

# Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize *R-r* complex

Elsbeth L.Walker<sup>1</sup>, Timothy P.Robbins<sup>2</sup>, Thomas E.Bureau<sup>3</sup>, Jerry Kermicle<sup>4</sup> and Stephen L.Dellaporta<sup>1,5</sup>

<sup>1</sup>Yale University, Department of Biology, New Haven, CT 06520–8104, <sup>3</sup>Department of Genetics, University of Georgia, Athens, GA 30602 and <sup>4</sup>Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA

<sup>2</sup>Present address: The Cambridge Laboratory, John Innes Centre, Colney Lane, Norwich, NR4 7UH, UK

<sup>5</sup>Corresponding author

Communicated by E.S.Coen

***R-r* controls the production of anthocyanin pigment in plant parts and the aleurone layer of seeds through the production of a family of related transcriptional activating proteins of the helix–loop–helix type. The *R-r* complex comprises a series of repeated, homologous components arranged in both direct and inverted orientations. These include the *P* component, a simple *R* gene that confers pigmentation of plant parts, and the *S* subcomplex that consists of a truncated inactive *R* gene called *q*, and two functional *R* genes, *S1* and *S2*, that pigment the aleurone. The *S* genes are arranged in an unusual inverted head-to-head orientation. The identity of each functional component was confirmed by microprojectile bombardment of intact maize tissues with cloned genomic DNA and by analysis of *in vivo* mRNA populations. Sequence analysis suggests that the *S* subcomplex was derived through the rearrangement of a simple *P*-like progenitor element. At the rearrangement breakpoints, features typical of the CACTA family of transposable elements were found. The location and arrangement of these CACTA element sequences implies that this element may have mediated the chromosomal rearrangements that led to the formation of the *R-r* complex. The unusual structure of *R-r* explains much of the meiotic instability of the complex.**  
*Key words:* anthocyanin/maize/*R-r* complex/transposable elements

## Introduction

Anthocyanins are the red and purple pigments produced in specific tissues of higher plants. In maize, the tissue-specific pattern of anthocyanin deposition is determined by individual members of the *R* gene superfamily. This family includes a large number of simple and complex *R* alleles, the chromosomally displaced *R* genes *Sn* and *Lc*, and alleles of the functionally related *B* locus (reviewed by Dooner *et al.*, 1991). *R* was cloned by transposon tagging (Dellaporta *et al.*, 1988) and several members of the *R* superfamily have been isolated by cross-hybridiza-

tion, including *R-r*, *Lc*, *Sn* and *B* (Chandler *et al.*, 1989; Ludwig *et al.*, 1989; Perrot and Cone, 1989; Robbins *et al.*, 1991; Tonelli *et al.*, 1991; Consonni *et al.*, 1992). *R* genes encode nearly identical myc-homologous, helix–loop–helix proteins (Ludwig *et al.*, 1989; Perrot and Cone, 1989; Consonni *et al.*, 1992) that are capable of activating transcription from promoters of structural genes in the anthocyanin biosynthetic pathway (Ludwig *et al.*, 1989; Goff *et al.*, 1990). *B* and *Lc* genes are sufficient to confer pigmentation to most maize tissues when these genes are expressed from a constitutive viral promoter (Ludwig *et al.*, 1989; Goff *et al.*, 1990), indicating that the pattern of anthocyanin deposition conferred by a given *R* gene is due to the tissue-specific expression of *R*, not to functional differences in *R* proteins.

One of the best studied members of the *R* superfamily is the *R-r* complex. This complex pigments the aleurone, coleoptile, seedling leaf tip, roots and anthers of mature plants. *R-r* is meiotically unstable, giving rise to spontaneous derivatives in which either all plant pigmentation (coleoptile, leaf tip, roots and anthers) or all seed pigmentation (aleurone) is lost (Stadler, 1942). The independent loss of plant and seed coloring capacity indicates the presence of two genetic elements at *R-r*, with one responsible for plant and the other for seed pigmentation. The elements, referred to as *P* and *S*, for plant and seed, behave genetically as though they reside on a tandem duplication that can undergo displaced synapsis. Unequal crossing over between *P* and *S* leads to loss of one member of the duplication (Stadler and Nueffer, 1953).

Figure 1 illustrates the meiotic instability of the *R-r* complex. Derivatives of *R-r* have been observed at an overall frequency of  $6.2 \times 10^{-4}$  (Dooner and Kermicle, 1971). Approximately one tenth of these are *P*-loss derivatives, termed *R-g* (*R* = colored aleurone, *-g* = green plant); the remaining nine tenths are *S*-loss derivatives, termed *r-r* (*r* = colorless aleurone, *-r* = red plant). *R-g* and *r-r* derivatives are recovered both on parentally and non-parentally marked chromosomes. Derivatives that are recovered on chromosomes with a non-parental combination of flanking markers are referred to as crossover or CO derivatives. These derivatives occur at a frequency of  $3.93 \times 10^{-4}$  for *r-r* and  $0.25 \times 10^{-4}$  for *R-g*. However,  $\sim 3/10$  of *r-r* derivatives and  $3/5$  of *R-g* derivatives were borne on parentally marked chromosomes (Dooner and Kermicle, 1971). These derivatives could arise through intrachromosomal recombination between *P* and *S*, which would result in the loss of one *R-r* component. The derivatives would then be genetically and molecularly indistinguishable from CO derivatives of the same type. However, many of these derivatives behave genetically as complex alleles composed of two genetic elements with one of the elements being non-functional, i.e. *p S* or *P s* (Stadler and Emmerling, 1956; Dooner and Kermicle,

1974; Kermicle and Axtell, 1981; Demopoulos, 1985). In order to simplify the molecular genetic analysis of *R-r*, Robbins *et al.* (1991) designated the term 'NCO' for those derivatives of *R-r* which retain a genetic duplication. This definition of NCO derivatives differentiates between alleles caused by unequal crossing over and those caused by other mechanisms, and will be used for the purposes of this study. The high frequency of NCO derivative formation,  $2 \times 10^{-4}$ , is not adequately explained by the tandem duplication model for *R-r* structure, which only predicts meiotic instability due to unequal crossover events.

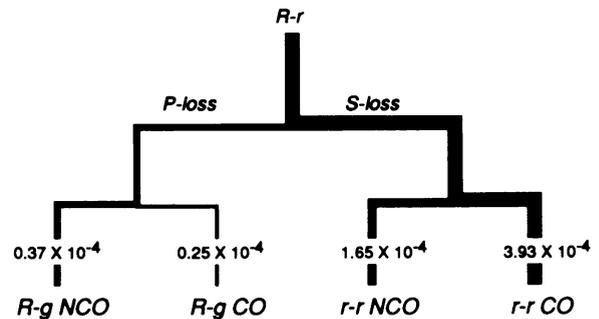
Molecular analysis indicates that the *R-r* complex is not a simple duplication (Robbins *et al.*, 1991). Two distinct regions of *R-r* were identified by a combination of cloning and genomic mapping: a simplex *P* gene and a region we refer to as the *S* subcomplex that contains multiple *R*-hybridizing sequences. The *S* subcomplex was shown to contain a non-functional truncated *R* component, called *q* and one or more *S* genes. Both *R-g* and *r-r* types of CO derivatives result from displaced synapsis and crossing over of homologous elements within the *R-r* complex (Robbins *et al.*, 1991). Crossing over can take place between *P* and *S* within homologous transcribed portions and lead to the loss of *S* function, thus leaving a *P*-only CO derivative. *P* can also pair with the non-functional gene fragment, *q*, resulting in the loss of *P* coding sequences and retention of the *S* subcomplex. NCO derivative alleles were shown to be the result of intrachromosomal rearrangements involving specific regions of the affected component. For example, of the five *R-g* NCO derivatives that have been examined, four have rearrangements localized within a 4 kbp region (Robbins *et al.*, 1991). Each *r-r* NCO derivative examined to date (Robbins *et al.*, 1991 and this work) has a rearrangement localized within a 5 kbp region of *S*.

