

CR1 – a dispersed repeated element associated with the *Cab-1* locus in tomato

R. Bernatzky,^{1,2} E. Pichersky,^{3,4} V. S. Malik³ and S. D. Tanksley^{1*}

¹Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853, USA (*author for correspondence); ²Present address: Department of Plant Sciences, University of Massachusetts, Amherst, MA 01003, USA; ³Laboratory of Cell Biology, The Rockefeller University, New York, NY 10021, USA; ⁴Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA

Received 21 October 1987; accepted in revised form 20 January 1988

Key words: repeated DNA, tomato, genetic mapping, transposition, chlorophyll a/b-binding polypeptide

Abstract

Cab-1 is a complex genetic locus in tomato consisting of four clustered genes encoding chlorophyll a/b-binding polypeptide. Southern blot analysis of total tomato DNA with genomic clones corresponding to the *Cab-1* locus has revealed the presence of a repetitive element in the 3 kb spacer regions between two of these genes. This repetitive element, named CR1, has been characterized via sequencing, genetic mapping and hybridization to related solanaceous species. Results indicate that there are as many as 30 copies of this element in the tomato genome and that most, if not all, are found at independent loci. Sites corresponding to 12 of the repeats have been located on different regions of chromosomes 2, 4, 5, 7, 10 and 11. A 1.6 kb *Pst*I-*Eco*RI fragment from the *Cab-1* locus containing the element was sequenced and found to be 75% AT-rich. No open reading frames larger than 150 bp were detected. Several imperfect inverted repeats flanked by direct repeats could be found at the ends of the element. This arrangement is reminiscent of known transposons. Southern hybridization analysis indicates that multiple copies of CR1 exist in all species of the genus *Lycopersicon* as well as in *Solanum lycopersicoides* and *S. tuberosum* (potato), but not in eggplant, pepper, petunia, *Datura* or tobacco. Melt-off experiments indicate that members of the CR1 family in the tomato genome are more closely related to one another than to homologous members in the genomes of *S. lycopersicoides* or *S. tuberosum*, suggesting some type of concerted evolution.

Introduction

It is well established that most of the DNA of eukaryotes is noncoding and often repetitive. The repetitive DNA of animals ranges from highly repeated tandem arrays (satellite DNA) to highly repeated interspersed sequences such as the *Alu* [18] and L1 (*Kpn*) [1] families. Tandemly repeated DNA is thought to fluctuate in amount primarily through the process of unequal exchange. In contrast, inter-

spersed families are believed to be transcribed and their amplification and dispersion may occur through an RNA intermediate [30].

In plants, much attention has been given to the study of highly repeated sequences which may occur as satellite DNA localized to heterochromatic areas, such as centromeres, telomeres or knobs [27, 11, 19], or interspersed throughout the chromosomes [14]. The study of these highly repeated sequences in related species has not only advanced our under-

standing of the evolution of this class of DNA but also of the species which contain those sequences [22, 34, 2, 20, 13, 11, 9, 21].

Unlike the case with highly repeated DNA, little information is available about the organization of low copy repeat sequences in plants. It is not known if members of this class of DNA exist in tandem copies or whether they follow rules of evolutionary change similar to those of highly repeated sequences. Because of their low copy number it is difficult to study their chromosomal distribution by *in situ* hybridization. Sequence divergence and mechanisms of turnover have not been established for this class and their taxonomic distribution has not been studied.

Recently, we reported results from studies of the organization of nuclear genes encoding the chlorophyll a/b-binding (CAB) polypeptides of photosystem II in the thylakoid membranes of chloroplasts [28, 35]. The PSII CAB polypeptides are encoded by approximately 10 genes organized in 5 distinct chromosomal loci. Two of these loci, *Cab-1* and *Cab-3*, are complex, each possessing at least four CAB genes [28, 35]. We have now examined the intergenic spacer regions in these two complex loci. While most spacer regions are single copy, we have discovered a repetitive element in the *Cab-1* locus which is a member of an interspersed, repeat family referred to as the CR1 family. The nucleotide sequence, genomic organization and phylogenetic distribution of the CR1 family is described in this paper.

Methods

Isolation of DNA and Southern blot analysis. DNA was isolated by a modified procedure of Murray and Thompson [25] as reported in Bernatzky and Tanksley [4]. The nylon membrane employed for Southern analysis was either Zetabind (AMF Cuono) or GeneScreen Plus (NEN). Blotting and hybridization were according to Bernatzky and Tanksley [5]. Hybridization solutions contained 5% high- M_r dextran sulfate and $5\times$ SSC and were carried out at 68 °C. Final washes were at 68 °C and contained $1\times$ SSC and 0.1% SDS.

Preparation and screening of *L. esculentum* genom-

ic library. Lambda clone preparation and isolation of sequences from *L. esculentum* (cv. T5) that were homologous to CAB cDNA fragments are reported elsewhere [28]. Fragments were subcloned into either pUC 9 or pUC 18. Plasmid DNA was isolated according to Birnboim and Doly [7].

Genomic reconstructions. Slot blot procedures to estimate copy number were according to Rivin [29]. Hybridization conditions were the same as for the Southern blots.

Genetic mapping. DNA fragments that were visualized by Southern analysis were mapped by the method of restriction fragment length polymorphism in an F2 progeny from a cross between *L. esculentum* (cv. VF36) and *L. pennellii* as outlined by Bernatzky and Tanksley [6]. Linkage analysis was performed with 200 previously mapped DNA markers and 18 isozyme markers [6, 33].

