Theor Appl Genet (2003) 108:121–130
DOI 10.1007/s00122-003-1382-1

ORIGINAL PAPER

G. P. Bernet · M. J. Asíns

Identification and genomic distribution of *gypsy* like retrotransposons in *Citrus* and *Poncirus*

Received: 15 March 2003 / Accepted: 13 June 2003 / Published online: 22 August 2003

Abstract Transposable elements might be importantly involved in citrus genetic instability and genome evolution. The presence of *gypsy* like retrotransposons, their heterogeneity and genomic distribution in *Citrus* and *Poncirus*, have been investigated. Eight clones containing part of the POL coding region of *gypsy* like retrotransposons have been isolated from a commercial variety of *Citrus clementina,* one of the few sexual species in *Citrus*. Four of the eight clones might correspond to active elements given that they present all the conserved motifs described in the literature as essential for activity, no in-frame stop codon and no frame-shift mutation. High homology has been found between some of these citrus elements and retroelements within a resistance-gene cluster from potato, another from *Poncirus trifoliata* and two putative resistance polyproteins from rice. Nested copies of *gypsy* like elements are scattered along the *Citrus* and *Poncirus* genomes. The results on genomic distribution show that these elements were introduced before the divergence of both genera and evolved separately thereafter. IRAPs based on *gypsy* and *copia* types of retrotransposons seem to distribute differently, therefore gypsy based IRAPs prove a new, complementary set of molecular markers in *Citrus* to study and map genetic variability, especially for disease resistance. Similarly to *copia*-derived IRAPs, the number of copies and heterozygosity values found for *gypsy* derived IRAPs are lower in *Poncirus* than in *Citrus aurantium*, which is less apomictic and the most usual rootstock for clementines until 1970.

Keywords Genetic variability · IRAP · Resistance genes · Apomixis · Citrus improvement

Communicated by C. Möllers

G. P. Bernet · M. J. Asíns (✉)
Instituto Valenciano de Investigaciones Agrarias,
Apdo. Oficial, 46113 Moncada, Valencia, Spain
e-mail: mjasins@ivia.es
Tel.: +34-96-3424067
Fax: +34-96-3424001

Introduction

Citrus is one of the most important fruit crops in the world. Citrus species are diploid (2n=18) trees with hesperidium fruits, and seeds often with two or more nucellar embryos which are genetically identical to the seed parent. Nucellar embryony (a type of apomictic reproduction) has very important consequences for evolution, breeding and the culture of citrus fruit trees (Asíns et al. 2002). Citrus is almost universally propagated by budding onto (nucellar) rootstocks to ensure yield uniformity. Bud mutations arise often in citrus (Raghuvanshi 1962; Cameron and Frost 1968) and are generally detected by the growers themselves in branches of trees showing altered horticultural traits, such as maturity and flowering time or fruit characteristics. Transposons can clearly inactivate genes by integration or by causing methylation in the region where they are located. They might also contribute to agronomic variation (maturation date, flesh color) by increasing allelic diversity or by changing the regulation of gene expression. Nevertheless, up to now, no experimental data supports transposon activity as a source of bud mutations.

Based on the domain structure in the POL region, LTR retrotransposons are divided into two groups, the Ty1/*copia* type (*pro-int-rt-rh* from 5′ to 3′) and the Ty3/*gypsy* type (*pro-rt-rh-int*). Asíns et al. (1999) investigated the presence of *copia*-like retrotranposons in citrus. They found that these elements were quite abundant throughout the citrus genome and very heterogeneous for the *rt* domain. Polymorphisms based on *copia*-like elements (RFLPs and IRAPs) have been found distinguishing groups of varieties within *Citrus sinensis* (Asíns et al. 1999), *Ctrus clementina* (Bretó et al. 2001) and *Ctrus limon* (Bernet et al. 2003). Moreover, polymorphisms based on these elements are more abundant than those based on primers of random sequence or simple sequence repeats (Bretó et al. 2001).

Gypsy type retrotransposons are the most similar elements to retroviruses. Both of them are organized in the same manner differing mainly in the infective

capability of retroviruses which critically depends on a third open reading frame (ENV) encoding envelope-glycoproteins (Frankel and Young 1998). *Gypsy* like retrotransposons containing ENV-like domains have been already reported even in the plant kingdom (Vicient et al. 2001b) where retroviruses are thought to be lacking. Until recently, little was known about the Ty3/*gypsy* group of elements in plants (Smyth et al. 1989; Purugganan and Wessler 1994). Nowadays, it seems clear that Ty3/*gypsy* like retrotransposons appear to be broadly distributed among plants in multiple families like the Ty1/*copia* group (Chavanne et al. 1998; Suoniemi et al. 1998; Friesen et al. 2001; Shcherban et al. 2001; Feschotte et al. 2002). There seems to be no difference for activity between both either types. This activity is low (Grandbastien 1998; Vicient et al. 2001a; Echenique et al. 2002; Feschotte et al. 2002) and low to middle repetitive LTR retrotransposons are more frequently found in EST collections of maize than the very high copy number elements (Meyers et al. 2001). More importantly, there are evidences that biotic and abiotic stresses are related to an increment of their activity (Hirochika et al. 1996; Echenique et al. 2002; Feschotte et al. 2002). Although integration sites for most mammalian and *Drosophila* retroelements appear to be distributed more or less randomly in the genome, a clear bias in the site-selection choice has also been observed, at least for some yeast retrotranposons. Thus, Ty3 elements integrate almost exclusively upstream of genes transcribed by RNA polymerase III (Chalker and Sandmeyer 1992). Therefore, the distribution of Ty1/*copia* like and Ty3/*gypsy* like retrotransposons on the citrus might be different. The investigation of their heterogeneity, activity and genomic distribution in apomictic perennial species might contribute to our understanding about the evolution of their genomes. This knowledge might also suggest new-ways to improve cultivated citrus. Hence, the objectives of the present paper are to investigate the presence of *gypsy* like retrotransposons, in *Citrus* and *Poncirus*, their heterogeneity and their genomic distribution.

Materials and methods

Plant materials

All citrus plants analyzed belong to the Citrus germplasm bank at IVIA. Varieties "Fino", "Doblefina" and "Loretina" from *C. limon* (L.) Burm f., *C. sinensis* (L.) Osb. and *C. clementina* Hort. ex Tan., respectively, were selected for amplification of gypsy like elements. Similarly, DNA from *C. limon* "Verna", *C. sinensis* "Ricolate", *C. clementina* "Marisol" and *Poncirus trifoliata* (L.) Raf. "Flying Dragon" were used for Southern blot analysis. A segregating population derived from the cross between *Citrus aurantium* L. "Afín Verna" and *P. trifoliata* "Flying Dragon" (A×Pa) consisting of 66 hybrids was genotyped for IRAPs based on four *gypsy* like elements (the most different ones among each other) to study their genomic distribution. This progeny had been previously used to obtain genetic linkage maps of the parental species (Ruiz and Asíns 2003)

Isolation of *gypsy* like fragments

Citrus genomic DNA extractions were carried out from 1 g of leaf tissue according to Dellaporta et al. (1983) with minor modifications (Ruiz et al. 2000). DNA concentration was estimated using a 6105 spectrophotometer (Jenway).

