Theor Appl Genet (1999) 99:503–510 © Springer-Verlag 1999

M.J. Asíns · A.J. Monforte · P.F. Mestre
E.A.Carbonell

Citrus and *Prunus copia*-like retrotransposons

Received: 10 December 1998 / Accepted: 28 December 1998

Abstract Many of the world's most important citrus cultivars ("Washington Navel", satsumas, clementines) have arisen through somatic mutation. This phenomenon occurs fairly often in the various species and varieties of the genus.The presence of *copia*-like retrotransposons has been investigated in fruit trees, especially citrus, by using a PCR assay designed to detect copia-like reverse transcriptase (RT) sequences. Amplification products from a genotype of each the following species *Citrus sinensis*, *Citrus grandis*, *Citrus clementina*, *Prunus armeniaca* and *Prunus amygdalus*, were cloned and some of them sequenced. Southern-blot hybridization using RT clones as probes showed that multiple copies are integrated throughout the citrus genome, while only 1–3 copies are detected in the *P. armeniaca* genome, which is in accordance with the *Citrus* and *Prunus* genome sizes. Sequence analysis of RT clones allowed a search for homologous sequences within three gene banks. The most similar ones correspond to RT domains of *copia*-like retrotransposons from unrelated plant species. Cluster analysis of these sequences has shown a great heterogeneity among RT domains cloned from the same genotype. This finding supports the hypothesis that horizontal transmission of retrotransposons has occurred in the past. The species presenting a RT sequence most similar to citrus RT clones is *Gnetum montanum*, a gymnosperm whose distribution area coincides with two of the main centers of origin of *Citrus* spp. A new C-methylated restriction DNA fragment containing a RT sequence is present in navel sweet oranges, but not in Valencia oranges from which the former originated suggesting, that retrotransposon activity might be, at least in part, involved in the genetic variability among sweet orange cultivars. Given that retrotransposons are quite abundant throughout the citrus genome, their activity should be investigated thoroughly before commercializing any transgenic citrus plant where the transgene(s) is part of a viral genome in order to avoid its possible recombination with an active retroelement. Focusing on other strategies to control virus diseases is recommended in citrus.

Key words Genetic variability · Fruit trees · Plant mobile elements · Horizontal transmission · Genetic transformation.

Communicated by P.M.A. Tigerstedt

M.J. Asíns (✉) · A.J. Monforte · P.F. Mestre · E.A. Carbonell
Instituto Valenciano de Investigaciones Agrarias (IVIA), Apartado Oficial, 46113 Moncada, Valencia, Spain
e-mail: mjasins@ivia.es

Introduction

Many of the world's most important cultivars have arisen through somatic mutation. The citrus industry of the world is highly dependent on varieties such as "Washington Navel", "Valencia", "Shamouti", "Pera", "Hamlin" oranges, "Marsh" grapefruit, easy-pealing mandarins such as satsumas and clementines, and "Eureka" lemon, most of which originated as bud mutations (Spiegel-Roy and Goldschmidt 1996). This process, however, goes on uncontrolled by man, with those of value becoming available only through chance discoveries and the astuteness of their finders in recognizing their potential and bringing them to fruition. Genetic improvement in *Citrus* spp. by hybridization has been much hampered because of heterozygosity, reproduction by nucellar embryony and juvenility. Consequently, it has been largely the result of the selection of naturally occurring somatic mutants.

Authors studying bud mutations frequently report about the "reversal" of the bud mutation to its "parent" type. Most described examples of this phenomenon can be explained by the chimerical nature of mutations, others can not (Mendel 1981). Another very interesting feature in the bud mutations of citrus is the occurrence of "parallel" mutations, i.e. the same mutations occur in different species and varieties: including variegated leaves, "willow" leaves, "pink-flesh" fruit, proliferation-"navels", ribbed and corrugated fruits, etc. Does this indicate that in many cases of bud variations, the same genes or the same somatic aberrations are involved? Is it

a change in the pattern of DNA methylation through the genome? Is it the activity of transposable elements?

Since the first *Ty-copia* elements were detected in plants they have been found across a broad phylogenetic spectrum and in all major lineages of plants including the Chlorophyta, Bryophyta, Pteridophyta, as well as the Gymnospermae (Voytas et al 1992; Kamm et al. 1996). Most of these *Ty*1-*copia* elements were identified by using a PCR assay designed to detect *copia*-like reverse transcriptase gene sequences, given that the reverse transcriptase (RT) region of the *pol* gene is the most highly conserved sequence of the retroelements (Doolittle et al. 1989). Most eukaryotic retrotransposons move only sporadically in the genome of their hosts. 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 for yeast retrotransposons. Thus, *Ty3* elements integrate almost exclusively upstream of genes transcribed by RNA polymerase III. There are indications that retrotransposons are responsible for a significant fraction of spontaneous mutations in plants (Hirochika 1995). Retrotransposons are silent in normally propagated plants but become active under stress conditions (tissue culture and protoplast formation, wounding, virus infections) through activation of transcription (Hirochika 1995).

In *Citrus* orchards, plants are pruned every year and are normally (and naturally) infected with viruses, among them citrus tristeza virus (CTV) which replicates abundantly in sweet orange and mandarin cultivars. In this context, natural gene mutations occur frequently in the somatic cells which produce fruits and new shoots often resulting in varied fruit chimeras (Fig. 1) and bud mutations or bud sports, affecting vegetative growth and which may or may not in time produce new fruit types.

The objectives of the present paper are to investigate the presence of "copia-like" retrotransposons in fruit trees, mainly *Citrus* spp., their genomic distribution, their heterogeneity and their contribution to bud mutations and citrus natural genetic variability.

