mapping

The initial mapping population contained 151 samples on which 52 SSR primer pairs were run, of which 37 were from the laboratory of Dr. Claire Lanaud at CIRAD (CIRAD-Biotrop, Montpellier, France), 15 were designed on site at SHRS, and six were RGHs used as markers. Samples from five trees were eliminated, two due to excessive missing data, and the remaining three appeared to be duplicates of another sample, in which case only one of the pair was retained. All markers were tested for goodness of fit to the expected 1:2:1 segregation ratio using a chi-squared test with 2 degrees of freedom. Nine SSR markers had chi-squared values with significance levels exceeding 5.99, corresponding to a Type I error rate of 0.05. The SSR markers eliminated did not map in a perceptible group, but were well distributed throughout the genome. One RGH (RGH2) had a chi-squared value slightly exceeding the desired acceptance level, but was retained nonetheless in order to find its map position. The map was created using JoinMap 3.0 (Van Ooijen and Voorrips 2001) software, using the Kosambi mapping function to convert recombination fractions into map distances (Kosambi 1944). Ten linkage groups (LG 1–LG 10, corresponding to the ten chromosomes of *T. cacao*) were created, nine of which contained all markers at a LOD score of 10 or higher; LG 6 contained one marker (mTcCIR25) which was linked to the group with a LOD score of 5.0, all other markers were linked to the group with a LOD of 10 or higher. LG 2 contained one marker (mTcCIR44) for which there was insufficient recombination information currently to place within the group, hence it was dropped from the map. Linkage groups were renamed to correspond to the linkage groups assigned by Lanaud for the CIRAD primers as they appear in GenBank. One primer, mTcCIR13, mapped to LG 7 in this data, but is given as being on LG 2 by Lanaud in GenBank; and one marker, mTcCIR1, was not found in GenBank, but mapped to the same linkage group as mTcCIR26, which is given as being on LG 8.

Results

RGH sequence library

Using the degenerate primers previously described, we amplified the DNA from seven different cultivars of cacao and created a plasmid library of the cloned amplicons. We picked 605 individual colonies and screened 40 of them by sequencing. Of the first 40 plasmids, 23 had a 412-bp insert that proved to be a fragment of chloroplast DNA. To remove the chloroplast DNA clones from the library, we screened the library by colony hybridization with a ^{32}P-labeled chloroplast DNA fragment and reduced the size of the library to 325 clones. From the 325 remaining clones, 180 inserts of approximately 520 bp or greater were sequenced, as we had discovered no RGH sequences in smaller fragments. Of the 180 inserts sequenced, 74 unique amino-acid translation sequences had significant homology to disease resistance genes as defined by a BLASTp search of GenBank.