In order to investigate the presence of *gypsy* like retrotransposons in Citrus spp. the PCR-based strategy and degenerate primers described by Suoniemi et al. (1998) were used. These primers, designed to match conserved residues from *rt* (forward) and *int* (reverse) domains of gypsy like retoelements, and DNA from three different *Citrus* species (*C. limon, C. clementina* and *C. sinensis*), were used for PCR. The amplified fragments containing *gypsy* like fragments were expected to be of approximately 1.6 kb, but according to the results obtained for different members of the plant kingdom might range from 2.0 to 0.8 kb. (Suoniemi et al. 1998). Diverse amplification conditions were tested in order to reduce PCR artifacts. Amplification reactions with the best results consisted of 300 ng of template DNA, 1 × supplied reaction buffer, 1.5 mM of $MgCl_2$, 100 µM of each dNTP, 0.5 µM of each degenerate primer, 1.25 u of *Taq* (Eco*Taq*, Ecogen) and sterile water up to 25 µl. Each reaction was overlaid with 25 µl of mineral oil and amplified in a PTC-100 thermal cycler (MJ Research) under the following conditions: an initial step at 94°C for 5 min; 30 cycles of 1 min at 94°C, 1 min at 47°C and 2 min at 72°C; and a final step at 72°C for 10 min. PCR products were mixed with 6 µl of 5 × loading buffer (50% v/v glycerol, 1 × TAE, 10% v/v saturated bromophenol blue, and 0.2% w/v xylene cyanole) and visualized with ethidium-bromide staining after electrophoresis in 0.8% agarose-TAE gels. In some cases, electrophoresis using 10% polyacrylamide sequencing-type gels followed by silver staining according to Ruiz et al. (2000) was used to improve band resolution. Amplification products were extracted from gels, eluted in sterile water and re-amplified to verify its isolation. Purified PCR products from *C. clememtina* "Loretina" were cloned into the pGEM-T Easy Vector System (Promega). To evaluate diversity, inserts from white colonies were digested with three restriction endonucleases (*Hin*dIII, *Xba*I and *Eco*RV) analyzed by 10% polyacrylamide sequencing-type gel electrophoresis.

DNA sequencing and sequence analysis

Both strands of selected clones were sequenced by the IBMCP (Instituto de Biologia Molecular y Celular de Plantas, Valencia) sequencing service. Given that the size of the selected clones was too large to read the entire sequence of the inserts, two sequences per clone (RT-side and INT-side), corresponding to the forward and reverse primers, were obtained and analyzed separately. Sequence analysis, alignments and putative translations were performed using SEQUENCHER (Gene Codes Corporation) and OMIGA (Accelrys Inc.) computer programs. Homology searches were done using the online service of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast), and TBLASTX and BLASTP searching tools. Statistical significance is represented by the E-value (p-value and E-value are nearly identical when $E<0.01$). Sequence similarities were graphically represented by dendograms using the p-distance and the neighbor-joining aggregation method, and bootstrapping with 500 replicates implemented in MEGA 2.1 (Kumar et al. 2001). Primer design to obtain probes and IRAPs by PCR was conducted using the PRIME program of the University of Wisconsin Genetics Computer Group (GCG) software package. GCG and SEQUENCHER programs were accessed through the Bioinformatics Service of the University of Valencia.

Southern-blot analysis

Genomic DNA (15 µg) from the above specified *Citrus* species and *P. trifoliata* was digested with four different 6-cutter restriction enzymes (*Eco*RI, *Hin*dIII, *Dra*I and *Bam*HI). The fragments were size-fractionated by 0.8% agarose-gel electrophoresis and trans-

ferred onto a positively charged nylon membrane (Hybond-N+, Amersham Pharmacia Biotech). To check probe specificity, approximately 5 ng of DNA from four representative *Citrus gypsy* like fragments (each representing one nucleotide-sequence homology group) were denatured and dot-blotted onto another nylon filter. These same fragments were digoxigenin-labeled using the DIG DNA labeling kit (Roche) and used as probes for hybridization, which was carried out under high stringency conditions achieved at 68°C for 14 h and by washing the membranes twice in 2 × SSPE, 0.1% SDS for 10 min at 68°C and twice again in 0.5 × SSPE, 0.1% SDS for 10 min at 68°C. The hybridization signals were subsequently detected with the ECF chemifluorescent substrate (Amersham Pharmacia Biotech), and analyzed on a STORM 860 optical scanner (Molecular Dynamics).

Development of IRAP markers and linkage analysis

Four primer pairs for IPAP markers (Kalendar et al. 1999) were designed from four distinct *Citrus gypsy* like sequences (C1, C2, C8 and C11). Each pair is based on the RT and INT regions of the sequence, facing outward. Primers are available from IVIA, upon request. Amplification reactions for 25 µl, final volume, contained 300 ng of template DNA, 1 × supplied reaction buffer, 100 µM of each dNTP, 0.12 µM of each primer and 1 u of *Taq* (Netzyme, N.E.E.D.). Reaction mixtures were amplified under the following conditions: an initial step at 95°C for 5 min; 45 cycles of 1 min at 95°C, 1 min at 44°C, a ramp of +0.3°C per s to 72°C and 3 min at 72°C; and a final step at 72°C for 8 min. Amplification products were resolved by polyacrylamide-gel electrophoresis and visualized by silver staining as above. IRAP segregation data in A×Pa family was analyzed using JOINMAP 3.0 (Van Ooijen and Voorrips 2001) with a linkage criterion of LOD generally above 4.0, a recombination fraction of 0.5 and Kosambi mapping function for linkage analysis. The population was analyzed as the "Cross-pollinator" population type with no previous knowledge of the linkage phase of the markers. Nomenclature of the linkage groups follows that described by Ruiz and Asíns (2003). New IRAPs were named as the *gypsy* like clone (C1, C2, C8 or C11) followed by a number that indicates the size in base pairs of the segregating band.

Results

Isolation and characterization of *Gypsy* like sequences in *Citrus*

Every citrus species tested yielded the same weak banding pattern consisting in three bands of approximately 2,200 bp, 1,650 bp and 1,550 bp (coded as A, B and C respectively in Fig. 1). *C. clementina* bands were more intense, so we tried to isolate them from agarose and Polyacrylamide-gels but the re-amplification of the largest fragment (A) was unsuccessful while B and C fragments were easily purified and subsequently used for cloning experiments. Only one clone from fragment B was obtained while those from fragment C were abundant. Polyacrylamide-gel electrophoresis showed slight size differences among the C clones. Variation among C clones was confirmed by endonuclease-restriction analysis since four different restriction patterns were found for C clones. All C clones showed the same restriction patterns except for C1, C8 and C11 that were unique. The only B positive clone (B9) showed the same restriction patterns as C1. Along with four clones representative of the common restriction pattern (C2, C12, C13 and C14), clones C1, C8, C11 and the B clone were also sequenced.