Materials and methods

Plant materials

All citrus plants analyzed, except cultivar "Caracara", belong to the *Citrus* germplasm bank at the IVIA and their genetic variability has already been studied by isoenzymatic markers (Herrero et al. 1996a, b). They are mature plants of similar age, growing in containers in a screen-house and free of virus and virus-like pathogens (Navarro et al. 1988). The following *Citrus* species were studied *C. clementina* Hort. ex Tan. (clementine mandarin), *C. sinensis* (L.) Osb. (sweet orange), *C. unshiu* (Mak.) Marc. (satsuma mandarin), *C. medica* L., *C. deliciosa* Ten., *C. temple* Hort. ex Tan., *C. tangerina* Hort. ex Tan., *C. tachibana* (Mak.) Tan., *C. aurantium* L. (sour orange), *C. myrtifolia* Raf., *C. madurensis* Lour. (calamondin), *Fortunella margarita* (Lour.) Swing. and *Poncirus trifoliata* (L.) Raf.. Several cultivars of the three first species were included: two satsuma cultivars ("Precoz" and "Tardia Picaña"), six clementine cultivars ("Marisol", "Arrufatina", "Fina", "Nules", "Hernandina", "Oroval"); six sweet oranges: "Navelina", "Navelate", "Washington Foyos" and "Caracara" from the navel group and, "Salustiana" and "Valencia Late" from the Valencia group. The cultivar "Caracara" was studied in the field collection because some branches were variegated. Several varieties of *P. armeniaca* L. ("Bergeron", "Rouge Rousillon", "Goldrich", "Canino", "Polonais" and "Stark Early Orange") and *P. amygdalus* Batsch (the spanish "Ramillete" and the italian "Tuono"), kindly provided by the Department of Fruit Breeding, CEBAS (Murcia), Spain, were also studied. Two tomato accessions: *Lycoperiscon esculentum* Mill. var. "Madrigal" and *Lycopersicon pimpinellifolium* (Jusl.) Mill., line 1, were used as negative controls in the Southern-blot analysis.

Fig. 1 Variegated calamondin fruits

Molecular analysis

Genomic DNA extractions followed the method of Dellaporta et al. (1983) with some modifications. Amplification reactions consisted of buffer [10 mM Tris-HCl, 50 mM KCl (Eurobiotaq)]; 1.5 mM $MgCl_2$ (Eurobiotaq); 100 μM of dNTPs (25 μM of each one); 0.2 μM of each degenerate primer (Hirochica and Hirochika, 1993); 1 u of *Taq* (Eurobiotaq); 150 ng of DNA and sterile water up to 25 μl. PCR was conducted in a MJ-PTC-100™ thermal controller with 96 wells 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 and 2 min. at 72°C; and a final extension step at 72°C for 8 min. The results of the amplifications were visualized after electrophoresis in 2% agarose TAE gels with ethidium bromide staining.

Amplification products (a wide band around 280 bp) from some genotypes were cloned using the "TA cloning kit" from Invitrogen. Amplification products and putative reverse transcriptase (RT) clones were used as probes in hybridizations to Southern blots of digested genomic DNA from plants. RT clones were named by the letters Cl (from *C. clementina*), Gr (from *C. grandis*), Si (from *C. sinensis*), Al (from the almond cultivar "Ramillete") or Ap (from the apricot cultivar "Stark Early Orange"), followed by the number of the clone. Probe labelling, hybridization and washing conditions (maximum stringency) were as described in Monforte et al. (1996).

The methylation status of genomic domains specified by the RT probes was analyzed using two restriction enzymes with the same recognition site (GATC) but differing in inhibition by methylation: *Sau*3AI (which is inhibited when C is methylated) and *Dpn*II (which is not inhibited when C is methylated). Plant DNA digestions using these restriction enzymes were run on 2% agarose gels.

Sequencing reactions were performed using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin El-

mer). Amplification reactions consisted of 100 ng of DNA, 3 pmols of each primer and 8 µl of reaction mix in a 20 µl final volume. PCR was conducted in a Perkin Elmer 2400 thermal controller with 96 wells under the following conditions: an initial step at 96°C for 1 min; 30 cycles of 10 s at 95°C, 5 s at 50°C and 4 min at 60°C. Reactions were loaded on a sequencing gel and read by the automatic sequencer ABI 373.

Phylogenetic analysis of RT sequences.

The evolutionary distance of the proportion of amino-acid differences (p distance) has been chosen because it has smaller variance than other more complex distances and is recommended when the number of nucleotides examined is not very large (<500) (Nei 1996).

The Neighbor-Joining method (NJ) was used for constructing phylogenetic trees. This method is a simplified version of the Minimum-Evolution method for inferring a bifurcating tree. The Minimum-Evolution tree is generally identical or close to the NJ tree when the number of sequences to be compared is relatively small (Nei 1996). To test the reliability of the NJ tree, Felsenstein's bootstrap test (in Nei 1996) was employed.

Computer-based amino-acid similarity searches of the GeneBank, DNA Data Bank of Japan, and EMBL databases were performed with the TFASTA search program of the University of Wisconsin Genetics Computer Group (GCG) software package accessed through the BioScience Computing Resource at the University of Valencia.

Multiple sequence alignment was performed using Clustal V (Higgins et al 1992). Translation to amino-acid sequences, calculation of the matrix distance, NJ tree construction and bootstrap tests with 500 replications were all performed by using MEGA (Kumar et al. 1993).

Results

Genomic distribution of RT sequences

PCR amplifications with degenerate primers yielded in most cases a single wide band on agarose gels (Fig. 2), of around 280 bp, which is the expected size of the RT domain of the *copia*-like retrotransposons. Amplification products were obtained when using genomic DNA from *C. sinensis* cv "Washington Foyos", *C. medica* var. Ethrog, *C. clementina* cv "Fina", *C. grandis* cv "Pink", *Prunus amygdalus* cvs "Ramillete" and "Tuono", and *P. armenica* cvs "Stark Early Orange", "Velazquez", "Priana" and "Goldrich". Amplification products from the citrus cultivars, the "Ramillete" almond and the "Stark Early Orange" apricot were cloned and some of them sequenced.

