Mol Gen Genomics (2003) 269: 406–419
DOI 10.1007/s00438-003-0849-8

ORIGINAL PAPER

M. Rossi · P. G. Araujo · F. Paulet · O. Garsmeur
V. M. Dias · H. Chen · M.-A. Van Sluys · A. D'Hont

Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane

Received: 7 November 2002 / Accepted: 7 April 2003 / Published online: 6 May 2003

Abstract A large sugarcane EST (expressed sequence tag) project recently gave us access to 261,609 EST sequences from sugarcane, assembled into 81,223 clusters. Among these, we identified 88 resistance gene analogs (RGAs) based on their homology to typical pathogen resistance genes, using a stringent BLAST search with a threshold e-value of e^{-50}. They included representatives of the three major groups of resistance genes with NBS/LRR, LRR or S/T KINASE domains. Fifty RGAs showed a total of 148 single-dose polymorphic RFLP markers, which could be located on the sugarcane reference genetic map (constructed in cultivar R570, $2n = \sim 115$). Fifty-five SSR loci corresponding to 134 markers in R570 were also mapped to enable the classification of the various haplotypes into homology groups. Several RGA clusters were found. One cluster of two LRR-like loci mapped close to the only disease resistance gene known so far in sugarcane, which confers resistance to common rust. Detailed sequence comparison between two NBS/LRR RGA clusters in relation to their orthologs in rice and maize suggests their polyphyletic origins, and indicates that the degree of divergence between paralogous RGAs in sugarcane can be larger than that from an ortholog in a distant species.

Keywords Sugarcane · Polyploid · Genetic mapping · Resistance gene analogs (RGAs) · Nucleotide binding site/leucine rich repeat (NBS/LRR)

Communicated by M.-A. Grandbastien

The first two authors contributed equally to this paper

M. Rossi · P. G. Araujo · V. M. Dias · M.-A. Van Sluys
Departamento de Botanica, Instituto de Biociencias,
Universidade de São Paulo, Rua do Matão 277,
055080-090 SP, São Paulo, Brazil

F. Paulet · O. Garsmeur · A. D'Hont (✉)
UMR1096, CIRAD, TA 40/03, Avenue Agropolis,
34398 Cedex 5, Montpellier, France
E-mail: dhont@cirad.fr
Fax: +33-4-67615605

H. Chen
Yunnan Sugarcane Research Institute,
Yunnan Academy of Agricultural Sciences,
Eastern Lingquan Road 363,
661600, Kaiyuan, Yunnan, P.R. China

Introduction

Pathogen recognition is the first step in the process that triggers plant resistance responses, and is usually mediated by single dominant resistance genes (*R* genes). Each of these gene products interacts directly or indirectly with the product of a corresponding avirulence (*Avr*) gene in a pathogen (Flor 1971; Keen 1990). Many R proteins from different dicot and monocot plant species, which confer resistance to a wide variety of pathogens, share several conserved motifs (Hammond-Kosack and Jones 1997). Based on the deduced structure of their products, *R* genes can be classified into three main groups. Members of the major group encode proteins with a nucleotide binding site (NBS) domain followed by a leucine rich repeat (LRR) region. This group can be further sub-divided into two classes based on the nature of its N terminal region. Proteins in the first class show homology to Drosophila Toll or the human interleukin receptor (TIR); whereas proteins in the second class have a coiled coil (CC) domain (Hammond-Kosak and Jones 1997; Pan et al. 2000). The second group of *R* genes encodes proteins with only LRRs (Dixon et al. 1998). The third group, represented exclusively by *Pto* (Martin et al. 1993), displays a serine-threonine kinase (S/T KINASE) domain. In addition, one example of an R protein with an LRR followed by a S/T KINASE has been reported (Song et al. 1995). As new genes with novel motifs are cloned, new classes of *R* genes are emerging; this is the case for the genes *RPW8* (Xiao et al. 2001) and *Rpg1* (Brueggeman et al. 2002).

Isolation of *R* genes has historically involved map-based cloning or transposon tagging, both of which are very labor-intensive and expensive strategies. The common features shared by R proteins have led to new cloning strategies. Since the late 90s, PCR primers corresponding to highly conserved amino acid sequences of the NBS domain have been used to amplify resistance gene analog (RGA) fragments from various plant species (Wang et al. 2001). Many of these RGAs appear to be linked to previously described resistance loci or QTLs. However, *R* genes are often members of multigene families, frequently organized in clusters, and this PCR strategy generally fails to identify the functional *R* genes within a given cluster (Graham et al. 2000).

Sugarcane is an economically important crop. However, analysis of its genome has lagged behind, compared to other important grass species, mostly due to its genetic complexity. Modern sugarcane cultivars are highly polyploid (2n = 100–130), derived from interspecific hybridizations between the domesticated sugar-producing species *Saccharum officinarum* L. (2n = 80) and the wild species *S. spontaneum* L. (2n = 40–128). They thus represent a particular challenge for breeding, genetics and gene cloning purposes (Butterfield et al. 2001; D'Hont and Glaszmann 2001; Grivet and Arruda 2001).

The Brazilian Sugarcane EST Sequencing Project (SUCEST) database, with 291,689 EST sequences, provides an invaluable source of information (Vettore et al. 2001). In this paper, we report the results of a search for resistance gene analogs (RGAs) using the SUCEST database. We have mapped these RGAs on the sugarcane reference genetic map (Grivet et al. 1996; Hoarau et al. 2001 and unpublished data), in order to investigate their genomic distribution and their relationship with disease resistance loci in sugarcane. Fifty-five single-sequence-repeat (SSR) loci were also mapped to allow the classification of the different haplotypes into homology groups. In addition, we compared the sequences of various members of two NBS/LRR resistance gene clusters to those of their orthologs in rice and maize.

Materials and methods

SUCEST database and sequence analysis

The SUCEST database encompasses 291,689 EST sequences derived from 37 different sugarcane cDNA libraries constructed from total RNAs isolated from various tissues, developmental stages and stress conditions including pathogen inoculated seedlings (Vettore et al. 2001). A total of 261,609 sequences have been grouped into 81,223 clusters based on an analysis with the phrap fragment assembly program. Results of comparisons between cluster consensus sequences and GenBank data were available for homology searches (Telles et al. 2001).

The 81,223 clusters were screened to identify RGAs. "NBS-LRR" and "disease resistance" were used as keywords, and *Mi-1.2* (gi3449380), *Rpm1* (gi963017), *RPS2* (gi549979), *Xa21* (gi1122443), *Prf* (gi1513144), *Pto* (gi430992), *Cf-2.1* (gi1184075), *N* (gi558887), *L6* (gi862905), *M* (gi1842251), *Pti1* (gi3668069), *RPR1* (gi4519936), *I2* (gi4689223), *Hcr2–5D* (gi7488988), *Hs1* $^{pro-1}$ (gi1850968), *b5* (gi2792210) and *Rp1-D* (gi5702196) coding sequences as key genes. The genes were chosen to represent a broad range of plants, pathogen specificities, and R protein structures known at the time the searches were carried out. To avoid spurious hits due to the enormous amount of data, a very stringent expectation value of e^{-50} or better was used.

Plant material

The progeny analyzed in this study consisted of 112 individuals obtained by the self-fertilization of cultivar R570; this is a subset of the population used to build an AFLP genetic map by Hoarau et al. (2001). R570 is a rust-resistant cultivar developed by CERF (Center d'Essai de Recherche et de Formation, Réunion). Rust resistance phenotypes were determined in the field on the island of Réunion, using natural infection as described in Daugrois et al. (1996).

Restriction Fragment Length Polymorphism (RFLP) analysis

The 55 selected clone sequences were amplified by PCR with universal primers (T7, T3, SP6). The PCR products were purified with the GFX PCR DNA and Gel Band Purification Kit (Amersham Pharmacia Biotech) and radioactive random priming labeling was carried out with the Megaprime DNA Labeling System (Amersham Pharmacia Biotech). Genomic DNA extraction, Southern blotting, and hybridizations were performed as previously described by Grivet et al. (1996). The enzymes used for DNA digestion were *Hin*dIII, *Sst*I, *Dra*I and *Eco*RV.

Simple Sequence Repeat (SSR) analysis

The progeny was analyzed with 76 SSRs developed at CIRAD in collaboration with Génoscope (Evry, France) from an enriched library made with DNA from the cultivar R570, and these markers were localized on a reference RFLP map (in preparation). The primers were end-labeled with [γ-^{33}P]ATP, and amplification was performed in an MJ Research PTC 100 Thermal Cycler in 20-µl reaction mixtures containing 50 ng of sugarcane DNA, 0.2 mM dNTP mix, 2 mM $MgCl_2$, 50 mM KCl, 10 mM TRIS-HCl (pH 8.3), each primer at 0.2 µM, and 1 U of *Taq* polymerase (Eurobio). The samples were denatured at 94°C for 5 min and subjected to 35 cycles of 94°C for 1 min, 46°C–55°C (depending on the SSR primer sequence) for 45 s, and 72°C for 30 s, followed by an extension step for 10 min at 72°C. After the addition of 20 µl of loading buffer (98% formamide, 10 mM EDTA, bromophenol blue, xylene cyanol), the amplified products were denatured at 92°C for 3 min, and 4 µl of each sample was loaded onto a 5% polyacrylamide gel with 7.5 M urea and electrophoresed in 0.5% TBE buffer at 55 W for 1 h 40 min. The gel was dried for 30 min at 80°C and exposed for 4 days to X-ray film (Fuji RX).

Marker scoring, analysis and map construction

Each segregating RFLP and SSR band was scored independently as a dominant marker (presence vs. absence) and the following nomenclature was adopted; for RGAs: RGA, followed by three digits indicating the EST clone number, then three letters indicating the enzyme used to reveal the marker and a letter indicating the marker; for SSRs: mSSCIR (microsatellite, *Saccharum* Spp, CIRAD), followed by the number of the SSR, and then the letter 'm' followed by a number indicating the marker. Since sugarcane is highly polyploid, only single-dose markers (Wu et al. 1992) were used for map construction. Such markers show a segregation ratio that is not significantly different (by the $\chi 2$ test) from 3:1 (presence:absence) at $P = 0.05$ (Grivet et al. 1996).