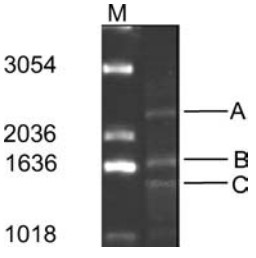

Fig. 1 Agarose-gel electrophoresis of amplification products using genomic DNA from *C. clementina* "*Loretina*" and *gypsy* degenerate primers. *M*: molecular-weight markers in base pairs

Sequence searches in NCBI databases revealed, in every case, high significant similarities (E-values ranging from e^{-98} to e^{-42}) to Ty3/*gypsy* type retrotransposons from different plant species. Additionally, some of our sequences strongly matched with a resistance-gene cluster from potato and several kafirin clusters from *Sorghum*. A representative pool of closely related sequences were aligned with our Citrus putative *gypsy* clones. A neighbor-joining tree based on nucleotide sequence alignment was constructed for both RT-side and INT-side sequences, resulting in very similar representations of relationships (Fig. 2) that reflect the same four groups obtained by restriction analysis, one including C2, C12, C13 and C14, another with C1 and B9, and the other two groups with only one sequence each (C8 and C11). Percentages of nucleotide identities between clones were also almost identical for RT and INT sides. Identities above 95% were found within groups, while between groups they ranged from 63% (C2 and C1 groups) down to 45%, between C2 and C8 groups. Citrus RT and INT nucleotide sequences were closely related to different members of the Ty3/*gypsy* class of retrotransposons, suggesting they are part of the POL coding region of citrus *gypsy* elements. Although all of them were isolated from the same clementine variety, three main groups are distinguished: C8 related to the *Monkey gypsy* like element from *Musa* (Balint-Kurti et al. 2000), C11 related to the *Reina* and *IFG7 gypsy* like elements from maize and *Pinus radiata*, respectively, and the rest, the C2 group plus C1 and B9, related to the *gypsy* like retrotransposons *CaRep* and *Dea1* (Thomsom et al. 1998) from *Cicer* and *Ananas*, respectively. Given that *Reina, Dea1* and *IFG7* are *gypsy* like plant elements lacking the *env* (envelope) domain (Vicient et al. 2001b) it could be assumed that our clementine elements lack it too, in spite of the indirect evidences contributed by these authors supporting the presence of such a domain in *C. sinensis* elements.

The potential correct frame for translation of both *rt* and *int* domains was inferred using the TBLASTX search tool. The resulting amino-acid sequences revealed several stop codons for B9, C1, C8 and C11, while no stop codon was found interrupting the putative coding region of the C2, C12, C13 or C14 clones. Again, the predicted translation products of RT and INT domains showed high similarities to Ty3/*gypsy* polyproteins from the

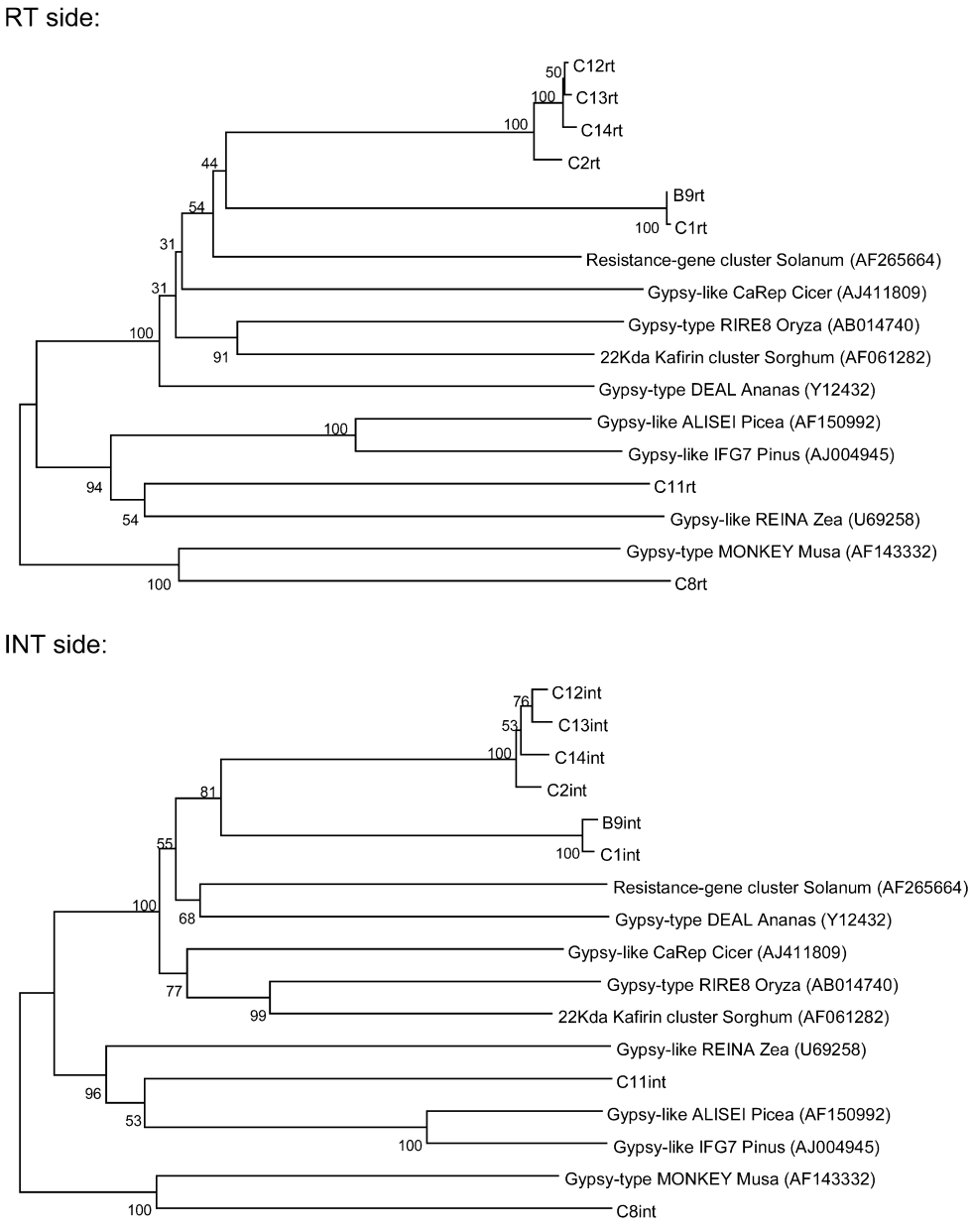

Fig. 2 Dendogram of Citrus *gypsy* like sequences and related sequences from NCBI databases. The tree is based on nucleotide sequence alignments. *Numbers* adjoining the branches indicate bootstrap values (percentages from 500 replicates) based on the p-distance estimate of sequence distance and neighbor-joining aggregation method. Numbers in brackets correspond to NCBI sequence accession numbers

NCBI protein databases. Matches with plant-disease resistance polyproteins and Sorghum kafirin cluster polyproteins were also found. The alignment of some of these proteins with our *Citrus* putative translation products is shown in Fig. 3. RT and INT translated sequences for B9, C1, C8 and C11 present not only stop codons but also frame-shift mutations (insertions and deletions). The putative amino-acid sequences from C2, C12, C13 and C14 show the blocks of residues widely described as universal in the Ty3/*gypsy* group and no stop codon. For these clones the translation of the RT domain begins with the invariant DD motif of the reverse transcriptase active site and, approximately 40 residues downstream, the highly conserved SKCEF block, including the invariant lysine (K) of RNA-dependent polymerases, is found (Suoniemi et al. 1998, citing Barber et al. 1990). Characteristic residues of the Rnase H domain are also present in these citrus clones since the conserved DAS motif containing a key aspartate (D) active-site could be identified (Springer and Britten 1993; Chavanne et el. 1998) followed by a glutamate (E) residue which is essential for RNAse H catalysis.