PCR reactions using DNA from *L. esculentum* cv "Madrigal", *L. pimpinellifolium* Line 1 or *P. trifoliata* var. "Flying Dragon" yielded no visible amplification product.

Hybridization of Suthern blots containing genomic DNA digestions from Citrus spp. with most putative RT clones from *Citrus*, showed multibanded patterns at high stringency conditions (Figs. 3 and 4), suggesting multiple copies were integrated throughout the citrus genome. *P. trifoliata* DNA also hybridized with RT clones from *Citrus* spp. showing a different but also multibanded pattern (Fig. 3). *L. eculentum* DNA did not hybridized with citrus RT clones.

A *P. armeniaca* RT clone used as probe showed only 1–3 bands when hybridized to Suthern blots containing DNA digestions from *P. amygdalus* and *P. armeniaca* cultivars (Fig.3). The signals were very weak, indicating a low number of copies per genome (1–3 copies). More

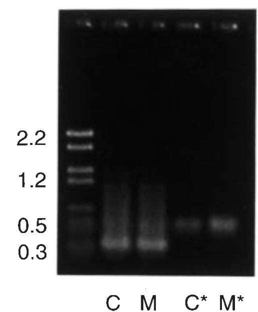

Fig. 2 Ethidium bromide-stained gel showing the amplification products obtained from *C. medica* (*M*) and *C. clementina* (*C*) using the degenerate RT primers. * Digoxigenin-labeled products

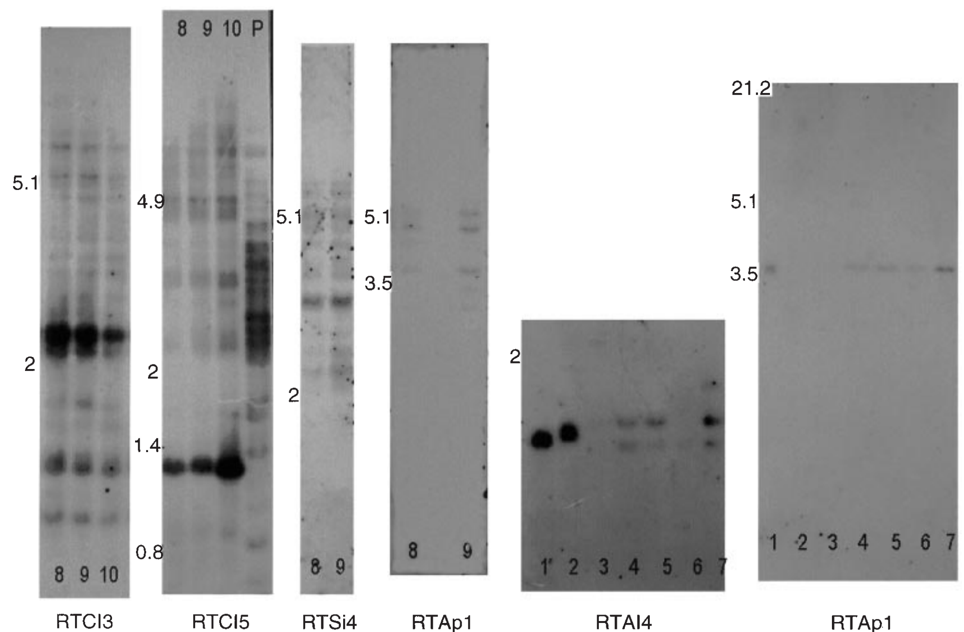

Fig. 3 Southern-blot analysis using RT clones Cl3, Cl5, Si4, Ap1 and Al4. *Lanes 1 and 2* correspond to DNA digestions from almond cultivars "Ramillete" and "Tuono", respectively. *Lanes 3–7*, apricot cultivars "Bergeron", "Rouge Rousillon", "Goldrich", "Canino" and "Polonais", respectively. *Lanes 8–10* correspond to clementine cultivars "Hernandina", "Nules" and "Fina". *Lane P* is *P. trifoliata* var. "Flying Dragon". Fragment sizes in kb are indicated to the left of each autoradiograph