In this report, we present new evidence pertaining to the structure, origin and meiotic instability of the *R-r* complex. Molecular cloning of additional *R* components has shown that the *S* subcomplex consists of two functional *S* genes, *S1* and *S2*, in addition to the previously identified non-functional, truncated component, *q* (Robbins *et al.*, 1991). The *S1* and *S2* components are arranged in an unusual head-to-head orientation. Sequence analysis of the relevant regions of each *R* component suggests that the rearrangement of a simplex *P*-like progenitor element resulted in the formation of the three components of the *S* subcomplex. DNA sequences at the rearrangement breakpoints suggest the involvement of a transposable element in the formation of the *S* subcomplex. Remnants of the element which remain at the rearrangement breakpoints have important implications for the formation of the complex, the expression of the *S* genes and the meiotic instability of the *R-r* complex as a whole.

## Results

### Isolation of *R-r* genomic clones

To determine the molecular organization of the *R-r* complex, genomic clones of *R-r* components were isolated from partial *Sau3A* genomic lambda libraries screened with probes representing the 5' non-coding region (pR-nj:1), the coding portion (R-sc:323I), the 3' untranslated



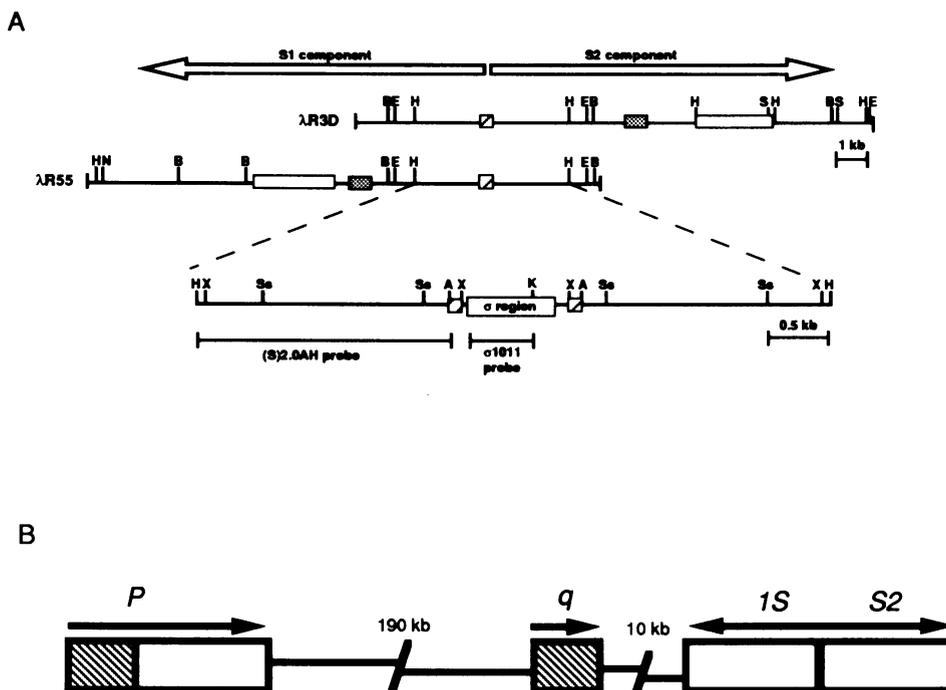
**Fig. 1.** Meiotic instability of *R-r*. *R-r* gives rise to *P*-loss (*R-g*) and *S*-loss (*r-r*) derivative alleles at the frequencies shown (Dooner and Kermicle, 1971). Derivative formation may be the result of crossing over [crossover (CO) alleles] or may be the result of another mechanism [non-crossover (NCO) alleles]. Line thickness indicates the relative frequency of each derivative type.

region (4.7H2.5) and the 3' flanking region (4.7E1.8) of the *R* gene unit (see Materials and methods for details on probes). In addition to the previously cloned regions of *R-r* (Robbins *et al.*, 1991), several new genomic clones were obtained in this way. Restriction maps and Southern hybridization of *R* probes to one of these genomic clones are summarized in Figure 2A. The maps have been aligned to show overlaps between the novel genomic clone,  $\lambda$ R55, and the previously reported clone,  $\lambda$ R3D, which is part of the *S* subcomplex (Robbins *et al.*, 1991). A second genomic clone,  $\lambda$ R68 (not shown), also overlaps with  $\lambda$ R55 and  $\lambda$ R3D.

Mapping of these genomic clones revealed an unusual organization of this part of the *S* subcomplex. The  $\lambda$ R55 and  $\lambda$ R68 clones hybridize to the coding region and the 3' untranslated region probes but extend in a direction opposite to that of the  $\lambda$ R3D clone. Isolation of these two overlapping clones confirms that this unusual structure is not the result of a rearrangement during cloning. Thus, it appears that there are two potentially intact genes in opposite orientations within the *S* subcomplex. This was later confirmed by microprojectile bombardment and RT-PCR experiments described below. We designate these two *S* genes as *S1* and *S2*; the relative orientation and position of each gene on the chromosome is shown above the maps in Figure 2A. Previously isolated genomic clones, together with the new clones reported here, represent a total of 60 kbp of *R-r* DNA that has been isolated and characterized.

### Molecular organization of the *S* subcomplex

Southern hybridization analysis of the 5 kbp *HindIII* region of the *S* subcomplex, summarized in Figure 2A, revealed the existence of two discrete *R* regions which hybridized to the 5' probe, pR-nj:1 (Figure 2A, hatched boxes), indicating the presence of two potential transcription start sites separated by <1 kbp of intervening DNA. Note that the restriction sites in this region are perfectly symmetrical, again indicating an inverted structure for *S1*. This symmetry breaks down toward the central portion of the 5 kbp region—the unique *KpnI* site shown in Figure 2A is located ~200 bp off center. This asymmetry allowed the orientation of the 5 kbp region in the *S* subcomplex within *R-r* to be determined. The two distinct fragments that hybridize to the 5' non-coding probe pR-nj:1 flank a



**Fig. 2.** (A) *R-r* genomic clones. The restriction sites are abbreviated as: H = *Hind*III; B = *Bam*HI; E = *Eco*RI; Ss = *Sst*II; S = *Sal*I; N = *Not*I; X = *Xba*I; and K = *Kpn*I. The clone  $\lambda$ R3D was reported previously (Robbins *et al.*, 1991). Genomic clones of *S* are aligned by common restriction sites. Shown below is a close up of a 5 kb region of *S* which contains the  $\sigma$  region. Southern hybridization results with the *R* probes pR-nj:1, R-sc:323I, and 4.7H2.5 are indicated by hatched, gray, and open boxes, respectively. New probes generated from the region, (S)2.0AH and  $\sigma$ 1011, are shown expanded below. (B) The overall structure of the *R-r* complex. The four components of *R-r* (*P*, *q*, *S1* and *S2*) are shown. Arrow direction indicates the relative orientation of each component. The promoter regions are indicated as hatched boxes; the coding portions are shown as open boxes. The distances between *R* components was determined by CHEF analysis (data not shown).