Two out of the three distinct conserved motives within the integrase domain of Ty3/*gypsy* retrotransposons are also present (Malik and Eickbush 1999). The N terminal and the central region could be identified in the *Citrus* clones (Fig. 3). The N-terminal integrase subdomain contains an $H_6H_{29}C_2C$ motif forming a zinc-finger structure implicated in binding to LTR retrotransposon sequences (Khan et al. 1991). The integrase central-core subdomain begins with the highly conserved GLLQPLPI motif (Suoniemi et al. 1998) and continues with the

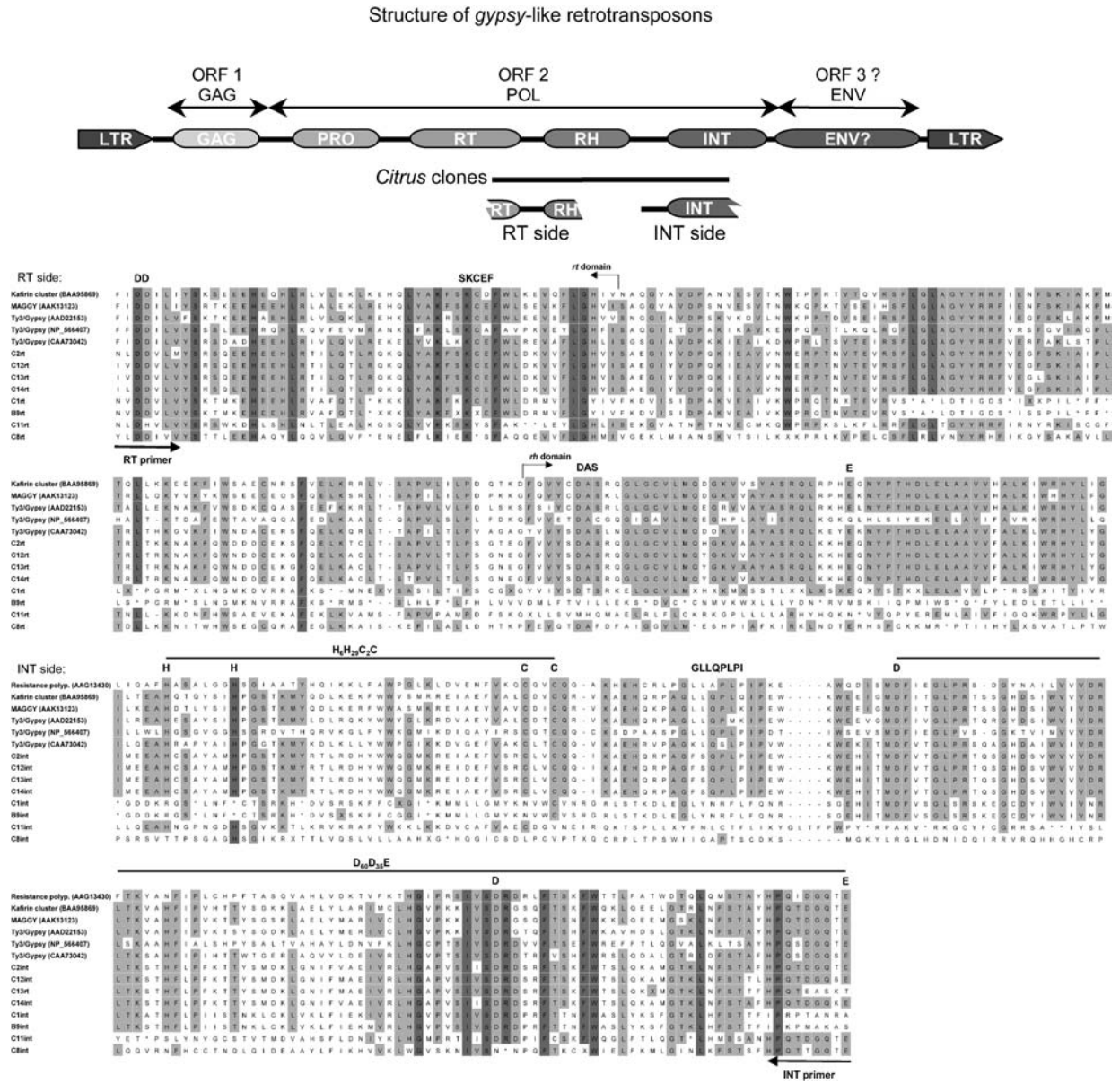

Fig. 3 Typical structure of *gypsy* like retrotransposons and predicted translation products of *Citrus gypsy* like sequences aligned with different proteins from NCBI databases. Gaps and stop codons are indicated as –and * respectively

catalytic $D_{60}D_{35}E$ motif essential for enzymatic activity (Kulkosky et al. 1992). The C-terminal subdomain is out of the region we have cloned.

Given that C2, C12, C13 and C14 *Citrus* clones (from now-on named the C2 group) present all the conserved motifs described in the literature as essential for *gypsy* like retrotransposon activity, they might correspond to a family of active *gypsy* elements. On the other hand, the sequence analysis of B9, C1, C8 and C11 clones shows several mutations disrupting the appropriate reading frame suggesting that these elements, if functional, are non-autonomous.

Genomic distribution of Citrus *gypsy* like elements

Internal specific primers were designed for the C1, C2, C8 and C11 sequences, each representing one group of citrus *gypsy* like elements. The resulting dig-labeled PCR amplification products were used as probes for genomic Southern-blot hybridizations under high-stringency conditions. RFLPs obtained using the C1 probe are shown in Fig. 4. Similar results were achieved when dig-labeled C2, C8 and C11 were used as probes. A very simple banding pattern was observed in all cases, suggesting a small number of copies of each clone in the *Citrus* and *Poncirus* genomes. *C. sinensis* and *C. clementina* hybridization patterns are almost identical. A few differences arise when genomic DNA from *C. limon* is hybridized.