DNA sequencing. DNA sequencing was according to Maxam and Gilbert [24].

Results

Location of repeated sequences in relation to the Cab-1 and Cab-3 gene clusters

In order to determine the complexity of the non-coding regions in the *Cab-1* and *Cab-3* gene clusters, fragments of the lambda clones were subcloned and used as probes against Southern blots of total tomato DNA digested with *EcoRI*. These blots were washed at low stringency ($1\times$ SSC, 68 °C). Figure 1 illustrates the hybridization pattern obtained for some of these subclones. The majority of the DNA that surrounds the coding sequences was found to be unique (single copy). However, one segment from the *Cab-3* cluster and two from the *Cab-1* cluster were determined to contain repeated sequences.

The *Cab-3* repeat (8A0.9) is located within a 900 bp *SstI-EcoRI* fragment that also contained a truncated Cab coding sequence ([28], Fig. 1). This repeat is 2.2 kb 3' to the *Cab-3* gene cluster. As a probe this sequence produced a smear of hybridizing

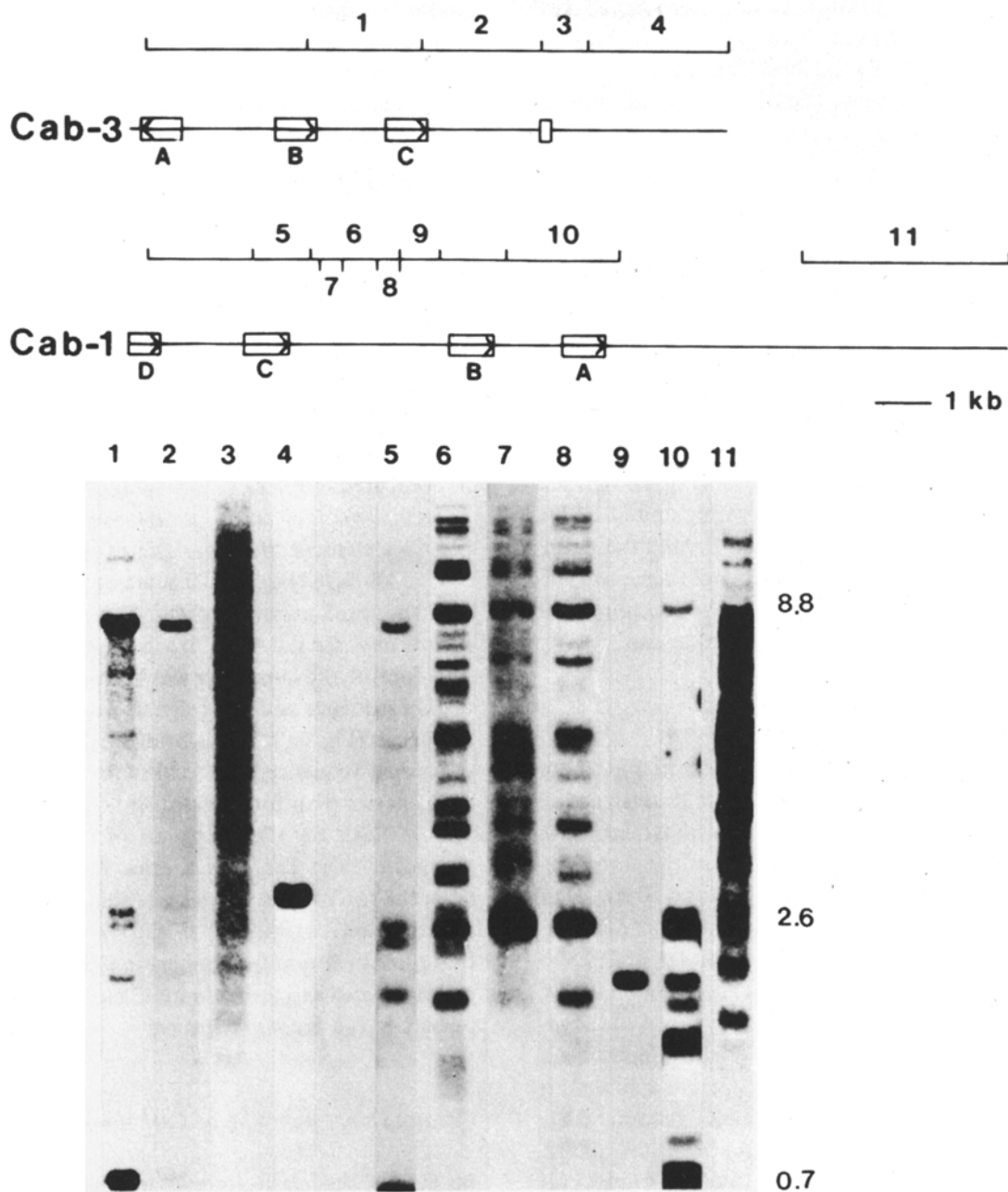


Fig. 1. Regions of the *Cab-1* and *Cab-3* loci investigated by Southern analysis. Boxes indicate coding regions and arrows, direction of transcription [28]. Lines above the loci refer to sequences tested on Southern blots. Sample autoradiographs are shown below for those regions marked with numbers. Each lane contained 3 μ g total tomato DNA digested with *Eco*RI and separated on 0.9% agarose. Fragment sizes in kb are shown at the right.

fragments showing very few discrete bands.