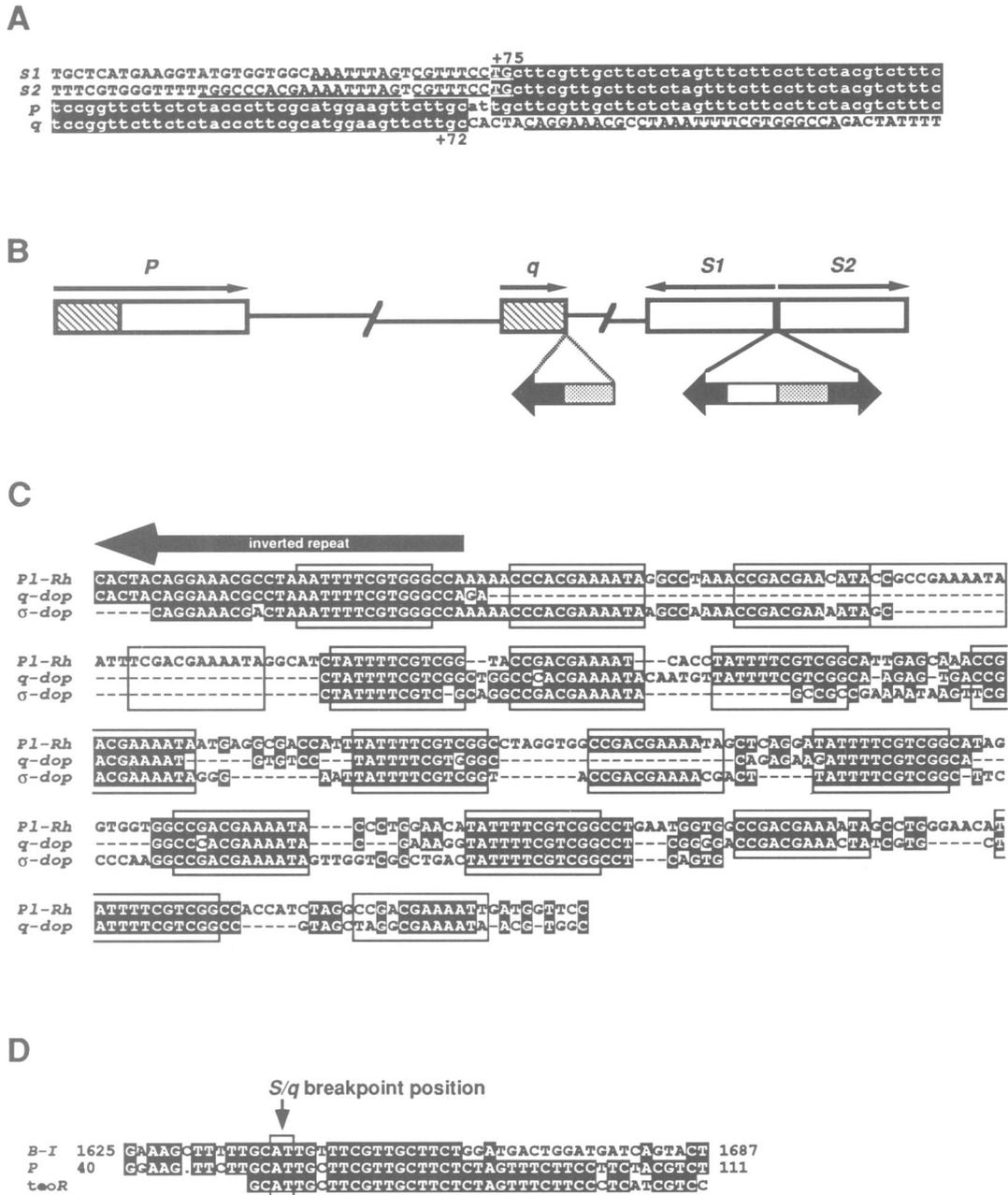
small central region which we call  $\sigma$ . The  $\sigma$  region does not hybridize with 5' probes from the *P* component, and thus appears to consist of sequences that are unrelated to these components. It is interesting to note that the *q* component is highly homologous to the 5' and promoter fragments of *P*, yet is truncated so that it shares no similarity with *P* coding sequences. Thus, the *S* components are similar to the *P* coding region, but not the *P* promoter region, while the *q* component is similar only to the *P* promoter and upstream sequences, but not the downstream coding sequences of *P*.

The organization of the *S* subcomplex within *R-r* was determined based on the structure of derivative alleles, and is summarized in Figure 2B. The *S2* region was previously shown to participate in unequal crossing over events with the *P* gene placing *S2* in direct orientation relative to *P* (Robbins *et al.*, 1991). *S1* must lie proximal to *P* and distal to *S2* because the 10 kbp *Hind*III and 4.5 kbp *Bam*HI fragments that characterize the *S1* gene (see Figure 2A) are lost following unequal crossing over between *S2* and *P* (Robbins *et al.*, 1991). *S1* is inverted with respect to *S2* and likewise inverted with respect to *P*, therefore it is designated within the complex as *IS*. The *q* component of *R-r* was previously shown to lie distal to *P* and proximal to *S2* (Robbins *et al.*, 1991). The physical association between *S1* and *S2* places *q* distal to *P* and proximal to both *S1* and *S2*. In summary, these data indicate the organization of the *R-r* complex to be *P-q-IS-S2*. CHEF gel analysis (E.Walker, unpublished observations) has shown that the distance between *q* and *S1* is fairly small, ~10–20 kbp. The *P* component, however,

is located much farther from the *S* subcomplex, 190 kbp proximal.

#### The interrelationship of *R-r* elements

As noted above, hybridization data from *R* genomic clones suggested that the truncated *q* and functional *S* genes each seem to consist of only one distinct portion of the *P* component: the *S* genes are similar to the transcribed portion of *P*, while the *q* component is similar only to the promoter portion of *P*. These data suggested that the *S* and *q* components might represent reciprocal products of a chromosomal rearrangement, and that the breakpoints of this rearrangement would lie close to or within the 5' non-translated leader region of each component. To examine this possibility, the sequences of the appropriate regions of *P*, *q*, *S1* and *S2* were determined. Within the sequences shown (Figure 3A), the upstream (5') regions of *P* and *q* are identical. Sequences of *P* and *q* extending upstream of the region shown in Figure 3A also show near perfect similarity (1 bp mismatched) over an additional 180 bp that were sequenced (not shown). This upstream similarity extends at least another 4.5 kbp based on restriction site conservation between the cloned regions (Robbins *et al.*, 1991 and data not shown). The 3' portions of the *P* and *q* components share no significant homology. Thus, as predicted by hybridization analysis, *q* represents a non-functional component with a truncation 72 bp downstream of the predicted *P* transcription start site [inferred from the position of the transcription start site of the highly related gene, *Lc* (Ludwig *et al.*, 1989)]. Conversely, Figure 3A shows that it is the downstream