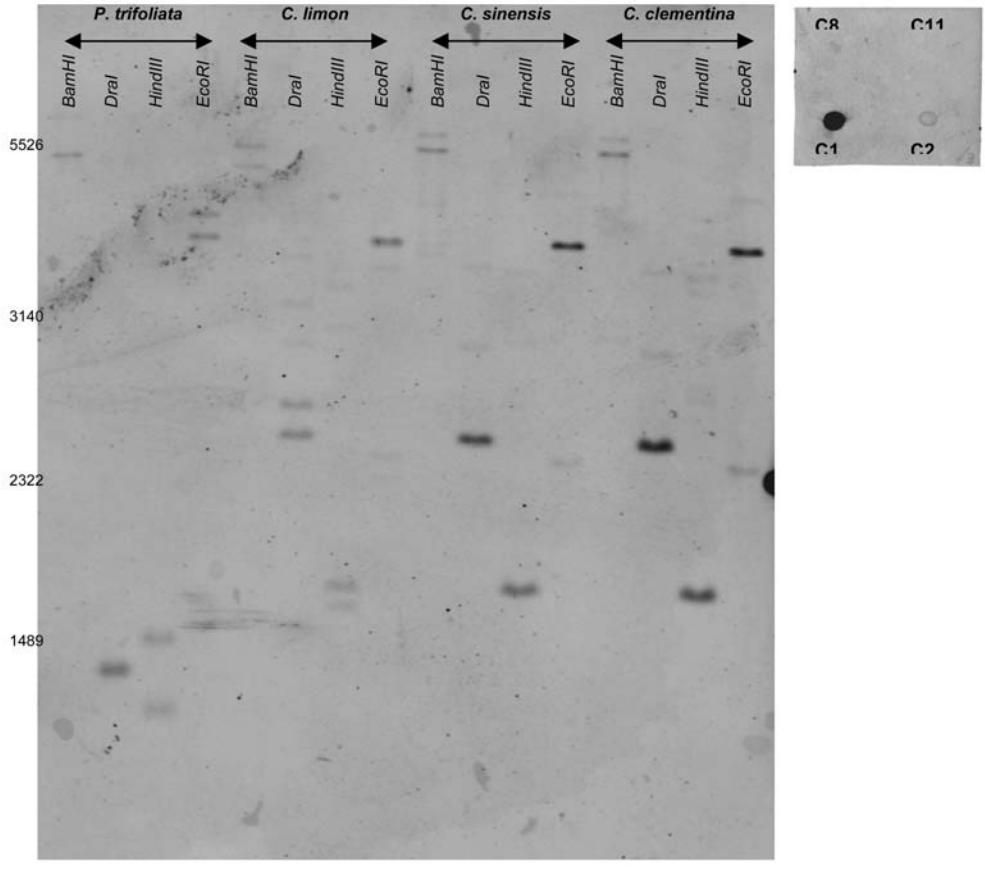

Fig. 4 Southern-blot analysis of digested genomic DNA from different species probed with C1. *Numbers to the left* correspond to molecular-weight markers (MWM IV, Roche) in base pairs. *To the right*, dot blot of clones also probed with C1 to check probe specificity

The most distinct pattern corresponds to the most distantly related species, *P. trifoliata*.

IRAP markers based on these *gypsy* like sequences (C1, C2, C8 and C11) were also developed (Fig. 5). Given that LTR termini can be quite distant from the RT and INT regions where outward primers anneal, and most IRAP bands are smaller than 1 kb, they must correspond to nested retroelements. Not all IRAPs segregated as dominant markers (presence *versus* absence of band). Two slightly different fragments segregate in *C. aurantium* for C11_410, C11_900, C2_800, C2_170 and C1_200 indicating they behave as co-dominant markers. In *P. trifoliata*, only C2_800 behaves similarly. Bands at C2_800, segregating in both parental species, behave as alleles at the same locus. Differences between allelic bands are revealed by differences in the mobility of heteroduplexes (C11_410 and C2_800 at each parent) or by slight size-differences (2–5 bp) between them (C11_900, C2_170, C1_200 and C2_800 of the parents). Four dominant IRAPs segregating at *C. aurantium* presented significantly distorted segregation ratios: C11_510 (χ^2=5.3), C1_250 (χ^2=4.6), C11_875 (χ^2=12.5) and C8_200 (χ^2=22.3). The direction of the distortion is towards the absence of bands for both C11 IRAPs, while for the others the direction is towards their presence.

All IRAP markers were mapped using a progeny derived from the cross *C. aurantium*×*P. trifoliata* that had been already genotyped for 120 markers (Ruiz and Asíns 2003). Linkage groups where these new IRAPs map, are shown in Fig. 6. Only five IRAPs were found to segregate in *P. trifoliata*, two of them co-segregate and do not join to any linkage group. On the other hand, all 20 IRAP markers segregating in *C. aurantium* gametes could be added to an existing linkage group, resulting in an extended and better-resolved linkage map. Most of the *gypsy* retrotransposon-based markers (15 out of 25) map without clustering to other markers, seven of them map at positions where ESTs are located (CR markers are microsatellites derived from EST sequences) and only one (C11_1000) joins a cluster of *copia* retrotransposon-based markers. Therefore, IRAPs based on both type of retrotransposons seem to distribute differently throughout the genome providing a new, complementary set of molecular markers. Data related to the number of bands, the heterozygosity percentage per sequence and parental species is presented on Table 1. The total number of bands obtained from *copia*-like elements is twice and four-times larger than the number of bands obtained from *gypsy* like elements in *C. aurantium* and *P. trifoliata*, respectively, suggesting they are less abundant than *copia*-like elements. Similarly to *copia* derived IRAPs, *P. trifoliata* presents lower heterozygosis percentages than *C. aurantium*. Contrary to the total number of bands, average heterozygosity for *gypsy* derived IRAPs was two-fold greater than that obtained from *copia*-derived IRAPs.

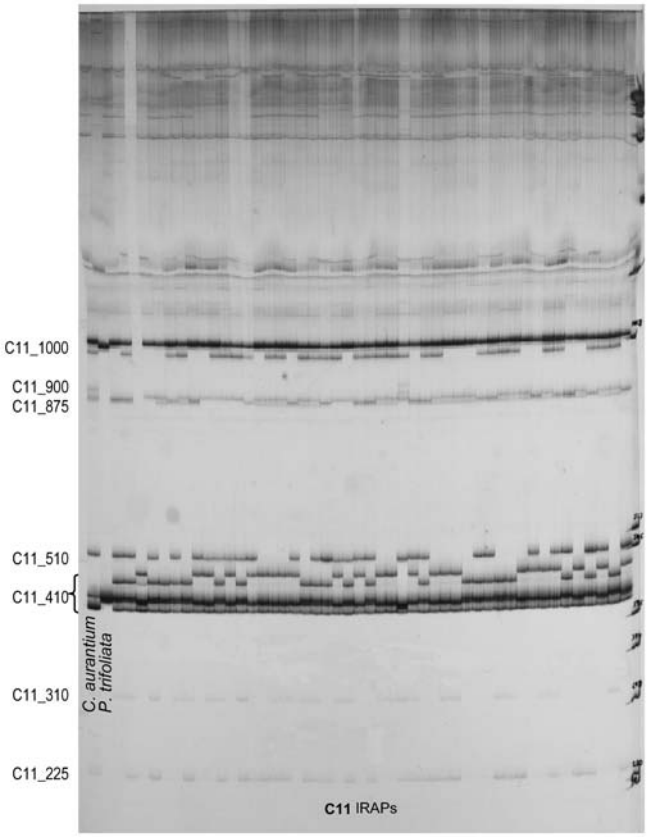

Fig. 5 IRAP banding patterns derived from the *gypsy* sequence C11 for hybrids belonging to A×Pa family. Segregating IRAP loci are shown at the left side

Discussion

We have described the isolation, analysis and genomic location of eight representative sequences isolated from *C. clementina* that show homology to retroelements of the Ty3/*gypsy* group. Four of them (C2 group) might correspond to active elements since they carry no in-frame stop codon nor frame-shift mutations.

