Theor Appl Genet (1990) 80:385–389

Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato

J. C. Miller* and S. D. Tanksley

Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA

Received February 13, 1990; Accepted May 3, 1990
Communicated by F. Salamini

Summary. Since the construction and use of RFLP genetic maps depends on the ability of cloned sequences to detect polymorphism, we have attempted to determine conditions under which maximum levels of polymorphism can be detected. Forty cloned nuclear sequences from three different libraries (cDNA, EcoRI genomic, and Pstl genomic) were hybridized to total DNA from 149 plants representing eight species of the tomato genus *Lycopersicon*. Five different restriction enzymes were employed in this study. We examined the relationship between polymorphism (number of restriction patterns) and clone size, restriction enzyme, size of hybridizing restriction fragments, and clone source (library). We found no relationship between clone size (ranging from 0.4 to 5.3 kb) and polymorphism. There was a strong positive relationship ($r^2 = 0.79$) between polymorphism and the average size of the fragments produced by each restriction enzyme. cDNA clones hybridized to larger fragments compared to genomic clones. cDNAs also detected significantly more polymorphism (approximately 25% more) than genomic clones – possibly indicating high levels of sequence variability in introns and/or areas flanking coding regions.

Key words: RFLP – Tomato – Probe source – Probe length – Restriction enzymes

Introduction

Genetic maps based on RFLP markers have been constructed for a variety of plant species (for review, see Helentjaris and Burr 1989; Tanksley et al. 1989). But if a given clone does not detect allelic differences (i.e., RFLPs) in the population or cross being examined, it cannot be mapped. In addition, the utility of an RFLP-based map, once constructed, continues to depend on the ability of the clones to detect polymorphism. In most studies, it would be desirable to detect higher levels of polymorphism. Unfortunately, little is known about the circumstances determining levels of detected polymorphism in plants.

We recently completed an RFLP-based phylogenetic study of the tomato genus *Lycopersicon* (Miller and Tanksley 1990). Fragments generated by five different restriction enzymes were examined at 40 loci. Such a large data set seemed ideally suited to test factors that might be related to the detection of restriction fragment length polymorphism. The objective of this study was to determine whether the ability to detect polymorphism (defined as number of different restriction patterns) is significantly related to the following parameters: (1) restriction enzyme, (2) average genomic fragment size detected, (3) clone source (library), and (4) clone size.

Materials and methods

Plant DNA

Total DNA was extracted from leaves of 149 *Lycopersicon* plants according to procedures described by Bernatzky and Tanksley (1986a). A total of 53 different accessions were assayed, chosen to represent the geographic diversity within each species. The species were as follows: *L. esculentum* Mill. (23 accessions, 2 plants of each), *L. pimpinellifolium* (Jusl.) Mill. (10 accessions, 2 plants), *L. cheesmanii* Riley (4 accessions, 2 plants), *L. parviflorum* Rick, Kes., Fob., & Holle (1 accession, 4 plants). *L. chmielewskii* Rick, Kres., Fob., & Holle (4 accessions, 4 plants), and *L. pennellii* (Corr.) D'Arcy, *L. hirsutum* Humb. & Bonpl. (4 accessions, 5 plants), and *L. peruvianum* (L.) Mill. (3

* Present address: Department of Agronomy, University of Wisconsin, Madison, WI 53706, USA

accessions of each, 5 plants). Fewer plants per accession were examined from the self-compatible (and therefore less variable) species than the self-incompatible species. All accessions are identical to those used in Miller and Tanksley (1990), with the exception of *L. peruvanium* accession LA2150, which was excluded from the present study because of questions about its taxonomic classification.

