Theor Appl Genet (1996) 92: 425–435

P.-C. Liou · F. G. Gmitter Jr. · G. A. Moore

Characterization of the *Citrus* genome through analysis of restriction fragment length polymorphisms

Received: 27 September 1994 / Accepted: 21 July 1995

Abstract Studies on the nature of restriction fragment length polymorphisms (RFLPs) were undertaken to characterize the *Citrus* genome. This type of analysis has not been carried out with any other perennial crop. *Citrus reticulata* Blanco cv Clementine, *C. xparadisi* Macf. cv Duncan, and an F_1 hybrid (LB 1-21) were used to determine what probe/enzyme combinations revealed polymorphisms in Southern analysis, and a backcross family (LB 1-21×'Clementine') of 65 randomly selected hybrid seedlings was used for some analyses. A majority (73%) of the clones examined from a *Pst*I genomic library appeared to detect single-copy sequences based on RFLP banding patterns, while clones from a cDNA library revealed a lower percentage of single copy sequences. When hybridization stringencies were lowered, 21% of the genomic clones examined revealed greater copy numbers. *Pst*I digestion of 'Duncan' DNA indicated abundant methylation, so the relatively high frequency of multiple-copy sequences observed at moderate stringency cannot be attributed to a lack of methylation of the *Citrus* DNA. The polymorphisms in banding patterns observed primarily resulted from insertions and/or deletions rather than from base substitutions, and a model is presented to account for the varying patterns obtained from individual probes with different restriction enzymes. Finally, a model for transposon activity in *Citrus* is proposed, based on observations made during the course of these studies.

Communicated by F. Salamini

P.-C. Liou[1] · F. G. Gmitter Jr.
Citrus Research and Education Center,
Institute of Food and Agricultural Sciences, University of Florida,
700 Experiment Station Road, Lake Alfred, FL 33850, USA

G. A. Moore (✉)
Horticultural Sciences Department,
Institute of Food and Agricultural Sciences, University of Florida,
Gainesville, FL 32611, USA

Present address:
[1] Taiwan Agricultural Research Institute, 189 Chung-cheng Road,
Wu-feng 41301 Taichung, Taiwan, Republic of China

Key words RFLP · Molecular markers · Fruit breeding

Introduction

Molecular biological technology has provided new approaches for plant breeding and genetics studies in plants. A molecular marker linkage map can be constructed for a crop of interest and subsequently used for locating and tagging desirable genes, for marker-assisted selection, and for map-based cloning (Tanksley et al. 1989; Reiter et al. 1992; Martin et al. 1993). Several types of molecular markers may be used in map construction, including restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), and isozymes (Tanksley et al. 1989; Williams et al. 1990). Molecular-marker maps, especially ones based on RFLPs, have also been used to characterize genome organization in a few annual plant species, in particular tomato (*Lycopersicon esculentum* Mill.) (Bernatzky and Tanksley 1986; Zamir and Tanksley 1988; Miller and Tanksley 1990), rice (*Oryza sativa* L.) and other grasses (McCouch et al. 1988; Ahn and Tanksley 1993; Bennetzen and Freeling 1993; Moore et al. 1993).

The genus *Citrus* can be readily and profitably studied using molecular techniques. Species in this genus are economically important crops throughout tropical and subtropical regions of the world, where they have been cultivated for thousands of years (Soost and Cameron 1975). *Citrus* species are diploids with relatively few chromosomes (2n=2x=18) and small genomes, e.g., 1C=0.38–0.62 pg (Soost and Cameron 1975; Guerra 1984; Arumuganathan and Earle 1991). Breeding and isozyme studies have shown that most *Citrus* species are heterozygous and that the genus is highly polymorphic (Torres 1983; Torres et al. 1985; Roose 1988). Most species are interfertile, and interspecific as well as some intergeneric hybrids can be obtained (Soost and Cameron 1975). However, a number of factors have made *Citrus* breeding a slow and difficult process. These include crossing barriers such as (self-and cross-incompatibility, sterility, and nucellar embryony),

long juvenile periods, high heterozygosity, the quantitative inheritance of most characters, and large individual plant sizes (Soost and Cameron 1975). Despite active and vigorous *Citrus* breeding programs, very few cultivars have been produced via standard breeding methods; most commonly grown cultivars have been derived from bud mutations or well-adapted superior seedlings (Hodgson 1967; Soost and Cameron 1975). The above factors have also hindered genetic studies, resulting in a poor understanding of *Citrus* genetics. Thus, the use of molecular methods could lead to valuable genetic insights and increased breeding efficiency.

Until recently, little information on *Citrus* molecular biology has been available (Roose 1988; Durham et al. 1992; Jarrell et al. 1992; Cai et al. 1994). The studies described in the present paper were done in conjunction with the development of an RFLP linkage map in *Citrus* (Liou 1990; Durham et al. 1992). The major objective was to examine the nature of RFLPs observed and thus to begin a characterization of the *Citrus* genome. Specific features of interest included: frequencies of single-copy, multiple and repetitive sequences, mechanisms underlying the observed polymorphism, and evidence suggestive of transposon activity.

Materials and methods

Plant materials

Three *Citrus* lines were used for the first evaluations of probe/enzyme combinations that revealed polymorphism in Southern analysis. They were: (1) *Citrus reticulata* cv Clementine (called 'Clementine' hereafter); (2) *C. xparadisi* cv Duncan (called 'Duncan' hereafter); and (3) a hybrid of 'Clementine'×'Duncan' designated LB 1-21. For some analyses, a backcross (LB 1-21×'Clementine') progeny of 65 randomly selected hybrid seedlings was used. Probe/enzyme combinations revealing polymorphism between the female parent (LB 1-21) and the male parent ('Clementine') were used in progeny analysis. All plant materials were maintained at the Citrus Research and Education Center, University of Florida.

Sources of probes

The clones employed in this study were the same ones used for mapping (Durham et al. 1992). To provide genomic clones, a *Pst*I-derived *Citrus* genomic library was constructed with the DNA of 'Temple' tangor, a presumed natural hybrid of *C. reticulata* and *C. sinensis* L. Osb. (Liou 1990; Durham et al. 1992); clones selected from this library were designated as pgCit probes. cDNA clones from two sources were available for this study. Two probes, designated pcPt001 and pcPt002, were from a cDNA library prepared with mRNA isolated from cold-acclimated 'Pomeroy' trifoliate orange [*Poncirus trifoliata* (L.) Raf.] (Durham et al. 1992). Probes designated pRLc were from a cDNA library of mRNA isolated from leaves of rough lemon (*C. jambhiri* Lush) (Jarrell et al. 1992), and were provided by M.L. Roose, University of California, Riverside. All of the cDNA clones had previously been shown in mapping studies to reveal polymorphism in *Citrus*.