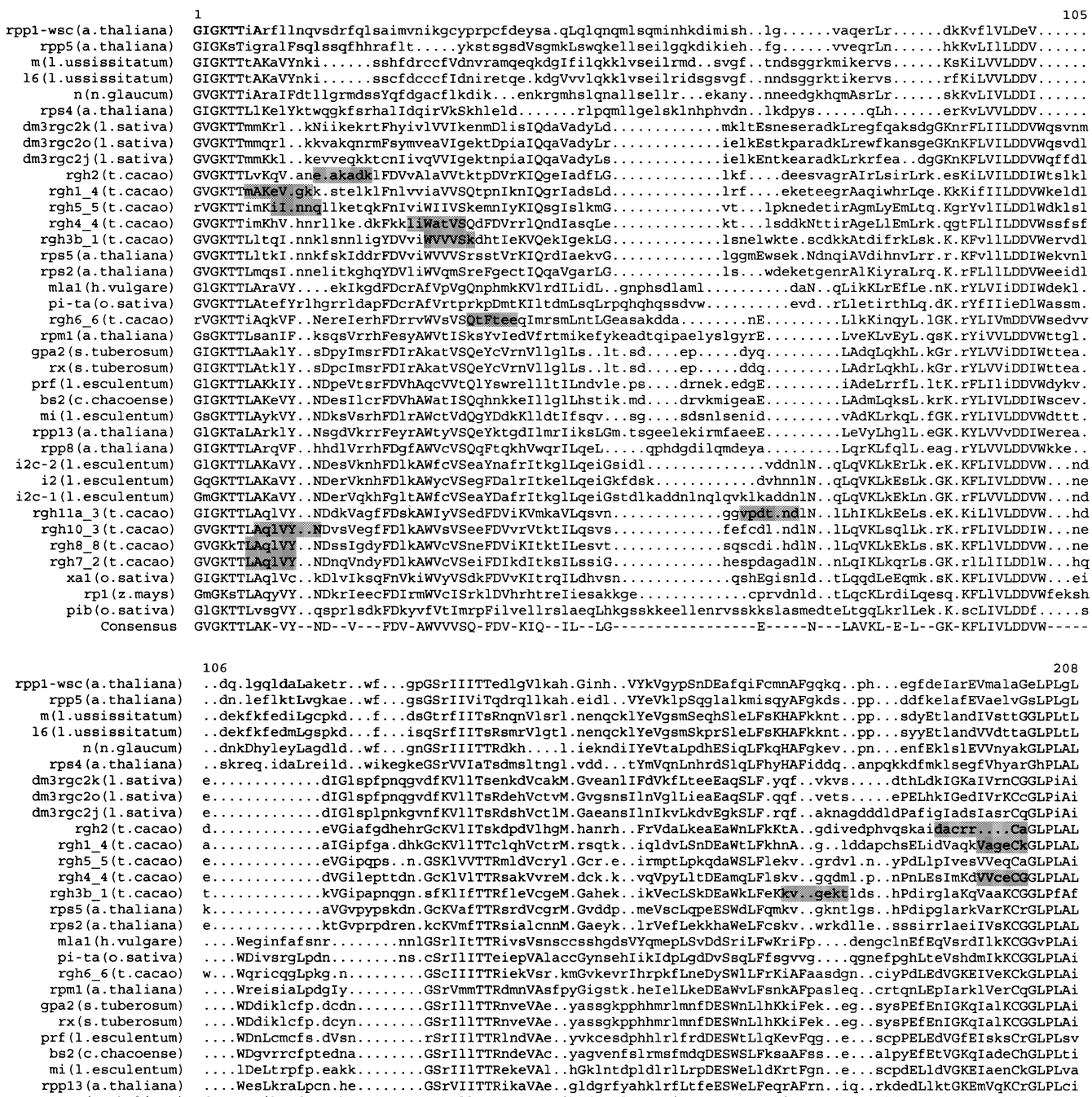


Fig. 1 Alignment of the amino-acid sequences for the NBS region of known resistance genes from plants and resistance gene homologues from *T. cacao*. A minimum of ten identical amino-acids are required for the consensus. Accession numbers for sequences are listed below in alphabetical order. Sequences were aligned in GCG PileUp using a PAM250 substitution matrix. Sites of fluorescently labelled SSCP primer sequences are highlighted for RGH. Reverse primers for RGH5 and RGH6 are beyond the aligned region

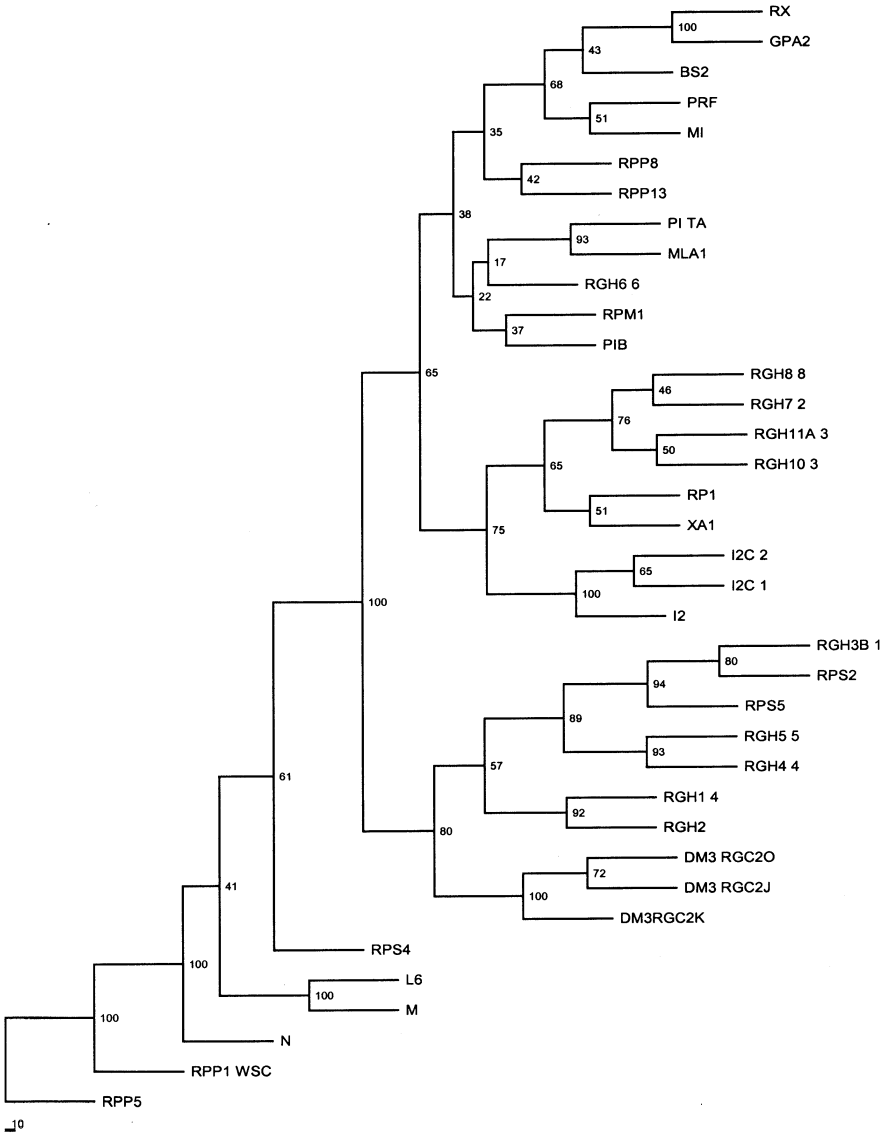


Fig. 2 Consensus tree of neighbor-joining analysis of the aligned amino-acid sequences of resistance gene homologues from *T. cacao* and resistance genes from plants. Numbers on branches are bootstrap values from 100 bootstrap calculation. Branch lengths are proportional to genetic distance