Clones

Plant DNA was digested with EcoRI, HindIII, DraI, EcoRV, and XbaI (Bethesda Research Laboratories and New England Biolabs), and the resulting fragments were then separated by electrophoresis on 0.9% agarose gels, and then blotted onto Genescreen Plus hybridization membrane as per Bernatzky and Tanksley (1986b). The filters were probed with radiolabelled insert (Feinberg and Vogelstein 1983) from each of 40 different nuclear clones. These clones were from three different libraries cloned from DNA of *L. esculentum* cv VF36. Fifteen were single-copy cDNA clones (0.4–1.25 kb) derived from leaf RNA (Bernatzky and Tanksley 1986c), 10 were single cope EcoRI random genomic clones (1.5–5.3 kb size range), and 15 were single-copy PstI (1.5–3.0 kb size range) random genomic clones (Miller and Tanksley 1990). CD72 and CD77 were not scored for XbaI because the high molecular weight fragments were not well separated, and were not included in the statistical analyses that required complete data sets. Filters were hybridized at 65 °C (5 × SSC), and washed to 1 × SSC stringency, 65 °C, and autoradiographed as described by Bernatzky and Tanksley (1986b).

Calculations

The fragments detected by each clone were assigned molecular weights by comparing them to lambda-HindIII and ΦX174-HaeIII standards included on each gel. The fragments seen in each individual for each probe-enzyme combination were analyzed with dBaseIII software by Ashton-Tate on an IBM PC/AT computer. Specifically, the following parameters were tabulated for each probe-enzyme combination: (1) the number of different-sized genomic fragments detected over all 149 plants, (2) polymorphism, defined as the number of different restriction fragment patterns, or combinations of fragments, found across the 149 plants, and (3) the average size of the hybridizing fragments. In this study, "polymorphism" refers to variation observed among individuals with respect to restriction patterns.

The method whereby statistics were generated is exemplified by clone TG45. Across all 149 plants, TG45 detected six different-sized fragments on EcoRV-cut DNA, and these fragments occurred in six different combinations or restriction patterns. The number of different genomic fragments found on XbaI-cut DNA was also six, but these six fragments occurred in eight different restriction patterns. The average size of these XbaI fragments for that probe was 7.8 kb. In this manner, parameters were estimated based on fragment data pooled from 149 plants across all probes.

Statistical analyses were done on an Apple MacIntosh II with Statview 512$^+$ by Brainpower, Inc. One-way ANOVAs were used to test for significant differences between various enzymes or probe types (libraries) with regard to the previously defined parameters. Regressions were used to determine the significance of the relationships between polymorphism and average size of fragments detected.

Results and discussion

The majority of the clones detected between 7 and 35 different restriction patterns per enzyme, with the exception of CD17 which detected 60 patterns/enzyme (Fig. 1). CD17 proved to be unusual, and was dropped from the analyses because it was an outlier in every test (Fig. 1). This probe alone hybridized to 116 discrete fragments and produced 261 different restriction patterns across the five enzyme, while the other clones each detected from 22 to 83 fragments and from 26 to 147 restriction patterns. Although the probe detects several genomic fragments in each individual, these fragments cosegregate, and CD17 appears to map to a single genetic locus.

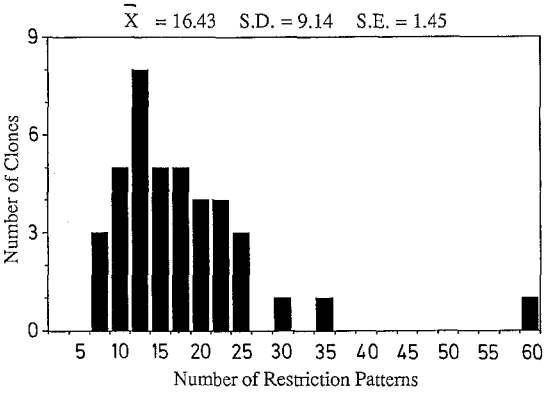

Fig. 1. Number of restriction patterns (per enzyme) detected by 40 clones probed across eight *Lycopersicon* species (149 plants total)

Comparisons among restriction enzymes

The five restriction enzymes used in this study fell into three different groups with respect to the average genomic fragment sizes generated (Table 1). DraI produced the smallest fragments (av 2.8 kb), EcoRI and HindIII were intermediate (6.6 and 6.3 kb), and EcoRV and XbaI had the highest values (8.6 and 8.9 kb). For all enzymes, the average size of the different genomic fragments was significantly higher than expected based on a GC content of 39% for tomato nuclear DNA as determined by HPLC (Messeguer et al. 1990; Table 1). DraI, EcoRI, and HindIII gave mean fragment sizes more than double the expected, and EcoRV and XbaI nearly triple. These deviations may be due to nonrandom distribution of nucleotides, and the relative frequences of di- or trinucleotide combinations may play a large role in determining actual frequencies of restriction sites.