**Fig. 3.** Rearrangement breakpoints in *R-r*. (A) Sequence comparison of the breakpoint regions of *P*, *q*, *S1* and *S2*. The sequences of *q*, *S1* and *S2* were aligned with the nucleotide sequence of *P*. Identical nucleotides are indicated by white letters. Twenty six and 16 bp repeat structures present adjacent to the breakpoints in *q*, *S1* and *S2* are indicated by underlining. Nucleotide positions given are relative to the start of transcription inferred from the highly related *Lc* gene (Ludwig *et al.*, 1989). (B) The organization of *doppia* sequences within the *R-r* complex. Shaded and open boxes represent the 5' and 3' regions of *R*. Below and expanded are *doppia* sequences which are present downstream of *q*, and within the  $\sigma$  region between *S1* and *S2*. *Doppia* sequences are found on the right of the breakpoint at *q*, on the right of the breakpoint at *S1*, and on the left of the breakpoint of *S2*. (C) Multiple alignment of elements from the  $\sigma$  region ( $\sigma$ -*dop*), *q* region (*q*-*dop*) and from *PI-Rh* (Cone *et al.*, 1993). The arrow above indicates the extent of the inverted repeat of *doppia*. White letters indicate identical nucleotides and open rectangles indicate the positions of subterminal repeats. Alignments were performed using the program PILEUP as part of the UWGCG computer program suite. The gap penalty was 2.0 and the gap length penalty was 0.3. Gaps were introduced for optimal alignment. (D) Sequence comparison of the *P* component of *R-r*, the maize *B* gene, and a teosinte *R* gene in the breakpoint region. White letters indicate sequence identity.

(3') regions of *S1* and *S2* that are identical to the *P* sequence. This perfect sequence identity between *P*, *S1* and *S2* extends for an additional 287 bp that were sequenced (not shown). The upstream portion of *P* (from -143 to +72) shares no sequence similarity to the *S* genes. Thus, the *S1*-*S2* promoter and 5' leader regions are dissimilar to those of *P*. The divergence in *S* and *P*

sequences occurs at position +75 of transcription. Based on restriction mapping and sequencing (Robbins *et al.*, 1991 and E.L. Walker, unpublished results), the similarity between *P* and the two *S* genes probably extends throughout their coding regions except for minor sequence polymorphisms.

The most striking observation from this analysis is that

the homology between *P* and *q* ends at nearly the precise position where the homology between the *P* and *S* components begins (see Figure 3A). The reciprocal nature of the *q* and *S* components and their extensive similarity to defined regions of *P* suggests a common mode of origin resulting from the rearrangement of an intact *P*-like ancestral element. The sequence comparison, shown in Figure 3A, of *P*, *q*, *S1* and *S2* at the *q/S* rearrangement breakpoint offers clues to the origin of the *R-r* complex. The overall organization of these sequences within the complex is illustrated in Figure 3B. An imperfect 26 bp inverted duplication is present just upstream and downstream of the breakpoints in *S2* and *q*, respectively (Figure 3A, underlined and Figure 3B); *S1* also shares the first 16 of these 26 nucleotides (Figure 3A, shown underlined, and Figure 3B). At *S1* and *S2*, this sequence begins exactly at the breakpoint, while at *q* the 26 bp sequence is displaced 5 bp downstream of the breakpoint (Figure 3A). The 33 bp just downstream of the breakpoint in *q* are identical to the terminal inverted repeat of an element of the CACTA family found in the *Pl-Rh* and *Pl-Bh* alleles of the maize *Pl* gene (Figure 3C; Cocciolone and Cone, 1993; Cone *et al.*, 1993). The sequences upstream of the breakpoint in *S2* are also related to this element, except that the first 5 bp of the terminal inverted repeat seem to be missing in *S2* (Figure 3C). Thus, while the *q* gene appears to contain an intact terminal inverted repeat sequence, the *S2* gene contains a truncated version of this element. Another important difference is that there is a single nucleotide polymorphism within the inverted repeat sequences at *S1* and *S2* which distinguishes these two from the terminal inverted repeat sequence present at *q* and *Pl-Rh*.

The sequence similarity between the element at *Pl-Rh* and the region upstream of the *S2* breakpoint extends beyond the terminal inverted repeat region, spanning a total of 197 bp of the  $\sigma$  region. Most of the sequence similarity observed is between a series of 12 bp repeated elements (Figure 3C, boxed sequences) that are present just downstream and upstream of breakpoints in *q* and *S2*, respectively. These 12 bp motifs are identical to those putatively identified as the subterminal repeat elements of the element present at *Pl-Rh*. The presence of inverted repeats with homology to the CACTA family of transposable elements, together with the presence of internal clustered short direct and indirect repeats characteristic of structures at transposable element ends, suggests the involvement of a transposable element in the formation of the *S* subcomplex of *R-r*. We call this element *doppia* (Latin: to duplicate). The reciprocal nature of *q* and *S2* sequences and breakpoints suggests that these components were derived by fracturing of a progenitor *P*-like component containing the *doppia* element at the breakpoint position.

Examination of the sequence of the *P* component near the breakpoint position gives no suggestion of *doppia* sequences nor of a transposable element footprint. To confirm this, we compared the sequence of *P* with those of the *B* locus of maize and an *R* gene cloned from teosinte. Had the the *P* component of *R-r* been visited by *doppia*, we would expect to have observed some sort of sequence variation at the target site. The result of this comparison is shown in Figure 3D. The sequences of each

of the genes are well conserved at the breakpoint region. Thus, the *P* component appears not to have been visited by *doppia*. However, it is possible that *doppia* was once present at *P*, but was excised precisely. In this case, no footprint would be observed.

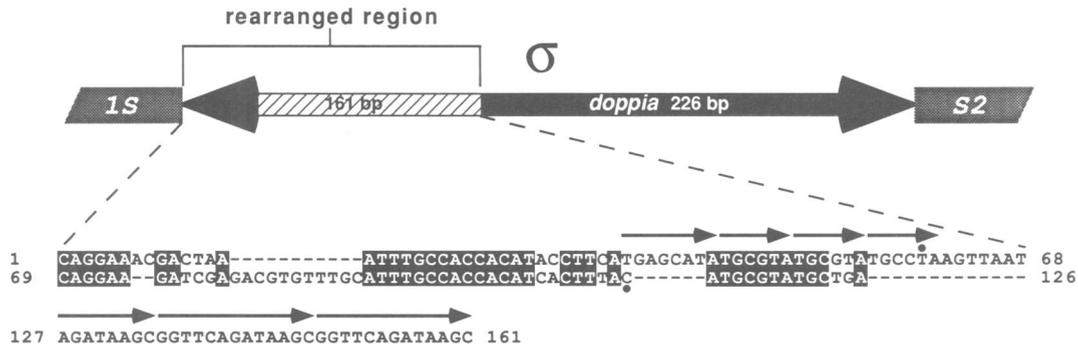
#### **Relationship of *S1* and *S2* components**

Several lines of evidence suggest that *S1* was derived from *S2*. The sequence upstream of the breakpoint at *S1* is identical to that upstream of *S2* over a span of 16 bp, as shown in Figure 3A. Thus, the 5 bp of the terminal inverted repeat that is missing in *S2* is likewise missing in *S1*. In addition, the single base pair polymorphism that distinguished the repeat element at *S2* from that at *q* is also present at *S1*. Therefore, *S1* represents an inversion of *R* sequences that includes coding sequences, the *S2/q* breakpoint and 16 bp of the terminal inverted repeat of the *S2* component. Thus, we can narrow the definition of the  $\sigma$  region to the 16 bp inverted repeat at *S1*, the 26 bp inverted repeat at *S2*, and the sequence that is flanked by them. The  $\sigma$  region thus is 387 bp long.