Restriction digests, electrophoresis, and Southern analysis

DNA isolation was as previously described (Durham et al. 1992). The DNA was digested with six 6-bp-recognition restriction enzymes (*Eco*RI, *Eco*Rv, *Bgl*II, *Hin*dIII, *Bam*HI and *Pst*I) according to manufacturers' instructions [Bethesda Research Laboratories, Gaithersburg, Md., or Boehringer Mannheim Biochemicals (BMB), Indianapolis, Ind.]. To construct filters, the digested DNA (2 µg per lane) was separated and transferred to nylon membranes as described (Durham et al. 1992).

Cloned inserts used as probes were isolated with the alkaline-lysis miniprep procedure (Maniatis et al. 1982), separated in 0.8% agarose gels, and purified using GENECLEAN (BIO 101 Inc., La Jolla, Calif.). A random primer DNA labeling kit (BMB) and -[^{32}P]-dCTP (New England Nuclear, Boston, Mass.) were used to label the probes (Durham et al. 1992). The prehybridization and hybridization procedures were according to Church and Gilbert (1984).

The basic method for washing the filters after hybridization was two washes with 0.5×SSC, 0.1% SDS, 65°C, 30 min; followed by another two washes with 0.1×SSC, 0.1% SDS, 65°C, 30 min. This was considered high stringency and has been estimated to detect homologies greater than 99% (Beltz et al. 1983; McCouch et al. 1988). Autoradiography was performed with Kodak X-omat AR X-ray film and one or two intensifying screens (Dupont Cronex Lightning-Plus) at –80°C for 1–7 days depending on the hybridization signal. In some instances, filters were first washed under more moderate stringency conditions (four washes with 0.5×SSC, 0.1% SDS, 65°C); this moderate stringency has been estimated to allow the detection of sequence homology between duplexes at a minimum level of 80% (Beltz et al. 1983; McCouch et al. 1988). The washed filters were then exposed to X-ray films with two intensifying screens per film. Once the films were developed, usually after 24–36 h, the same filters were washed under high-stringency conditions as above and re-exposed to compare the gene copy numbers and polymorphisms revealed by different stringency conditions. Also, in some instances, a higher-stringency washing procedure (0.05×SSC, 0.1% SDS, 65°C for the last two washes) was utilized to determine the stringency tolerance of *Citrus* DNA.

Calculations

The sizes of the fragments detected by each clone were determined by comparing them to molecular-weight standards. Filters with the DNA of 'Clementine', 'Duncan', and LB 1-21, were produced using various probe/enzyme combinations to determine which combinations revealed polymorphism, as well as the number and size of the hybridizing fragments detected among the three individuals. For construction of the linkage map (Durham et al. 1992), and for other analyses described below, data generated from the analysis of segregation of the entire set of markers (including RFLP and isozyme markers) in the backcross-progeny population were processed through LINKAGE-1 (Suiter et al. 1983).

Results and discussion

Gene copy number profiles determined under varying stringency conditions

A total of 97 genomic and cDNA clones were tested for copy number using genomic DNA from 'Clementine', 'Duncan', and LB 1-21 and Southern analysis with high-stringency washes. Four clones revealed apparent maternal inheritance and were excluded; consequently, 93 clones were classified as hybridizing to single-copy, multiple-copy, and repetitive sequences as follows: single-copy sequences were those where no more than two clear bands were present per lane and bands appeared to follow Mendelian inheritance patterns in 'Clementine', 'Duncan', and LB 1-21 (Fig. 1A); multiple-copy sequences had three to

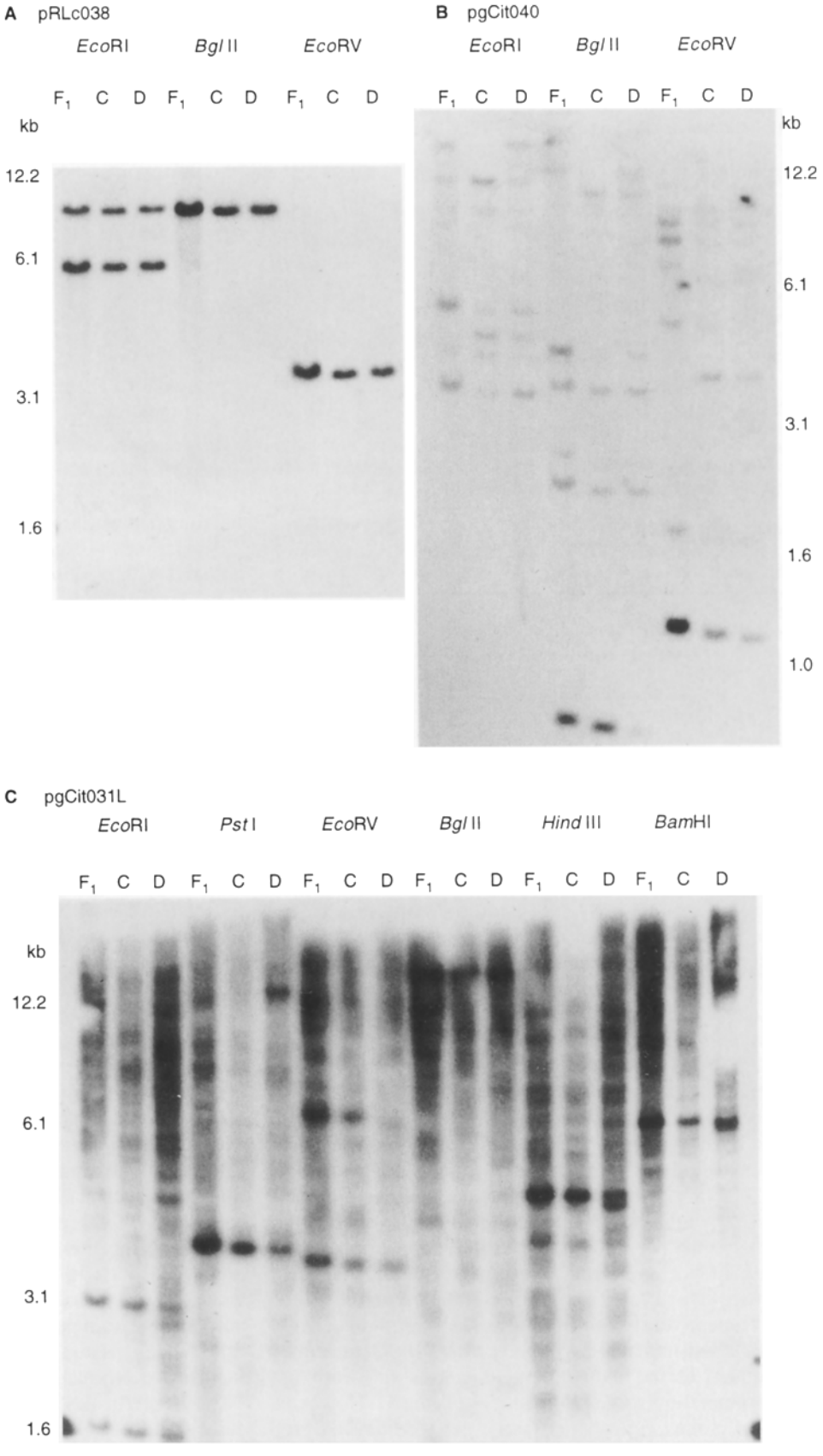

Fig. 1 Examples of filters probed with clones classified as revealing **A** single-copy (pRLc038), **B** multiple-copy (pgCit040), and **C** repetitive (pgCit031L) sequences after washing with high-stringency conditions (0.1×SSC, 65°C). In each set of three lanes (i.e., 1-2-3, 4-5-6, etc.) the DNA was digested with a different restriction enzyme. The *first lane* in each set of three=the F_1 LB 1-21 (F_1), the *second lane*='Clementine' (*C*), and the third lane='Duncan' (*D*). Molecular weights in kb are indicated