The hybridization patterns were much more discrete for the two repeat regions in or around the *Cab-1* locus. A moderately repeated sequence (3B4) was

found on a 3.8 kb *Eco*RI fragment that is approximately 3.5 kb 3' to the first coding sequence of the gene cluster. This clone hybridized to >50 fragments on a Southern blot. The second *Cab-1* repeat,

termed CR1, is confined to a 1.65 kb *EcoRI-PstI* fragment located in the 3 kb spacer region between the second (*Cab-1B*) and third (*Cab-1C*) genes. Upon probing restricted genomic DNA, CR1 was found to hybridize to 20–30 fragments (Fig. 1).

The CR1 repeat was chosen for further study because of the discrete hybridization pattern observed on genomic Southern blots and because it was the only repeat found in the intergenic spacer regions in the CAB gene clusters. In order to delimit the length of this repeat, subclones were made of both ends of the 1.65 kb fragment and used as probes on Southern blots. A 270 bp *DraI-EcoRI* fragment at the 3'-end as well as a 200 bp *Sau 3A* fragment that ends 40 bp from the 5'-end gave hybridization patterns that were similar to the entire CR1 clone (Fig. 1). This sets the minimum length of the repeating unit at 1.1 kb. Probing with flanking sequences beyond these subclones did not produce a discernable repeated pattern, indicating that the ends of the repeating unit may lie within these 5'- and 3'-fragments.

CR1 copy number estimation

The copy number of the CR1 sequence in the tomato genome was estimated by slot blot reconstruction analysis, under low stringency conditions (final wash at $1\times$ SSC 65°C) [29]. Quantities of the cloned sequence that correspond to 1, 10, 30 and 60 copies of the sequence along with a standard amount of tomato DNA were applied to a slot blot and hybridized to isolated CR1 insert. Hybridization signals were produced that correspond to 10–12 copies of the sequence in the tomato genome. A densitometric scan of the *EcoRI* Southern blot also suggests a minimum of 10 copies since the 2.6 kb *EcoRI* genomic fragment on which CR1 resides comprised 10% of the total signal (Fig. 1). However, a total of 30 *EcoRI* restriction fragments were found to hybridize with varying intensity on genomic Southern blots, suggesting that there may be as many as 30 related copies of the sequence in the genome hybridizing with different degrees of efficiency (Fig. 1).

CR1 in other solanaceous taxa

The location of a CR1 repeat within a cluster of a

conserved and universal plant gene (CAB) raises the possibility that the repeat sequence might itself be conserved. We therefore investigated the occurrence of the CR1 repeat in other related taxa. Southern blots were made with DNA from several varieties of cultivated *L. esculentum* as well as a number of accessions of each of the wild species of tomato. In all cases the repeat was found to be present in a similar number of *EcoRI* fragments (see representative examples in Fig. 2A). Other genera of the Solanaceae family were also analyzed. One accession each of *Solanum lycopersicoides*, *S. tuberosum*, *S. melongena*, *Capsicum annuum*, *C. chinense*, *Petunia hybrida*, *Datura innoxia* and *Nicotiana tabacum* were tested. Six μg of DNA was loaded for these genera as compared to 1.5 μg for the tomato species. Only *S. lycopersicoides* and *S. tuberosum* hybridized to the CR1 element after washing at 68°C with $1\times$ SSC, 0.1% SDS (Fig. 2B). It was noted with interest that the third member of the *Solanum* genus, *S. melongena*, did not hybridize to the CR1 fragment. As a control, the same filter was hybridized to a CAB coding sequence and good signals were obtained for all species (Fig. 2C). The hybridization signals from the repeat sequence were weaker for the two *Solanum* species than for tomato. At higher stringency ($0.2\times$ SSC, 68°C), nearly all of the signal was washed off the *Solanum* species while the signal from the two tomato species was similar to that observed at lower stringency (Fig. 2D). This indicates that the repeat family members in tomato are more closely related as a group to the CR1 than any of the repeats in the *Solanum* species.

Chromosomal mapping of CR1 insertion sites

Because of the relatively small number of hybridizing fragments on Southern blots it was possible to explore the genomic organization of this repeated family by the segregation of restriction fragments. CR1 fragments were mapped relative to DNA and isozyme markers of known position on the 12 chromosomes of tomato. An F2 population from an interspecific cross of *L. esculentum* and *L. pennellii* had been previously used for such mapping and was segregating for more than 200 molecular markers [6, 33]. The following restriction enzymes were chosen