Within  $\sigma$ , there are two distinct regions, as illustrated in Figure 4. The first is the region that shares homology with the element at *Pl-Rh*, and thus appears to have been derived from *doppia*. This region spans a total of 226 bp and is located adjacent to *S2*. The second region consists of 161 bp of sequence that is found neither in *q*, *Pl-Rh*, nor in the corresponding region of *P*. We refer to this region as 'rearranged', because it contains multiple repeated motifs as shown in Figure 4 that were probably derived through the rearrangement of a simpler sequence. Interestingly, there is a 54 bp sequence within this region (Figure 4, position 60–113, indicated by dots) with 85% identity to the *P* promoter region from –530 through –586 relative to the start of transcription (not shown). The  $\sigma$  region is likely to serve as the promoter region for both *S1* and *S2*, since it lies just 5' to the transcription initiation sites predicted from the sequence of an *S* cDNA clone (Perrot and Cone, 1989).

#### **Microprojectile bombardment assays of *R* components**

The structural analysis of the *S* subcomplex indicates that there are two potential *S* transcription units separated by the small  $\sigma$  region. It has not been possible to determine genetically whether both or only one of the two *S* components is functional because CO derivatives always lose both components and each *r-r* NCO derivative has a lesion that deletes the  $\sigma$  region and affects both *S* genes (see below). Therefore, to assess the functionality of *S1* and *S2* components, we used a microprojectile bombardment strategy similar to that developed for assessing expression of the *R*-related genes *B* and *Lc* (Perrot and Cone, 1989; Goff *et al.*, 1990). The genomic DNA inserts from three  $\lambda$  clones,  $\lambda$ R55—an intact *S1* clone,  $\lambda$ R3D—an intact *S2* clone, and  $\lambda$ R4.7—an intact *P* clone, were precipitated on gold particles and used to bombard intact aleurone or seedling tissue. These tissues were genetically competent to express anthocyanins except for a deficiency in *R* function (see Materials and methods). Figure 5 shows the result of this experiment. As a positive control, aleurone and coleoptile tissue were bombarded with an *Sn* cDNA clone (Consonni *et al.*, 1992) driven by the constitutive



**Fig. 4.** Sequence analysis of the  $\sigma$  region. The  $\sigma$  region situated between *S1* and *S2* is composed of *doppia* sequences (black box and arrowheads) and a rearranged region (hatched box). Truncation of the arrowheads indicates the positions of the five deleted nucleotides from each terminus. Below is shown the pairwise alignment between nucleotides 1–68 and 69–161 of  $\sigma$ : the rearranged region. This region was formed, in part, by an imperfect direct repeat. The first 15 bp of the rearranged region are nearly identical to the *doppia* terminus. Arrows above sequences indicate the positions of tandemly arranged repeats. Black dots indicate a region that shares 85% sequence similarity with the 5' flanking regions of the *P* component and the *Lc* gene (data not shown).

cauliflower mosaic virus (CaMV) 35S promoter. The 35S–*Sn* cDNA construct was able to direct anthocyanin biosynthesis in both aleurone and coleoptile tissues (Figure 5G and H). Both the *S1* and *S2* DNAs pigmented aleurone cells (Figure 5B and D). The *S1* and *S2* DNAs were not functional in any seedling tissue (Figure 5A and C). Thus, both *S1* and *S2* are potentially functional *S* genes capable of producing anthocyanin pigment in aleurone, but not in plant parts. The *P* gene pigmented coleoptile cells (Figure 5E). This cell type is normally pigmented by the *P* component of *R-r*. The *P* clone was not functional in aleurone (Figure 5F), a tissue not normally pigmented by the *P* component of *R-r*.

#### RT-PCR analysis of aleurone mRNA

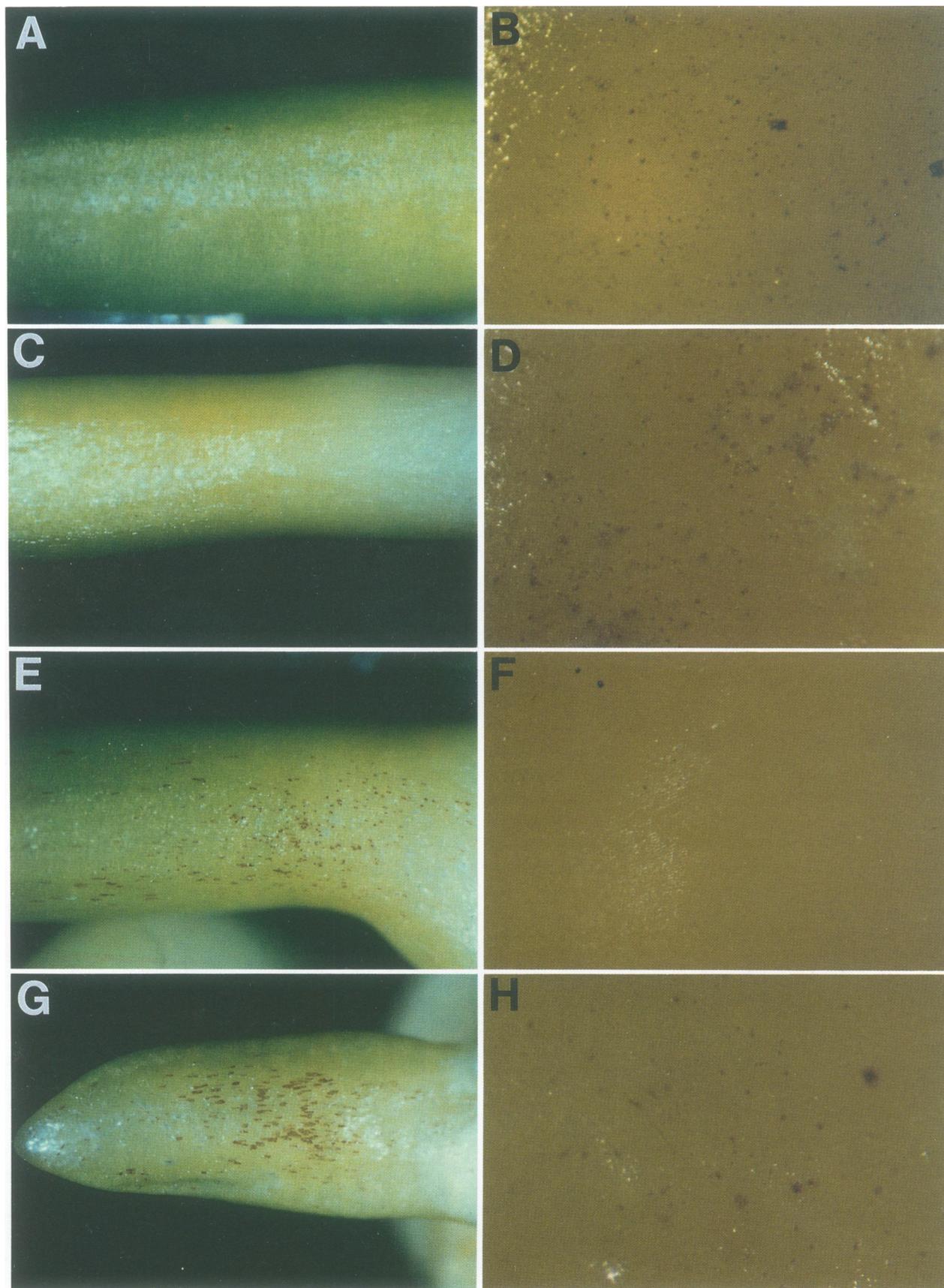
To extend the bombardment assay results indicating that both *S1* and *S2* are functional in aleurone, we used a RT-PCR assay to examine expression of mRNA from each *S* gene. In this assay, aleurone mRNA is reverse transcribed to cDNA using oligo(dT) as a primer and then PCR amplified using primers that span the last intron and the 3' non-coding portions of each *R* component. The amplified 3' segment of each *R* component exhibits several sequence polymorphisms which are occasionally associated with a restriction enzyme site polymorphism. By analyzing amplification products for diagnostic restriction sites, the presence of *P*, *S1* and/or *S2* messages can be determined unambiguously.