Nucleotide sequences of the cloned fragments revealed high homology to diverse well-characterized plant *gypsy* like retrotransposons (Figs. 2 and 3). The *in silica* predicted translation peptides were useful to detect a group of clones with no frame disruption and conserved motifs essential for autonomous retrotransposition activity (Fig. 3). The sequences of the other group of clones carried several stop codons and frame-shift mutations suggesting that these elements, if functional, are non-autonomous. Copies with frame-shifts and in-frame stop codons within the POL coding region have been reported for several retroelements of the gypsy like family (*Micropia*, Lankenau et al. 1988; *Cyclops*, Chavanne et al. 1998; *yoyo*, Zhou and Haymer 1998), and have been explained by the gradual accumulation of mutations during relatively long periods of silence after bursts of high transpositional activity (Chavanne et al. 1998). Since most isolated clones belong to the C2 group, a very homogeneous group, this could support the hypothesis that they have been amplified more recently than the others (Feschotte et al. 2002). Therefore, the *C. clementina* genome seems to contain potentially active *gypsy* like elements (the C2 group).

The sequences of the C2 group, C1 and B9, also showed high homology to elements within a resistance-gene cluster in potato that confers resistance to a virus and a nematode (Van der Vossen et al. 2000). Detailed analysis of this cluster showed conserved domains such as those corresponding to the GAG protein, reverse transcriptase and integrase core domains in the proper frames and order, to deduce the presence of a *gypsy* like retroelement between resistance genes. Recently, the complete sequence of the *P. trifoliata* citrus tristeza virus resistance gene locus (282,699 nucleotides) has become available (AF 506028) (Yang et al. 2003). This contig contains several putative disease-resistance genes similar to the rice *Xa21* gene, the tomato *Cf-2* gene and the

Table 1 Number of *gypsy* based amplification products (bands), number of (polymorphic) IRAP bands and heterozygosity values for each parent of the A×Pa family. Heterozygosity has been estimated as the percentage of polymorphic bands. Data from *copia* derived IRAPs was reported by Ruiz and Asíns (2002)

Primers based on clones	Parental	Bands	Polymorphisms	Heterozygosity (%)	
C1	*C. aurantium*	13	8	61.5	52.4
	P. trifoliata	8	3	37.5	
C2	*C. aurantium*	5	5	100.0	75.0
	P. trifoliata	3	1	33.3	
C8	*C. aurantium*	3	3	100.0	66.7
	P. trifoliata	3	1	33.3	
C11	*C. aurantium*	12	9	75.0	60.0
	P. trifoliata	3	0	0.0	
Total for *gypsy*-like elements	*C. aurantium*	33	25	75.8	60.0
	P. trifoliata	17	5	29.4	
Total for *copia*-like elements	*C. aurantium*	74	33	44.6	30.8
	P. trifoliata	69	11	15.9	
Average over *Gypsy*-based IRAPs	*C. aurantium*	8.25	6.25		
	P. trifoliata	4.25	1.25		
Average over *Copia*-based IRAPs	*C. aurantium*	18.5	8.25		
	P. trifoliata	17.25	2.75		

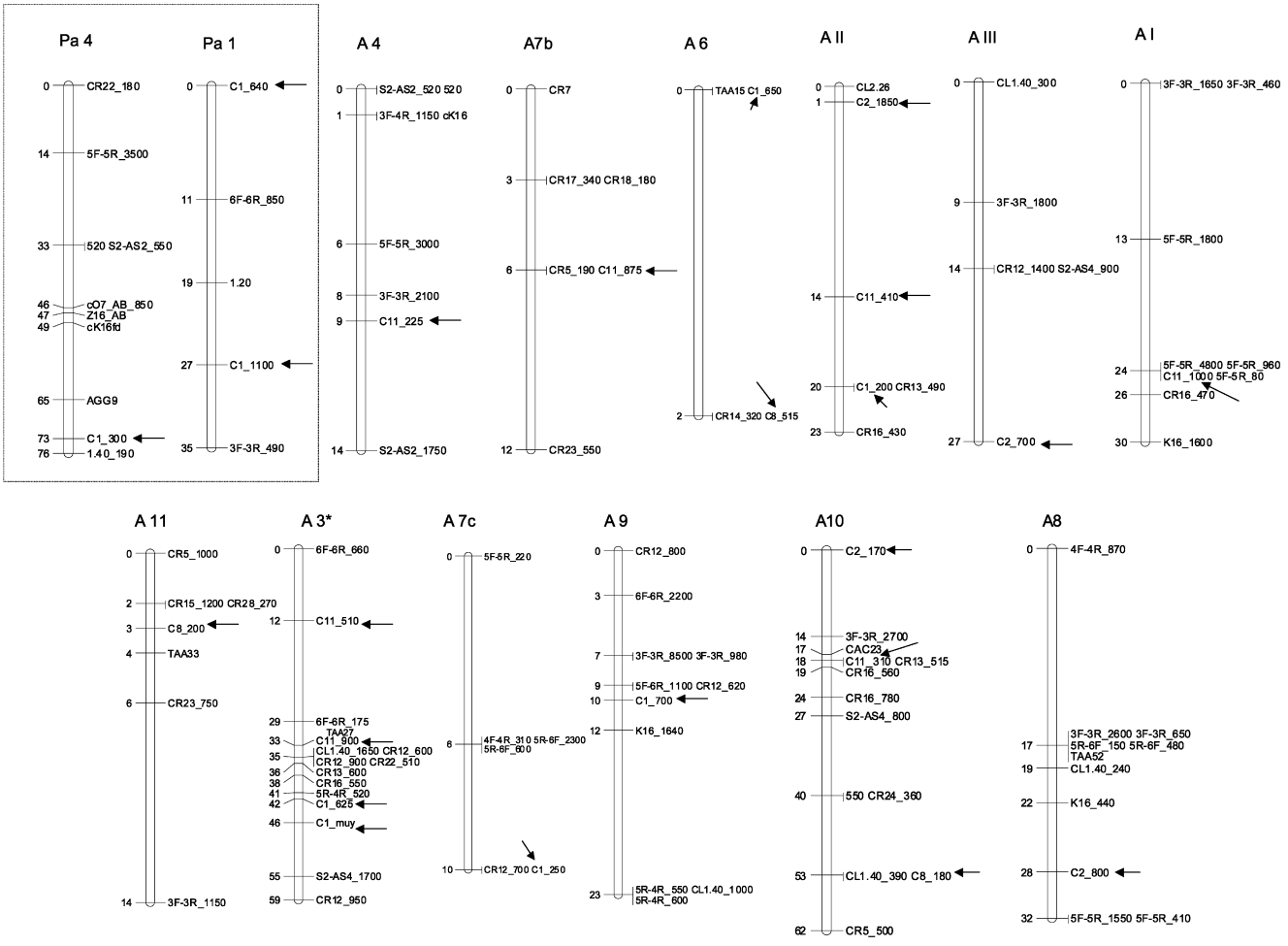

Fig. 6 Location of g*ypsy* derived IRAPs on the *C. aurantium* and *P. trifoliata* maps (Ruiz and Asíns 2003). Framed linkage groups belong to *P. trifoliata* map. Linkage groups are not scaled. *Asterisk* at linkage group A3 means that marker TAA27 is included but its position is not reliable. *Roman numbers* correspond to linkage groups that no homology has been found with respect to linkage groups of other maps. If by contrast, the linkage group is numbered with *Arabic numbers*, this indicates that this group presents two or more markers that are common to another linkage group of another map