The single-dose markers were added to the AFLP matrix (883 markers × 112 individuals) developed by Hoarau et al. (2001). The new map was built using MAPMAKER 3.0 (Lander et al. 1987). Marker grouping was performed by two-point analysis at a LOD score threshold of 5 and a recombination fraction threshold of 0.35. Co-segregation groups (CGs) were then ordered by multipoint analysis and the distances calculated using the Haldane function. For homology group VII, we had additional data and thus the map distances were calculated with data from 316 individuals. CGs were assembled into homology groups (HGs) based on (1) common RGA or SSR markers between CGs; and (2) common SSR and AFLP markers with a R570 map encompassing mainly RFLP markers (Grivet et al. 1996, and unpublished results). A minimum of two common markers was necessary for assembly of two CGs into the same HG. When a correspondence between HG and CG could be established between the two maps, we assigned the same name to them, a Roman numeral from I to VIII for the HG, and a number for the CG. Assigned CGs with no correspondence between the two maps were named with the number of the HG followed by a letter. CGs not assigned to a HG were named as U (unassigned) followed by a number.

Analysis of clusters of NBS/LRR-like RGAs

The full length sequences of eight NBS/LRR-like EST clones (RGA118, RGA281, RGA326, RGA185, RGA267, RGA162, RGA152 and RGA087) were obtained by primer walking. Nucleotide sequences were aligned using the program Sequence Navigator 1.0.1 for Macintosh. Sequence variability was estimated using Nei's measure of nucleotide diversity (π) and calculated with the program DnaSP (Rozas and Rozas 1997).

Results

Identifying RGAs in the SUCEST database

Key gene and keyword searches in the SUCEST database identified 88 clusters homologous to known pathogen resistance genes with an expectation cut-off value, for the best matching query, of e^{-50} or better. Twenty-two ESTs presented homology to genes encoding NBS-LRR resistance proteins, 13 showed homology to LRR-coding genes and 53 were S/T KINASE homologs. No TIR/NBS/LRR-like RGAs were identified, even though genes encoding these three domains (like *N, L6* or *M*) were used as key genes (Table 1). Matches to the NBS or LRR regions of these genes had poorer e-values than did CC/NBS/LRR genes.

A single clone per cluster was selected for further analysis. To increase the likelihood of obtaining full length mRNAs, we chose the most 5′ clone. After identity confirmation by sequencing, 55 of the 88 clones analyzed were selected for mapping. We excluded clones that were wrongly addressed, showed evidence of rearrangement or represented redundant information. Table 1 indicates, for the 55 ESTs, the corresponding cluster-consensus homology and the relevant protein domain (Genbank accession numbers: BQ803996 to BQ804049). Only the best hits against a known *R* gene or RGA are included in Table 1. Hence, not all clones listed show an e value of e^{-50} or better. A number of clusters in Table 1 are indicated as *Pti1* homologs. *Pti1* is not a resistance gene, but is a *Pto* interactor which shares 36.4% overall protein identity with it (Zhou et al. 1995).

The distribution of RGAs in the sugarcane genome

Fifty-five ESTs were tested on the self-progeny of cultivar R570; no polymorphisms were detected for three of them (RGA251, RGA231 and RGA176) with any of the four enzymes assayed. The other 52 ESTs produced 272 polymorphic markers (an average of 5.23 markers/probe) and, of these, 177 segregated as single-dose markers (3:1 ratio, average of 3.40 markers/probe) and could be used for mapping. Out of these 177 markers, 148 markers corresponding to 50 RGA clones, were localized on the AFLP map (Hoarau et al. 2001) while the others remained unlinked (Fig. 1). Seventy-six SSRs tested on the same progeny produced 170 single-dose markers, of which 134, corresponding to 55 SSRs, were localized on the map. SSR and RGA markers were used to assemble co-segregation groups (CGs) into homology groups (HGs) as described in Materials and methods. The map encompasses 128 CGs, of which 66 could be assigned to seven of the eight HGs in the reference RFLP maps (Grivet et al. 1996, and unpublished results). The RGA markers map on 59 of the 128 CGs. They are present in all seven identified HGs. Six RGAs map on HG I, seven on HG II, two on HG III, six on HG IV, seven on HG VI, two on HG VII and 16 on HG VIII. Alleles of the same RGA map mainly onto the same HG, with four exceptions: RGA142 and RGA526 map on HG IV and HG VIII, RGA258 maps on HG II and HG VI, and RGA149 maps on HG III and HG VI.

RGAs are not equally distributed along the chromosomes. RGAs that were not more than 5 cM apart were defined as members of a cluster. On this basis, we determined all cluster loci, referred to the basic genome complement, that contain different RGAs (Table 2). We identified four cluster loci with three to six different RGAs, and six cluster loci with two different RGAs. Clusters 1, 7 and 8, contain four, six and three RGAs, respectively, that map on several homologous chromosome segments of HG I and VIII. In these three cases, not all RGAs were mapped on all the homologous CGs. The distance between RGAs on each CG is variable, but in at least one CG the distance between each pair of consecutive RGAs is ≤ 5 cM.

Sixteen RGAs produced more than one marker on the same CG. The majority are clustered and have been identified with an asterisk in the CG column in Table 1. Some of these markers may be redundant, identifying the same allele due to the presence of a restriction site in the RGA sequence. However, since a few of them are separated by recombination events, we retained all of them on the map.

To date, the only pathogen resistance locus mapped in sugarcane is the common rust resistance gene located