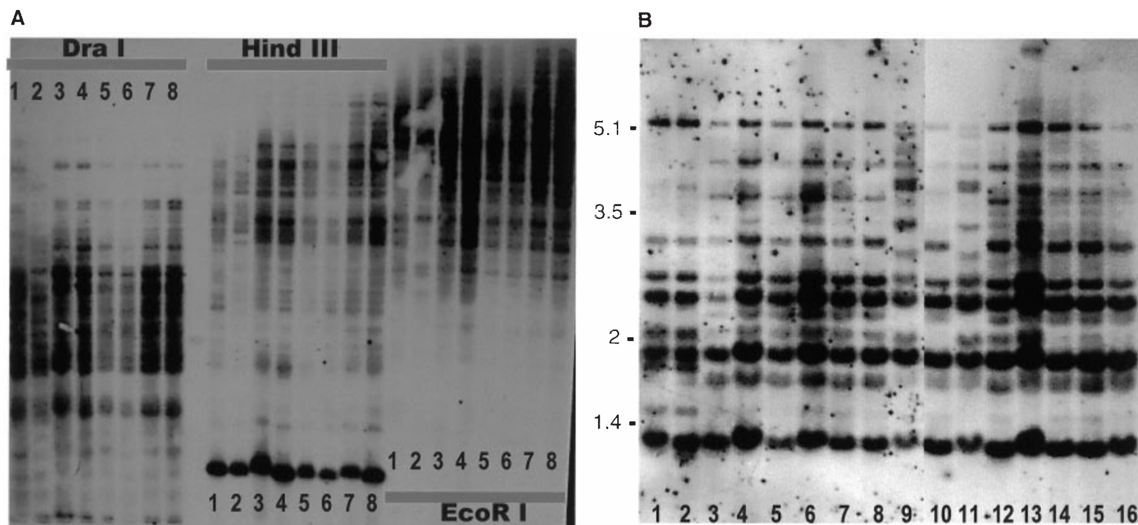

Fig. 4 Southern-blot analysis using RTGr5 (**A**) and RTCl6 (**B**) as probes. **A** *Dra*I, *Hind*III and *Eco*RI DNA digestions of *1*: *C. sinensis*, *2 C. medica*, *3–8* clementine cultivars “Oroval”, “Hernandina”, “Nules”, “Fina”, “Arrufatina” and “Marisol”. (**B**) *Hind*III digestions of 1–2 *C. unshiu* (satsuma mandarins), 3–8 clementine cultivars “Marisol”, “Arrufatina”, “Fina”, “Nules”, “Hernandina” and “Oroval”, *9 and 11 C. deliciosa*, *10 C. temple*, *12 C. tangerina*, *13 C. tachibana 14—16 C. sinensis* (sweet oranges) cultivars “Navelate”, “Navelina” and “Washington Foyos”

```
RTAl4       AFLNGELIEE IYMKQSDGFV SKGEENLVCK LQKSIYGLKQ ASRQWYLKFD EVVKSQGFIDNPLDECIYMK -FNGRNFIFM LL
RTCl3       ...H.D.E.. ...I.LC..R VA.K..HM.R .I..L..... SP....KR.. QFIQG.K.TRSEH.H.V.FR RLPDGA..YL ..
RTCl5       ...H...E.. ...LELE..T ET.K.....R .N..L.DP.. .PMC..KR.. SFIM.F.YNRLSS.H.V.Y. R.EDND.VIL ..
RTCl6       ...H...Q.. FV.Q.PE.Y. TS.K.DCI.L FK.PL...N. SP.H...... NFMITHA.NRCNY.C.V.FT ETGRGGM.YL .V
RTGr6       T..H.K.E.. ...L.PG..A ET.K.....R .N..L.V... .P.C..KR.. SFIM.L.YNRLSS.H.A.Y. R.EDND..IL ..
RTSi4       P..MVI.MKK SIXN.P.... LP.N.KN... .V..L..... .PK..HE... S.IL.H..KH.NA.K...F. ..MNDFGVII CW
RTSi9       ...HR..E.. ...L.PE..A XT.N.....R .N..L..... .PGC..KR.. SFIM.L.YNRLSS.H.A.Y. R.EDNV..IL M.
ASGPOL2     .....D.N.. V..E.PE..I LP.N.RK... .I..L..... .PK..HE... S.IL.Y..KY.SA.K...S. ..TDKYGVLV C.
PINCOPIAA   ...H.AIK.. V.VE.PL..E VQDRDTY..R .K.AL..... .P.A.NERM. SYLMKL..TRSNA.PNL.F. VVE.KPL.LV .Y
PTNCOPIAA   ...H...D.Q ...H.PK..M IQ.K.DH..L .K..L..... SP....KR.. TFMVGNDYCRSKF.S.V.HR KLLDGS.VYL ..
GNGCOPIA    ...H.D.E.. ...L.PE..A EE.K.....R .N..L..... .P.C..KR.. SYIM.L.YNRLNA.PYT.F. R.DED-..IL ..
CYRCOPIA    ...H.D.NDD ...E..E... IR.KK..IYR .N..L..... ...Y..KR.. SFMM.L..SRCEV.NF..F. RY.DDSL.IL ..
LIRCOPIAC   ..FH.D.E.. ...H.PT.Y. AP.K..K..R .K..L..... .P....K... SFMSGN.YRRCHA.H.C.L. K.DTS-Y.IL ..
MZEPOL2     .......D.. ...E.P.... V..Q.SK... .L..L..... .PK..HE... TTLT.A..AI.EA.R.V.YR .CG.GEGVIL C.
```

Fig. 5 Comparison of amino-acid sequences of RT clones with most similar amino-acid sequences. For abbreviations see the text (Results)

bands were obtained when Southern blots contained DNA from *C. clementina* but not as many as when using citrus RT clones as probes (Fig. 3). A *P. amygdalus* RT clone (RTAl4) produced strong signals for DNA from cultivars of this species but with only one band, while DNA from apricot cultivars produced in 1–3 bands (Fig. 3). Therefore, citrus RT sequences are more abundant in citrus genomes than apricot and almond RT sequences are through the *Prunus* genome. *Citrus* spp. present copies homologous to an almond RT sequence although in a small number than the citrus RT sequences.

Heterogeneity of RT sequences

Some putative RT fragments were sequenced and their sequences compared to homologous sequences searched for within three gene banks to check their nature (Fig. 5). These sequences have been submitted to EMBL. All of them correspond to RT domains of *copia*-like retrotransposons from unrelated plant species: *Gnetum montanum* (gb/M94476), *Cycas revoluta* (gb/M94473), *Liriodendron tulipifera* (gb/M94479), *Platanus occidentalis* (gb/M94485), and *Pinus coulteri* (gb/M94488). The closest sequences to RTAl4 and RTSi4 correspond to *pol* genes for the reverse transcription of *Asparagus officinalis* (dbj/D12834) and *Zea* mays (dbj/D12831).

Genetic relationships among all these sequences are graphically represented in the NJ tree based on the proportion of amino-acid differences (Fig. 6). Citrus RT sequences fall into three groups. Moreover, citrus RT sequences isolated from the same species (*C. sinensis* or *C. clementina*) do not cluster together. Results from bootstrapping tests show that some unions are very consistent such as RTCl5 with RTGr6 and RTSi9 (99% replications in bootstrap tests), and this group with GNGCOPIA from *G. montanum* (97%); or RT Si4 with ASGPOL2 from *A. officinalis* (98%). These results reflect great heterogeneity among RT sequences within each *Citrus* species genome.