Arabidopsis thaliana Rps2 gene (Yang et al. 2001). Between resistance genes, it also contains conserved domains corresponding to four gypsy like elements; the sequence of two of them presents high homology (over 80%) to our C2 clone, another presents 79% homology to our C8 clone and the third one is not related to ours. Not as easy to explain is the case of two proteins directly submitted to NCBI databases (without a reference paper) and described as the "putative plant disease resistance polyprotein" from *Oryza sativa* (database accessions AAM51835 and AAG13430). Both amino-acid sequences showed high similarities to the putative translations of our *Citrus* fragments (the integrase side only). The analysis of those polyprotein sequences yielded regions with strong homologies to conserved (total or partial) *gypsy* domains, such as the integrase core domain and the CHROMO (chromatin organization modifier) domain (Malik and Eickbush 1999).

Many transposable elements have been found at resistance-gene clusters (Richter and Ronald 2000; Wicker et al. 2001). Members of a resistance gene family are often arranged as tandem-direct repeats, which is consistent with their origin through gene duplication and their continued evolution through unequal exchange. In plants, it has long been hypothesized that transposable elements play a role in the reconstruction of genomes in response to environmental stresses such as tissue culture, irradiation or pathogen infection (McClintock 1984; Wessler et al. 1995). Transposable-element insertion into and excision from regulatory and coding regions can change the coding capacity and expression patterns of the gene (McDonald 1995; Wessler et al. 1995; Marionette and Wessler 1997). Additionally, movement of transposable elements may result in further allelic diversity, either by disrupting genes, or by influencing recombination or chromosomal rearrangements. Under the gene-for-gene model proposed by Flor (1956), plant and pathogen genes involved in the interaction are subject to different evolutionary forces. Since virulence is recessive, a simple loss-of-function mutation in the avirulence gene of the

pathogen allows it to become virulent on the host, while the plant must gain a new resistance function to counter new pathogen biotypes or species (Richter and Ronald 2000). Therefore, the presence of factors that increase allelic diversity at resistance genes must be selectively advantageous for the plant. In support of this hypothesis, studies of the maize resistance locus *rp1* revealed that recombination of flanking markers was associated with the creation of novel resistance phenotypes (Richter et al. 1995).

As suggested by Fedoroff (2000), a genetic mechanism, in addition to polyploidization, transposition and duplication, that drives plant genome expansion might be the preferential-transmission of tandemly repeated sequences (including retroelments) through the gametes. Our data on the direction of the distortions does not support this hypothesis, since only half the number of IRAP with distorted segration ratios showed a bias towards the presence of band.

Polymorphisms based on the C2 sequence are proportionally the most-abundant. This makes the heterozygosity percentage reach the highest value in *C. aurantium* for IRAPs based on this element. Addionally, it shows the highest proportion of co-dominant IRAPs. Most IRAPs in *Citrus* are dominant, i.e., segregation corresponds to presence versus absence of the band (Ruiz and Asíns 2003). An IRAP is co-dominant when the amplified inter-retrotransposon fragments from both homologous chromosomes in a heterozygote differ in size or sequence. Since the allelic differentiation (mutation ratio) should be slower for non-methylated (potentially active) retrotransposons than for methylated (inactivated) retrotransposons, the high proportion of co-dominant C2-IRAPs might be considered as another indirect evidence re-inforcing the hypothesis that C2 sequences correspond to active (not methylated in certain genomic positions, at least) *gypsy* like retrotransposons.

Both Southern-hybridization analysis and mapping of *gypsy based* IRAPs indicate that nested copies of these elements are scattered along the *Citrus* and *Poncirus* genomes. It is also clear that these elements were introduced before the divergence of both genera and evolved separately thereafter. All IRAPs map in different positions when linkage maps of both genomes are compared. Southern-hybridization analysis and mapping IRAPs based on both *copia*-like (Asíns et al. 1999; Ruiz and Asíns 2003) and *gypsy* like retrotransposons, also show that they are more abundant and polymorphic through the *Citrus* genome than through the *Poncirus* genome. If the thesis that the mechanisms that control transposition are a reflection of the more general capacity of eukaryotic organisms to detect, mark and retain duplicated DNA through repressive chromatin structures (Fedoroff 2000) is true, then we could say that this control capacity is greater in *P. trifoliata* than in *C. aurantium*. Another non-exclusive possibility is related to their reproductive system. Hickey (1982) predicted that loss of sex would result in a population free of transposable elements by preventing their spread. Recently, Arkhipova and Meselson (2000) reported that the pattern of occurrence of *LINE*-like and *gypsy* like retrotransposons in sexual and ancient asexual taxa supports Hickey's thesis. In our case, both species are highly apomictic. Nevertheless, it could be said that the degree of apomixis evaluated as a function of the mean percentage of zygotic (sexual) seedlings (Asíns et al. 2002) is higher for *P. trifoliata* than for *C. aurantium*. It is important to point out again that all *gypsy* like retrotransposons we have isolated come from one of the few sexual species in *Citrus, C. clementina*. Clementines are much more closely related to *C. aurantium* than to *P. trifoliata* and they were mostly grafted on *C. aurantium* trees until a few decades ago. Therefore, the abundance of retrotransposons in *Citrus* could be due not only to a more recent common sexual ancestor (vertical transmission) but also to horizontal transmission, or even infection of the rootstock by the grafted variety given that, at least, some *Citrus gypsy* like retroelements might bear not only active domains for retrotransposition but also ENV-like domains, conferring infective capability, as Vicient et al (2001b) indirectly deduced. Concerning horizontal transmission, a *gypsy* like retrotransposon named *yoyo* has been isolated from the Mediterranean fruit fly, *Ceratitis capitata* (Zhou and Haymer 1998), but it shows no significant homology to any of the clementine *gypsy* like elements.

In conclusion, a largely heterogenous set of *gypsy* like elements has been isolated from the clementine cultivar 'Loretina' that recently originated by somatic natural mutation. Since *gypsy based* IRAPs present a different genomic distribution compared to that of *copia*-based IRAPs, they constitute a new, complementary set of molecular markers that are available to study cultivar diversity and follow the variation of agronomic traits in segregant progenies derived from *Citrus*. Among these traits, their application to locate disease resistance gene clusters seems specially promising.

Acknowledgements This work was supported in part by grants from Conselleria de Cultura, educació i Ciència (GPB), Instituto Nacional de Investigaciones Agrarias (SC99-047) and Ministerio de Ciencia y Tecnología (AGL2002-02395). We are grateful to Dr. Luis Navarro for allowing us the use of the Citrus Germplasm Bank at IVIA.