Since differences between enzymes were found with respect to polymorphism and average fragment size, we explored the relationship between these two variables and found them to be positively correlated ($r^2 = 0.791$, slope $= -0.581$, SE $= 0.172$) (Fig. 2). Thus, the average size of the genomic fragments generated by different enzymes is positively correlated with the amount of polymorphism detected on DNA with that enzyme. Such a

Table 1. Averages by enzyme for number of different restriction patterns, number of different fragments detected across the genus, and average size of these fragments, averaged across 39 probes

Restriction enzyme	No. restr. patt.		No. fragments		Fragment length (kb)		
	Mean[a]	SE	Mean	SE	Exp.[b]	Obs.	SE
*Dra*I	8.587a	0.778	8.026	0.653	1.242	2.77a	0.113
*Eco*RI	11.008ab	1.394	8.897	0.725	3.039	6.647b	0.244
*Hin*dIII	9.174ab	0.918	8.359	0.665	3.039	6.300b	0.242
*Eco*RV	11.341ab	1.041	9.615	0.673	3.039	8.586c	0.279
*Xba*I	12.455b	1.614	9.486	0.601	3.039	8.901c	0.300

[a] Values followed by the same letter (a, b, c) were not found to be significantly different at the 0.05% level according to Fisher Protected LSD (Snedecor and Cochran 1980)

[b] Based on GC content of 39% for tomato nuclear DNA (Messeguer et al. 1990) and random ordering of nucleotides

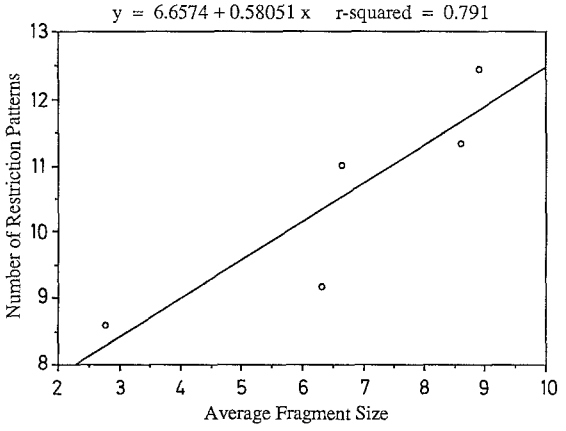

Fig. 2. Regression of average molecular weight of fragments detected on DNA cut with five different enzymes against the average number of restriction fragment patterns detected on that DNA (data presented in Table 1)

Table 2. Averages by library for number of different restriction fragment patterns, number of different fragments detected across the genus, and average size of these fragments, averaged across the probe-enzyme combinations within each library

Library	No. unique restr. patt. detected		No. unique genomic fragments detected		Av length hybridizing genomic frag. (kb)	
	Mean[a]	SE	Mean	SE	Mean	SE
cDNA	12.527a	1.185	9.309	0.546	7.468a	0.545
*Pst*I	9.867b	0.521	9.067	0.410	6.213ab	0.436
*Eco*RI	8.464b	0.814	7.980	0.621	5.457b	0.491

[a] Values followed by the same letter (a, b, c) were not found to be significantly different at the 0.05% level according to Fisher Protected LSD (Snedecor and Cochran 1980)

result would be expected if polymorphisms are generated by insertions/deletions (or other DNA rearrangements). The larger the fragment size, the more likely that it will encompass an insertion/deletion. A similar correlation between fragment size and polymorphism has also been reported for rice (McCouch et al. 1988). Moreover, the enzymes that gave large fragments in tomato and hence detected larger amounts of polymorphism (EcoRV and XbaI) are the same enzymes that yield larger fragments and greater polymorphisms in rice (McCouch et al. 1988). These results suggest that selective use of enzymes generating larger fragments can result in increased detection of polymorphism. In this study, the enzyme producing the largest fragments (XbaI) detected almost 50% more unique restriction patterns than the enzyme producing the smallest fragments (DraI).