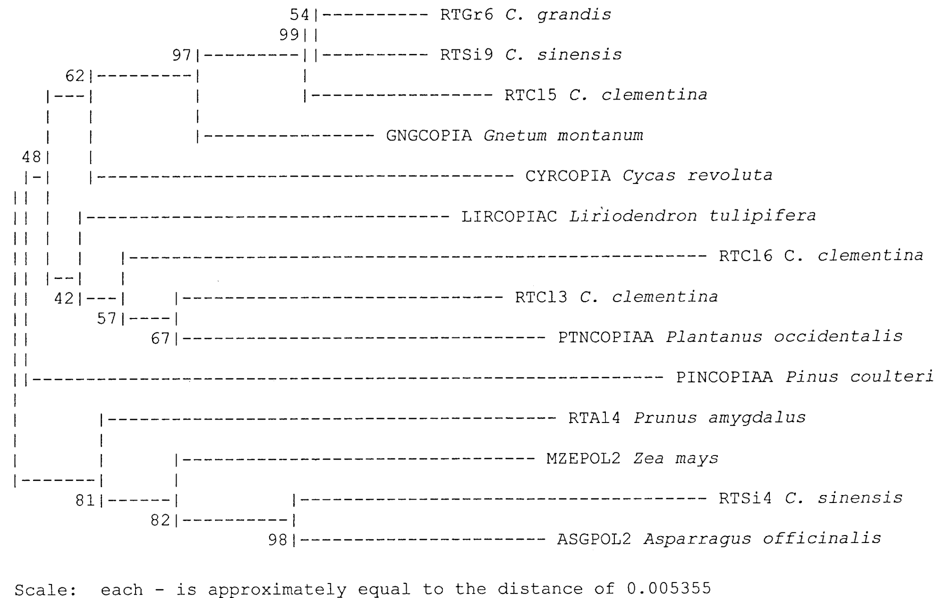

Fig. 6 Phylogenetic analysis of RT sequences using the p distance (Nei 1996) and the Neighbor-Joining method for constructing phylogenetic trees. *Numbers* at unions correspond to the percentage of replications where the grouping was obtained (Bootstrap test with 500 replications)

Genetic diversity among *Citrus* species revealed by the genomic distribution of RT sequences

The genomic organization of RTGr5 and RTCl6 was analyzed by Southern-blot analysis of several *Citrus* species (Fig. 4). Strong signals were observed in all digests (especially in *C. tachibana*) showing that these *copia*-like retrotransposons are highly repeated within their genomes. These digests revealed no clear difference among *C. sinensis* or *C. clementina* cultivars. By contrast, *Prunus* RT clones revealed RFLPs among apricot and almond cultivars (Fig. 3).

The hybridization pattern in *Sau*3AI digests revealed several bands of high molecular size that were absent in the *Dpn* II digests, suggesting abundant C methylation at genomic regions including RT sequences (Fig. 7a). *Citrus* spp. DNA digestions with the isoschizomer *Dpn*-II revealed important differences among sweet orange cultivars at genomic regions containing RTCl6. Thus, all navel cultivars assayed show an additional band (band c in Fig. 7b) compared to cultivars from the Valencia group. Within this latter group two differences (arrows in Fig. 7b) are observed between "Salustiana" and "Valencia Late" cultivars from the Valencia group. All these differences observed with *Dpn*II were not apparent using *Sau*IIIAI; thus, the relevant restriction sites must be C-methylated.

*Hin*dIII and *Dpn*II digests revealed differences in the hybridization patterns among the *Citrus* species using the RTCl6 probe. These differences are found between the satsuma and clementine mandarins. While satsuma hybridization patterns are somehow similar to the navel

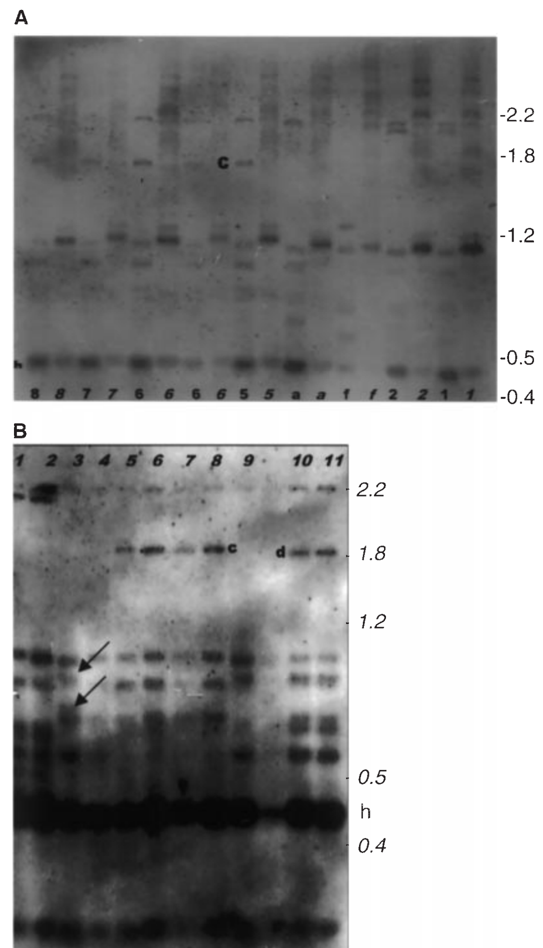

Fig. 7A, B Southern-blot analysis using RTCl6 as probe. DNA digestions with enzymes that recognize the same sequence (GATC) but is inhibited by C-methylation, *Sau*IIIAI (lanes in *italics*), or is not inhibited by C-methylation (*Dpn*II). **A** Alternative digestions of DNA from the same genotype with both enzymes. **B** Only *Dpn*II DNA digestions. *Lanes 1–2 C. madurensis*, *2* variegated mutant; *3–8* sweet orange cultivars, from the Valencia group, "Salustiana", "Valencia Late" and, from the navel group,"Caracara", Variegated branch from "Caracara", "Navelina" and "Navelate", respectively; *9 C. myrtifolia*; *10–11 C. unshiu*; *f Fortunella margarita* and *a C. aurantium*. *Arrows* point differences between "Salustiana" and "Valencia Late"

oranges, clementine hybridization patterns are almost identical to those shown by *C. tangerine* DNA digests and similar to those shown by sour orange (*C. aurantium*) and Valencia oranges (Fig. 4 and 7).