Table 1 Characteristics and map location of the 55 EST-RGA studied

RGA	EST clone	Phrap cluster	Cluster consensus homology				Map position[c]		
			Gene[a]	Species[b]	e-value	Domain	HG	CG (bands mapped)	Copies
482	SCSFSB1102C02.g	SCJLRT1020F05.g	*Cf-2.1*	Lp	2e-78	LRR	VI	VI2, U5	9
183	SCEQRZ3024A05.g	SCEQRZ3024A05.g	*Xa21*	Ol	1e-08	LRR	I	I6, I5	~25
131	SCCCST1C06H04.g	SCSGLR1045B05.g	*Xa21*	Ol	2e-75	LRR	VIII	VIII2	6
386	SCRFLV1037B07.g	SCVPRT2081F09.g	*Xa21*	Ol	1e-67	LRR	II	II8, II6	7
327	SCQGAM2108A06.g	SCQGAM2108A06.g	*Xa21*	Os	4e-65	LRR	–	–	–
366	SCRFAD1117H03.g	SCMCRT2102E02.g	*Xa21*	Ol	5e-62	LRR	II	II9, II10, II12	10
173	SCEQRT2028B11.g	SCEQRT2028B11.g	*Xa21*	Os	5e-53	LRR	VIII	VIII5(3)*, VIII15(2)*, VIII16, VIII2	11
024	SCAGFL3024G01.g	SCSGRT2064H06.g	*Hcr2-5D*	Le	2e-34	LRR	I	I7, Ia, I6, I5(2)*	7
149	SCEPRZ3048F07.g	SCCCLR1001A03.g	*Xa21*	Ol	1e-84	LRR	III, VI	III4, VI3, VI5	11
137	SCEPAM1051A02.g	SCEPAM1051A02.g	*Hcr2-2A*	Lp	1e-31	LRR	VII	VII1a, VIIa(3)*	~16
019	SCCCLR1066C04.g	SCCCLR1066C04.g	*Hcr2-0A*	Le	2e-16	LRR	VII	VII14, VIIa	~24
441	SCSBAD1126D07.g	SCSBAD1126D07.g	*I2C-2*	Le	2e-43	NBS/LRR	IV	IV1(2)*, IV2(2)*	11
251	SCJFSB1010F08.g	SCEQSB1C01F10.g	*I2C-2*	Le	5e-22	NBS/LRR	–	–	–
088	SCCCCL3080E03.g	SCCCCL3080E03.g	*Xa1*	Os	9e-32	NBS/LRR	–	U7	~13
118	SCCCNR2002A04.g	SCJLRZ1019B10.g	*RPR1*	Os	0.0	NBS/LRR	VIII	VIIIa, VIII5, VIII15(2)*, VIII2	~16
016	SCAGAM2124A05.g	SCAGAM2124A05.g	*Pib*	Os	5e-25	NBS/LRR	–	U4	~15
057	SCBFSB1045F08.g	SCCCAM2002B02.g	*I2*	Le	2e-46	NBS/LRR	I	Ia, I9(2)*, Ib	12
039	SCBFAD1089A09.g	SCBFAD1089A09.g	*PIC17*	Zm	3e-42	NBS/LRR	VIII	VIII2, VIII9, VIII11	9
272	SCJLLB2079H05.g	SCSGAD1006H08.g	*RPP1-WsA*	At	3e-18	NBS/LRR	VI	VI3, VI2, VIb, VIa, U36	7
542	SCUTST3130D10.g	SCCCFL2002E03.g	*Hv1LRR2*	Hv	3e-35	NBS/LRR	VIII	VIIId, VIIIe	6
196	SCEZAM2031E03.g	SCEZAM2031E03.g	*Gpa2*	St	2e-35	NBS/LRR	VI	VI2, VI1a(2)*, VIa, VIb	~12
145	SCEPLB1044F11.g	SCEPLB1044E11.g	*TMV*	At	8e-72	NBS/LRR	II	IIa	6
162	SCEQAD1016E12.g	SCEQAD1016E12.g	*Rp1-D*	Zm	9e-80	NBS/LRR	–	U11(3)*	~12
326	SCQGAM2029B12.g	SCMCST1052G12.g	*RPR1*	Os	7e-73	NBS/LRR	VIII	VIII1, VIII2, VIIIa(2)*, VIIIb, VIII5	8
267	SCJLFL3019A12.g	SCJLFL3019A12.g	*RPR1*	Os	5e-63	NBS/LRR	VIII	VIIIb, VIII5	~17
185	SCEQRZ3090E04.g	SCRLRZ3043E10.g	*RPR1*	Os	1e-61	NBS/LRR	VIII	VIII1, VIII2, VIIIa, VIIIb, VIII5	8
281	SCJLRZ1018G01.g	SCJLRZ1018G01.g	*RPR1*	Os	2e-59	NBS/LRR	VIII	VIII1, VIII2, VIIIa(2)*, VIIIb, VIII5	10
087	SCCCCL3080B03.g	SCCCCL3080B03.g	*Rp1-D*	Zm	1e-99	NBS/LRR	–	U11	4
152	SCEPRZ3130D02.g	SCEPRZ3130D02.g	*Rp1-D*	Zm	2e-83	NBS/LRR	–	U11, U51	13
313	SCMCSD1063E01.g	SCCCLR1072H06.g	*Pto*	Le	5e-99	S/T KINASE	I	I6(2)*, I9(2)*	11
488	SCSFSD1066C01.g	SCEQRT2096A03.g	*Pto*	Le	1e-98	S/T KINASE	VIII	VIII11, VIIIg, U58	8
231	SCJFLR1071E05.g	SCJFLR1071E05.g	*Pti1*	Le	4e-59	S/T KINASE	–	–	–
258	SCJFST1047F02.g	SCRFLB1056D11.g	*Pti1*	Le	8e-75	S/T KINASE	II, VI	II9, II10, VI2	8
184	SCEQRZ3089A06.g	SCJFRZ1007B02.g	*Pti1*	Le	2e-61	S/T KINASE	IV	IV1(2)	~13
526	SCUTFL1064A02.	SCVPLB1015A04.g	*Pto*	Le	8e-58	S/T KINASE	IV, VIII	IV1, VIII3, VIII4	9
083	SCCCAM2097B01.g	SCMCCL6052C09.g	*Pto*	Le	3e-55	S/T KINASE	I	I7, I4, Ic	~11
406	SCRLLV1051H06.g	SCRULB1062B02.g	*Pto*	Le	3e-55	S/T KINASE	VIII	VIII2	6
142	SCEPFL3088D07.g	SCQGST1032C12.g	*Pto*	Le	2e-53	S/T KINASE	IV, VIII	IV1, VIII3, VIII4	~11
169	SCEQHR1079G02.g	SCBFLR1046A11.g	*Pti1*	Le	1e-159	S/T KINASE	II	II9, II10, II8	11
342	SCQGSB1080F03.g	SCJFST1012D01.g	*Pto*	Le	4e-56	S/T KINASE	VIII	VIII2, VIII11(2)	10
396	SCRLFL1010A02.g	SCMCLR1053B12.g	*Pto*	Le	5e-50	S/T KINASE	VI	VIc, VId, U51	6
335	SCQGFL8014H09.g	SCQSLR1061E06.g	*Pto*	Le	6e-50	S/T KINASE	VIII	VIII2, VIII11(2)*	14
372	SCRFFL1030E07.g	SCCCRZ2C03F07.g	*Pti1*	Le	1e-131	S/T KINASE	IV	IV1(2)*	8
012	SCACRZ3109D03.g	SCJFRT1059H08.g	*Pti1*	Le	1e-96	S/T KINASE	II	IIb, IIa, II6, U27	10
031	SCAGFL8043B10.g	SCJFRT2057D03.g	*Pti1*	Le	2e-71	S/T KINASE	–	U43	~12
275	SCJLRT1013B11.b	SCCCRZ3002D04.g	*Pti1*	Le	2e-69	S/T KINASE	II	II9, IIb, IIa, II7	~10
082	SCCCAM2003A02.g	SCCCAM2003A02.g	*Pti1*	Le	2e-66	S/T KINASE	VIII	VIII1(2)*	~12
125	SCCCSB1002G04.g	SCRFLB1053B04.g	*Pti1*	Le	2e-64	S/T KINASE	I	I4, I9, I8	8
533	SCUTSB1075A03.g	SCCCCL4007D06.g	*Pti1*	Le	3e-58	S/T KINASE	IV	IV1, IV2	~20
116	SCCCLR1C06B05.g	SCRFLR1012B04.g	*Pti1*	Le	8e-57	S/T KINASE	–	–	–
523	SCUTFL1058C03.g	SCEZRZ3016C11.g	*Pti1*	Le	2e-53	S/T KINASE	III	IIIb, IIIa	7
371	SCRFAM2127C10.g	SCEZLB1009D01.g	*Pti1*	Le	1e-51	S/T KINASE	–	U3, U46	8
176	SCEQRT2096C09.g	SCJFRT1007C05.g	*Pti1*	Le	1e-51	S/T KINASE	–	–	–
367	SCRFAD1118B11.g	SCRLRZ3039D01.g	*Pti1*	Le	6e-51	S/T KINASE	VI	VId, VI11	13
129	SCCCST1C02D03.g	SCJFST1015D08.g	*Pti1*	Le	3e-50	S/T KINASE	–	U55, U46	8

[a]Accession No.: *Cf-2.1*, gi7489083; *Xa21* (O.l.), gi7434424; *Xa21* (O.s.), gi2130082; *Hcr2-5D*, gi7488988; *Hcr2-2A*, gi3894389; *Hcr2-0A*, gi3894385; *I2C-2*, gi7489066; *Xa1*, gi7489454; *RPR1*, gi4519936; *Pib*, gi6172381; *I2*, gi4689223; *PIC17*, gi3982626; *RPP1-WsA*, gi3860163; *Hv1LRR2*, gi5669782; *Gpa2*, gi5911745; *TMV*, gi9757959; *Rp1-D*, gi5702196; *RPR1*, gi4519936; *Pto*, gi626010; *Pti1*, gi:3668069

[b]Species of origin: At, *Arabidopsis thaliana*; Hv,*Hordeum vulgare*; Le,*Lycopersicon esculentum*; Lp, *Lycopersicon pimpinellifolium*; Ol, *Oryza longistaminata*; Os,*Oryza sativa*; St,*Solanum tuberosum*; Zm, *Zea mays*

[c]HG, homology group; CG, co-segregation group; Copies, number of bands detected with the enzyme used for mapping. When there are several markers derived from the same RGA on the same CG, the number of markers is indicated in *parentheses*; if they are clustered, this is indicated by an *asterisk*

in CG VII1a (Asnaghi et al. 2000). Two LRR RGAs (RGA137 and RGA019) map on HG VII. Alleles of RGA019 map on CGs VIIa and VII14. Alleles of RGA137 map on CG VIIa, clustered with RGA019, and on CG VII1a some 5.2 cM from the rust resistance gene.

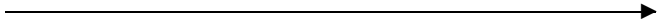


Fig. 1 Locations of the 148 RGA markers (*shaded*) on the genetic map of the sugarcane cultivar R570. The map encompasses 1123 markers, including AFLP and SSR (mSSCIR) markers, assembled into 128 Cosegregation Groups and seven Homology Groups (*numbered boxes*). Genetic distances in centiMorgans are indicated on the *left*. The rust resistance gene is indicated on CG VII.1

Characterization of the NBS/LRR RGA clusters

Two NBS/LRR RGA cluster loci were identified. Cluster 10 is located on CG U11 and contains three RGAs (RGA162, RGA152 and RGA087) with homology to the maize rust resistance gene *Rp1-D* (Collins et al. 1999). Cluster 7 is located on HG VIII on six homologous CGs (VIII1, VIII2, VIIIa, VIIIb, VIII.5 and VIII.15) and includes five NBS/LRR RGAs (RGA118, RGA281, RGA326, RGA185 and RGA267) with homology to the rice gene *RPR1*, which is responsible for probenazole-induced resistance to rice blast disease (Sakamoto et al. 1999).

Analysis of the full-length sequences of eight RGA clones revealed that almost all cDNAs seem to be incomplete at the 5′ end when compared to rice *RPR1* and maize *Rp1-D*, due possibly to an incomplete reverse transcriptase reaction. There were two exceptions: RGA118 from the *RPR1*-like cluster and RGA162 from the *Rp1-D* -like cluster. Figures 2 and 3 show the derived protein sequence alignments for *RPR1*-like and *Rp1-D* -like clusters, respectively, and indicate the NBS and LRR domains as well as their conserved motifs. Clones RGA162 and RGA152, from the *Rp1-D* -like cluster, appear to be pseudogenes, as they have stop codons in amino acid positions 655 and 233, respectively. RGA267, from the *RPR1*-like cluster, also has stop codons at positions 330 and 335 (Table 3). Although there is no difference between RGA162 and RGA152 at the amino acid level, the cDNAs are not derived from the same gene because they have sequence differences in the 3′ non coding region (data not shown) and they map 1.7 cM apart (Fig. 1).

With the aim of evaluating the divergence between and within these two NBS/LRR cluster loci, we calculated the sequence variability. For inter-cluster comparison, we aligned the part of the nucleotide sequence encoding the NBS domain of RPR1, RGA118, Rp1-D and RGA162 which were the only full length clones (amino acids 223 to 624 of RPR1 with amino acids 248 to 454 of Rp1-D). We chose this domain because it is the domain most conserved between *R* genes and outside of this region there is no significant alignment between RPR1 and Rp1-D. Since some RGA clones are incomplete, and do not include the NBS domain, it was impossible to align this region for intra-cluster analysis. Thus, we aligned part of the LRR nucleotide sequence (amino acids 557 to 901 of RPR1 for the *RPR1*-like cluster with amino acids 1008 to 1292 of Rp1-D for the *Rp1-D*-like cluster). This region corresponds to the most variable region in the *R* genes. Despite the fact that the comparison involved a variable region for intra-cluster analysis and a conserved region for inter-cluster analysis, the intra-cluster diversity at the nucleotide level (0.10 ± 0.04 for the *Rp1-D*-like cluster and 0.22 ± 0.03 for the *RPR1*-like cluster) appeared lower than the inter-cluster value (0.42 ± 0.1). This allowed the separation of these sugarcane RGAs into two clearly distinct groups: the *RPR1*-like group and the *Rp1-D*-like group.