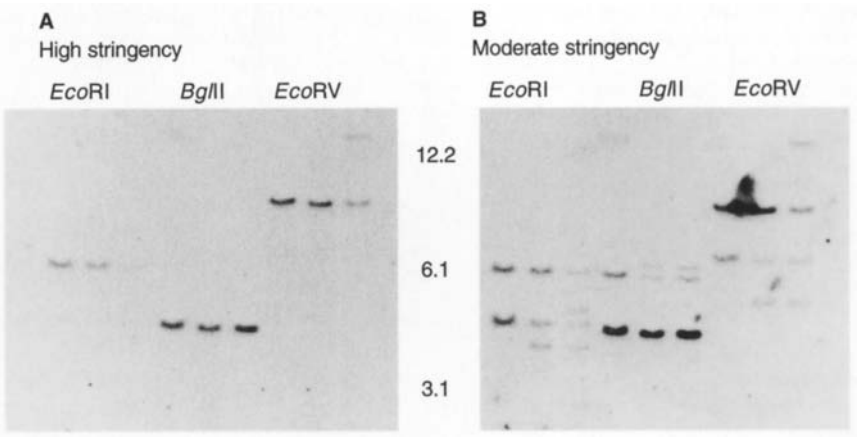

Fig. 2 A, B The genomic clone pgCit020 revealed a single-copy sequence after high-stringency washing of the filter (**A**), but multiple bands after washing under moderate conditions (**B**). Polymorphism was apparent only after moderate-stringency washing. In each set of three lanes (i.e., 1-2-3, 4-5-6, etc.) the DNA was digested with a different restriction enzyme. The *first lane* in each set of three=the F_1 LB 1-21, the *second lane*='Clementine', and the *third lane*='Duncan'. Molecular weights in kb are indicated

Table 1 Numbers and percentages of clones hybridizing to single-copy, multiple-copy, and repetitive sequences after Southern analysis and high-stringency washing conditions (0.1×SSC, 65°C)

Clone type tested	Number clones copy	No. (%) single copy	No. (%) multiple sequences	No. (%) repetitive sequences
cDNA	19	12 (63)	6 (32)	1 (5)
Genomic	74	54 (73)	14 (19)	6 (8)
Total	93	66 (71)	20 (21)	7 (8)

ten bands present per lane (Fig. 1B), and repetitive sequences had either more than ten bands per lane or smeared signals with no discrete bands (Fig. 1C).

With high-stringency washes, the *Pst*I genomic clones revealed 73% single-copy, 19% multiple-copy, and 8% repetitive sequences (Table 1). The genome size of *Citrus* has been estimated to be approximately 0.8–1.2 pg/2C (Guerra 1984; Arumuganathan and Earle 1991), similar to that of *O. sativa* and similar or somewhat smaller than *L. esculentum* (McCouch et al. 1988; Arumuganathan and Earle 1991). When the copy number profile of the *Citrus* *Pst*I genomic library was compared to the *Pst*I genomic library of rice (McCouch et al. 1988) and the random-sheared library of tomato (Zamir and Tanksley 1988), also determined under high-stringency conditions, the *Citrus* clones had a slightly lower percentage of single-copy clones (85% and 78% in rice and tomato, respectively), a greater or equal percentage of multiple-copy clones (12% and 18% in rice and tomato, respectively), and a greater percentage of repetitive clones (3% and 4% in rice and tomato, respectively). Surprisingly, the group of cDNA clones studied revealed a greater percentage of multiple-copy sequences and a lower percentage of single-copy sequences than the *Pst*I genomic clones (Table 1). However, the cDNA clones were not randomly selected; they had previously detected polymorphism in other *Citrus* populations (Jarrell et al. 1992). Further, the number of cDNA clones was not large enough to expect that the observed percentages are truly representative of *Citrus* cDNA libraries in general.

To determine the effects of stringency conditions on the banding patterns produced following Southern analysis, 24 *Citrus* clones (15 genomic and nine cDNA clones) were assayed with both moderate- and high-stringency washes (Fig. 2). Five (three genomic and two cDNA clones) of the 24 studied clones (21%) revealed greater copy numbers under moderate-, rather than high-, stringency conditions. Only 9 of the 15 *Citrus* genomic clones analyzed at moderate stringencies appeared to be single copy in nature. While this is a low number of clones from which to draw a conclusion, this frequency was nearer to that of the single-copy sequences found in rice under similar conditions (58%) than the frequency found in tomato (92%) (McCouch et al. 1988; Miller and Tanksley 1990).

There are at least two potential explanations for why a relatively high frequency of multiple-copy sequences was detected in *Citrus* genomic DNA under both high and moderate stringencies. The genomic clones used in this study were derived from digestion with *Pst*I, a methylation-sensitive restriction endonuclease, and most likely represented less-methylated regions of the *Citrus* genome. Since coding sequences are often hypomethylated, regions of plant genomes that are less methylated may be enriched for coding regions and thus represent single-copy sequences (Burr et al. 1988; McCouch et al. 1988; Kesseli et al. 1994). However, repeated sequences of different plant genomes may also be characterized by different degrees of *C*-methylation (McCouch et al. 1988; Messeguer et al. 1991; Wu and Tanksley 1993). When *C*-methylation occurs less frequently in repeated DNA, such DNA might be included more frequently in a *Pst*I genomic library and the library would be less biased toward single-copy sequences. Alternatively, the high percentage of clones hybridizing to multiple sequences could be due to the existence of multigene families or duplications in the *Citrus* genome (Beltz et al. 1983; Jarrell et al. 1992). Multiple-copy sequences that are detected only at more moderate stringencies may be due to the presence of members of multigene families that cross-hybridize in Southern analysis because of partial sequence homologies (Beltz et al. 1983).