To study the possible relationship between the genetic instability of citrus and the change in the genomic distribution of retrotransposons, two variegated forms and their original clones were studied. One of these forms is maintained virus-free at the Citrus germplasm bank (*C. madurensis*) while the other appeared spontaneously in two branches of a tree in the field (the navel orange "Caracara"). Only differences regarding the relative intensity of bands of larger molecular size compared to the common band h, of lower molecular size, for *Dpn*II and also *Sau* 3AI digests of normal and variegated forms of *C. madurensis* were found (Fig. 7a, lanes 1 and 2).

Discussion

Genomic distribution of RT sequences

The amount of DNA in an unreplicated haploid cell (the C value) is relatively constant within a species. However, in higher plants it is particularly variable between species, ranging over nearly three orders of magnitude from 100 Mb (*Arabidopsis thaliana*) to nearly 100,000 Mb (*Fritillaria* species) (Smyth 1991). And yet the structural and developmental complexity of plant species with the lowest amounts of DNA per cell is not fundamentally different from those with the highest. The number of genes translated overall in the mature tobacco plant is estimated to be around 60,000 (Kamalay and Goldberg 1980). Thus, 75 Mb of DNA sequence (1.25 kb long on average, processed transcripts) is required for the total coding capacity of translated genes, close to the possible coding capacity of the *Arabidopsis* genome. The *Citrus* genome (0.62 pg/1 C, Guerra 1984) is larger than the Prunus genome which is quite small, 0.3 pg/1 C, around twice the size of *Arabidopsis* genome (Dickson et al. 1992).

Reannealing experiments early showed that most genomes carry a range of different copy numbers:

1. tandem repeated satellite DNAs; these are usually transcriptionally inactive, and large blocks are preferentially located in the constitutive heterochromatin (C-bands).
2. Dispersed repeats. Most of them are mobile elements, especially retrotransposons in plants.

The *Citrus* genome contains many more copies of RT domains than the *Prunus* genome. Therefore *copia*-like retrotransposons seem to be more abundant in the citrus genome which is larger than the *Prunus* genome. These results agree with those recently reported by Barakat et al. (1998) on the comparative genome organization of *A. thaliana* and *Grammineae*; the large genomes of the latter (from 415 Mb in rice to 5300 Mb in Barley) comprise many gene-empty regions, where transposons are abundant, separating gene clusters. Gene organization of the *Prunus* genome might be similar to that of the *Arabidopsis* genome. Therefore, *Prunus* genome could be a useful model for the molecular study of fruit trees.

Retrotransposons are a class of dispersed middle repetitive sequences and have contributed to the genetic diversity of their host species. Three mechanisms are involved: transposition, homologous recombination between retrotransposons or LTRs, and frequent mutation of methylated cytosine to thymine in retrotransposon sequences. Due to these characteristics, retrotransposon probes have been used as efficient RFLP markers in some plant species (Fukuchi et al. 1993), although this has not been the case for others such as *Pinus* spp. (Kamm et al. 1996). Genetic variability within *Prunus* spp. has been found using *Hind*III digests of genomic DNA and RT clones as probes, while within-species variability has not been found in *C. unshiu* or C. clementina. These results agree with previous studies on the distribution of genetic variability in *Citrus* and *Prunus* using isozymatic markers (Badenes et al. 1996; Herrero et al. 1996a) and the different improvement methods used for these species (selection of mutations and selection after hybridization, respectively). On the other hand, differences among *C. sinensis* cultivars have been found using a RT clone as a probe in Southern-blot analysis. Nearly all commercially important sweet orange cultivars originated through mutations altering horticultural characters, mostly fruit traits (Hodgson 1967). Consequently, many sweet orange cultivars can be distinguished only by fruit traits. Roose (1988) found no variability among 33 cultivars that belong to three groups, i.e. Valencia, blood and navel oranges (based on fruit traits), regarding isozyme and RFLP profiles. Similarly, Herrero et al. (1996a) found very little isozymatic variability within sweet oranges. Recently, using ISSR (inter-simple sequence repeat) markers that have more discriminating power than isozymes or RFLPs, Fang and Roose (1997) have found that all cultivars have almost the same ISSR fingerprints. Consequently, it has been surprising to find differences among sweet orange cultivars for hybridization patterns of *Dpn*II digests using a RT clone. The main difference (band c in Fig. 7) distinguishes navel from Valencia oranges, contrary to the ISSR differences that seem to affect individual cultivars but not the horticultural groups within *C. sinensis*.

Similarities in restriction patterns for RT probes agree with known genetic relationships among *Citrus* species as defined by Herrero et al (1996b). Thus, *C. tangerina* presents the same restriction pattern as *C. clementina*; sweet oranges and *C. temple* present very similar restriction patterns whereas satsuma and clementine mandarines show important differences in their restriction patterns at genomic regions containing RTCl6. Regarding the *Dpn*II fragments that hybridize to RTCl6, Valencia cultivars, sour orange and clementine mandarines are much alike while navel oranges and satsuma mandarines are very similar to each other. This finding reinforces the idea of a separate origin for both types of mandarines. Therefore, in *Citrus* spp., the distribution pattern of RTCl6 reflects

their phylogeny. A likely interpretation of these patterns is that the retrotransposon containing RTCl6 may have infected a common ancestor of the lineage and subsequently undergone sporadic bursts of amplification in different branches. One of these bursts might be related to the origin of navel oranges. Since the differential *Dpn*II fragments are not seen using *Sau*IIIAI, this implies that the corresponding restriction sites are now C-methylated. DNA methylation is a frequently reported mechanism controlling transposable elements (Finnegan et al. 1998).