Comparison among libraries

Despite the fact that the cDNA clones on average were shorter (0.4–1.3 kb) than the genomic clones (1.5–5.3 kb), they were found to hybridize to genomic fragments larger (on average) than those hybridizing to the genomic clones (Table 2). The average size of fragments detected by cDNAs was 7.5 kb versus 6.2 kb and 5.5 kb, respectively, for PstI and EcoRI genomic clones. Differential GC content (between coding and non-coding regions) may account for some or all of this difference between the cDNA clones and the genomic clones. Because of requirements of codon usage, coding regions are likely to have a higher GC content than the rest of the genome. A recent survey of coding and spacer regions in the tomato genome bears out this assertion (Messeguer et al. 1990). All of the restriction enzymes used in this study have a high proportion of A/T in their recognition sites. The higher GC content in coding regions of the genome (and possibly different use of di- and trinucleotides) makes it less likely that recognition sites for these enzymes will occur in coding regions, thus resulting in larger restriction fragments.

cDNA clones detect more polymorphism. Coding regions are more conserved than other sequences in the tomato

nuclear genome (Zamir and Tanksley 1988). Nonetheless, cDNA clones detected more polymorphism than either PstI or EcoRI genomic clones (Table 2). For example, the cDNA clones tested in this study detected approximately 25% more unique restriction patterns than clones from the PstI library and almost 50% more than clones from the EcoRI library. Similar observations regarding differences between cDNA and genomic clones have also been made in studies with lettuce (Landry et al. 1987) and lentils (Havey and Muehlbauer 1989). Because of the conserved nature of coding regions, it seems likely that the majority of the polymorphisms detected with cDNA clones are occurring either outside of the coding region (5′ or 3′) or in the introns. Data presented in this report, however, do not allow us to discriminate between these possibilities.

EcoRI versus PstI clones. The genomic clones came from two different libraries: those constructed with PstI and those made with EcoRI-cut DNA. PstI recognizes the sequence 5′-CTGGA/G-3′, but will not cleave if the cytosine at the 5′ end is methylated (Nelson and McClelland 1987). Ninety-two percent of the nuclear clones in this library were found to be single-copy (Miller and Tanksley 1990). EcoRI will cut in methylated as well as unmethylated regions, and only 33% of this library was single-copy.

There is much evidence indicating that coding sequences or regions upstream from coding regions are often hypomethylated (Bird 1987). If hypomethylated regions in tomato do contain coding sequences, then it is reasonable to expect that the PstI library would also be enriched for gene sequences. The fact that statistics for this library place it intermediate between the cDNA and EcoRI libraries with respect to average fragment size and restriction patterns detected suggests that, while the PstI library may be enriched for transcribed regions, it does not represent the identical pool of sequences as the cDNA library and, thus, must contain additional single-copy nuclear sequences (Table 2). Preliminary results from Northern blots also indicate that the PstI library is qualitatively different from a cDNA library and does not represent highly expressed sequences (R. Messeguer and S.D. Tanksley, unpublished data).

Effect of insert size. No significant association was found between insert size, in the range of 0.4 to 5.3 kb, and polymorphism. Insert size was found to be negatively associated with average fragment size across the eight species ($r^2 = 0.156$, slope $= -0.711$, SE $= 0.279$). This is probably due to the cDNA clones, which have already been established as being smaller yet hybridize to fragments much larger than themselves. Removal of these clones from the data set annulled the significance of the relationship.

It is possible that the range in sizes of clones used in this study was not sufficient to detect a relationship between probe length and polymorphism. If nucleotide substitution or insertion/deletion were sources of RFLPs, one would expect that longer probes would detect more polymorphism since they would be able to detect insertions/deletions over a larger section of the chromosome as well as to assay more restriction sites for base substitution. However, in order to significantly increase the portion of the genome assayed by a clone, it would be necessary for the clone to be significantly larger than the average size of the hybridizing genomic fragments generated by the restriction enzyme used. In this study, the average clone size (<2 kb) is much smaller than the average size of the hybridizing genomic fragment size generated by all enzymes (6.5 kb).