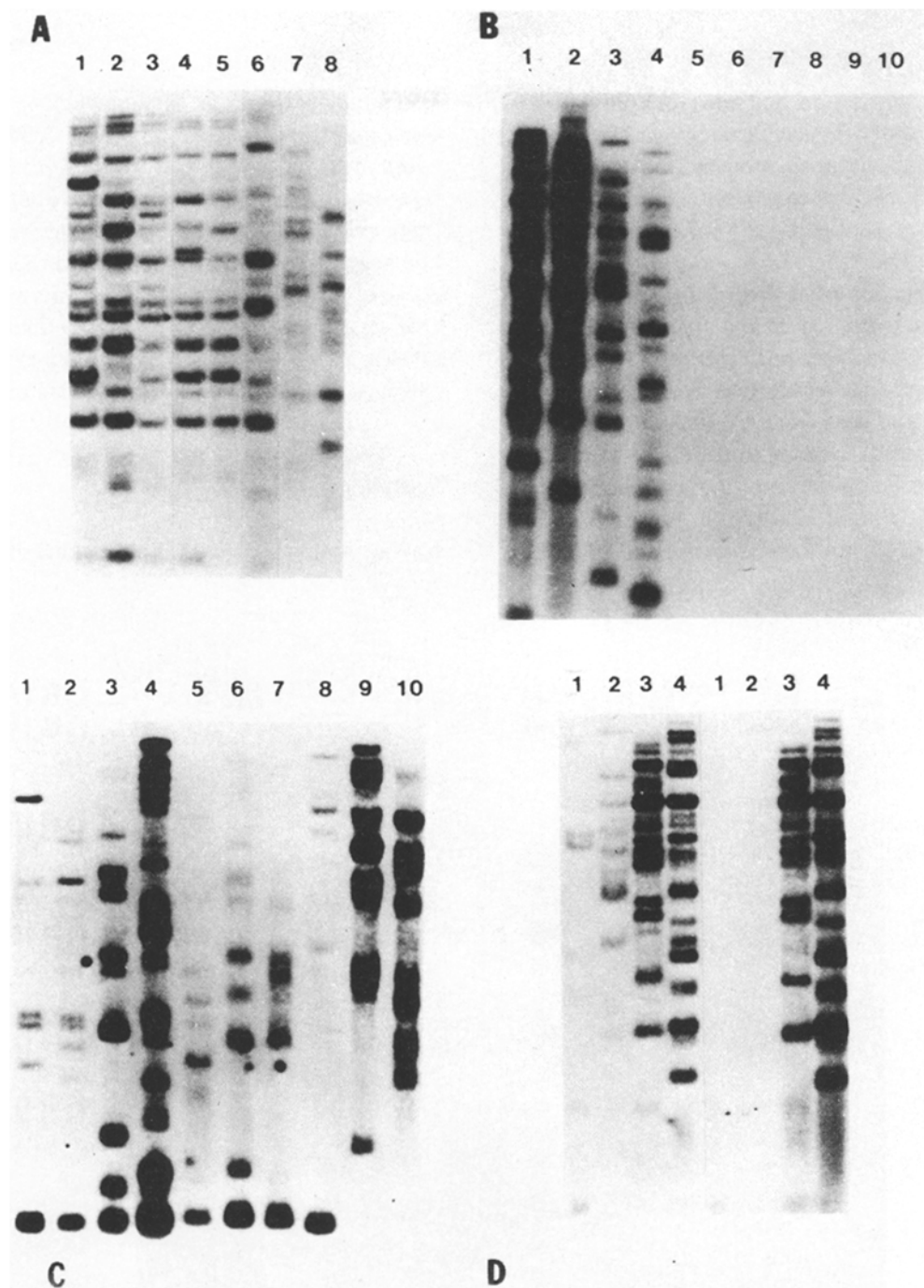


Fig. 2. Southern analysis of tomato cultivars and other species of the Solanaceae. DNA from each was digested with *Eco*RI and probed with CR1.

(A) *L. esculentum* cultivars include 1) VF36, 2) New Yorker, 3) Spring Set VF. Wild accessions are 4) *L. esculentum* var. *cerasiforme* LA1511, 5) *L. esculentum* var. *cerasiforme* LA1455, 6) *L. pimpinellifolium* LA1579, 7) *L. pennellii* LA1940 and 8) *L. peruvianum* LA371. Each lane contained 3 μ g DNA.

(B) 1) *L. esculentum* cv. VF36, 2) *L. pennellii* LA716, 3) *Solanum lycopersicoides*, 4) *S. tuberosum* (potato) cv. Katadin, 5) *S. melongena* (eggplant) cv. Black Beauty, 6) *Capsicum annuum* (pepper) CA133, 7) *C. chinense* CA4, 8) *Datura innoxia*, 9) *Petunia hybrida* cv. Mitchell, 10) *Nicotiana tabacum*. Lanes 1 and 2 contained 1.5 μ g DNA. The autoradiograph is overexposed for the tomato species.

(C) Same filter as in (B) but stripped of the CR1 probe and hybridized to a CAB coding sequence.

(D) Lanes 1 and 2 are 6 μ g of *S. tuberosum* and *S. lycopersicoides* respectively, and lanes 3 and 4 are 1.5 μ g of *L. pennellii* and *L. esculentum*. Each was digested with *Eco*RI and probed with CR1. At the left the filter was washed at a final stringency of 1 \times SSC, 0.1% SDS at 68 $^{\circ}$ C and at the right the final wash was at 0.2 \times SSC, 0.1% SDS at 68 $^{\circ}$ C.

to identify restriction fragment length polymorphisms necessary for genetic mapping: *DraI*, *BstNI*, *PstI* and *MspI*. Forty-six progeny DNA were digested with each of these enzymes and probed with the CR1 insert. A sample of the *DraI*-digested progeny is shown in Fig. 3.

The segregation of 16 fragments generated from the 4 enzymes were scored and analyzed for linkage with the previously mapped markers. Seven of the mapped fragments were derived from the *L. esculentum* parent and nine from *L. pennellii*. Five of the sixteen fragments mapped to the *Cab-1* locus while twelve others mapped to positions on chromosomes 2, 4, 5, 6, 7, 10 and 11 (Fig. 4). Two pairs of loci (*CRIC* and *CRID* on chromosome 5 and *CRIE* and

CRIF on chromosome 7) mapped close to one another and each showed only one recombinant between them. Two other loci, *CRIB* (*DraI* *L. esculentum* fragment) and *CRIM* (*BstNI* *L. pennellii* fragment) are possibly allelic since they map close to one another on chromosome 11 and there were no progeny plants in which both fragments were absent. Overall the genetic analysis indicates that members of this repeat family are distributed on different chromosomes throughout the genome of tomato.