Horizontal transmission

Considering the amino-acid sequences of RT domains cloned from *Citrus* spp., all of them cluster with copia-like retrotransposon reverse transcriptases or pol genes for the reverse transcription of plant species. Retrotransposon sequences amplified by PCR display varying degrees of sequence heterogeneity (Flavell et al. 1992; Voytas et al. 1992; Hirochika and Hirochika 1993; Smyth 1993). In fact, the *Ty*1-*copia* group retrotransposons from plants are unusually heterogeneous when compared to their counterparts in animals and lower eukaryotes (Flavell et al. 1992). A large heterogeneity among RT sequences from *C. clementina* or *C. sinensis* has also been observed. Thus, RTSi9 is more similar to a RT clone from *G. montanum* than to RTSi4. Results from bootstrap tests support these sequence similarities among RT clones from divergent host species because of the high repeatability of associations of some RT citrus sequences with the GNCCOPIA of *G. montanum* or the ASGPOL2 of *A. officinalis*. This similarity between RT sequences from very divergent hosts is rendered striking by the fact that reverse transcriptases are among the most variable proteins known (Xiong and Eickbush 1990). Additionally, the observed base substitution rate for a retrotransposon (*Ty*1 of yeast) is 2.5×10^{-5} bp per replication cycle (Gabriel et al. 1996). Thus, even a few cycles of insertion, transcription, reverse transcription and re-insertion would lead to a rapid divergence in the sequence of retrotransposons. All these observations suggest that horizontal movement of retrotransposons has occurred regularly in the past, even between hosts in different kingdoms (Flavell et al. 1992). Viruses are obvious candidate vectors. Among them, plant caulimovirus are of interest because, while they are double-stranded DNA viruses, their replication involves a RT step. In fact, it seems likely that the ancestor of the caulimoviruses ”captured” the RT of an LTR retrotransposon relatively recently (Xiong and Eickbush 1990). There are examples indicating that the genomes of infectious agents may sometimes capture retrotransposon sequences from their host cells. If this is the case, a common habitat for plant species among which horizontal transmission of retrotransposons has occurred is to be expected. Noteworthy, the species presenting a RT sequence most similar to citrus RT clones is *G. montanum* (only 22.5% amino-acid substitutions, considering RTGr6). The distribution area of this gymnosperm species is from Central to South East Asia (Markgraf 1930), two of the main centers of diversity and origin of actual *Citrus* spp. (Vavilov 1951). *G. montanum* is a woody vine whose broad leaves superficially resemble those of angiosperms while its conducting cells in the wood include open-ended pipes known as vessels, which is also characteristic of angiosperms. Therefore, although both species are not phylogenetically closely related, they have shared common habitats.

Genetic instability of citrus

Computer-assisted database searches using several *copia*-like retroelements as query sequences have revealed that ancient, degenerate retrotransposon insertions are found in close proximity to 21 previously sequenced plant genes, suggesting that these elements may be involved in gene duplication and the regulation of gene expression (White et al. 1994; Bureau et al. 1996). Thus, it appears that transposable element insertions have been important contributors to the establishment of novel patterns of transcription for a variety (up to 100 found so far) of plant genes, and could then be significant effectors of evolutionary change (Labrador and Corces 1997).

Numerous copies of *copia*-like elements exists in citrus spp. *Copia*-like elements are highly expressed in *Drosophila* and have been shown to be the causative agent of many spontaneous mutations (Bingham and Zachar 1989). In contrast, plant *copia*-like retrotranposons are transcribed at low levels under normal conditions and have been found to be responsible for only a few mutations. However, several stress conditions have been reported to activate them through activation of transcription, such as tissue culture, protoplast formation, crown gall tumors induced by *Agrobacterium tumefaciens*, wounding, pathogen infections as well as microbial elicitors and abiotic factors known to induce the plant defense response (Hirochika 1995; Moreau-Mhiri et al. 1996; Pearce et al. 1996; Vernhettes 1997).

Genetic transformation of citrus cultivars via *A. tumefaciens* has been envisaged as a quick method for citrus improvement to overcome the main difficulties found by classical methods in most woody plants. May this improvement methodology speed up undesired evolutive processes? It has been postulated that retroviruses are relative newcomers to the world, having perhaps only arisen after the advent of mammals (in Doolittle et al. 1989). There are many known avian retroviruses but these are often closely related to mammalian viruses, suggesting that these viruses have crossed the species boundaries between birds and mammals (Doolittle et al. 1989; Xiong and Eickbush 1990). Going further, Doolittle et al. (1989) speculate that retroviruses have not colonized plants (among other phyla) because their env genes cannot mediate such a huge jump, enabling the plant-virus interaction to take place. We have shown in the present paper that retrotransposons have propagated profusely through the citrus genome, that horizontal transmission has most likely occurred and that they are involved in genetic variabili-

ty among sweet orange cultivars. The relative importance or contribution of transposon activity to the genetic instability of citrus remains to be studied. Meanwhile, and given that citrus is an abundant perennial crop, it does not seem advisable to commercialize transgenic citrus plants where the transgene(s) is part of a viral genome (like coat protein genes) as a means to control virus diseases, without carefully studying the possibilities suggested by Xiong and Eickbush (1990) that new types of viruses may evolve either by the capture of RT sequences from retrotransposons by pre-existing viruses, or else by these transposable elements acquiring additional genes and becoming a virus.

Acknowledgements The authors are grateful to the late Dr. Efigenio García from the Department of Fruit Breeding, CEBAS (Murcia) for providing the *Prunus* cultivars and the BioScience Computing Resource at the University of Valencia. This work was supported by research grants from the Conselleria de Cultura, Educación y Ciencia (GV-3114/95), CICYT (AGF97–0417) and European Commission (ERBICT98310).