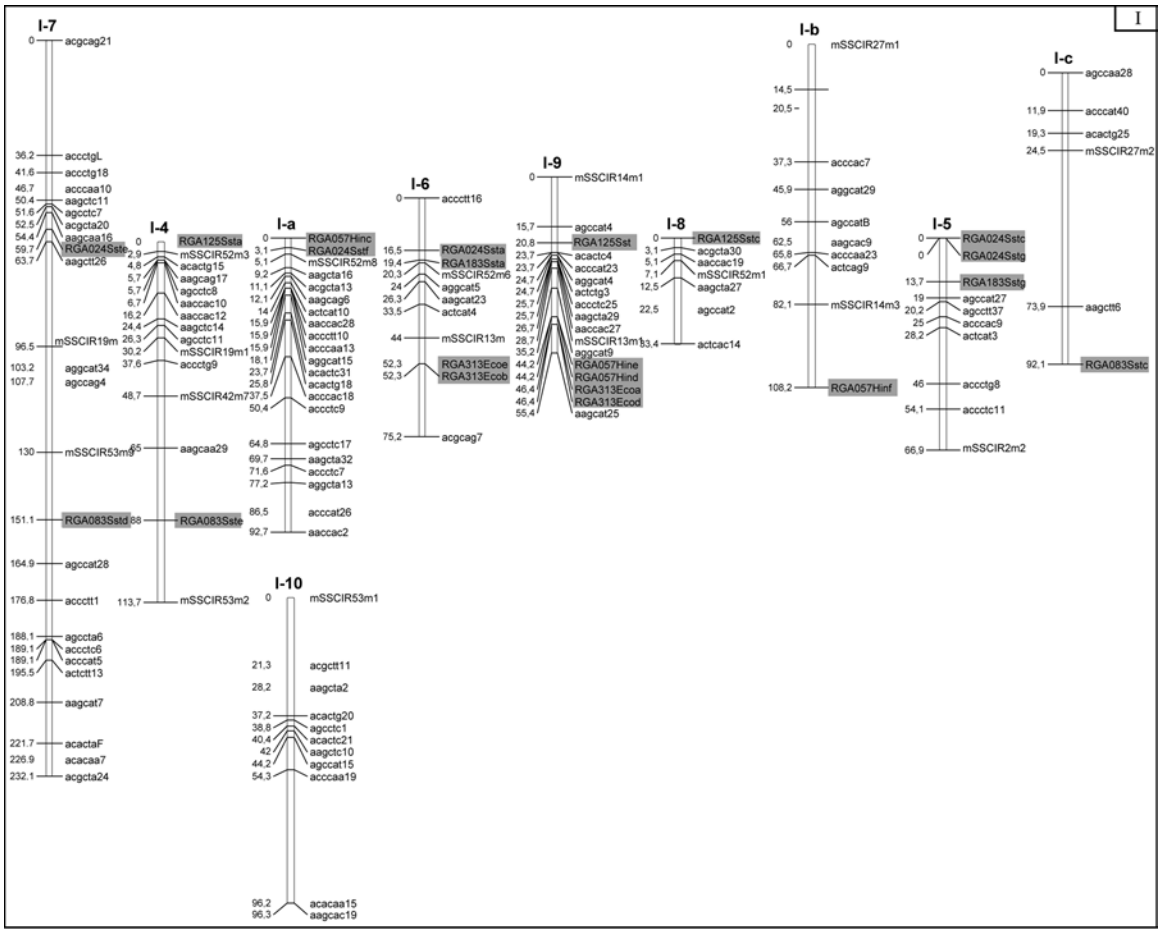
Discussion

The discovery of common sequence motifs between plant resistance genes has led to their use to develop candidate gene approaches for identifying resistance genes and analyzing their distribution in plant genomes. In this study, we have exploited the sugarcane EST database assembled in the course of the SUCEST project for both purposes.

Among the 81,223 phrap clusters comprising the 261,609 EST sequences, we have identified 88 clusters that are highly similar to *R* genes, using stringent screening procedures. Examples of RGAs encoding proteins with the three classical domains present in *R* genes (NBS/LRR, LRR and S/T KINASE) were found. No TIR/NBS/LRR-like RGAs were identified, supporting the hypothesis that this class of *R* genes has undergone divergent evolution in grasses and dicots (Pan et al. 2000, Goff et al. 2002).

We have mapped 148 markers representing 50 RGAs on the AFLP genetic map of the sugarcane cultivar R570 (Hoarau et al. 2001). Since sugarcane cultivars are highly polyploid and heterozygous, these RGAs were mapped simultaneously on several haplotypes. The SSR markers enabled us to relate the RGA mapping data to the RFLP map of R570 (consisting of approximately 1000 RFLP markers; Grivet et al. 1996, and unpublished results) and thus to organize the different haplotypes into homology groups. This will also allow the comparison of the distribution of RGAs in sugarcane to that in other species of Gramineae (Glaszmann et al. 1997, Dufour et al. 1997).

R genes are frequently reported to occur in clusters (Michelmore and Meyers 1998). In the *Arabidopsis* genome, 33% of the *R* genes are organized in pairs and 36% in clusters of three to nine members (The *Arabidopsis* Genome Initiative 2000). In sugarcane, 16 of the 50


I
I-7
0 acgcag21
36,2 accctgL
41,6 accctg18
46,7 acccaa10
50,4 aagctc11
51,6 agcctc7
52,5 acgcta20
54,4 aagcaa16
59,7 RGA024Sste
63,7 aagctt26
96,5 mSSCIR19m
103,2 aggcat34
107,7 agccag4
130 mSSCIR53m3
151,1 RGA083Sstd
164,9 agccat28
176,8 accctt1
188,1 agccta6
189,1 accctc6
189,1 acccat5
195,5 actctt13
208,8 aagcat7
221,7 acactaF
226,9 acacaa7
232,1 acgcta24
I-4
0 RGA125Ssta
2,9 mSSCIR52m3
4,8 acactg15
5,7 aagcag17
5,7 agcctc8
6,7 aaccac10
16,2 aaccac12
24,4 aagctc14
26,3 agcctc11
30,2 mSSCIR19m1
37,6 accctg9
48,7 mSSCIR42m2
65 aagcaa29
68 RGA083Sste
113,7 mSSCIR53m2
I-a
0 RGA057Hinc
3,1 RGA024Sstf
5,1 mSSCIR52m8
9,2 aagcta16
11,1 acgcta13
12,1 aagcag6
14 actcat10
15,9 aaccac28
15,9 accctt10
15,9 acccaa13
18,1 aggcat15
23,7 acactc31
25,8 acactg18
37,5 acccac18
50,4 accctc9
64,8 agcctc17
69,7 aagcta32
71,6 accctc7
77,2 aggcta13
86,5 acccat26
92,7 aaccac2
I-6
0 acccttt16
16,5 RGA024Ssta
19,4 RGA183Ssta
20,3 mSSCIR52m6
24 aggcat5
26,3 aagcat23
33,5 actcat4
44 mSSCIR13m
52,3 RGA313Ecoe
52,3 RGA313Ecob
75,2 acgcag7
I-9
0 mSSCIR14m1
15,7 agccat4
20,8 RGA125Sst
23,7 acactc4
23,7 acccat23
24,7 aggcat4
24,7 actctg3
25,7 accctc25
25,7 aagcta29
26,7 aaccac27
28,7 mSSCIR13m1
35,2 aggcat9
44,2 RGA057Hine
44,2 RGA057Hind
46,4 RGA313Ecoa
46,4 RGA313Ecod
55,4 aagcat25
I-8
0 RGA125Sstc
3,1 acgcta30
5,1 aaccac19
7,1 mSSCIR52m1
12,5 aagcta27
22,5 agccat2
33,4 actcac14
I-b
0 mSSCIR27m1
14,5
20,5
37,3 acccac7
45,9 aggcat29
56 agccatB
62,5 aagcac9
65,8 acccaa23
66,7 actcag9
82,1 mSSCIR14m3
108,2 RGA057Hinf
I-5
0 RGA024Sstc
0 RGA024Sstg
13,7 RGA183Sstg
19 agccat27
20,2 agcctt37
25 acccac9
28,2 actcat3
46 accctg8
54,1 accctc11
66,9 mSSCIR2m2
I-c
0 agccaa28
11,9 acccat40
19,3 acactg25
24,5 mSSCIR27m2
73,9 aagctt6
92,1 RGA083Sstc
I-10
0 mSSCIR53m1
21,3 acgctt11
28,2 aagcta2
37,2 acactg20
38,8 agcctc1
40,4 acactc21
42 aagctc10
44,2 agccat15
54,3 acccaa19
96,2 acacaa15
96,3 aagcac19
II
II-9
0 RGA366Hina
9,3 mSSCIR41m1
13,8 mSSCIR56m7
16,4 acgcta17
19,1 accctg26
21,8 accctg25
23,6 accctg16
33,6 aagctc4
38,4 aggctg9
38,4 acgctt44
38,4 aggcat8
40,1 agccat6
42,8 mSSCIR35m
47,4 accctg11
54,1 aggcat19
59,9 RGA258Draa
82,4 acactaK
130 RGA169Hinc
146,3 acactt30
152 accctc16
157,2 aggcta19
176,2 mSSCIR34m2
188,8 aagctt37
196,1 acactc7
203,5 aggcat6
205,7 aagcta13
207,7 agccag17
210,9 RGA275Ecod
221,3 agccaa14
II-12
0 RGA366Hind
8,2 mSSCIR35m3
II-10
0 actcag11
5 aagcac4
10 mSSCIR41m2
23,3 RGA366Hinc
46,7 mSSCIR35m2
68,2 RGA258Drab
87,8 acccttt5
110, RGA169Hinf
II-6
0 actcat34
9,4 agccag16
25,4 RGA386Drac
33,3 acccat13
34,5 actctt1
55,5 mSSCIR34m1
63,8 mSSCIR25m2
81,4 aggcat24
82,4 acccat12
92,9 actcac6
95,8 aggcta6
109, RGA012Ecoa
II-8
0 actcac7
12,5 mSSCIR50m
24,9 aggctc4
31,1 agcctt20
35,1 mSSCIR25m1
49,3 mSSCIR34m2
73,1 acactt28
75,2 acccac13
82,2 RGA386Draa
86 RGA169Hina
99,3 acgctt35
118, mSSCIR48m2
119, aaccac24
120, mSSCIR69m2
122, aaccac5
150, acacta19
II-a
0 acacta14
30,5 accctgC
39,7 acgcag19
58,7 acactg21
63,3 aggctc5
66,4 accctc5
82,1 acgctt10
87,9 agccaa21
92,5 agccaa1
115, accctc10
116, acccaa29
118, actcag5
120, agccat12
125, RGA275Ecoe
130 RGA012Ecof
143, RGA145Ecoe
II-b
0 acgcagB
29,9 acccac3
47 acgctt6
47,8 accctc12
48,6 aggctg3
48,6 aagcaa10
54,8 RGA275Ecoc
59,2 RGA012Ecoc
77,1 agccaa25
90,4 aagcac18
II-11
0 mSSCIR48m3
0 mSSCIR56m
II-7
0 RGA275Ecof
10,9 aagcag5
13,7 acactt14
13,7 actcac9
15,6 aagcag37
18,4 aggctaC
21,2 acgctt8
21,2 aagcaa9
26 acccaa31
27,9 aagcag12
32,8 mSSCIR39m2