The 74 RGH nucleotide sequences were aligned in GCG using PileUp and from the alignment a pairwise distance matrix was created. Sequences were grouped into categories using a criterion of greater than 80% nucleotide identity (uncorrected genetic distance) and 11 categories of sequences resulted. Representatives of each category of RGHs were translated, the amino-acid translations trimmed and aligned with the protein sequences from known disease resistance genes (Fig. 1), except for RGH9 which had no sequences with an uninterrupted reading frame. The aligned amino-acid sequences were analyzed as described in Materials and methods and a consensus tree is presented (Fig. 2). Some categories of cacao RGHs showed greater similarity with resistance genes from other plants than with each other. For example, RGH3 was more similar to Rps2 from *Arabidopsis thaliana* than to RGHs 1, 2, 4 and 5 which were in the same cluster. RGH 6 was clustered with Pi Ta, Mla1 and Pib, all resistance genes from monocots. RGH 7, RGH 8, RGH 10 and RGH 11 are clustered with Xa1 of *Oryzae sativa* and Rp1 of *Zea mays*, and the I2 and I2C genes of *Lycopersicon esculentum*. The resistance genes containing the TIR motif in the NBS region (Rpp1, Rpp5, N, M, L6 and Rps4) (Hammond-Kosack and Jones 1997) are clustered separately from all the *T. cacao* RGHs. Thus, all the currently identified RGHs are of the non-TIR type.

To identify the sequences as RGHs, we translated the sequences and used the BLAST algorithm to compare the amino-acid sequences to sequences in the GenBank database. The nucleotide sequences and the amino-acid sequences had several interesting and unique features (Table 2).

The RGH3 category was subdivided into RGH3a, b and c. RGH3a had three sequences, all sharing a 32 nucleotide deletion that caused a frameshift such that there were no uninterrupted reading frames after the deletion. RGH 3b and 3c are similar to Rps5 and Rps2 of

Table 2 The 11 categories of RGH sequences from *T. cacao*, the number of sequences in each, their original degenerate amplification primers, and unique features of each group, if any

Category	Number	Forward primers	Reverse primers	Unique features
1	6	PLOOPGA or RG1	GLPL6	Seven nucleotide substitutions among alleles, all leading to non-synonymous amino-acid substitutions (five of seven substitutions in 1st position). The Sca6 allele contains a stop codon
2	1	PLOOPGA	GLPL6	
3	12	RG1	RG2	RGH3a has a 32-nt deletion. RGH3b_3 has a 24 amino-acid frameshift. The amino-acid sequence of 3c subgroup has a stop codon in same position for six members
4	10	RG1	RG2	
5	9	PLOOPGA or RG1	GLPL6	Extra 250 nucleotides on the 3' end due to an alternate priming site for GLPL6. Additional 3'amino-acid sequences match with the RGH sequence in GenBank
6	6	PLOOPGA	GLPL6	Extra 240 nucleotides on the 3' end (see RGH 5) and 40 nucleotides shorter on the 5' end. Five amino-acid frameshift in RGH 6_5, 25 aa frameshift in RGH 6_6
7	2	PLOOPGA	GLPL6	RGH7_1 has a 27-nt deletion
8	12	PLOOPGA or RG1	GLPL6	
9	3	PLOOPGA	GLPL6	No uninterrupted reading frame
10	3	PLOOPGA	GLPL6	
11	11	PLOOPGA	GLPL6	11a subgroup has an extra 60 nucleotides at the 5' end due to an alternate priming site for PLOOPGA

A. thaliana (Fig. 2). RGH 3b_3 (AF402713) has a single nucleotide insertion that causes a frameshift for 24 amino-acids until a single nucleotide deletion brings it back into frame. A BLASTp search with the 24 amino-acid frameshift region as a query gave no hits in GenBank. A BLASTp search of the same region in the other reading frame (RGH 3b_1, AF402727) did match sequences in GenBank but none were associated with resistance genes. Three of the substitutions unique to RGH 3b_3 in this region prevent stop codons in the shifted frame. RGH 3c contains six sequences and is distinguished by a number of substitutions that define the group and by a stop codon. This stop codon at position 189 in the aligned amino-acid sequences is avoided by a substitution in position 1 of the codon for RGH 3b_1 and RGH 3b_2 and a substitution in position 3 of the codon for RGH 3b_3. The stop codon is in a portion of the sequence where all the 3b and 3c members are in the same translation frame.

RGH 5 contains nine sequences and possibly represents two loci, as cloning and sequencing of RGH 5 fragments from Sca6 generated more than two allelic sequences (AY050263–AY050268). The RGH 5 sequences are longer by 235 nucleotides at the 3' end as the GLPL6 primer annealed to a region further downstream. The extra 73 amino acids matched well with resistance genes in a BLASTp search. No stop codons were found in any of the sequences.

RGH 6 also had an additional 240 nucleotides on the 3' end due to an alternate priming site for the GLPL6 primer as in RGH5. RGH 6_5 (AF402743) has a single nucleotide insertion that causes a frameshift for five amino acids before an additional insertion of two nucleotides brings it back into the original frame. RGH 6_6 (AF402759) has a single nucleotide insertion that causes a 25 amino-acid frameshift from aa 222 until the end of the molecule (beyond the aligned amino-acid sequences in Fig. 1). When the 25 amino acids in the frameshifted region were used as a query in a BLASTp search, no hits were found in GenBank. The same 25 amino-acids from the non-frameshifted region in the other RGH 6 sequences had a $3 \times e^{-04}$ expectation score with a disease resistance protein.