Sequence analysis of CR1

Having established the dispersed nature of this small

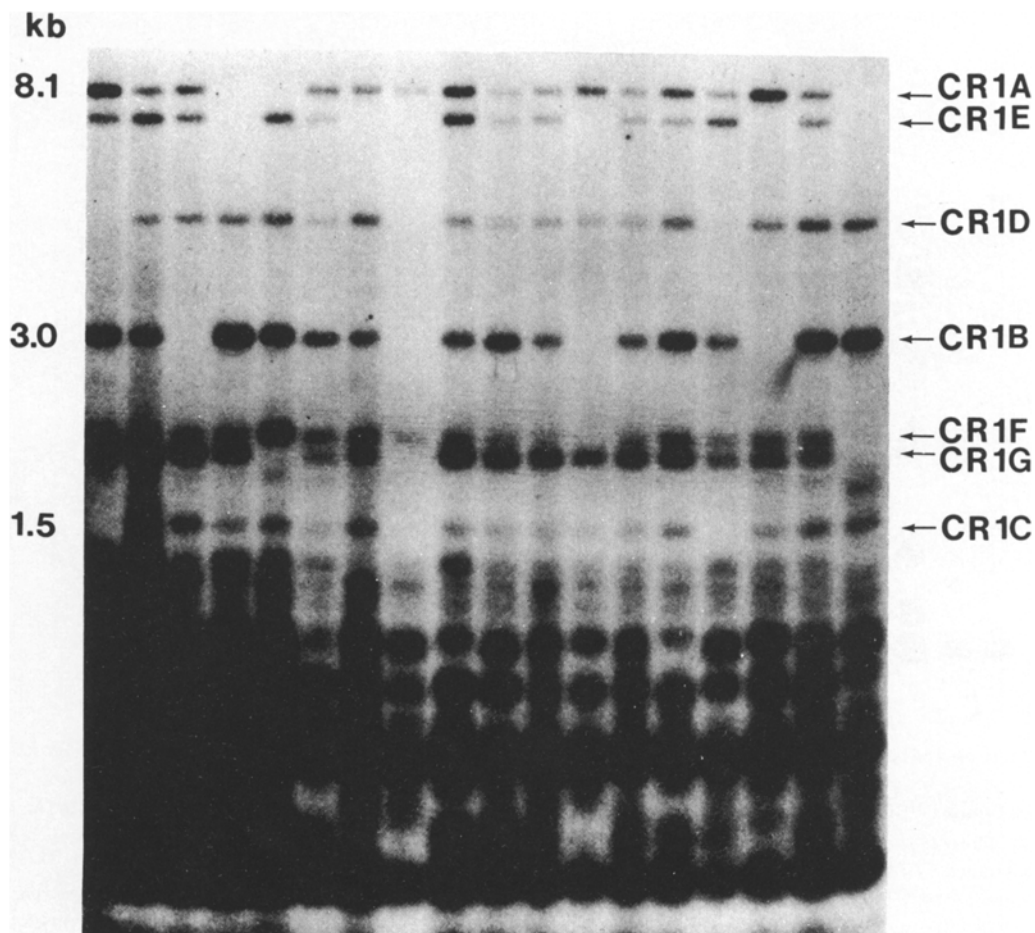


Fig. 3. Sample of F2 progeny from a cross between *L. esculentum* and *L. pennellii*. 3 μ g DNA from each was digested with *DraI*. CR1 related fragments that were segregating and that map to specific chromosomal locations are shown at the right. Fragment size scale is shown at the left.

repetitive family, the cloned CR1 fragment was sequenced to possibly gain insight into how the sequence has become distributed throughout the genome.

The sequence of the CR1 clone is shown in Fig. 5. There is no evidence that more than one copy of the repeat sequence is contained on the CR1 clone. Adenine and thymine nucleotides are in abundance, constituting 75% of the sequence. There is a direct duplication of 21 base pairs (90% homology) at nucleotides 471 and 492. This may have resulted from recombination within a misaligned *Cab-1* gene cluster during meiosis.

The sequence was analyzed for inverted repeats that are characteristic of the termini of transposable elements (see [26] for review). As previously discussed, the ends of the repeat unit are expected to lie within the terminal 5' and 3' subclones and the

```

10      20      30      40      50      60
CTGCAGTAA TTTTCGATTA TCACGTACAG TGGTGTAGAG TGATCATTAT GACTAATGTC

70      80      90      100     110     120
TCTTCATGAC CATGAGATTA ACTAAGAAAA GTCTATTTCAT GGATCTATGT TAAAAGAAGA
-----> A ----->
130     140     150     160     170     180
AAAAACGATG GTGGTACTG GCGCTAGTTC CTTTGGATTC TTTGAAAAAT GTCAACAAGT

190     200     210     220     230     240
GTGTTTGAA TTATCAGCAT AAACATGAAG GTCAAAACTT AAAAGTTGAA TTACCATTTT
-----> B -----> C ----->
250     260     270     280     290     300
AGCCTTATAC TAAATTAATA TATTATAAAT TAATTIAATTA ATTAATATAT AAATAATGCTT

310     320     330     340     350     360
AAATTTGTAA TTAAGTTATG ATGACTTGTG ATAACCATGG CATGAGGCAT AGCATATGAT
-----> D ----->
370     380     390     400     410     420
CACTTATTGT TAAAATAATG TAGGAGAATG AATAGTTCTA AAAATAATTA ATTATAAAGT

430     440     450     460     470     480
TATTACGAAA ACGTATTAGA GAAATTTTTC ATTTCICTAT CIGATCAGTT TTAGGTGTAA

490     500     510     520     530     540
TTTTGTCATA ATGTGTATTA ATACCACACC TCGACTATCT CATCATTAAAT ATTCACACCT

550     560     570     580     590     600
TCATAATCTT TCCACACTCT CAAGAAGTTG ATGGCACITC TTGGATCACT TAATTTGTAA

610     620     630     640     650     660
TGTTTTTATC TTCTATTAGC ATTTTCCTAT ATTATATATT ATACTTTCTC ATGTAAATCAA