The results of this study are illustrated in Figure 6. The cDNA from homozygous *R-r* or *R-g:1* aleurone was amplified and examined for diagnostic restriction sites. The homozygous *R-r* material could potentially express mRNA from three *R* genes: *P*, *S1* and *S2*. *R-g:1* is a CO derivative of *R-r* which retains only the *q*, *S1* and *S2* components and so only has the potential to express the two *S* genes. Controls for this study were genomic clones representing the 3' ends of *P*, *S1* and *S2*. PCR amplification products from the genomic clone controls (Figure 6A, lanes 1, 4 and 5) are larger than the products from cDNA (Figure 6A, lanes 2 and 3) because of the presence of intron sequences within the amplification products. The expected restriction enzyme cleavage patterns for each *R* gene are shown in Figure 6B. Restriction enzyme *Ava*II cuts *S2* and *P*, but not *S1*; restriction enzyme *Afl*III cuts

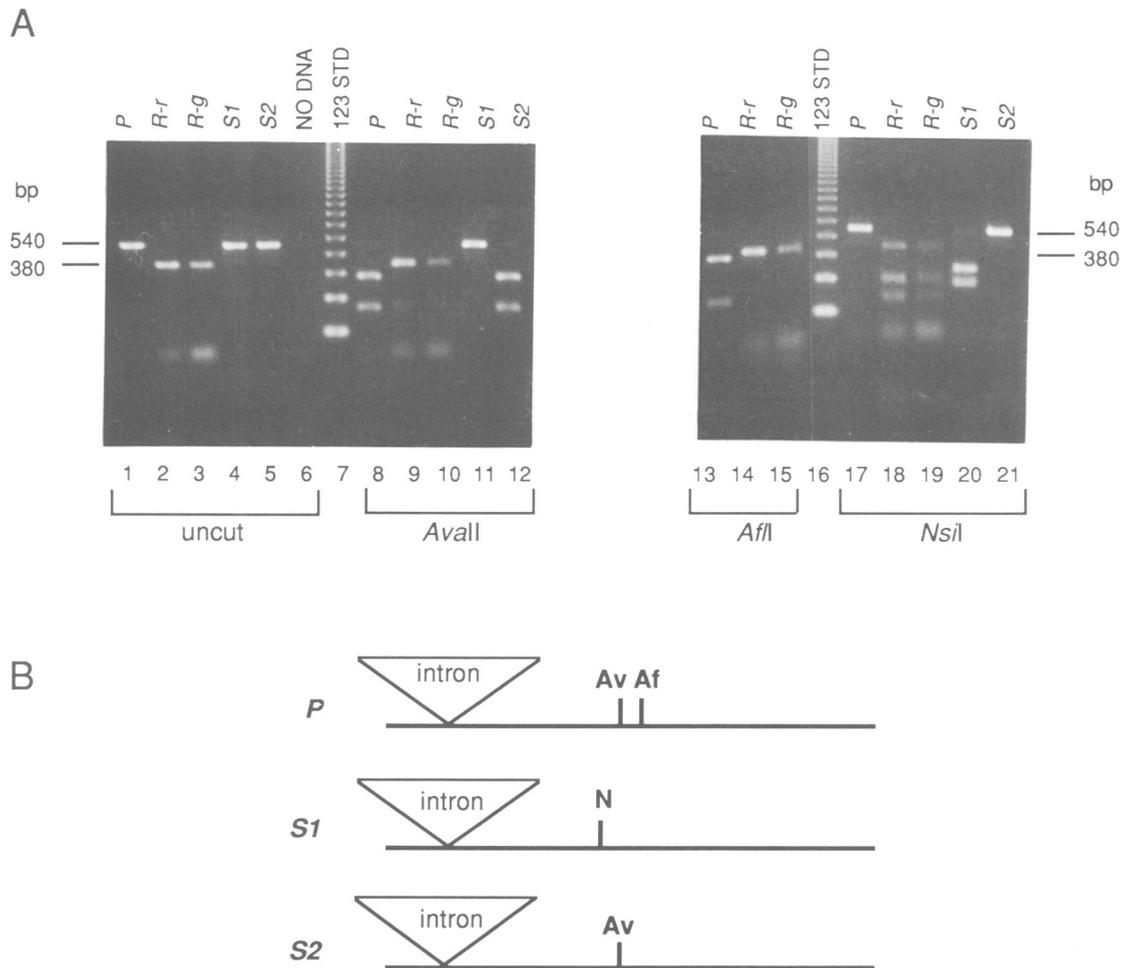
only *P*; restriction enzyme *Nsi*I cuts only *S1*. Figure 6A, lanes 8–12, shows the result of *Ava*II digestion of the amplified products. In both *R-r* and *R-g* aleurone cDNA populations (Figure 6A, lanes 9 and 10), both cut and uncut species are present. While the cut species from *R-r* aleurone cDNA could represent either *P* or *S2*, the cut species from *R-g* aleurone cDNA can only represent *S2*, because the *P* gene is not present in the genome. The uncut species represents *S1*. Thus, both *S1* and *S2* are expressed in aleurone mRNA. Digestion with *Afl*III (Figure 6A, lanes 13–15), indicates that the *P* gene is not significantly expressed in aleurone because there are no visible digestion products from *R-r*-amplified cDNA (Figure 6A, lane 14). The restriction pattern produced by *Nsi*I (Figure 6A, lanes 17–21) confirms the conclusion reached from the *Ava*II digest that both *S1* and *S2*—the cut and uncut species, respectively, are present in cDNA from *R-g* aleurone (Figure 6A, lane 19). It was also noted that the product from *S1* was more abundant than that from *S2*. Since primers match both gene sequences exactly and the amplification products are similar in size, this difference may indicate that the steady-state level of mRNA from *S1* is higher than that of *S2* in the aleurone. In summary, RT-PCR analysis indicates that both *S* genes are expressed in the aleurone, although *S1* mRNA levels may be higher than *S2*.