With regard to clone size, it should be pointed out that longer probes do provide the benefit of requiring shorter exposure times during autoradiography, which is one of the more time-consuming aspects of RFLP mapping. This is especially relevant to cDNA clones. While these clones, on average, detect more polymorphism than genomic clones, they are also shorter. As a result, the exposure time with autoradiography had to be significantly increased in order to obtain a comparable signal to the genomic clones (on average an extra 1–2 days of exposure compared to genomic clones). cDNA clones are limited in length by the size of gene transcripts, whereas genomic clones are only limited in size by the length of single copy sequences in the genome. Tomato has a large proportion of single copy DNA, and stretches of 10–20 kb or more of single-cope DNA are not uncommon (Zamir and Tanksley 1988; A.H. Paterson and S.D. Tanksley, unpublished results).

Conclusions

Detecting sufficient levels of polymorphism is an important issue in the use of RFLP maps in plant breeding and genetics. This is especially true for self-pollinated crops like the cultivated tomato, which tend to harbor less polymorphism at the DNA level than their outcrossing counterparts (Miller and Tanksley 1990; Helentjaris et al. 1985). In this study we have addressed some of the variables potentially affecting the amount of polymorphism detected by RFLP probes in tomato.

With respect to restriction enzymes, we report that there is a positive relationship between the average genomic fragment size produced with different enzymes and the level of polymorphism detected. In other words, enzymes that cut less frequently tend to detect more polymorphism in the tomato genome. This finding is consistent with a similar finding in rice and supports the idea that insertions/deletions or other DNA rearrangements

account for a significant amount of RFLP variation in plants (McCouch et al. 1988). Since tomato and rice represent divergent extremes in flowering plants (dicots versus monocots), these results may be applicable to other plant taxa.

We also conclude that cDNA clones detect significantly more variation than random genomic clones (i.e., EcoRI clones) or genomic clones selected on the basis of hypomethylation (i.e., PstI clones). That cDNA clones and PstI clones behave differently in this regard indicates that PstI clones are not coincidental with coding regions even though coding regions (and surrounding regions) may also be undermethylated (Bird 1987). cDNA clones detect significantly larger genomic restriction fragments, which may reflect different nucleotide composition (e.g., greater GC content), which result in less frequent restriction sites for the enzymes employed in this study.

Finally, we found no correlation between the size of the probe (ranging from 400 to 5,300 bp) and the amount of polymorphism detected by each clone. While longer probes should theoretically detect more polymorphism, since they cover larger segments of the genome, this effect should not be observed until the clone size exceeds the size of the average genomic fragment generated by the restriction enzymes being utilized. In this study, the longest clone (5,300 bp) was shorter than the average size of the genomic fragments (6,500 bp) hybridizing to the clones surveyed.

Acknowledgements. This work was supported by the United States Department of Agriculture CRGO, the National Science Foundation, and the Binational Agricultural Research and Development fund (BARD no. US-388-87) to SDT. Thanks to Dr. A. Paterson for help with statistical analyses and to Drs. S. McCouch, M. Bonierbale, and G. Martin for advice and comments and to Steffie David for typing the manuscript.

References

Bernatzky R, Tanksley SD (1986a) Genetics of actin-related sequences in tomato. Theor Appl Genet 72:314–321

Bernatzky R, Tanksley SD (1986b) Methods for detection of single or low copy sequences in tomato on southern blots. Plant Mol Biol Rep 4:37–41

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

Bird A (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

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

Helentjaris T, Burr B (eds) (1989) Development and application of molecular markers to problems in plant genetics. Curr Commun Mol Biol, Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegman S (1985) Restriction fragment polymorphisms as probes for plant diversity, and their development as tools for applied plant breeding. Plant Mol Biol 5:109–118

Landry BS, Kesseli R, Leung H, Michelmore RW (1987) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (*Lactuca sativa*, L.). Theor Appl Genet 74:646–653

McCouch S, 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 M, Steffens J, Tanksley S (1990) Characterization of the level, target sites, and inheritance of cytosine methylation in tomato nuclear DNA. (Submitted)

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet (in press)

Nelson M, McClelland M (1987) The effect of site-specific methylation on restriction-modification enzymes. Nucleic Acids Res 15:219–230

Snedecor GW, Cochran WG (1980) Statistical Methods. Iowa State University Press, Ames/IA

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

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