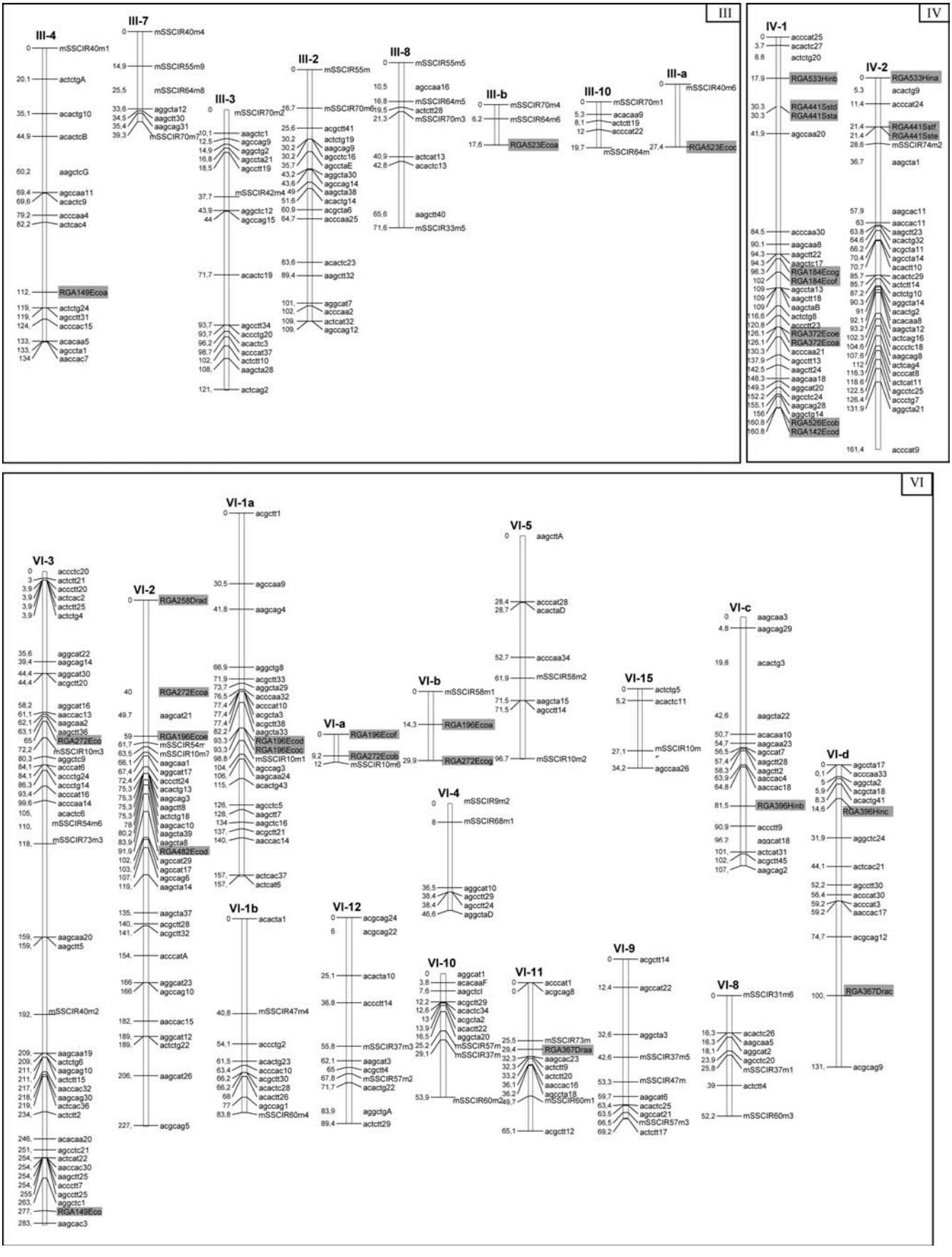

III
III-4
0 mSSCIR40m1
20,1 actctgA
35,1 acactg10
44,9 acactcB
60,2 aagctcG
69,4 agccaa11
69,6 acactc9
79,2 acccaa4
82,2 actcac4
112, RGA149Ecoa
119, actctg24
119, agcctt31
124, acccac15
133, acacaa5
133, agccta1
134 aaccac7
III-7
0 mSSCIR40m4
14,9 mSSCIR55m9
25,5 mSSCIR64m8
33,6 aggcta12
34,5 aagctt30
35,4 aagcag31
39,3 mSSCIR70m7
III-3
0 mSSCIR70m2
10,1 aagctc1
12,5 agccag9
14,9 aggctg2
16,8 agccta21
18,5 agcctt19
37,7 mSSCIR42m4
43,9 aggctc12
44 agccag15
71,7 acactc19
93,7 agcctt34
93,7 accctg20
96,2 acactc3
98,7 acccat37
102, actctt10
108, aagcta28
121, actcag2
III-2
0 mSSCIR55m
16,7 mSSCIR70m6
25,6 acgctt41
30,2 actctg19
30,2 aagcag9
30,2 agcctc16
35,7 agcctaE
43,2 aggcta30
43,6 agccag14
49 aagcta38
51,6 acactg14
60,9 acgcta6
64,7 acccaa25
83,6 acactc23
89,4 aagctt32
101, aggcat7
102, acccaa2
109, actcat32
109, agccag12
III-8
0 mSSCIR55m5
10,5 agccaa16
16,8 mSSCIR64m5
19,5 actctt28
21,3 mSSCIR70m3
40,9 actcat13
42,8 acactc13
65,6 aagctt40
71,6 mSSCIR33m5
III-b
0 mSSCIR70m4
6,2 mSSCIR64m6
17,6 RGA523Ecoa
III-10
0 mSSCIR70m1
5,3 acacaa9
8,1 actctt19
12 acccat22
19,7 mSSCIR64m
III-a
0 mSSCIR40m6
27,4 RGA523Ecoc
IV
IV-1
0 acccat25
3,7 acactc27
8,8 actctg20
17,9 RGA533Hinb
30,3 RGA441Sstd
30,3 RGA441Ssta
41,9 agccaa20
84,5 acccaa30
90,1 aagcaa8
94,3 aagctt22
94,3 aagctc17
96,3 RGA184Ecog
102 RGA184Ecof
109 agccta13
109 aagctt18
109 aagctaB
116,6 actctg8
120,8 accctt23
126,1 RGA372Ecoe
126,1 RGA372Ecoa
130,3 acccaa21
137,9 agcctt13
142,5 aagctt24
148,3 aagcaa18
149,3 aggcat20
152,2 agcctc24
155,1 aagcag28
156 aggctg14
160,8 RGA526Ecob
160,8 RGA142Ecod
IV-2
0 RGA533Hina
5,3 acactg9
11,4 acccat24
21,4 RGA441Sstf
21,4 RGA441Sste
28,6 mSSCIR74m2
36,7 aagcta1
57,9 aagcac11
63 aaccac11
63,8 aagctt23
64,6 acactg32
66,2 acgcta11
70,4 agccta14
70,7 acactt10
85,7 acactc29
85,7 actctt14
87,2 actctg10
90,3 aggcta14
91 acactg2
92,1 acacaa8
93,2 aagcta12
102,3 actcag16
104,6 accctc18
107,6 aagcag8
112 actcag4
116,3 acccat8
118,6 actcat11
122,5 agcctc25
126,4 accctg7
131,9 aggcta21
161,4 acccat9
VI
VI-3
0 accctc20
3 actctt21
3,9 accctt20
3,9 actcac2
3,9 actctt25
3,9 actctg4
35,6 aggcat22
39,4 aagcag14
44,4 aggcat30
44,4 acgctt20
58,2 aggcat16
61,1 aaccac13
62,1 aagcaa2
63,1 aagctt36
65 RGA272Ecoc
72,2 mSSCIR10m3
80,3 aggctc9
84,1 acccat6
84,1 accctg24
86,3 accctg14
93,4 acccat16
99,6 acccaa14
105, acactc6
110, mSSCIR54m6
118, mSSCIR73m3
159, aagcaa20
159, aagctt5
192, mSSCIR40m2
209, aagcaa19
209, actctg6
211, aagcag10
211, actctt15
217, aaccac32
218, aagcag30
219, actcac36
234, actctt2
246, acacaa20
251, agcctc21
254, actcat22
254, aaccac30
254, aagctt25
254, accctt7
255 agcctt25
263, aggctc1
277, RGA149Ecoc
283, aagcac3
VI-2
0 RGA258Drad
40 RGA272Ecoa
49,7 aagcat21
59 RGA196Ecoe
61,7 mSSCIR54m7
63,5 mSSCIR10m7
66,1 aagcaa1
67,4 aggcat17
72,4 accctt24
75,3 acactg13
75,3 aagcag3
75,3 aagctt8
75,3 actctg18
78 aagcac10
80,2 aagcta39
83,9 aagcta8
91,9 RGA480Ecod
102 agccat29
103 agccat17
107 agccag6
119 aagcta14
135 aagcta37
140 acgctt28
141 acgctt32
154 acccatA
166 aggcat23
166 agccag10
182 aaccac15
189 aggcat12
189 actctg22
206 aagcat26
227 acgcag5
VI-1a
0 acgctt1
30,5 agccaa9
41,8 aagcag4
66,9 aggctg8
71,9 acgctt33
73,7 aggcta29
76,5 acccaa32
77,4 acccat10
77,4 acgcta3
77,4 acgctt38
82,2 aagcta33
93,3 RGA196Ecod
93,3 RGA196Eccc
98,8 mSSCIR10m1
104, agccag3
106, aagcaa24
115, acactg43
126, agcctc5
128, aagctt7
134 aagctc16
137, acgctt21
140, aaccac14
157, actcac37
157, actcat6
VI-1b
0 acacta1
40,8 mSSCIR47m4
54,1 accctg2
61,5 acactg23
63,4 acccac10
66,2 acgctt30
66,2 acactc28
68 acactt26
77 agccag1
83,8 mSSCIR60m4
VI-a
0 RGA196Ecof
9,2 RGA272Ecob
12 mSSCIR10m6
VI-12
0 acgcag24
6 acgcag22
25,1 acacta10
36,8 accctt14
55,8 mSSCIR37m3
62,1 aagcat3
65 acgctt4
67,8 mSSCIR57m2
71,7 acactg22
83,9 aggctgA
89,4 actctt29
VI-b
0 mSSCIR58m1
14,3 RGA196Ecoa
29,9 RGA272Ecog
VI-4
0 mSSCIR9m2
8 mSSCIR68m1
36,5 aggcat10
38,4 agcctt29
38,4 agcctt24
46,6 aggctaD
VI-10
0 aggcat1
3,8 acacaaF
7,6 aagctcl
12,2 acgctt29
12,6 acactc34
13 acgcta2
13,9 acactt22
16,5 aggcta20
25,2 mSSCIR57m
29,1 mSSCIR37m
53,9 mSSCIR60m2
VI-5
0 aagcttA
28,4 acccat28
28,7 acactaD
52,7 acccaa34
61,9 mSSCIR58m2
71,5 aagcta15
71,5 agcctt14
96,7 mSSCIR10m2
VI-11
0 acccat1
0 acgcag8
25,5 mSSCIR73m
29,4 RGA367Draa
32,3 aagcac23
32,3 actctt9
33,2 actctt20
36,1 aaccac16
36,2 aggcta18
49,7 mSSCIR60m1
65,1 acgctt12
VI-15
0 actctg5
5,2 acactc11
27,1 mSSCIR10m
34,2 agccaa26
VI-9
0 acgctt14
12,4 agccat22
32,6 aggcta3
42,6 mSSCIR37m5
53,3 mSSCIR47m
59,7 aagcat6
63,4 acactc25
63,5 agccat21
66,5 mSSCIR57m3
69,2 actctt17
VI-c
0 aagcaa3
4,8 aagcag29
19,8 acactg3
42,6 aagcta22
50,7 acacaa10
54,7 aagcaa23
56,5 agccat7
57,4 aagctt28
58,3 aagctt2
63,9 aaccac4
64,8 aaccac18
81,5 RGA396Hinb
90,9 acccttg
96,2 aggcat18
101, actcat31
102, acgctt45
107, aagcag2
VI-8
0 mSSCIR31m6
16,3 acactc26
16,3 aagcaa5
18,1 aggcat2
23,9 agcctc20
25,8 mSSCIR37m1
39 actctt4
52,2 mSSCIR60m3
VI-d
0 agccta17
0,1 acccaa33
5 aggcta2
5,9 acgcta18
8,3 acactg41
14,6 RGA396Hinc
31,9 aggctc24
44,1 actcac21
52,2 agcctt30
56,4 acccat30
59,2 acccat3
59,2 aaccac17
74,7 acgcag12
100, RGA367Drac
131, acgcag9