Table 2 Polymorphisms detected by different restriction enzymes

Restriction enzyme[a]	Genomic library[b]	cDNA library[b]
Among 'Clementine', 'Duncan', and their hybrid LB 1-21		
*Eco*RI	45	40
*Bgl*II	43	50
*Eco*RV	42	30
*Hin*dIII	38	14
*Bam*HI	29	29
*Pst*I	21	29
Any enzyme	71 (36.3)	65 (32.0)
Between 'Clementine' and LB 1-21		
*Eco*RI	36	10
*Bgl*II	30	40
*Eco*RV	36	20
*Hin*dIII	41	0
*Bam*HI	23	14
*Pst*I	12	0
Any enzyme	60 (29.7)	45 (14.0)
Between 'Clementine' and 'Duncan'		
*Eco*RI	42	40
*Bgl*II	40	45
*Eco*RV	34	30
*Hin*dIII	35	14
*Bam*HI	29	29
*Pst*I	23	29
Any enzyme	64 (33.8)	50 (31.2)

[a] With restriction enzymes *Eco*RI, *Bgl*II, and *Eco*RV, 77 genomic and 20 cDNA clones were tested; with *Hin*dIII, *Bam*HI, and *Pst*I, 34 genomic and 7 cDNA clones were tested

[b] Percent polymorphic clones (average)

McCouch et al. (1988) suggested that their rice *Pst*I library was less biased toward single-copy sequences than ones constructed from maize or tomato because *C*-methylation was less frequent in rice than in the other species. To test this hypothesis, they compared digestion of rice and tomato DNA with methylation-sensitive restriction enzymes, such as *Pst*I, and *C*-methylation-insensitive enzymes, such as *Eco*RI. Digestion with *Eco*RI cleaved the DNA of both species into a wide range of fragment sizes uniformly distributed following agarose-gel electrophoresis. In contrast, *Pst*I digestion of tomato DNA left a large amount remaining at high molecular weight, while *Pst*I-digested rice DNA showed a distribution of fragment sizes intermediate between that produced by *Pst*I digestion of tomato DNA, and rice and tomato DNA digestion with *Eco*RI. In the present study, when 'Duncan' grapefruit DNA was digested with *Eco*RI and *Pst*I and separated on a 1.5% agarose gel, the result was comparable to that observed with tomato DNA, i.e., a substantial portion of the *Pst*I-digested citrus DNA remained as high-molecular-weight fragments (data not shown). Thus it appears that the *Citrus* genome contains abundant methylated DNA and the relatively high frequency of multiple-copy sequences revealed at moderate stringencies is not due to the lack of methylation in repeated DNA.

Differences in the ability of different enzymes and clones to detect RFLPs

A survey using 97 clones and six restriction enzymes was carried out to determine which combinations of restriction enzymes and clones would reveal polymorphism among 'Clementine', 'Duncan', and their F_1 hybrid LB 1-21 (Table 2): 68 clones (70%) revealed polymorphism among all three genotypes; 59 clones (61%) were polymorphic between 'Clementine' and 'Duncan'; and 55 clones (57%) revealed polymorphism between 'Clementine' and LB 1-21. The frequencies at which RFLPs were detected by different restriction enzymes were variable, with a tendency toward higher frequencies with *Eco*RI, *Bgl*II, *Eco*RV and *Hin*dIII, and lower ones with *Bam*HI and *Pst*I.

The cDNA clones used in this study cannot be regarded as random clones; they were selected because they had previously revealed polymorphism and it was expected that they would detect polymorphism here at a high frequency. However, the percentage of genomic clones detecting polymorphism was greater than that of the cDNA clones (Table 2). This was also observed during linkage analysis of a separate intergeneric *Citrus* hybrid population (Durham et al. 1992). The apparent superiority of *Pst*I genomic clones over cDNA clones for detecting polymorphism in *Citrus* contrasts with tomato (Miller and Tanksley 1990), lentil (Havey and Muehlbauer 1989), and lettuce (Landry et al. 1987), where cDNA clones detected significantly more polymorphism.

Features of RFLP in *Citrus*: evidence for insertions and/or deletions

Restriction polymorphisms in plants have been attributed to two causes: the alteration of a restriction site by point mutation or gross alterations in DNA structure such as insertions, deletions, or rearrangements (Burr et al. 1983). An RFLP resulting from base substitution should affect the restriction pattern of only one enzyme. Conversely, an RFLP caused by insertion or deletion should be generated by any enzyme that has a restriction site near the affected area. In this case, the probabilities of different enzymes detecting polymorphism with a given probe would not be independent if the restriction sites for different enzymes along a piece of genomic DNA are independent and do not overlap (Burr et al. 1983; McCouch et al. 1988; Kesseli et al. 1994).

Southern analysis showed that 58% (32/55) of the clones polymorphic between 'Clementine' and LB 1-21 were polymorphic with more than one restriction enzyme. To study this phenomenon in more detail, the probability that a given enzyme detected polymorphism with a given probe was regressed on the number of other enzymes detecting polymorphism with the same probe, as had been previously done for rice (McCouch et al. 1988). Southern analyses of 39 clones tested with six different restriction enzymes were used for this study. The R^2 values ranged from 0.94 to 0.59, with three larger than 0.80 (*Eco*RV,