```

```

670     680     690     700     710     720
GTTATATTTA TTTTTTTTAT ATGAGTTTGT TATGCTCAAT GTTTTATTTA TCTTATTTGA

730     740     750     760     770     780
ACTCATCTAA ATTTGTGAAT ATTTCAACTA CAATAAAATTT TTTAAATTTG ACITTTTTTTT

790     800     810     820     830     840
ATGAAGATAC ATTAAGAAAT GAGGTTAATA TTTCATAGTG ATGATGCAGT AAGTGAACCA

850     860     870     880     890     900
ACACAGATTG AACTGTTTCA AGTGGTATCA ATTTTCAGGT AATAAATCAT TTTTCCTTAT

910     920     930     940     950     960
ATGCAATTAA TTATTTTTGT GTAAGTTGAT ATGTATTATA TAAAATTTTG GTTTGATTTT

970     980     990     1000    1010    1020
TAAATAAAAA TTATTTTATA GAAGTTGAGA CGATACACTT AATTGGATTA TCCAAAAAAG

1030    1040    1050    1060    1070    1080
GAATCTTTTA CACAAGGACA AGAAGAAAGA CTTCGTGCTT ATCACTGCAT TTCATGACAA

1090    1100    1110    1120    1130    1140
TCAATGGAAT ATCTTACCAG AAAAAAGTT GATATAATTT TGTACAAGTG AATCCAACCTC

1150    1160    1170    1180    1190    1200
ACTAACAAAA AAATTAAGA AAGAAAAAAA ATGAACATGA AGAATAATAA AATCAAGATA
-----> DR ----->
1210    1220    1230    1240    1250    1260
AAGAGAATCA ATGGTGTCTT ATAAGTTGAT GAACCACTGT AACTTCATTA TTTTITGAAT
----->
1270    1280    1290    1300    1310    1320
TTACGAGGAA ATGAATTATT GATGAAATTT GTGAAATGTA CATATATACC TTTGGATCGG

1330    1340    1350    1360    1370    1380
TAACAAATAT GATTGAAAGT AGTATTTTTT CACGGTATAA TAGCTAATTC TTGATATATA

1390    1400    1410    1420    1430    1440
TATTGTTTAA ATTTATTATT TGCCTACAAA ACATTCTATA AAGTATTTTA GGATAAATGT

1450    1460    1470    1480    1490    1500
GCAACGCACG TTTTCGAAAA CTAGTGTGTG TATAAAATCC AAAAACCAAT ATCTTGACTT
-----<
1510    1520    1530    1540    1550    1560
AATGTGAATT AGAACTTGTG AAATAACTAC ATTACTCCAT ATGACCATAT GTATTATGAT
-----< B' -----<
1570    1580 D' 1590    1600    1610    1620
CTGACTCAAT CTGCTTATG CTCATACCAT AATTTTACA AGGGGATCA AAATATGAAA

1630    1640 A' 1650    C'
-----< A' -----< C' -----<
AAAAAATAAA TAGACCCATA AAGAAGCAAG GGGAAATTC

```

Fig. 5. Nucleotide sequence of the fragment that contains the CR1 repeat. This fragment was isolated from the *Cab-1* locus (region 6 of Fig. 1). A–D = sets of inverted repeats flanked by direct repeats found at ends of CR1 repeat element. DR = 21 bp direct repeat found in middle of sequence.

search for short repeats was confined to these areas (0–300 bp at the 5'-end and 1300–1658 bp at the 3'-end). Based on characteristics of known plant transposons, a minimum length was set at 9 bp and a requirement of at least 80% homology. At these criteria 66 potential inverted repeats were found in the areas examined. Inverted terminal repeats of plant transposable elements are also known to be flanked by shorter direct repeats of target DNA. Considering these additional criteria only 4 of the potential inverted repeats were found to be flanked by direct perfect repeats of 3–6 bp (Fig. 5). It is possible then that one of these pairs of terminal repeats represents the remnants of a transposable element. However, no open reading frames longer than 150 bp were found within the CR1 sequence, suggesting that if this were a transposable element it has been silenced and that its sequence has accumulated many base changes.

Discussion

We report here the isolation and characterization in tomato of a low copy repeat sequence, CR1. One copy of this element is located between two of the major chlorophyll *a/b*-binding coding sequences in the *Cab-1* gene cluster on chromosome 2. Reconstruction experiments and Southern analysis suggest that there are as many as 30 copies of this repeat in the genome of tomato. There are no additional members of the CR1 repeat family associated with other CAB coding sequences of the *Cab-1* or *Cab-3* clusters based on hybridization of the repeat with lambda clones that contain these clusters. The other 11 mapped members are also not tightly linked with the other CAB loci – *Cab-2*, *Cab-4*, or *Cab-5*. Although three of the repeat members were on chromosome 7, along with *Cab-4*, they all showed recombination with the *Cab-4* locus.