RGH 7 contained only two sequences which are easily aligned at the nucleotide level. RGH 7_1 has a 27 nucleotide deletion when compared to RGH 7_2 and to other related resistance genes.

Identification of specific RGH categories

We designed primers specific to an individual sequence in each RGH category. These primers were used to amplify the DNA from the original library cultivars. We found that each pair of gene-specific primers except those for the RGH9 category amplified a band of the expected size in each of the seven *T. cacao* cultivars that were used to prepare the library (data not shown). Thus, none of the RGH categories are specific to a single cultivar. Because no amplification products were detected for the RGH9 category specific primers, it was not further analyzed.

SSCP Analysis of RGH1 in an F_2 Population

Initial amplification of the genomic DNA from the cultivars used to make the library showed no length

differences for the RGH 1 amplicon. However, as differences had been noted in the RGH 1 sequences in the library, we designed fluorescently labelled primers and analyzed the amplicons by SSCP. We used these primers to amplify the genomic DNA from an F_2 population of cacao. To our knowledge, the F_2 population of trees from Brazil we are studying is the only existing F_2 population of cacao. The F_2 was made by selfing TSH516, a member of the F_1 population created in Trinidad by crossing Sca6 with ICS1.

In Fig. 3, the first panel is the amplified clone of RGH1 (AF402715) and represents the pattern for a single sequence. The blue peak represents the mobility of the strand labelled with the forward primer and the green peak is the mobility of the strand labelled with the reverse primer. In this particular instance, there is only one major peak for each strand, suggesting that there is only one major stable conformation for each strand. In the second panel, the amplification products of ICS1 are shown. There are two distinct blue peaks which represent the two alleles of RGH1 but only a single major green peak. There are two blue peaks for Sca6 with mobilities different from the ICS1 alleles. In the F_1, two blue peaks are seen, one coincident with one of the ICS1 alleles and one coincident with one of the Sca6 alleles. Only three patterns of peaks were found in the 146 individuals of the F_2 population tested which represent inheritance of only the ICS1 allele, only the Sca6 allele or both.

To verify the SSCP results, alleles were amplified and cloned from the parents, the F_1 (TSH516) and the homozygous F_2 individuals P4 and P20. The sequence differences for the alleles are listed in Table 3. Sequencing of ten clones from each source with the TcRGH1for405 and TcRGH1rev405 primers suggests that RGH1 is a single locus and that we had identified all of the alleles of that locus. No other sequences were observed. Although the two Sca6 alleles differed by two nucleotide substitutions, the mobility of the electromorphs was much more similar than the two ICS1 alleles which differed by only a single substitution. All of the nucleotide substitutions in the alleles of RGH 1 resulted in amino-acid substitutions. Importantly, the T for C substitution in Sca6 B produced a stop codon (TGA). Thus, the individuals homozygous for the Sca6 allele in the F_2 population should not produce a full-length RGH 1 protein.