Fig. 1 (Contd.)

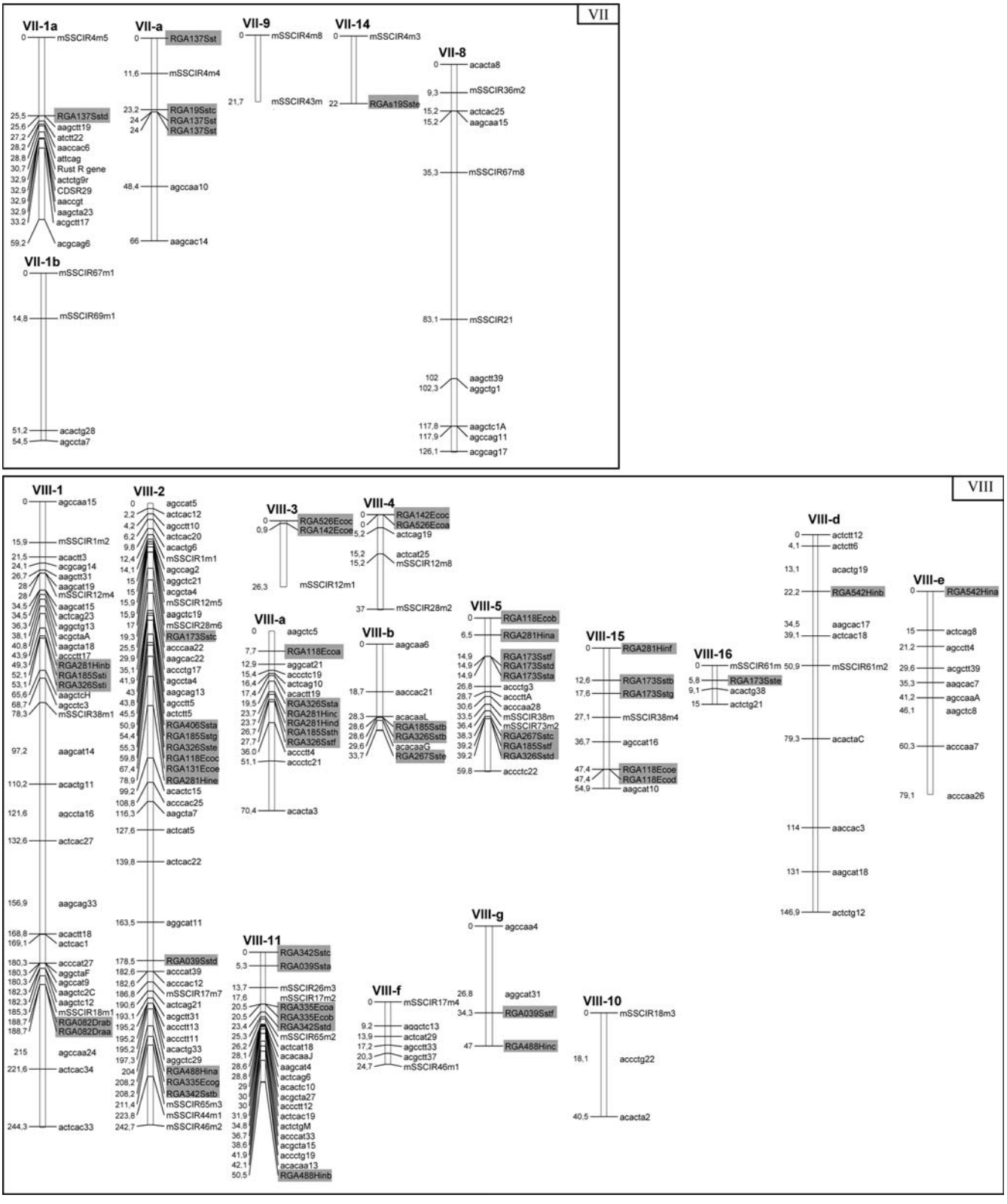


Fig. 1 (Contd.)

mapped RGA loci are organized in four clusters containing three to six different RGAs, while 12 are in pairs (Table 2). The RGAs that belong to the same cluster were not all mapped on every homologous CG. This is probably due to the constraints of mapping in polyploids (only single dose markers can be mapped) but could also be a consequence of gene losses that may have been part of the rapid and extensive genome changes