The CR1 element is not unique to the genus *Lycopersicon*. This sequence hybridized with similar complexity to *Solanum lycopersicoides* and *S. tuberosum* (potato). These *Solanum* species are closely related to the *Lycopersicon* genus and so the presence of the repeat in these species is not surprising. The absence of this sequence from *Capsicum*

(pepper) is noteworthy since a complex locus orthologous to the tomato *Cab-1* locus is present in pepper and has been genetically mapped (Tanksley *et al.*, in preparation). This locus forms part of a conserved linkage block of markers that exhibit the same linear order as tomato. This evidence suggests, but does not prove, that the CR1 repeat has transposed into the *Cab-1* gene cluster some time after the divergence of tomato and pepper.

The *Solanum* species hybridized with less intensity to the CR1 repeat than did tomato. However, this difference in intensity is hard to evaluate. The Southern blots contained four times as much DNA for the *Solanum* species than for tomato but little is known about the genome size of *S. lycopersicoides* and *S. tuberosum*. Nonetheless, it is significant that under more stringent wash conditions (0.2× SSC, 68°C) the probe was removed from the DNA of *Solanum* species but was still hybridized to DNA of the tomato species (Fig. 2B). This indicates that the majority of the CR1 family members in tomato are more closely related to one another than they are to any of the sequences in the *Solanum* species. If these sequences were present before the separation of these genera from a common ancestor, then divergence of the sequence within a lineage should be as great as between lineages. Therefore some form of concerted evolution is acting on this family of repeats. It is possible that these sequences have continued to be duplicated and deleted since these species separated. Such a process would tend to make the sequence within a plant lineage more similar.

An interesting result from the hybridization of this repeat to other genera of the Solanaceae family is the lack of hybridization to *Solanum melongena* (eggplant). This result was also observed for the other two CAB-associated repeat sequences, 8A0.9 and 3B4. These two repeats also hybridized to *S. lycopersicoides* and *S. tuberosum* but did not hybridize to *S. melongena* or to the other species tested (data not shown). The CR1, 9A0.9 and 3B4 repeats are not related since they do not cross hybridize to one another. Therefore, despite their presence in or around CAB loci, they are independent repeats. The hybridization experiments using each of these repeats all suggest tomato is more closely related to potato than either are to eggplant. It would be interesting to test

other *Solanum* species with these and other non-coding DNA sequences to determine if they can provide useful information on the relationships within this large genus.

The duplication and dispersion of repeated sequences in plants has been known for some time [14] but the mechanisms for such events are not well understood. A model of unequal exchange is generally applied to the amplification of tandemly repeated sequences. This is not a likely mechanism for the CR1 repeats since it is not tandemly repeated at the *Cab-1* locus. The majority of other *EcoRI* fragments observed on Southern blots do not exceed the intensity of the fragment that was cloned and therefore other integration sites about the genome probably also contain only a single copy of the repeat. A mechanism of duplication without unequal exchange could result entirely from a process of replicative transposition or transposition followed by chromosome assortment. If the sizes of the transposed segments are small then the duplications might not interfere with normal chromosome pairing. As discussed in Shapira and Finnerty [31] this type of process has likely produced duplications in yeast and fungi.

The data presented in this paper suggest that the CR1 sequence has been transposed into the *Cab-1* locus relatively recently. While it is possible that CR1 is an active transposable element, the nucleotide sequence does not provide strong evidence to support this. Neither strand shows significant open reading frames and the size of the element (ca. 1.5 kb) is shorter than most autonomous transposons that have been characterized in plants [26]. There are, however, inverted repeats (>80% homology) with short direct flanking repeats at the end of the element that are characteristic of transposable elements. Whether any of these correspond to the true endpoints of the repeat can be determined only when additional independent CR1 repeats are sequenced and compared with the one reported in this paper.

Another class of movable elements is represented by the highly repeated and dispersed sequences of animals and have been termed retrotransposons [30]. These sequences are transcribed into RNA which is proposed to be involved with their transposition. At present we do not know if the CR1 se-

quence is transcribed. However, the CR1 sequence data do not fit the consensus sequence of poly A sequences at the 3'-ends of such elements. The CR1 sequence contains A-rich sequences; however, they are not confined to one end of the fragment.

The pattern of sequence dispersion found for the CR1 family is similar to that of duplicated coding sequences previously found in plants. A number of isozyme-coding genes have been found to be duplicated on different chromosomes in diploid plant species [16, 17, 32]. In tomato, a genetic study of a number of cDNA clones has revealed that approximately 40% of all coding regions are duplicated and that many of the duplicate copies are on different chromosomes [6]. These duplications generally involve relatively short sequences since closely linked, flanking markers are usually not duplicated. It is possible that the process that led to the duplication and dispersal of the CR1 repeat family is also responsible for dispersed gene duplications in tomato and other plant species.

Acknowledgements

Research supported by NSF grant DCB-8517773 to SDT. EP was a recipient of an NIH post-doctoral fellowship. Thanks to Steffie David for typing the manuscript.