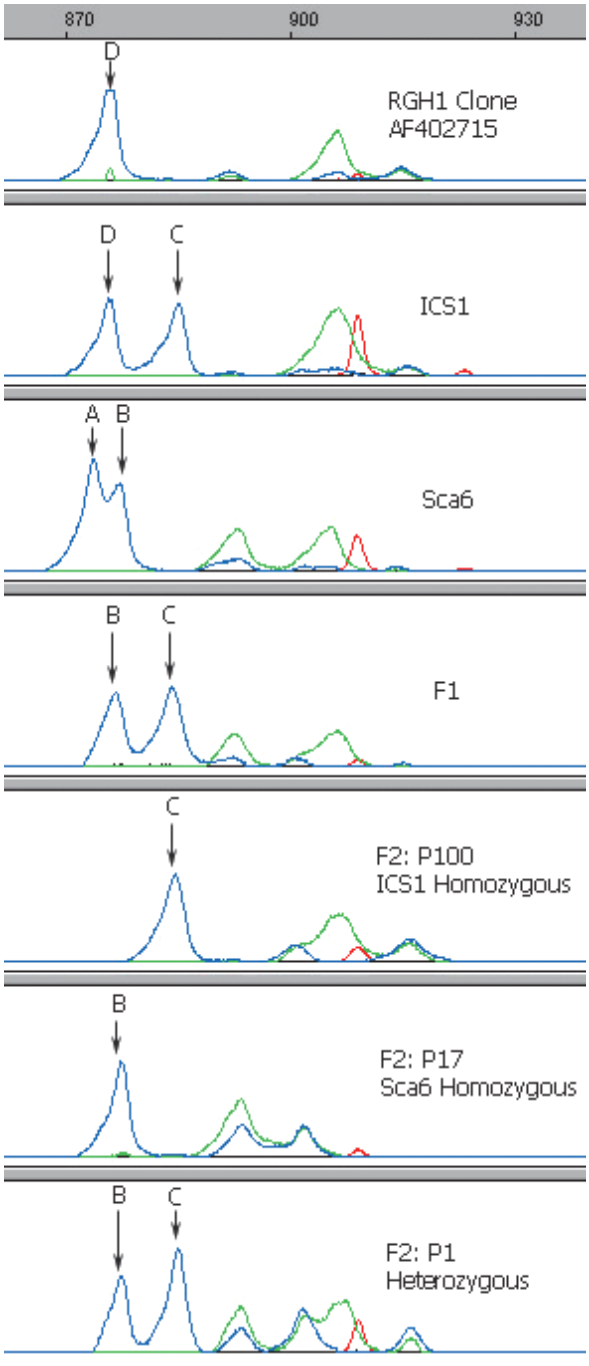
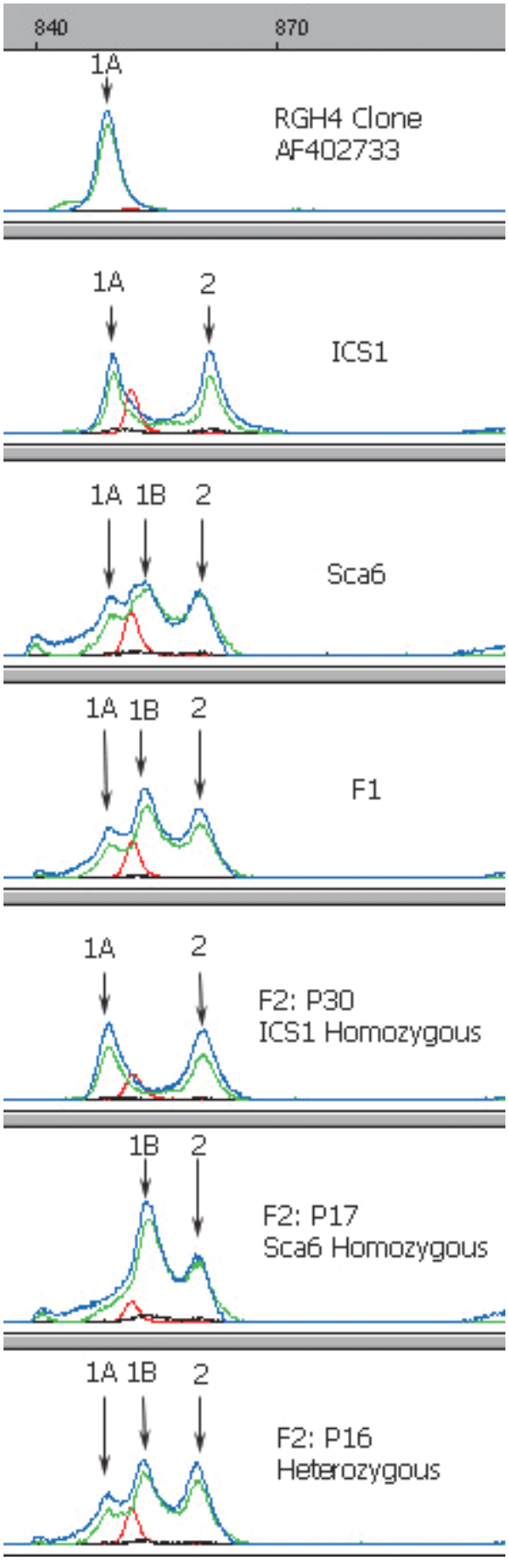


Fig. 3 Electropherograms of SSCP analysis of RGH 1 amplicons from *T. cacao*. Green peaks are strand labelled with TcRGH1for437FAM. Blue peaks are strand labelled with TcRGH1rev437HEX. Clone is RGH1_2 (AY402720). Sca6, ICS1, F1 (TSH516) and P1, P17 and P100 are the parents, F_1, and three individuals of the F_2 population respectively. Red peaks are added molecular-weight standards. Electropherograms are aligned by size (mobility). Scale at top of figure refers to a calculated molecular weight which is related to retention time (scan number)

Mapping of RGH in the Brazilian F_2 population

Using the sequence specific primers designed to each of the RGH categories, we initially analyzed by SSCP the amplified DNA from the parents, F_1 and eight individuals from the F_2 population. RGH1 and 2 were simple patterns with one peak per strand. The F_1 was heterozygous and segregation in the F_2 was easily scored. RGH4 had a slightly more complicated pattern (Fig. 4). The cloned DNA representing a single sequence had a single peak for both the upper and lower strands and the peaks were



coincident. ICS1 had two peaks, which suggested it was heterozygous at the RGH4 locus. However, Sca6 had three peaks, two of them identical in mobility with those of ICS1. The F_1 also had the same three peaks as Sca6 with a slightly different pattern of peak heights. Only three patterns of peaks were found in the F_2 population (Fig. 4): a pattern identical to ICS1, a pattern identical to the F_1 and a pattern with two of the three peaks from Sca6, which we designated as homozygous Sca6.

RGHs 5, 7 and 11 had more complicated patterns with the clones representing a single sequence having multiple peaks for each strand (data not shown). However, the F_1 was heterozygous for these patterns and only three easily scored patterns representing individuals homozygous for the Sca6 allele, homozygous for the ICS1 allele or heterozygous were found in the 146 F_2 individuals. The parents, F_1 and the F_2 individuals tested were homozygous for RGH6 and RGH10. The patterns for RGH3 and RGH8 were too complex for analysis (data not shown).

Construction of linkage groups for RGHs

Figure 5 describes ten linkage groups for the CIRAD microsatellite markers, the SHRS microsatellite markers and the RGHs. These linkage groups have been numbered to correspond with the linkage groups constructed for a separate mapping population of cacao (Risterucci et al. 2000). RGHs 1, 4 and 5 mapped to linkage group 7. RGHs 2, 7 and 11 were mapped to linkage groups 9, 10 and 1 respectively. RGHs 4 and 5 shared a great deal of identity at both the nucleotide and amino-acid sequence levels and map within 4 cM of each other. RGHs 1 and 2 which also shared significant identity at the nucleotide and amino-acid sequence levels segregated independently. The chi-square values for goodness of fit to a 1:2:1 segregation in the F_2 population are presented in Table 4. Only RGH2 had a skewed segregation at the 5% confidence level.