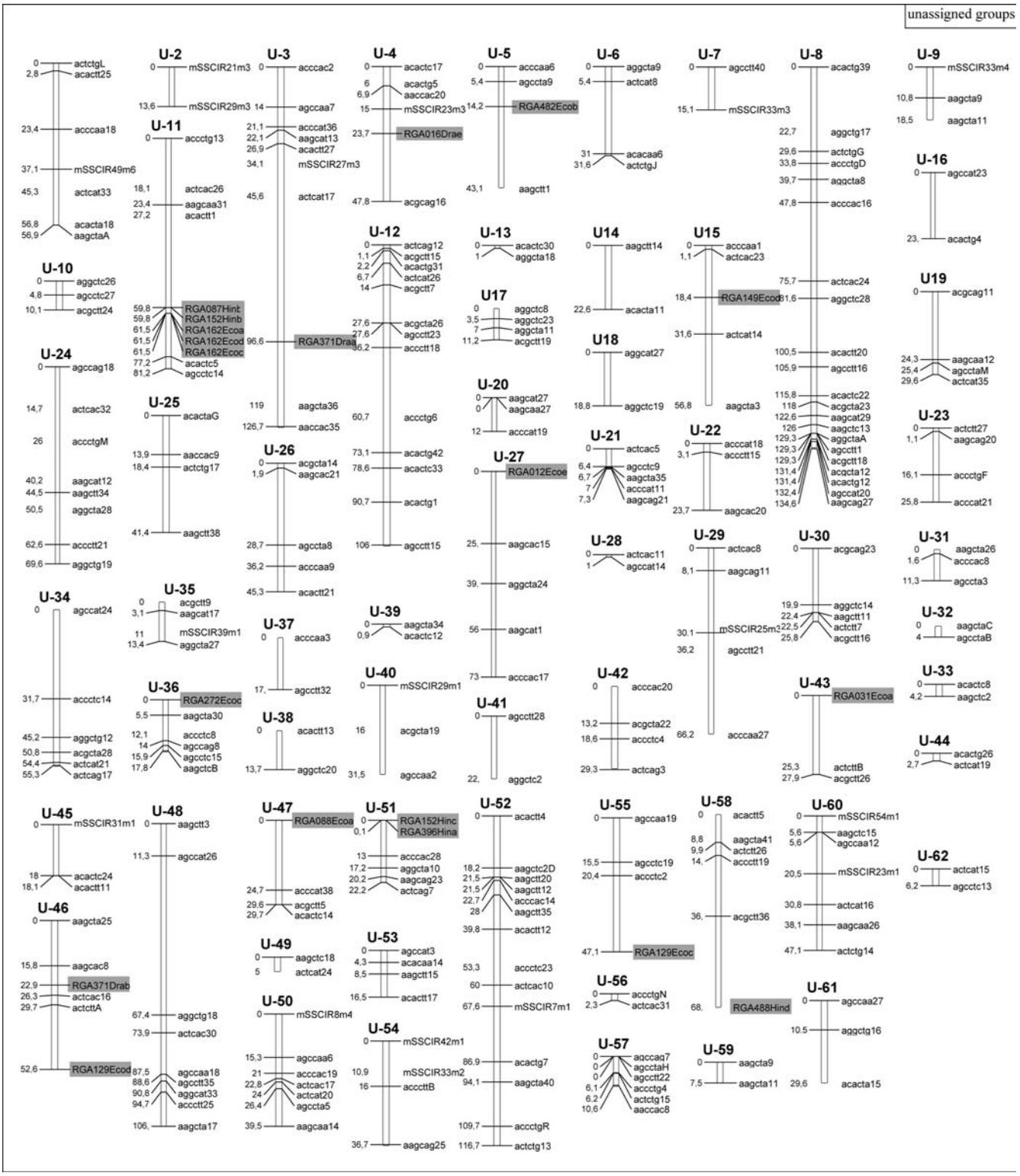


Fig. 1 (Contd.)

after polyploidization (Wendel 2000; Feuillet et al. 2001).

There is evidence that *R* gene clusters may contain functionally related genes that are not necessarily similar at the sequence level. This is the case for the tomato *Pto* cluster, which encodes five related kinases and a NBS/LRR (*Prf*) protein also required for resistance to *Pseudomonas syringae* (Salmeron et al. 1996). In this study, we also observed a few clusters containing RGAs with different protein domains (Table 2).

Table 2 Clusters containing different RGAs

Cluster No.[a]	RGA clones RGA domains	Cluster size (cM)	Map position	
			HG	CG
1*	057 (NBS/LRR), 024 (LRR), 183 (LRR), 313 (S/T KINASE)	–	I	Ia; Ib; I5; I6; I7;I9
2	275 (S/T KINASE), 012 (S/T KINASE)	4.7; 4.4	II	IIa; IIb
3	386 (LRR), 169 (S/T KINASE)	3.8	II	II8
4	526 (S/T KINASE), 142 (S/T KINASE)	0	IV	IV1
5	526 (S/T KINASE), 142 (S/T KINASE)	0.9; 0	VIII	VIII3; VIII4
6	019 (LRR), 137 (LRR)	0.8	VII	VIIa
7*	118 (NBS/LRR), 185 (NBS/LRR), 267 (NBS/LRR), 281 (NBS/LRR) 326 (NBS/LRR), 406 (S/T KINASE)	–	VIII	VIII1; VIII2; VIIIa; VIIIb; VIII5; VIII15
8*	335 (S/T KINASE), 342 (S/T KINASE), 488 (S/T KINASE)	–	VIII	VIII2; VIII11; VIIIg
9	152 (NBS/LRR), 396 (S/T KINASE)	0.1	–	U51
10	087 (NBS/LRR), 152 (NBS/LRR), 162 (NBS/LRR)	1.7	–	U11

[a]This number refers to the basic genome. Clusters 1, 7 and 8 (indicated by the *asterisks*) consist of several RGAs that map on several homologous chromosome segments, the distance between RGAs on each CG is variable but at least in one CG the distance between each pair of consecutive RGAs is less 5 cM or less. Not all RGAs were mapped on each CG (see Fig. 1 and text for details)

It is noteworthy that all the RGAs homologous to rice RPR1 map together in cluster 7, and all the NBS/LRR RGAs homologous to maize Rp1-D map together in cluster 10. Sequence comparison of these NBS/LRR RGAs with the respective references in rice or maize reveals that the members of a given sugarcane cluster are more similar to the alien reference (RPR1 or Rp1-D) than to members of the other sugarcane NBS/LRR locus. This observation suggests the existence of a common ancestral gene for rice *RPR1*and the sugarcane *RPR1*-like cluster, and for maize *Rp1-D* and the sugarcane *Rp1-D*-like cluster.

In addition, protein sequence alignments of the RPR1-like group (Fig. 2), as well as nucleotide sequence analysis (data not shown), show that sugarcane RGAs are not always more similar to each other than to the corresponding rice ortholog. This phenomenon of greater distance between paralogous than between orthologous sequences has already been highlighted by others (Michelmore and Meyers 1998; Feuillet et al. 2001), and led Michelmore and Meyers (1998) to propose a model for resistance cluster evolution called ''birth and death''.

Many authors have reported linkages between RGAs and disease resistance loci or QTLs (Wang et al. 2001, Graham et al. 2000). This is particularly interesting for sugarcane since, due to its particularly complex genome, only one resistance gene has been localized so far (Daugrois et al 1996; Asnaghi et al. 2000). This major resistance gene, which confers resistance to common rust, has been located on R570 maps and is the focus of a map-based cloning approach (Asnaghi et al. 2000, and unpublished results). The present work identified an LRR cluster near the rust resistance locus, thus indicating the presence of RGAs in this genome region. In addition, the data generated in this study on the distribution of RGAs in the sugarcane genome will provide extremely valuable information for current efforts aimed at mapping resistance genes for other sugarcane diseases including leaf scald (Offmann, personal communication) and smut (Raboin et al. 2001).

Despite the success of the RGA approach to the identification of disease resistance loci, the challenge often remains in recognizing the functional gene within clusters. *R* gene clusters typically contain several related sequences, and even in the best studied cases, only for half of them has any specificity been demonstrated (Michelmore and Meyers 1998). With regard to this aspect, EST-RGA resources may have advantages, compared to PCR amplification of conserved motifs or ''candidate genes'' from genome sequencing data. The EST approach considers only expressed genes, thus eliminating many pseudogenes that cannot be transcribed. However, cDNAs with internal stop codons, indicative of non-functional protein, were also found in this study (RGA162, RGA152 and RGA267), and already reported by Vicente and King 2001. In polyploids, the formation of pseudogenes through accumulation of mutations may be a consequence of the reduction in selection pressure on genes that are present in several copies (Wendel 2000).

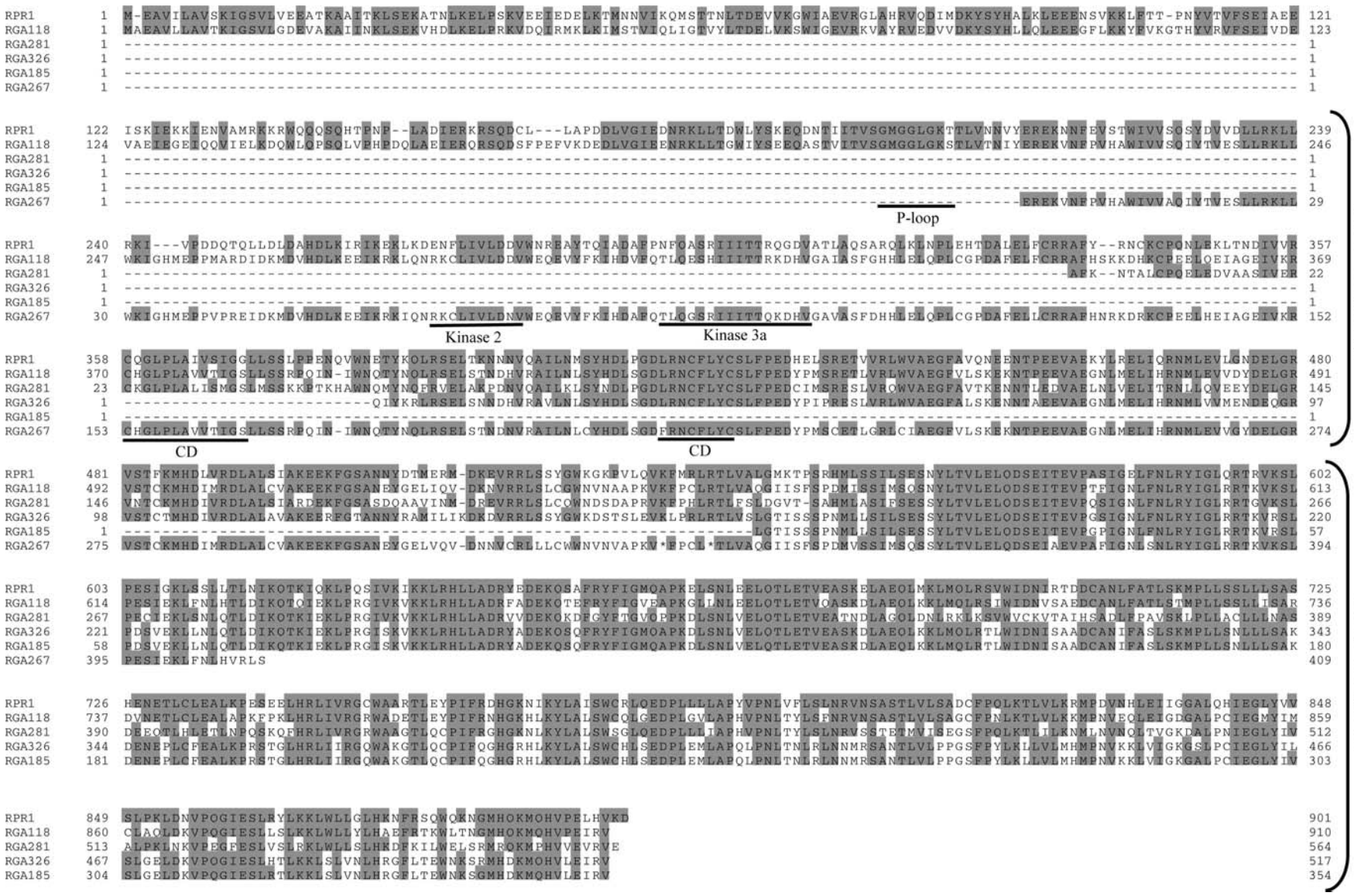


Fig. 2 Alignment of derived protein sequences encoded in the *RPR1*-like NBS/LRR-like RGA cluster on HG VIII. The *shaded* amino acids indicate sequence identity to the RPR1 protein of rice. NBS motifs (P-loop, Kinase 2 and Kinase 3a) and regions conserved between resistance gene products (CD) are *underlined*. Protein domains are indicated on the *right*