References

Arkhipova I, Meselson M (2000) Transposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci USA 97:14473–14477

Asíns MJ, Monforte AJ, Mestre PF, Carbonell-EA (1999) Citrus and Prunus *copia*-like retrotransposons. Theor Appl Genet 99:503–510

Asíns MJ, Garcia MR, Ruiz C, Carbonell EA (2002) Molecular markers for the genetic analysis of apomixis. In: Jain SM, Brar DS, Ahloowalia BS (eds) Molecular techniques in crop improvement. Kluwer Academic Publishers, Dordrecht, pp 266–281

Balint-Kurti PJ, Clendennen SK, Dolezelová M, Valárik M, Dolezel J, Beetham PR, May GD (2000) Identification and

chromosomal localization of the monkey retrotransposon in *Musa* sp. Mol Gen Genet 263:908–915
Barber AM, Hizi A, Maizel JVJ, Hughes SH (1990) HIV-1 reverse transcriptase: structure predictions for the polymerase domain. AIDS Res Hum Retrov 6:1061–1072
Bernet GP, Mestre PF, Pina JA, Asíns MJ (2003) Molecular discrimination of lemon cultivars. HortScience (in press)
Bretó MP, Ruiz C, Pina JA, Asíns MJ (2001) The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. Mol Phylog Evol 21:285–293
Cameron JW, Frost HB (1968) Genetics, breeding and nucellar embryony. In: Reuther W, Batchelor LD, Webber HJ (eds) The citrus industry, vol II. Division of Agricultural Science, University of California, Berkeley, pp 325–370
Chalker DL, Sandmeyer SB (1992) Ty3 integrates within the region of RNA polymerase-III initiation. Genes Dev 6:117–128
Chavanne F, Zhang DX, Liaud MF, Cerff R (1998) Structure and evolution of *Cyclops*: a novel giant retrotransposon of the Ty3/*Gypsy* family highly amplified in pea and other legume species. Plant Mol Biol 37:363–375
Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21
Echenique V, Stamova B, Wolters P, Lazo G, Carollo VL, Dubcovsky J (2002) Frequencies of Ty1-*copia* and Ty3-*gypsy* retroelements within the *Triticeae* EST databases. Theor Appl Genet 104:840–844
Fedoroff N (2000) Transposons and genome evolution in plants. Proc Natl Acad Sci USA 97:7002–7007
Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nature Rev 3:329–341
Flor HH (1956) The complementary genic systems in flax and flax rust. Adv Genet 8:29–54
Frankel AD, Young JA (1998) HIV-1: fifteen Proteins and an RNA. Annu Rev Biochem 67:1-25
Friesen N., Brandes A, Heslop-Harrison JS (2001) Diversity, origin and distribution of retrotransposons (*gypsy* and *copia*) in Conifers. Mol Biol Evol 18:1176–1188
Grandbastien MA (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3:181–187
Hickey DA (1982) Selfish DNA: a sexually-transmited nuclear parasite. Genetics 101:519–531
Hirochika H, Sugimoto K, Otsuki Y, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93:7783–7788
Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98:704–711
Khan E, Mack JPG, Katz RA, Kulkosky J, Skalka AM (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res 19:851–860
Kulkosky J, Jones KS, Katz RA, Mack JPG, Skalka AM (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol 12:2331–2338
Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics (submitted)
Lankenau DH, Huijser P, Jansen E, Miedema K, Hennig W (1988) *Micropia*: a retrotransposon of *Drosophila* combining structural features of DNA viruses, retroviruses and non-viral transposable elements. J Mol Biol 204:233–46
Malik HS, Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/*Gypsy* class of LTR retrotransposons. J Virol 73:5186–5190
Marionette S, Wessler SR (1997) Retrotransposon insertion into the maize waxy gene results in tissue –specific RNA processing. Plant Cell 9:967–978
McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801
McDonald JF (1995) Transposable elements—possible catalysts of organismic evolution. Trends Ecol Evol 10:123–126
Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 11:1660–1676
Purugganan MD, Wessler SR (1994) Molecular evolution of Magellan, a maize Ty3/gypsy like retrotransposon. Proc Natl Acad Sci USA 91:11674–11678
Raghuvanshi SS (1962) Cytogenetical studies in genus *Citrus*. IV. Evolution in genus *Citrus*. Cytologia 27:172–188
Richter TE, Ronald PC (2000) The evolution of resistance genes. Plant Mol Biol 42:195–204
Richter TE, Pryor AJ, Bennetzen JL, Hulbert SH (1995) New rust resistance specificities associated with recombination at the *Rp1* complex in maize. Genetics 141:373–381
Ruiz C, Asíns MJ (2003) Comparison between *Poncirus* and *Citrus* genetic linkage maps. Theor Appl Genet 106:826–836
Ruiz C, Bretó MP, Asíns MJ (2000) A quick methodology to identify sexual seedlings in citrus breeding programs using SSR markers. Euphytica 112:89–94
Shcherban AB, Vaughan DA, Tomooka N, Kaga A (2001) Diversity in the integrase coding domain of a *gypsy* like retrotransposon among wild relatives of rice in the *Oryza officinalis* complex. Genetica 110:43–53
Smyth DR, Kalitsis P, Joseph JL, Sentry JW (1989) Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the *gypsy* group of *Drosophila*. Proc Natl Acad Sci USA 86:5015–5019
Springer MS, Britten RJ (1993) Phylogenetic relationships of reverse transcriptase and RNAse H sequences and aspects of genome structure in the *gypsy* group of retrotransposons. Mol Biol Evol 10:1370–1379
Suoniemi A, Tanskanen J, Schulman AH (1998) *Gypsy* like retrotransposons are widespread in the plant kingdom. Plant J 13:699–705
Thomson KG, Thomas JE, Dietzgen RG (1998) Retrotransposon-like sequences integrated into the genome of pineaple, *Anannas comosus*. Plant Mol Biol 38:461–465
Van der Vossen E, Rouppe van der Voort J, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema W, Klein-Lankhorst (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, Voorips RE (2001) JoinMap 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands
Vicient CM, Jääskeläinen M, Kalendar R, Schulman A (2001a) Active retrotransposons are a common feature of grass genomes. Plant Physiol 125:1283–1292
Vicient CM, Kalendar R, Schulman A (2001b) Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res 11:2041–2049
Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821
Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf, Keller B (2001) Analysis of a contiguous 211-kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316
Yang Z-N, Ye X-R, Choi S, Molina J, Moonan F, Wing RA, Roose ML, Mirkov TE (2001) Construction of a 1.2-Mb contig including the cistrus tristeza virus resistance gene locus using a bacterial artificial chromosome library of *P. trifoliata* (L.) Raf. Genome 44:382–393
Yang Z-N, Ye X-R, Choi S, Molina J, Roose ML, Mirkov TE (2003) Sequence analysis of a 282-kilobase region surrounding the citrus tristeza resistance gene (*Ctv*) locus in *Poncirus trifoliata* L. Raf. Plant Physiol 131:482–492
Zhou Q, Haymer D (1998) Molecular structure of *yoyo*, a *gypsy* like retrotransposon from the Mediterranean fruit fly, *Ceratitis capitata*. Genetica 101:167–178