Discussion

We arbitrarily created the library from the amplicons of seven different cacao cultivars so that we would not have to produce and screen a large number of different libraries. Once we had identified the 74 RGH sequences, we grouped them into 11 categories not knowing if the categories represented loci or cultivar to cultivar differ-

Fig. 4 Electropherograms of SSCP analysis of RGH 4 amplicons from *T. cacao*. Green peaks are strand labelled with TcRGH4for411FAM, blue peaks are strand labelled with TcRGH4rev411HEX. Clone is RGH4_1 (AF402733). Sca6, ICS1, F_1, P16, P17 and P30 are the parents, F_1, and three individuals of the F_2 population respectively. Red peaks are added molecular-weight standards. Electropherograms are aligned by size (mobility). Scale at top of figure refers to a calculated molecular weight which is related to retention time (scan number)

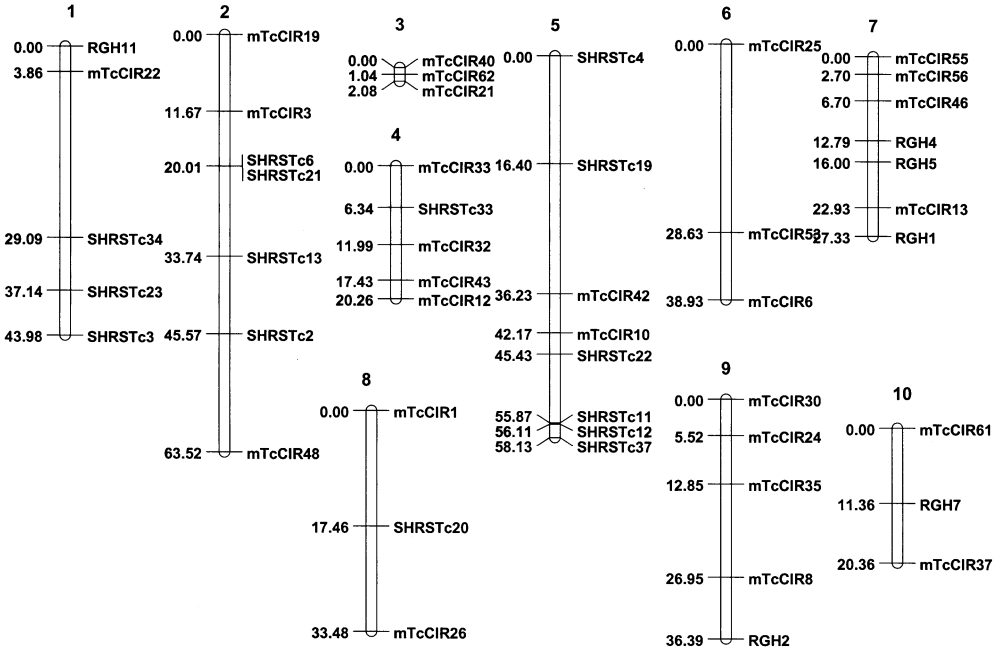


Fig. 5 Genetic linkage map of the Brazilian F_2 population

Table 3 Nucleotide and amino-acid substitutions in the RGH1 alleles. Nucleotide positions are in reference to primer RGH1for405 5' end as position 1. ICS1 D and RGH1 clone (AF402715) amplicons are identical in sequence

RGH1 allele	Pos 157	AA	Pos 238	AA	Pos 289	AA
Sca6 A (AY050259)	G	V	C	R	C	H
Sca6 B F_1 B (AY050260)	G	V	T	*	G	D
ICS1 C F_1 C (AY050261)	A	I	C	R	G	D
ICS1 D (AY050262) RGH1 clone (AF402715)	G	V	C	R	G	D

Table 4 Chi-squared values for RGH loci in the F_2 Population of *T. cacao.* Chi-squared values are from a 1:2:1 goodness of fit test with two degrees of freedom.

Gene name	χ^2	Prob. ($>\chi^2$)	No. with SCA6 allele	No. of heterozygotes	No. with ICS1 allele
RGH 1	3.2	0.201897	40	78	27
RGH 2	8.9	0.011679	23	89	33
RGH 4	3.1	0.212248	44	72	29
RGH 5	2.3	0.316637	43	72	30
RGH 7	5.2	0.074274	26	85	34
RGH 11	0.8	0.67032	32	77	36

ences. The presence of a representative of each category (except RGH9) in each cultivar is evidence that we have identified at least ten loci. For RGH1, the SSCP results and sequencing of the RGH 1 alleles suggest that it is a single locus. For RGH4, we interpret the presence of three peaks in Sca6 and the F_1, and the observation of a peak of the same mobility in all individuals analyzed was evidence for two loci, which we have labelled one and two in Fig. 4. We have not yet sequenced all the alleles from these individuals to verify both loci. It is interesting to note that from our library of cloned sequences for RGH4, the single-stranded mobility of the clone is different from that of the peak observed in all individuals of the Brazilian F_2 population. Our preliminary SSCP data and sequencing of the RGH 5 alleles suggests that there are at least two loci amplified by RGH 5 specific primers.