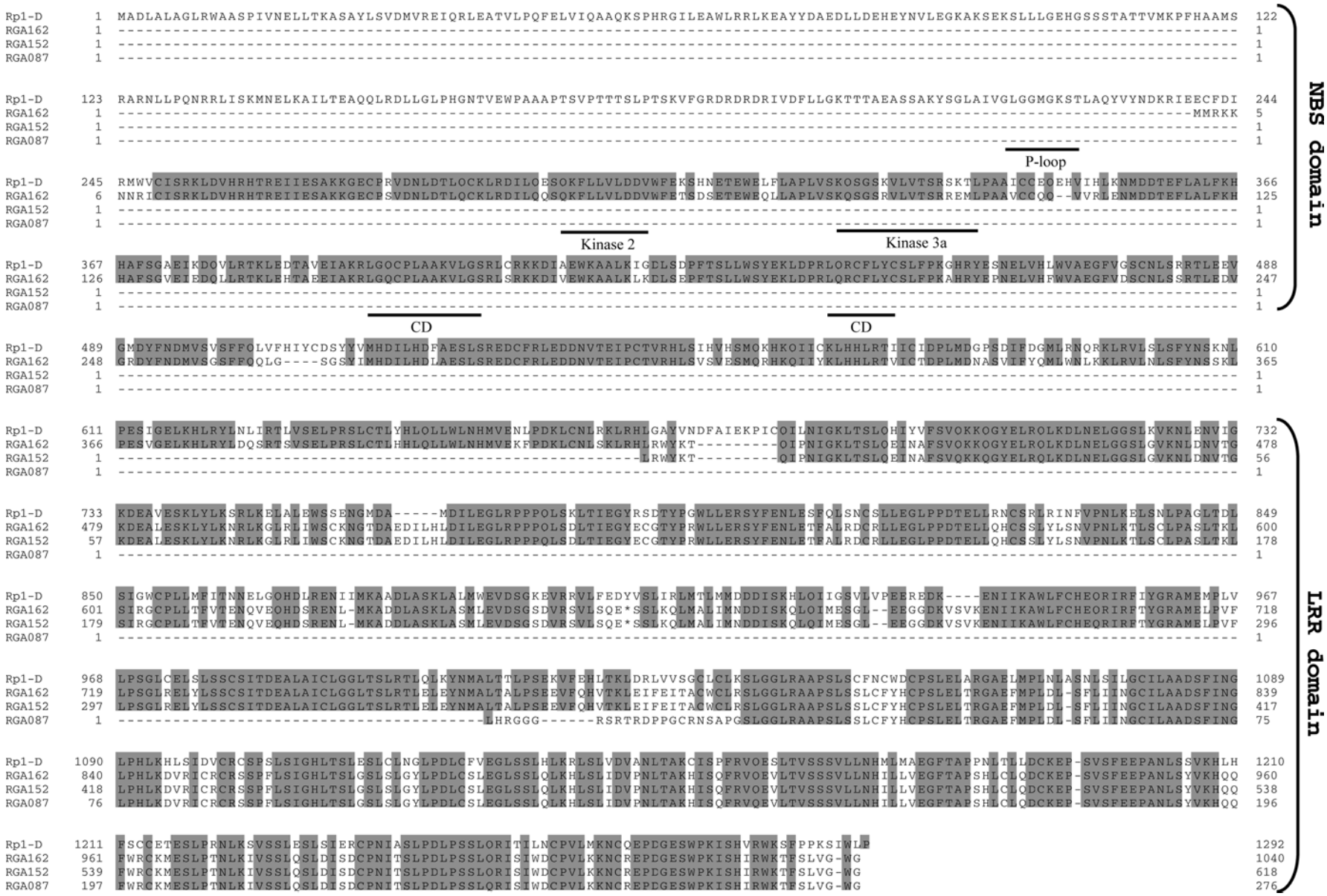


Fig. 3 Alignment of derived protein sequences encoded in the *Rp1-D* -like NBS/LRR-like RGA cluster on CG U11. The *shaded* amino acids indicate sequence identity to the Rp1-D protein from maize. NBS motifs (P-loop, Kinase 2 and Kinase 3a) and regions conserved between resistance gene products (CD) are *underlined*. Protein domains are indicated on the *right*

Table 3 Characteristics of *RPR1*- and *Rp1-D* -like RGAs

Clone name	Closest homolog	Size (nt)	Poly(A) tail	Size of putative protein product (aa)	Met Start Codon	Internal stop codon
RGA118	*RPR1* (901 aa)	3182 bp	+	910	+	–
RGA281	*RPR1* (901 aa)	1992 bp	+	564	–	–
RGA326	*RPR1* (901 aa)	1840 bp	+	517	–	–
RGA185	*RPR1* (901 aa)	1359 bp	+	354	–	–
RGA267	*RPR1* (901 aa)	1436 bp	+	409	–	+
RGA162	*Rp1-D* (1292 aa)	4012 bp	+	1040	+	+
RGA152	*Rp1-D* (1292 aa)	2462 bp	+	618	–	+
RGA087	*Rp1-D* (1292 aa)	1392 bp	+	276	–	–

Acknowledgements M. Rossi, P. G. Araujo and V. M. Dias were recipients of FAPESP fellowships. This work was partially supported by grants from FAPESP and CNPq (Brazil). We thank J. C. Glaszmann, L. Grivet, G. Piperidis, J. B. Morel and Cristina Juarez for helpful contributions.

References

Asnaghi C, Paulet F, Kaye C, Grivet L, Deu M, Glaszmann JC, D'Hont A (2000) Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. Theor Appl Genet 101:962–969

Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, Druka A, Steffenson B (2002) The barley stem rust-resistance gene *Rpg1*is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328–9333

Butterfield MK, D'Hont A, Berding N (2001) The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. Proc Soc Afr Sugarcane Technol Assoc 75:1–5

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H., Glaszmann JC, D'Hont A (1996) A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar ''R570''. Theor Appl Genet 92:1059–1064

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JDG (1998) The tomato *Cf-5*disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell 10:1915–1925

D'Hont A, Glaszmann JC (2001) Sugarcane genome analysis with molecular markers, a first decade of research. Proc Int Soc Sugarcane Technol 24:556–559

Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor Appl Genet 94:409–418

Feuillet C, Penger A, Gellner K, Mast A, Keller B (2001) Molecular evolution of receptor-like kinase genes in hexaploid wheat. Independent evolution of orthologs after polyploidization and mechanisms of local rearrangements at paralogous loci. Plant Physiol 125:1304–1313

Flor HH (1971) Current status of the gene-for-gene concept. Annu Rev Phytopathol 9:275–96

Glaszmann JC, Dufour P, Grivet L, D'Hont A, Deu M, Paulet F, Hamon P (1997) Comparative genome analysis between several tropical grasses. Euphytica 96:13–21

Goff AS, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 296:92–100

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

Grivet L, Arruda P (2001) Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol 5:122–127

Grivet L, D'Hont A, Roques D, Feldman P, Lanaud C, Glaszmann JC (1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. Genetics 142:987–1000

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

Hoarau JY, Offman B, D'Hont A, Risterucci AM, Roques D, Glaszman JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. Theor Appl Genet 103:84–97

Keen NT (1990) Gene-for-gene complementarity in plant-pathogen interactions. Annu Rev Genet 24:447–463

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

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earie E, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Pan Q, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50:203–213

Raboin LM, Offmann B, Hoarau JY, Notise J, Costet L, Telismart H, Roques D, Rott P, D'Hont A (2001) Undertaking genetic mapping of sugarcane smut resistance. Proc Soc Afr Sugarcane Technol Assoc 75:94–98

Rozas J, Rozas R (1997) DnaSP version 2.0: a novel software package for extensive molecular population genetic analysis. Comput Appl Biosci 13:307–311

Sakamoto K, Tada Y, Yokozeki Y, Akagi H, Hayashi N, Fujimura T, Ichikawa N (1999) Chemical induction of disease resistance in rice is correlated with the expression of a gene encoding a nucleotide binding site and leucine-rich repeats. Plant Mol Biol 40:847–855

Salmeron JM, Oldroyd GED, Rommens CMT, Scofield SR, Kim HS, Lavelle DT, Dahlbeck D, Staskawicz BJ (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86:123–133

Song W, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806

Telles GP, Braga MDV, Dias Z, Lin TL, Quitzau JAA, da Silva FR, Meidanis J (2001) Bioinformatics of the sugarcane EST project. Genet Mol Biol 24:9–15

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

Vettore AL, da Silva FR, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genet Mol Biol 24:1–7

Vicente JG, King GJ (2001) Characterization of disease resistance gene-like sequences in *Brassica oleracea* L. Theor Appl Genet 102:555–563

Wang Z, Taramino D, Yang D, Liu G, Tingey SV, Miao G, Wang G (2001) Rice ESTs with disease-resistance gene- or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol Genet Genomics 265:302–310

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor Appl Genet 83:294–300

Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, Turner JG (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by *RPW8*. Science 291:118–120

Zhou J, Loh YT, Bressan RA, Martin GB (1995) The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response.