#### Complex rearrangements of the $\sigma$ region cause loss of *S* function

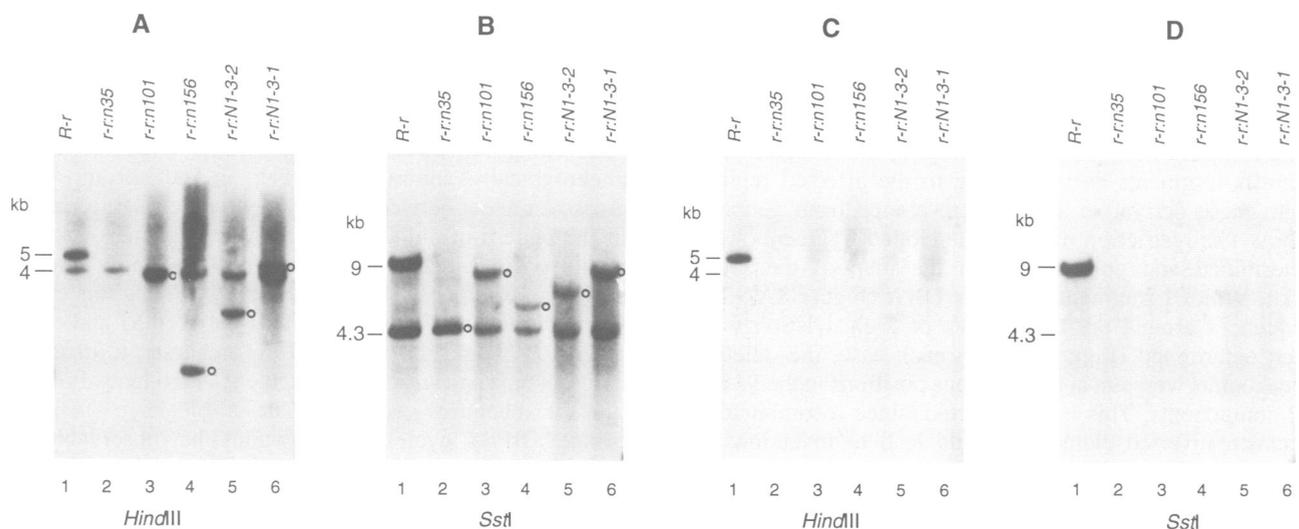
We have previously shown that the NCO derivatives associated with loss of *S* function (*r-r* NCO alleles) are caused by intrachromosomal rearrangements in a specific region within the 5 kbp *Hind*III fragment of the *S* subcomplex (Robbins *et al.*, 1991). All other regions of *R-r* remain structurally intact in these derivatives. Because there are two functional *S* genes, complete loss of *S* function is expected only if a rearrangement affects both genes. Southern blot analysis of each *r-r* NCO allele is shown in Figure 7. The probe used in these experiments, S2.0AH, is shown in Figure 2. The 4 kbp *Hind*III and 4.3 kbp *Sst*I fragments represent *P* (Robbins *et al.*, 1991) and were unaffected in the *r-r* NCO derivatives (Figure 7A and B). The probe also hybridizes to a single 5 kbp *Hind*III or 9 kbp *Sst*I fragment of the *S* subcomplex in



**Fig. 5.** Microprojectile bombardment assay of *R-r* genomic clones. DNA from  $\lambda$  clones,  $\lambda$ R4.7,  $\lambda$ R55 and  $\lambda$ R3D representing *S1* (A and B), *S2* (C and D) and *P* (E and F), respectively, was used to bombard intact seedling (A, C, E and G) and aleurone (B, D, F and H) tissue. A plasmid containing an *Sn* cDNA clone (Consonni *et al.*, 1992) under the control of the CaMV 35S promoter was used as a positive control for both tissues (G and H).



**Fig. 6.** Analysis of aleurone mRNA populations. (A) cDNA from aleurone homozygous for either *R-r* (lanes 2, 9, 14 and 18) or *R-g* (lanes 3, 10, 15 and 19) was subjected to PCR amplification using primers oR31A and oR32. Control DNA samples from genomic clones of *P* (lanes 1, 8, 13 and 17); *S1* (lanes 4, 11 and 20) and *S2* (lanes 5, 12 and 21) were likewise amplified, as was a no DNA control (lane 6). Samples were then digested overnight with restriction enzymes *Ava*II (lanes 8–12), *Afl*II (lanes 13–15) or *Nsi*I (lanes 17–21). Products were then separated on 2% agarose gels. (B) Restriction maps of the amplified region of each *R* gene.



**Fig. 7.** Genomic blot analysis of *r-r* NCO alleles. Genomic DNA from plants homozygous for *R-r* and *r-r* NCO derivatives *r-r:n35*, *r-r:n101*, *r-r:n156*, *r-r:NI-3-1*, and *r-r:NI-3-2* was digested with restriction enzymes *Hind*III (A and C) or *Sst*I (B and D), separated on agarose gels and transferred to filters for hybridization. The filters were hybridized with the probe S2.0AH (A and B), which detects the *S* and *P* components of *R-r*, washed by boiling, and hybridized with a second probe,  $\sigma$ 1011, (C and D) which is specific for the  $\sigma$  region between the *S1* and *S2* genes. Open circles indicate fragments with altered molecular weights.

**Table I.** Southern blot analysis of *r-r* NCO alleles

Derivative	<i>Hind</i> III (kb)	<i>Sst</i> I (kb)	Deletion
<i>R-r</i>	5.0	9.0	—
<i>r-r:n35</i>	0.7	4.3	4.3
<i>r-r:n101</i>	3.9	7.8	1.1
<i>r-r:n156</i>	1.6	5.6	3.4
<i>r-r:NI-3-1</i>	4.6	8.6	0.4
<i>r-r:NI-3-2</i>	3.0	7.0	2.0

Genomic Southern blots were hybridized with the probe S2.0AH which corresponds to the 5' untranslated region, first exon and part of the first intron of *S1* and *S2*. Data presented in this table refer only to hybridizing bands corresponding to *S1* and *S2*; bands corresponding to *P* (which are also detected by the S2.0AH probe) are not recorded here, but are present in unmodified form in *R-r* and all *r-r* NCO derivatives shown.

*R-r* DNA (Figure 7A and B, lane 1). Both *Hind*III and *Sst*I fragments are smaller in each derivative allele (Figure 7A and B, lanes 2–6), suggesting that the *S* subcomplex rearrangements are deletions. Probes hybridizing to *R* coding regions were used to confirm that the regions outside of the 5 kbp *Hind*III fragment, i.e. the 10 kbp *Hind*III fragment of *S1* and the 4 kbp *Hind*III fragment of *S2*, are intact in all derivatives (Robbins *et al.*, 1991) and data not shown). As summarized in Table I, the deletions present in these *r-r* alleles range in size from 0.4 kbp (*r-r:NI-3-1*) to 4.3 kbp (*r-r:n35*).