We have been able to turn the RGHs into useful genetic markers by taking advantage of SSCP. However, we have done the SSCP analysis in an F_2 population which has allowed us to score complicated patterns of peaks without having to identify each allele by sequencing. We are currently investigating the resolution of this method by genotyping the RGH categories in a broad and diverse collection of cacao cultivars to determine how many different alleles can be accurately detected. There appears to be no reasonable method to predict the number or mobility of peaks that a single sequence will generate under non-denaturing conditions. We are attempting to determine the optimum resolution of the method by analysis of our cloned RGH sequences at different temperatures and in complex mixtures. With these caveats, we still believe that this candidate gene strategy

coupled with SSCP provides a powerful method to add genetic markers of known (or supposed) function to the map of relatively unstudied organisms such as *T. cacao*.

The number of resistance genes or resistance gene homologues in a single plant has been estimated as high as 1% of the total expressed genes (Bergelson et al. 2001). In *Arabidopsis*, this is estimated at 182 genes arranged in 49 single-gene loci and 32 clusters of 2 to 12 genes. Conservatively, we may have so far identified only ten loci in cacao and more should be expected. RGHs are broadly divided into Toll/Interleukin Receptor (TIR) motifs containing and non-TIR containing groups (Hammond-Kosack and Jones 1997). Although only non-TIR RGHs have been observed in monocots, dicots should have both types of RGHs. Because all of the 74 sequences we have identified in cacao clustered with non-TIR containing resistance genes, there are probably other TIR-containing sequences that could be identified in cacao to increase the coverage of disease resistance clusters in the genome. We are currently designing primers from TIR-containing RGH sequences in GenBank to attempt to identify TIR-containing RGHs in cacao. Interestingly, no TIR-containing RGH sequences were identified in coffee (Noir 2001).

In our preliminary study of cacao RGHs, we aligned the nucleotide sequences and divided them into categories based on sequence similarity. We have aligned the translated nucleotide cacao sequences with protein sequences from known resistance genes. Our different RGH groups may only be distantly related to each other, as some of our RGH sequences were more closely related to RGH sequences from monocots than they were to other RGH sequences from cacao. Thus, any attempt to understand the evolution of the RGH sequences would require looking at a single RGH locus across a number of related species and genera in the same family. We have been able to detect RGH 1 amplicons (data not shown) in other *Theobroma* species and in *Herrania*, a closely related genus to *Theobroma* (Whitlock and Baum 1999). We are currently sequencing these RGH 1 amplicons from other *Theobroma* species and closely related genera to investigate the evolution of a single-copy RGH locus.

In other research on NBS/LRR-type resistance genes, the greatest amount of variability has been seen in the LRR region (Meyers et al. 1998b; Bergelson et al. 2001) and relatively little variability in the NBS region. We have observed in some of our RGH categories a surprising number of non-synonymous substitutions between alleles. For example, in RGH 1, all seven of the observed nucleotide substitutions led to non-synonymous amino-acid substitutions. Allelic differences in a coding region might be expected to occur most frequently in the codon third position and result in a synonymous substitution with no change in the amino-acid. Interestingly, the majority of the substitutions were in the first position and all substitutions led to amino-acid differences. One of the substitutions created a stop codon in the nucleotide binding site. This same stop codon substitution was found in the allele from the Sca6 parent that was inherited in the F_2 population segregating for resistance to witches' broom. Frameshifts and deletions have been reported for resistance gene analogs in soybean (Graham et al. 2000; Hayes et al. 2000). In some cases (Graham et al. 2000) these truncated or frameshifted transcripts were identified in cDNA, demonstrating that they are expressed and therefore are not pseudogenes.

To determine the correlation between SSCP pattern and number of alleles, we amplified, cloned and sequenced the amplicons for RGH 1 from the Sca 6, ICS 1, the F_1 (TSH516) and a number of F_2 progeny. In addition to identifying RGH 1 as a single locus, we discovered that one of the Sca 6 alleles had a stop codon in the reading frame that was not seen in any of the other RGH 1 alleles. As this allele had been sequenced from a number of individual clones that each had the stop codon, we doubt that this was either a sequencing error or an amplification error. Since Sca 6 is the cultivar most resistant to witches' broom, it would be quite interesting to determine if a non-functional allele provides some advantage in terms of resistance. In barley, the *mlo* allele that confers broad-spectrum powdery mildew resistance is a loss-of-function mutation (Hulbert et al. 2001). Currently, we are acquiring the phenotypic data from the F_2 population to determine if there is a correlation between the RGH 1 alleles and resistance to witches' broom. For individuals that are homozygous for the Sca6 allele, they should not be able to produce a functional copy of the RGH 1 protein. We are currently cloning and sequencing all alleles from all RGH categories in TSH516 to determine if other non-functional alleles occur in the F_2 population. In addition, we have obtained apices from F_2 individuals that are homozygous for the Sca6 RGH 1 allele that contains a stop codon to determine if RGH 1 mRNA is present.