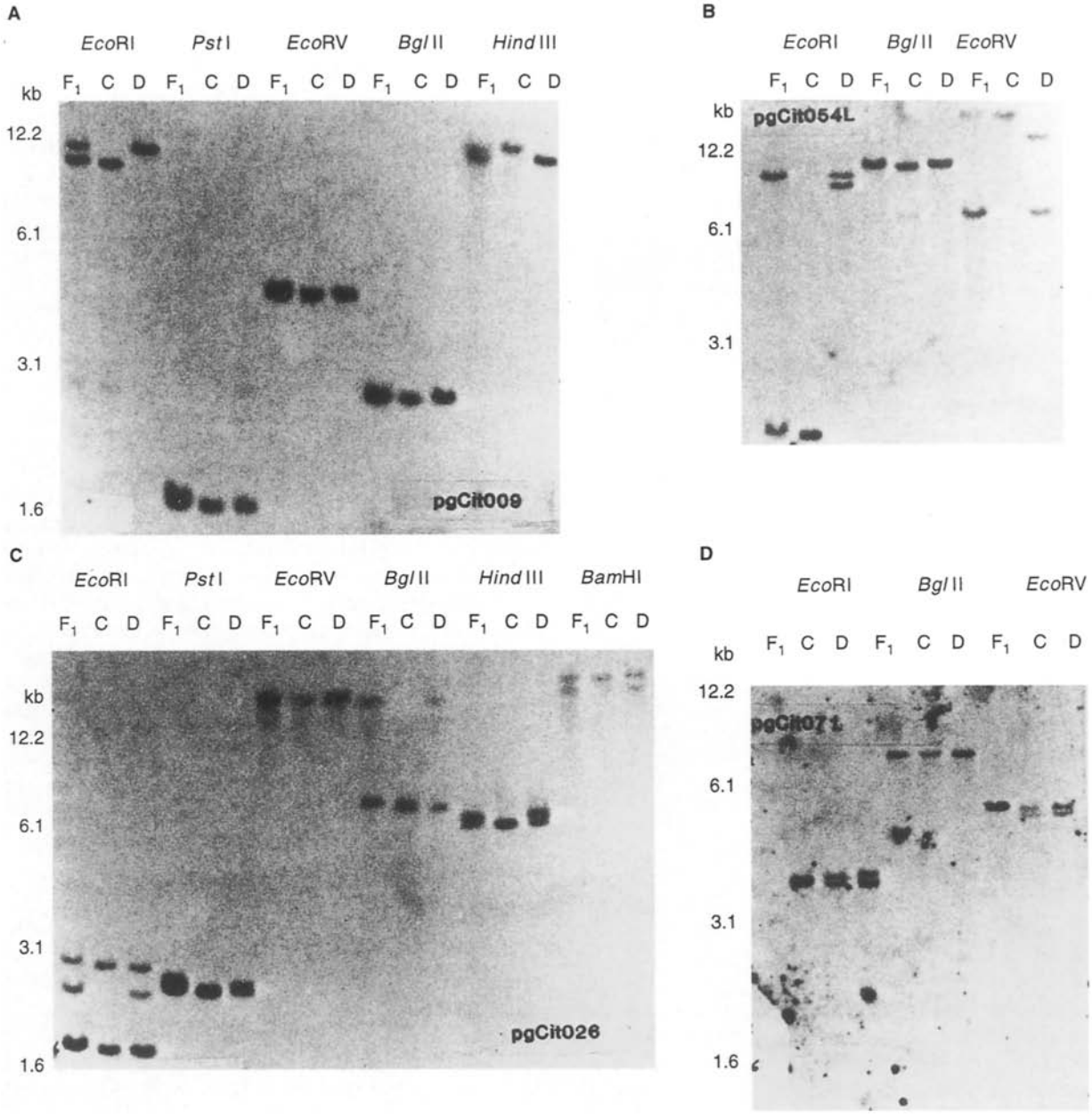

Fig. 3A–D Southern analysis of some clones revealed polymorphisms with more than one enzyme and different banding patterns with different enzymes. **A** pgCit009; **B** pgCit054L; **C** pgCit026; **D** pgCit071. In each set of three lanes (i.e., 1-2-3, 4-5-6, etc.) the DNA was digested with a different restriction enzyme. The *first lane* in each set of three=the F_1 LB 1-21 (F_1), the *second lane*='Clementine' (*C*), and the third lane='Duncan' (*D*). Molecular weights in kb are indicated

*Hin*dIII and *Bgl*II), two greater than 0.70 (*Eco*RI and *Pst*I), and with only *Bam*HI less than 0.60. These high R^2 values indicate that the probability of different enzymes detecting polymorphism with a given probe is not independent, and therefore insertion/deletion is a major mechanism for generating RFLPs in *Citrus*. The variation among R^2 values may be related to different frequencies of restriction sites among enzymes; large numbers of restriction sites increase the frequency of the restriction-pattern changes involved in the insertion/deletion events, and vice versa.

McCouch et al. (1988) presented a model comparing the expected restriction fragment patterns that would be generated by enzymes with different restriction sites and polymorphism due to a base substitution versus an insertion event. Their model presumed that two different enzymes had restriction sites flanking the insertion, but at different locations. The result with both enzymes was that, when the insertion was present, a larger restriction fragment was produced than when it was not, but fragment sizes varied between the two enzymes because they cut the DNA at different locations. Thus, fragment sizes varied with the restriction fragment used, but fragment patterns (i.e., a larger fragment being produced if an insertion was present) re-

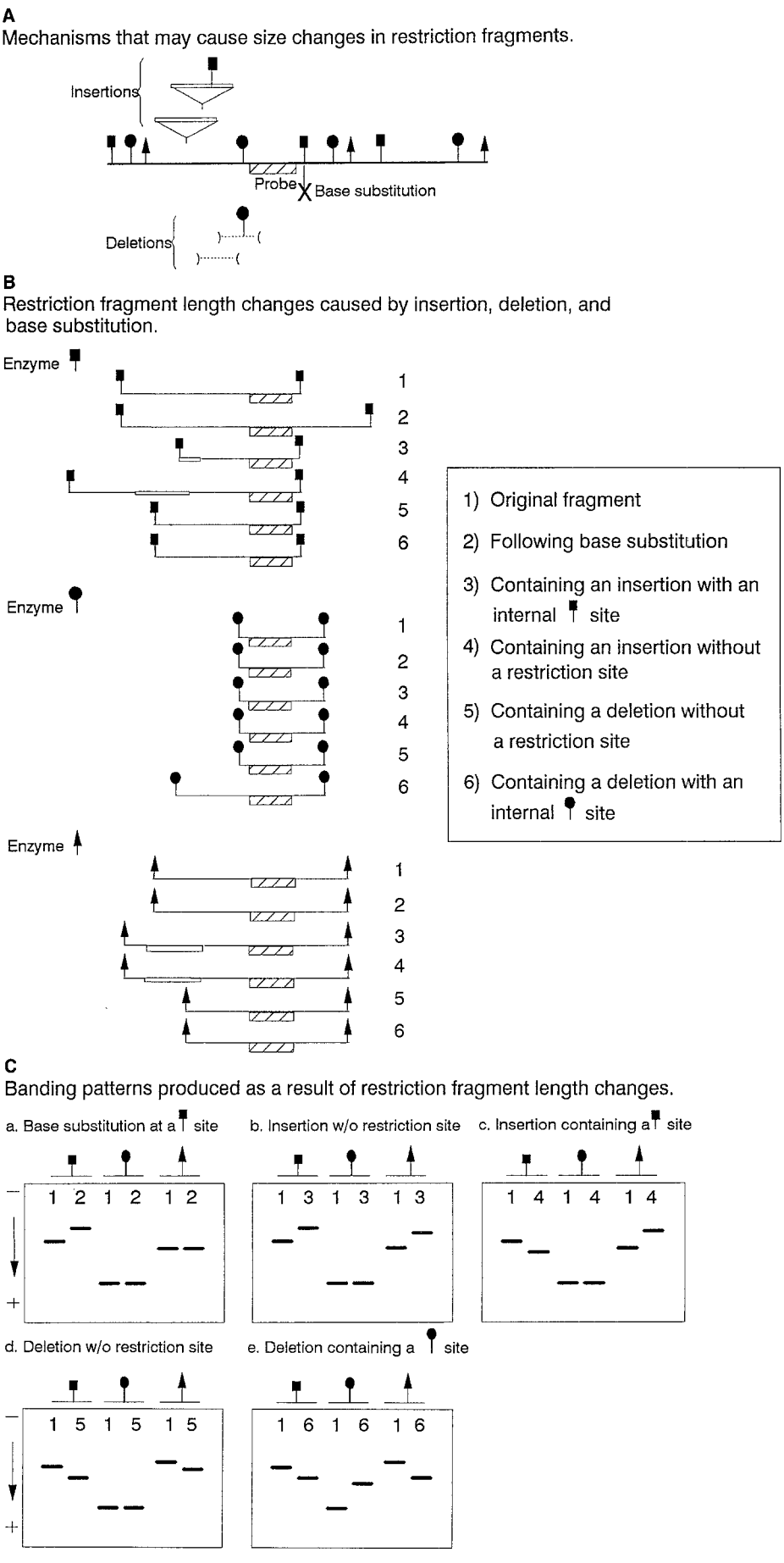

Fig. 4A–C Model depicting the expected restriction fragments generated by different mechanisms causing variability (base substitution, insertion, or deletion) and different restriction enzymes. As in the model of McCouch et al. (1988), a polymorphism caused by a base substitution is likely to detected with only one restriction enzyme, while polymorphisms due to insertions/deletions are apparent after digestion with more than one enzyme. In the present model, however, an insertion/deletion encompassing a restriction site can cause a change in restriction fragment length so that the fragment may be larger or smaller than the original fragment