References

1. Adams JW, Kaufman RE, Kretschmer PJ, Harrison M, Nienhuis AW: A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. *Nucl Acids Res* 8: 6113–6128 (1980).
2. Arnold ML, Appels R, Shaw DD: The heterochromatin of grasshoppers from the *Caledia captiva* species complex. I. Sequence evolution and conservation in a highly repeated DNA family. *Mol Biol Evol* 3: 29–43 (1986).
3. Barnes SR, James AM, Jamieson G: The organisation, nucleotide sequence, and chromosomal distribution of a satellite DNA from *Allium cepa*. *Chromosoma* 92: 185–192 (1985).
4. Bernatzky R, Tanksley SD: Genetics of actin-related sequences of tomato. *Theor Appl Genet* 72: 314–321 (1986).
5. Bernatzky R, Tanksley SD: Methods for detection of single or low copy sequences in tomato on Southern blots. *Pl Mol Biol Rep* 4: 37–41 (1986).

6. Bernatzky R, Tanksley SD: Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* 112: 887–898 (1986c).
7. Birnboim HC, Doly J: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl Acids Res* 7: 1513–1523 (1979).
8. Capesius I: Sequence of the cryptic satellite DNA from the plant *Sinapsis alba*. *Biochim Biophys Acta* 739: 276–280 (1983).
9. Dennis ES, Peacock WJ: Knob heterochromatin homology in maize and its relatives. *J Mol Evol* 20: 341–350 (1984).
10. Deumling B: Sequence arrangement of a highly methylated satellite DNA of a plant, *Scilla*: A tandemly repeated inverted repeat. *Proc Natl Acad Sci* 78: 338–342 (1981).
11. Deumling B, Greilhuber J: Characterization of heterochromatin in different species of the *Scilla siberica* group (Liliaceae) by in situ hybridization of satellite DNAs and fluorochrome banding. *Chromosoma* 84: 535–555 (1982).
12. Dover GA: Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111–117 (1982).
13. Flavell RB, Rimpau J, Smith DB: Repeated sequence DNA relationships in four cereal genomes. *Chromosoma* 63: 205–222 (1977).
14. Flavell RB, O'Dell M, Hutchinson J: Nucleotide sequence organization in plant chromosomes and evidence for sequence translocation during evolution. *Cold Spring Harbor Symposium on Qualitative Biology*, Vol. 45 (1981).
15. Ganai M, Riede I, Hemleben V: Organization and sequence analysis of two related satellite DNAs in cucumber (*Cucumis sativus* L.). *J Mol Evol* 23: 23–30 (1986).
16. Gottlieb LD: Conservation and duplication of isozymes in plants. *Science* 216: 373–380 (1982).
17. Gottlieb LD: Phosphoglucosyltransferase and isocitrate dehydrogenase gene duplications in *Layia* (Compositae). *Amer J Bot* 74: 9–15 (1987).
18. Houck CM, Reinhart FP, Schmid CW: A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* 132: 289–306 (1979).
19. Hutchinson J, Lonsdale DM: The chromosomal distribution of cloned highly repetitive sequences from hexaploid wheat. *Heredity* 48: 371–376 (1982).
20. Hwu HR, Roberts JW, Davidson EH, Britten RJ: Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. *Proc Natl Acad Sci* 83: 3875–3879 (1986).
21. Lima-de-Faria A, Arnason U, Widegren B, Essen-Moller J, Isaksson M, Olsson E, Jaworska H: Conservation of repetitive DNA sequences in deer species studied by Southern blot transfer. *J Mol Evol* 20: 17–24 (1984).
22. Martin G, Wiernasz D, Schedl P: Evolution of *Drosophila* repetitive-dispersed DNA. *J Mol Evol* 19: 203–213 (1983).
23. Martinez-Zapater JM, Estelle MA, Somerville CR: A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol Gen Genet* 204: 417–423 (1986).
24. Maxam AM, Gilbert W: Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol* 65: 499–560 (1980).
25. Murray M, Thompson WF: Rapid isolation of high molecular weight plant DNA. *Nucl Acids Res* 8: 4321–4325 (1980).
26. Nevers P, Shepherd NA, Saedler H: Plant transposable elements. *Adv Bot Res* 12: 103–203 (1986).
27. Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ: Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proc Natl Acad Sci* 78: 4490–4494 (1981).
28. Pichersky E, Bernatzky R, Tanksley SD, Breidenbach RB, Kausch AP, Cashmore AR: Molecular characterization and genetic mapping of two clusters of genes encoding chlorophyll a/b binding proteins in *Lycopersicon esculentum* (tomato). *Gene* 40: 247–258 (1985).
29. Rivin C: Analyzing genome variation in plants. *Meth Enzymol* 118: 75–85 (1986).
30. Rogers JH: Long interspersed sequences in mammalian DNA. Properties of newly identified specimens. *Biochim Biophys Acta* 824: 113–120 (1985).
31. Shapira SK, Finnerty VG: The use of genetic complementation in the study of eukaryotic macromolecular evolution: Rate of spontaneous gene duplication at two loci of *Drosophila melanogaster*. *J Mol Evol* 23: 159–167 (1986).
32. Tanksley SD: Organization of the nuclear genome in tomato and related diploid species. *Amer Natl* 130: S46–S56 (1987).
33. Tanksley SD, Miller J, Paterson A, Bernatzky R: Molecular mapping of plant chromosomes. *Stadler Genetics Symp* (in press) (1987).
34. Trick M, Dover GA: Unexpectedly slow homogenisation within a repetitive DNA family shared between two subspecies of Tsetse fly. *J Mol Evol* 20: 322–329 (1984).
35. Vallejos CE, Tanksley SD, Bernatzky R: Localization in the tomato genome of DNA restriction fragments containing sequences homologous to the rRNA (45 S), the major chlorophyll a/b-binding polypeptide and the ribulose biphosphate carboxylase genes. *Genetics* 112: 93–105 (1986).