Our investigation of RGHs in *T. cacao* appears to confirm what has been observed for other plants (Collins et al. 1998; Graham et al. 2000; Noir et al. 2001). There is a great diversity of RGHs and sequence polymorphism at individual loci that makes them useful genetic markers. At least two of our RGHs are clustered within 4 cM of each other, as has been seen for resistance genes in plants (Hammond-Kosack and Jones 1997). With regard to other types of molecular markers, we find RGHs at about the same rate as microsatellites in terms of the number of RGHs identified per clones sequenced. The potential advantage is that the RGHs may be closely linked to disease resistance genes. They may prove more useful as genetic markers for specific disease resistance genes than the more randomly located microsatellite sequences. However, although we were able to map six of the ten RGH categories in our Brazilian F_2 population, none of our RGH loci coincided with QTLs for *Phytophthora palmivora* resistance mapped in an F_1 population of cacao (Flament et al. 2001). Whereas microsatellite loci are identified by length differences between individuals, we have been able to use the sequence diversity in the RGHs to identify individual alleles by SSCP. As the alleles often differ by more than a single nucleotide substitution, we

are identifying alleles by mobility shifts that may not correlate to single nucleotide polymorphisms (SNPs). The ability to resolve alleles by SSCP on the ABI3100, a 16 capillary automated genetic analyzer, makes high throughput analysis of RGHs possible. Thus, we hope with this paper to demonstrate the utility and potential of SSCP for population genotyping and mapping. However, an F_2 population segregating for disease resistance or a large F_1 population made between a susceptible and a resistant parent are necessary to identify which of the RGH markers are linked to a particular disease resistance phenotype. In many tropical tree crops, it is the lack of such populations, rather than technical difficulty, that is the greatest impediment to the identification of useful genetic markers for marker-assisted selection.

Acknowledgements We acknowledge the financial support of Mars, Inc., Hackettstown, N.J., and especially Mr. John Lunde and Dr. Jeff Morgan.

References

Aarts MG, te Lintel HB, Holub EB, Beynon JL, Stiekema WJ, Pereira A (1998) Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol Plant Microbe Interact 11:251–258

Ahnert DE (2000) Use of QTLs for Witches' Broom resistance in Cocoa breeding. Proc Int Workshop on New Technologies and Cocoa Breeding, 16–17 October 2000, Kota Kinabalu, Sabah, Malaysia, pp 116–119

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. Science 292:2281–2283

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Interact 11:968–978

Deng C, Davis TM (2001) Molecular identification of the yellow fruit color (c) locus in diploid strawberry: a candidate gene approach. Theor Appl Genet 103:316–322

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20:758–760

Felsenstein J (1989) PHYLIP – phylogeny inference package (Version 3.2). Cladistics 5:164–166

Flament MH, Kebe I, Clement D, Pieretti I, Risterucci AM, N'Goran JAK, Cilas C, Despreaux D, Lanaud C (2001) Genetic mapping of resistance factors to *Phytophthora palmivora* in cocoa. Genome 44:79–85

Gentzbittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, De Labrouhe DT, Nicolas P (1998) Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L. Theor Appl Genet 96:519–525

Graham MA, Marek LF, Lohnes D, Cregan P, Shoemaker RC (2000) Expression and genome organization of resistance gene analogs in soybean. Genome 43:86–93

Hammond-Kosack K, Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Physiol Plant Mol Biol 48:575–607

Hayes AJ, Yue YG, Maroof MAS (2000) Expression of two soybean resistance gene candidates shows divergence of paralogous single-copy genes. Theor Appl Genet 101:789–795

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39:285–312

Hunter JR (1990) The status of cacao (*Theobroma cacao*, Sterculiaceae) in the Western Hemisphere. Econ Bot 44:425–439

Irizarry H, Rivera E (2002) Early yield of five cacao families at three locations in Puerto Rico. J Agric Univ PR 82:163–171

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93:11,746–11,750

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet PJ, Lagoda L (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol Ecol 8:2141–2152

Lande R (1992) Marker-assisted selection in relation to traditional methods of plant breeding. In: Stalker HT, Murphy JP (eds) Plant Breeding in the 1990's. CAB International, Wallingford, UK, pp 437–451

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998a) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW (1998b) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 10:1833–1846

Motamayor JC, Risterucci AM, Lopez AP, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity 89:380–386

Noir S, Combes MC, Anthony F, Lashermes P (2001) Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). Mol Genet Genomics 265:654–662

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Pires JL, Monteiro ED, Luz EDMN, Silva SDVM, Pinto LRM, Figueria A, Ahnert DE, Brugnerotto MIB (1996) Cocoa breeding for Witches' Broom resistance at CEPLAC, Bahia, Brazil. Proc Int Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement, 24–26 November 1996, Salvador, Brazil, pp 67–71

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH, Lanaud C (2000) A high-density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Shen KA, Meyers BC, Islam-Faridi MN, Chin DB, Stelly DM, Michelmore RW (1998) Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol Plant Microbe Interact 11:815–823

Smith TPL, Godtel RA, Lee RT (2000) PCR-based setup for high-throughput cDNA library sequencing on the ABI3700 automated DNA sequencer. BioTechniques 29:698–700

Sunnucks P, Wilson ACC, Beheregaray LB, Zenger K, French J, Taylor AC (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. Mol Ecol 9:1699–1710

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), version 4. Sinauer Associates, Sunderland, MA

van der Vossen EAG, van der Voort JNAM, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. Plant J 23:567–576

Van Ooijen JW, Voorrips RE (2001) JoinMap Version 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands

Whitlock B, Baum DA (1999) Phylogenetic relationships of *Theobroma* and *Herrania* (Sterculiaceae) based on sequences of the nuclear gene Vicilin. Systematic Bot 24:128–138