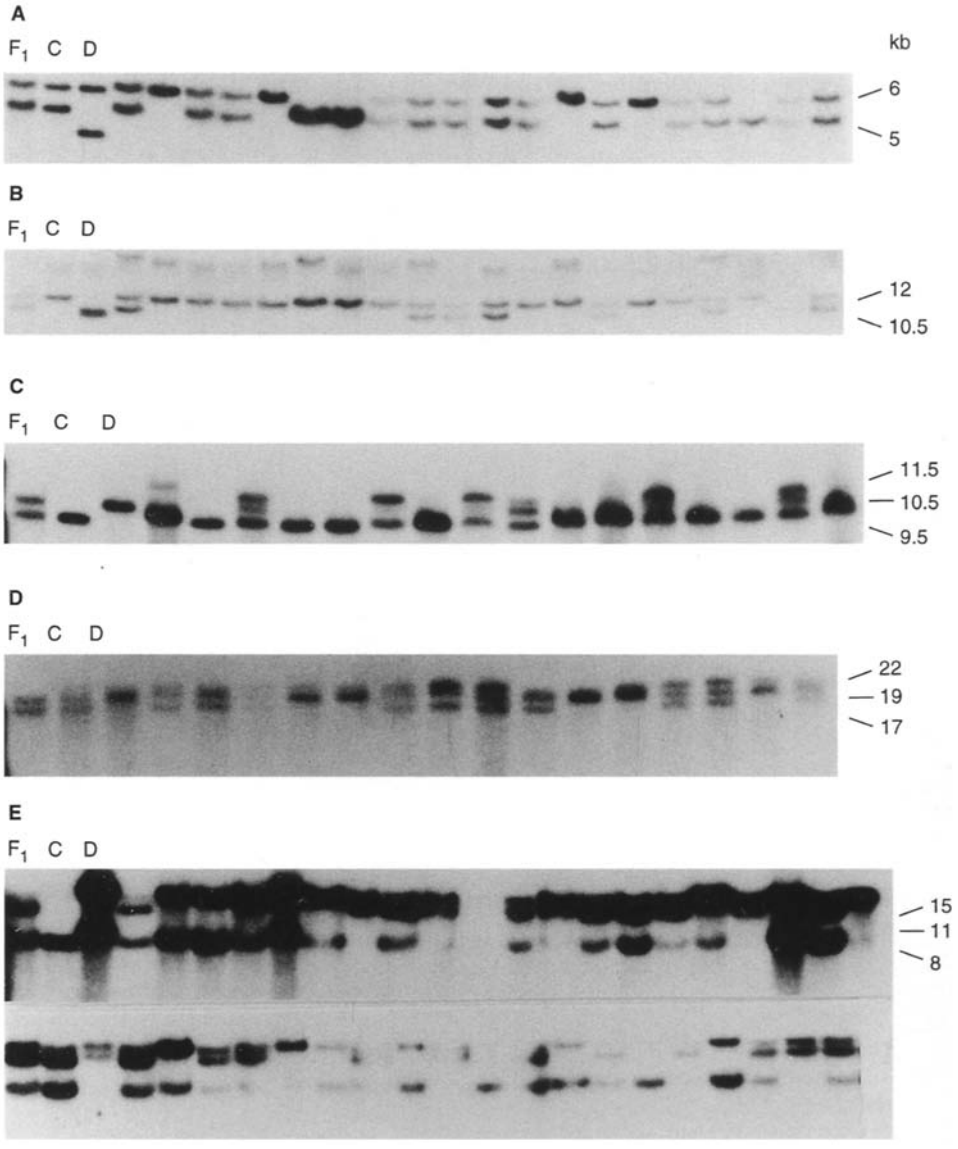

Fig. 5 Southern analysis of parent trees and part of the progeny population. When the filters were probed with **A** pgCit035 and **B** pgCit053, the progeny displayed bands segregating as expected from the parental phenotypes. On filters probed with **C** pgCit009, **D** pgCit046, and **E** pgCit048, the progeny banding patterns were not as predicted from parental phenotypes. On all filters, *lane 1* =the F_1 hybrid LB 1-21 (*F_1*), *lane 2*='Clementine' (*C*), *lane 3*= 'Duncan' (*D*), the remaining lanes=segregating BC_1 progeny of LB 1-21×'Clementine'. The 11.5-, 22-, and 15-kb bands revealed by pgCit009, pgCit046, and pgCit048 respectively are suggested to be analogous to the alleles designated a in the models depicted in Fig. 6, and the 10.5-, 19-, and 8-kb bands are analogous to the alleles designated a* in these models

mained the same. In our study, only one of 32 clones displayed this consistency of pattern. The other clones that revealed polymorphism with more than one enzyme yielded different banding patterns with different enzymes (Fig. 3) and were not satisfactorily explained by this model. An extended model based on that of McCouch et al. (1988) to explain the changes in banding patterns observed with many of the clones in this study is presented in Fig. 4. According to the model, both insertions and deletions may cause either increases or decreases of the restriction fragment length if there is a critical restriction site on the inserted or deleted fragment.

Many unusual examples of polymorphism can be explained by this model. For example, when clone pgCit009 was used to probe the genomic DNA of 'Clementine', 'Duncan', and their F_1 hybrid LB 1-21 following restriction with five different enzymes, varying banding patterns were revealed (Fig. 3A). The F_1 hybrid appeared to be heterozygous after digestion with both *Eco*RI and *Hin*dIII, but homozygous when the other three enzymes were used. However, in the *Eco*RI digestion, the hybridizing fragment obtained from the 'Duncan' parent was larger than that obtained from 'Clementine', while the opposite was true following *Hin*dIII digestion. These patterns could be caused by an insertion containing either an *Eco*RI or *Hin*dIII restriction site. This is illustrated by the model in Fig. 4Cc, where enzymes ↑ and ↑ could be *Eco*RI and *Hin*dIII, and enzyme ↑ could represent *Pst*I, *Eco*RV, or *Bgl*II. In a second example, the RFLPs detected by clone pgCit071 (Fig. 3D) could be caused by an insertion containing an *Eco*RI or *Eco*RV site, resulting in changes of fragment sizes among individuals following restriction with these two enzymes. In this case, the F_1 appears to be homozygous following restriction with all three enzymes, while both parents display heterozygous *Eco*RI and *Eco*RV restriction patterns but appear to be homozygous when restricted with