References

Badenes ML, Asíns MJ, Carbonell EA, Llácer G (1996) Genetic diversity in apricot (*Prunus armeniaca* L.) attending to plum pox virus resistance improvement. Plant Breed 115:133–139

Barakat A, Matassi G, Bernardi G (1998) Distribution of genes in the genome of *Arabidopsis thaliana* and its implications for the genome organization of plants. Proc Natl Acad Sci USA 95:10044–10049

Bingham P, Zachar Z (1989) Retrotransposons and the FB transposon from *Drosophila melanogaster*. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington D.C., pp 485–502

Bureau TE, Ronald PC, Wessler SR (1996) A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc Natl Acad Sci USA 93:8524–8529

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Dickinson EE, Arumuganathan K, Kresovich S, Doyle JJ (1992) Nuclear DNA content variation within the rosaceae. Am J Bot 79:1081–1086

Doolittle RF, Feng DF, Johnson MS, McClure MA (1989) Origin and evolutionary relationships of retroviruses. Quart Rev Biol 64:1–30

Fang DQ, Roose ML (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers Theor Appl Genet 95:408–417

Finnegan EJ, Gender, RK, Peacock WJ, Dennis ES (1998) DNA methylation in plants. Annu Rev Plant Mol Biol 49:223–247

Flavell AJ, Smith DB, Kumar A (1992) Extreme heterogeneity of the *Ty*1-*copia* group of retrotransposons in plants. Mol Gen Genet 231:233–242

Fukuchi A, Kikuchi F, Hirochika H (1993) DNA fingerprinting of cultivated rice with retrotransposon probes. Jpn J Genet 68:195–204

Gabriel A, Willems M, Mules EH, Boeke J (1996) Replication infidelity during a single cycle of *Ty*1 retrotransposition. Proc Natl Acad Sci USA 93:7767–7771

Guerra MS (1984) Cytogenetics of Rutaceae. II. Nuclear DNA content. Caryologia 37:219–236

Herrero R, Asíns MJ, Carbonell EA, Navarro L. (1996a) Genetic diversity in the orange subfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. Theor Appl Genet 92:599–609

Herrero R, Asíns MJ, Pina JA, Carbonell EA, Navarro L (1996b) Genetic diversity in the orange subfamily Aurantioideae. II. Genetic relatioships among genera and species. Theor Appl Genet 93:1327–1334

Higgins DG, Bleasby AJ, Fuchs R (1992) Clustal V improved software for multiple sequence alignament. Comput Appl Biosci 8(2):189–191

Hirochika H (1995) Regulation of plant retrotransposons and their use for genome analysis. Gamma Field Symp 34:77–91

Hirochika H, Hirochika R (1993) *Ty*1-*copia* group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 68:35–46

Hodgson RW (1967) Horticultural varieties of citrus. In: Reuther W, Webber HJ, Batchelor LD (eds) The citrus industry, vol 1. University of California Press, Berkeley, pp 431–591

Kamalay JC, Goldberg RB (1980) Regulation of structural gene expression in tobacco. Cell 19:935–946

Kamm A, Doudrick RL, Heslop-Harrison, Schmidt T (1996) The genomic and physical organization of *Ty*1-*copia*-like sequences as a component of large genomes in *Pinus ellioti* var. elliotti and other gymnosperms. Proc Natl Acad Sci USA 93:2708–2713

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetic analysis. University Park, Penn State University

Labrador M, Corces VG (1997) Transposable element-host interactions: regulation of insertion and excision. Annu Rev Genet 31:381–404

Markgraf F (1930) Monographie der gattung *Gnetum* In: Doctters Van Leeuwen WM, Von Faber FC, Beumee JGB (eds) Bulletin du Jardin Botanique. Serie III, volume X., livre 4. Jardin Botanique de Buitenzorg, pp 466–490

Mendel K (1981) Bud mutations in citrus and their potential commercial value. Proc Int Soc Citriculture, vol 1, pp 86–89

Monforte A, Asíns MJ, Carbonell EA (1996) Salt tolerance in *Lycopersicon* species. IV. Efficiency of marker-assisted selection for salt tolerance improvement. Theor Appl Genet 93:765–772

Moreau-Mhiri C, Morel JB, Audeon C, Ferault M, Grandbastien MA, Lucas H (1996) Regulation of expression of the tobacco Tnt1 retrotransposon in heterologous species following pathogen-related stresses. Plant Journ 9:409–419

Navarro L, Juarez J, Pina JA, Ballester JF, Arregui JM (1988) The citrus variety improvement program in Spain (CVIPS) after 11 years. In: Timmer LW, Garnsey SM, Navarro L, eds. Proc 10th Conf Inter Org Citrus Virologists, IOCV, Riverside, pp 400–406

Nei M (1996) Phylogenetic analysis in molecular evolutionary genetics. Annu Rev Genet 30:371–403

Pearce SR, Kumar A, Flavell AJ (1996) Activation of the *Ty*1-*copia* retrotransposons of potato (*Solanum tuberusum*) during protoplast isolation. Plant Cell Rep 15:449–953

Roose ML (1988) Isozymes and DNA RFLPs in citrus breeding and systematics. In: Goren R, Mendel K (eds) Proc 6th Int Citrus Congr. Balaban Publishers, Rehovot, Israel, vol 1, pp 155–165

Smyth DR (1991) Dispersed repeats in plant genomes. Chromosoma 100:355–359

Smyth DR (1993) Plant retrotranposons. In: Verma DPS (ed) Control of plant gene expression. CRC Press, Boca Raton, Florida, USA, pp 1–15

Spiegel-Roy P, Goldschmidt EE (1996) Biology of citrus. Cambridge University Press, Cambridge, UK

Vavilov NI (1951) Phytogeographic basis of plant breeding. The origin, variation, immunity and breeding of cultivated plants. Chron Bot 13:253–261

Vernhettes S, Grandbastien MA, Casacuberta JM (1997) In vivo characterization of transcriptional regulatory sequences involved in the defense-associated expression of the tobacco retrotransposon Tnt1. Plant Molecular Biology 35:673–679

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) Copia-like retrotransposons are ubiquitous among plants.Proc Natl Acad Sci USA 89:7124–7128

White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91:11,92–11,96

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences EMBO J 9:3353–3362