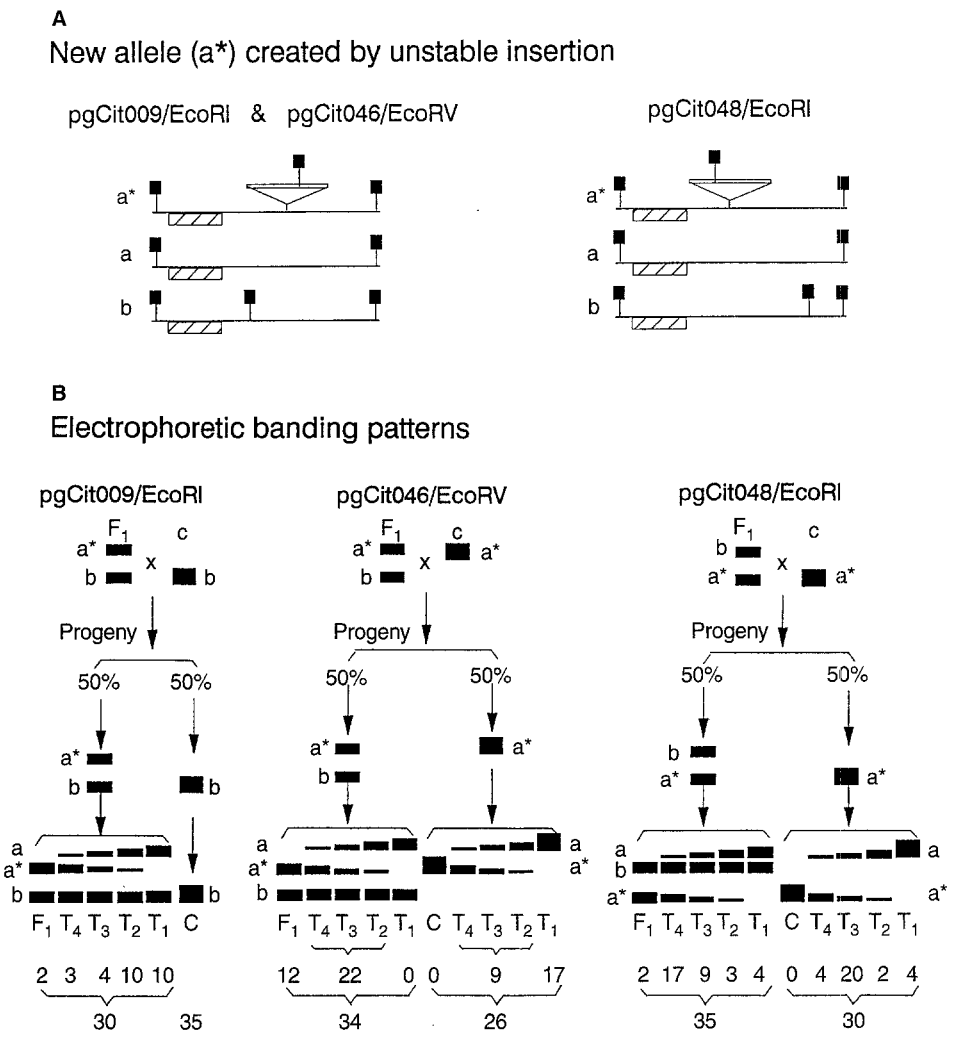

Fig. 6A, B Schematic models depicting the restriction sites and expected banding patterns proposed to be caused by transposon activity. **A** illustrates how a new allele a* might be created from allele a by an unstable insertion in three cases detected when progeny filters were probed with pgCit009/*Eco*RI, pgCit046/*Eco*RV, and PgCit048/*EcoRI*. Pertinent restriction sites are represented by ↑ and b is the other allele assumed to be present. **B** diagrams the electrophoretic banding patterns produced in each situation, with migration of DNA fragments in the gel shown from top to bottom. In each case, when the heterozygous F_1 LB 1-21 is crossed to the homozygous male 'Clementine' (*C*), the progeny are expected to segregate for the parental banding patterns in a 1:1 ratio; however, patterns consisting of multiple bands were obtained (see Fig. 5). In the models, F_1 and *C* designate the parental banding patterns and *T* represents patterns obtained following excision of a transposon. *T1* to *T4* represent excisions that occured very early in plant development (*T1*), so that almost all of the plant cells contain the a allele, to increasingly later times (*T4*), where tissues with mixtures of a and a* alleles are found. The *numbers* below the diagram are the actual numbers of progeny plants showing the various banding patterns in Southern analysis

*Bgl*II. Here enzymes ↑ and ↑ in Fig. 4Cc could represent *Eco*RI and *Eco*RV in this real situation, and enzyme ↑could represent *Bgl*II.

More than half of the clones where polymorphism was revealed by more than one enzyme (17/32, or 53%) could easily be explained by this model because in these cases only two or three alleles appeared to be involved. The patterns produced with the other 15 clones could also be explained with this model, but the explanations become more complex because three or more alleles appeared to be present among the three *Citrus* genotypes. In total, at least 18 clones appeared to recognize loci where more than two alleles were present in the banding patterns of the three plants surveyed. Thus, the frequency of multiple mutations, inferred from the frequency of loci with more than two alleles, was estimated as 33% (18/55), which is greater than the 5–10% observed in humans and *Drosophila* (Hudson 1989). *Citrus* plants are highly heterozygous and genetically unstable (Soost and Cameron 1975). A high frequency of insertion/deletions may be basic to this heterozygosity and genetic instability.

Unusual banding patterns suggestive of transposon activity

During segregation analysis of the backcross-progeny population of 65 seedlings from LB 1-21×'Clementine', some restriction-fragment banding patterns were observed that suggested that some insertions or deletions in *Citrus* may be associated with transposon activity. Three clones (pgCit009, pgCit046, and pgCit048) were found to produce banding patterns that did not follow expected Mendelian segregation. Filters of progeny trees probed with these clones displayed extra bands that were not expected based on the parental genotypes; 43% (28/65), 80% (48/60), and 97% (63/65) of the progeny had extra bands as revealed by pgCit009/*Eco*RI, PgCit046/*Eco*RV, and pgCit048/*Eco*RI, respectively (Fig. 5).

Several possible explanations were considered for these unexpected results. First, the unusual bands might have resulted from incomplete digestion of the progeny DNA; but this seems unlikely because the extra bands in different plants were of equivalent size when the same clone was

used. Second, the unusual restriction patterns might have been the result of methylation, especially if plants were under stress or in different developmental stages; but all plants were of the same age and apparently healthy. Third, the extra bands might have been caused by two overlapping loci. However, this possibility was not supported by the observed banding patterns, where an increase in intensity in one band from an individual occurred concurrently with a decrease in intensity in another band; this was the most striking aspect of the banding patterns of these clones. A hypothesis of transposon activity, depicted schematically in Fig. 6, was developed to explain this phenomenon.

The model assumes that the insertion in question is reversible and can excise independently in different somatic cells. The time during somatic-tissue development at which the excision occurs determines the banding patterns obtained from different individuals. An extremely early excision in one cell that gives rise to the entire sampled tissue would lead to complete disappearance of one of the bands expected from the parental phenotypes (band a* in Fig. 6) and would create a new band (band a). No excision, or one extremely late in development, would preserve the expected banding pattern and create no new band. Reversions that occurred between these two extreme situations would produce three bands (a*, a, and b in Fig. 6) with different dosage effects between two of them (a* and a) depending on the time of excision. The time when a reversion occurred can be estimated by studying the band dosages and patterns. The frequencies of progeny plants with extra bands actually detected by clones pgCit009/*Eco*RI, 046/*Eco*RV, and 048/*Eco*RI were quite different, implying that transposable activities were different in these three cases (Figs. 5, 6). Furthermore, clones pgCit009 and pgCit048 detected presumed transposable activity after DNA digestion with the same enzyme (*Eco*RI), but the insert sizes inferred from the data and the model (i.e., the differences in size between a and a*) were quite different in the two cases, implying that more than one transposon could be present in *Citrus*.

When the progeny were divided into groups based on their presumed "normal" genotypes (without excision, where only parental alleles appear) (Fig. 6), chi-square tests showed a good fit to the expected segregation ratio for all three loci. These results support the proposed hypothesis of transposable insertion and excision. Further, two of these three clones (pgCit009 and 048) have also been used in a mapping study with an intergeneric *Citrus* progeny population (Durham et al. 1992). In this population, no extra bands were revealed and inheritance was as predicted by Mendelian segregation. The different behavior of these markers with different progeny implies that the occurrence of presumed extra alleles in different populations is genotype dependent, and probe independent. This is also consistent with the transposon phenomenon and further supports the hypothesis.

Another question must be addressed here: why were these proposed transposable insertions not excised in the tissues of the parent trees that were collected for DNA isolation? Transposons may be autonomous or non-autonomous for insertion/excision activity, such as the well studied Ac-Ds system in maize (Fedoroff 1983). It is possible that some transposons in *Citrus* are non-autonomous, so that the transposon activity is genotype dependent and an activating factor is needed for transposition. Alleles caused from insertions are stable in genomes without activating factors. Hake et al. (1989) reported that the Ds insertion in maize could create banding patterns similar to those shown by clones pgCit009, pgCit046, and pgCit048 when the Ac element was introduced into the genome as the activator. In that study, progeny whose genome did not contain the activator showed only parental alleles in Southern analysis, and only some of the plants with Ac introduced into their genomes showed extra alleles. These results suggested that the existence of an activator inside the genome is necessary for the reversional excision of the non-autonomous element; however, excision may not occur even if the activator is present.

In conclusion, with the analysis of the nature of *Citrus* RFLPs, we have initiated a molecular characterization of an important fruit crop. The *Citrus* genome shares similarities with the genomes of tomato and rice. However, there is a very high degree of heterozygosity as evidenced by the extent of polymorphism observed in this and other studies (Durham et al. 1992; Jarrell et al. 1992; Cai et al. 1994). This polymorphism results primarily from insertions and/or deletions, and there are data suggesting that some of this variability may be due to the presence of transposons in the *Citrus* genome.

Acknowledgements This research was supported in part by USDA NRICGP Plant Genome grant 92-37300-7591 and a grant from the FDACS Citrus Production Research Advisory Council. Florida Agricultural Experiment Station Journal Series No. R-04592.

References

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90:7980–7984

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Beltz GA, Jacobs KA, Eickbush TH, Cherbas PT, Kafatos FC (1983) Isolation of multigene families and determination of homologies by filter hybridization methods. Methods Enzymol 100:266–285

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, colinearity and complementarity. Trends Genet 9:259–261

Bernatzky R, Tanksley SD (1986) Towards a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887–898

Burr B, Evola SV, Burr FA (1983) The application of restriction fragment length polymorphism to plant breeding. In: Stelow JK, Hollaender A (eds) Genetic engineering: principles and methods, vol 5. Plenum Press, New York London, pp 45–59

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

Cai Q, Guy CL, Moore GA (1994) Extension of the linkage map in *Citrus* using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. Theor Appl Genet 89:606–614

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Durham RE, Liou PC, Gmitter Jr FG, Moore GA (1992) Linkage of restriction fragment length polymorphisms and isozymes in *Citrus*. Theor Appl Genet 84:39–48

Fedoroff NV (1983) Controlling elements in maize. In: Shapiro JA (ed) Mobile genetic elements. Academic Press, New York, pp 1–58

Guerra, M dos S (1984) Cytogenetics of Rutaceae. II. Nuclear DNA content. Caryologia 37:219–226

Hake S, Vollbrecht E, Freeling M (1989) Cloning Knotted, the dominant morphological mutant in maize using Ds2 as a transposon tag. EMBO J 8:15–22

Havey MJ, Muehlbauer FJ (1989) Linkages between restriction fragment length, isozyme,and morphological markers in lentil. Theor Appl Genet 77:395–401

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

Hudson RR (1989) How often are polymorphic restriction sites due to a single mutation? Theor Popul Biol 36:23–33

Jarrell DC, Roose ML, Traugh SN, Kupper RS (1992) A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. Theor Appl Genet 84:49–56

Kesseli RV, Paran I, Michelmore RW (1994) Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. Genetics 136:1435–1446

Landry BS, Kesseli RV, Farrara B, Michelmore RW (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331–337

Liou P-C (1990) A molecular study of the citrus genome through restriction fragment length polymorphism and isozyme mapping. PhD dissertion, University of Florida, Gainesville

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

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

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Messeguer R, Ganal MW, Steffens JC, Tanksley SD (1991) Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. Plant Mol Biol 16:753–770

Miller JC, Tanksley SD (1990) Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato. Theor Appl Genet 80:385–389

Moore G, Gale MD, Kurata N, Flavell RB (1993) Molecular analysis of small grain cereal genomes: current status and prospects. Biotechnology 11:584–589

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89:1477–1481

Roose ML (1988) Isozymes and DNA restriction fragment length polymorphisms in citrus breeding and systematics. In: Goren R, Mendel K (eds) Proc 6th Int *Citrus* Cong. Margraf Scientific Books, Weikersheim, pp 57–67

Soost RK, Cameron JW (1975) Citrus. In: Janick J, Moore JN (eds) Advances in fruit breeding. Purdue University, West Lafayette, Indiana, pp 507–540

Suiter KA, Wendel JF, Case JS (1983) LINKAGE-1: A PASCAL computer program for the detection and analysis of genetic linkage. J Hered 74:203–204

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257–264

Torres AM (1983) Fruit trees. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, part B. Elsevier, Amsterdam, pp 401–421

Torres AM, Mau-Lastovicka T, Williams TE, Soost RK (1985) Segregation distortion and linkage of *Citrus* and *Poncirus* isozyme genes. J Hered 76:289–294

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

Wu K-S, Tanksley SD (1993) PFGE analysis of the rice genome: estimation of fragment sizes, organization of repetitive sequences and relationships between genetic and physical distances. Plant Mol Biol 23:243–254

Zamir D, Tanksley SD (1988) Tomato genome is comprised largely of fast-evolving, low-copy number sequences. Mol Gen Genet 213:254–261