In Figure 7C and D, the  $\sigma$ -specific probe,  $\sigma$ 1011, was used to determine whether the deletion covered the  $\sigma$  region which contains the putative promoter(s) for both *S* genes. None of the derivatives retained significant homology to  $\sigma$  in the altered restriction fragment. The *r-r* NCO derivatives, therefore, are caused by loss of the  $\sigma$  region and adjacent sequences between *S1* and *S2*.

Three *r-r* NCO derivatives representing small, medium and large deletions (*r-r:NI-3-1*, *r-r:NI-3-2* and *r-r:n156*) were chosen for further analysis. Each of these derivatives occurred as a single colorless kernel on an otherwise fully pigmented ear (Kermicle and Axtell, 1981). Derivative *r-r:n156* was derived from an *R-r* homozygote. The other two derivatives were from hemizygous *R-r* stocks in which *R-r* was paired against the 10B chromosome of translocation B-10A, which lacks the distal 2/3 of the long arm of chromosome 10, including the *R* region. The *Hind*III fragments corresponding to the affected regions from these derivative alleles were cloned from genomic DNA. The restriction maps of the cloned fragments were determined and compared with the map of the intact 5 kbp *Hind*III fragment from *R-r* DNA (Figure 8A). The sequences around the breakpoints of each deletion were then determined (Figure 8B). In each case, the deletion breakpoints were not at homologous positions in the *S1* and *S2* components. This is as expected, since recombination between inverted elements should lead to inversions of the intervening DNA, not deletions.

The deletion breakpoints of *r-r:n156* are located within the first introns of the *S1* and *S2* genes (Figure 8A). Between the breakpoints, 31 bp of filler DNA is inserted (Figure 8B). The origin of the filler DNA found in *r-r:n156* could not be determined; the sequence is not present within 500 bp upstream or downstream of the deletion breakpoints. The rearrangements found in both *r-r:NI-3-1*

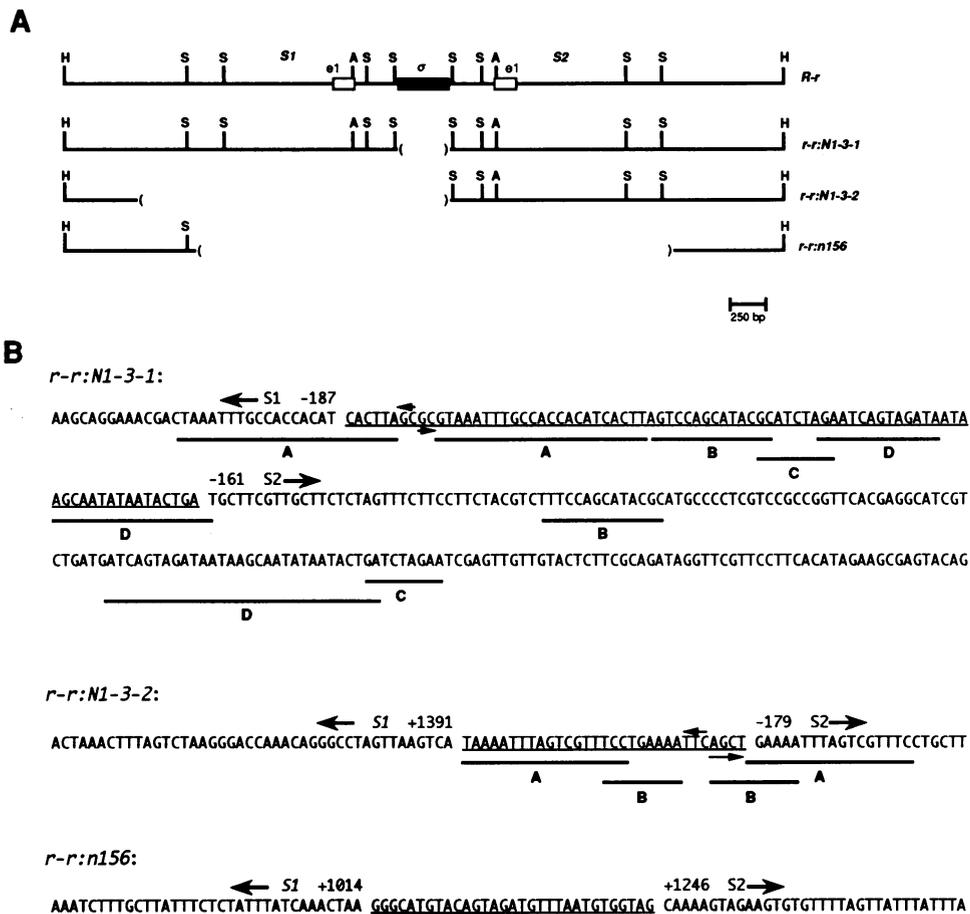
and *r-r:NI-3-2* were more complex. In both of these mutants, filler DNA was found between the deletion breakpoints. The filler sequences were derived from DNA located within 110 bp of the breakpoints (Figure 8B). The filler DNA present in *r-r:NI-3-1* was derived from a discontinuous series of four sequence motifs (designated A–D in Figure 8B) found both upstream and downstream of the deletion breakpoints. Two duplicated motifs were present in the filler DNA of the *r-r:NI-3-2* derivative. The motifs found in the filler DNA often overlap partially (e.g. motifs B and C of *r-r:NI-3-1*) and, in both derivatives, one of the motifs spans the deletion breakpoints (Figure 8B, motif A of *r-r:NI-3-1*; motif B of *r-r:NI-3-2*). In each derivative a small inverted repeat was also present in the filler DNA (Figure 8B, indicated by arrows). The right deletion breakpoints of both *r-r:NI-3-1* and *r-r:NI-3-2* are located within the right inverted repeat of *S2* (compare Figure 8B and Figure 4C).

## Discussion

### *The molecular organization of R-r explains its meiotic instability*

Studies presented here demonstrate that the *R-r* complex is made up of three functional genes *P*, *S1* and *S2* and a non-functional gene fragment, *q*, arranged structurally as *P-q-1S-S2* (the *S1* and *S2* genes are arranged in head-to-head orientation). The structural information that has been obtained for *R-r* can be used to explain the patterns of displaced synapses and crossing over events that lead to the formation of CO derivatives of *R-r*. Robbins *et al.* (1991) demonstrated that displaced exchanges occur between *P* and *S2* and between *P* and *q*, and found no evidence for other pairing arrangements, despite the presence of other *R*-homologous elements at *R-r*. In light of our present findings, it is clear that other regions of recombination were not detected for two reasons. First, the *q* element shares no extensive sequence homology with either of the two *S* genes, and therefore unequal crossing over between *q* and *S* genes would not be expected. Pairing between the two *S* genes could occur, but since they lie in inverted orientation to one another, and are functionally duplicate, intrachromosomal crossing over between them would result in an inversion of the intervening sequences which would be expected to be phenotypically neutral. Likewise, an intrachromosomal crossover event between *P* and *S1* would cause no change in the pattern of anthocyanin deposition of the plant and would go undetected, although the *q*-containing interval between *P* and *S1* would be inverted following this event. Interchromosomal crossover events between *S1* and either *S2* or *P* would be expected to generate a dicentric chromosome and an acentric fragment as products. Production of dicentrics would initiate a breakage–fusion–bridge (BFB) cycle which would be observable in progeny aleurones. Since kernels displaying *R* phenotypes indicative of a BFB cycle are rare, interchromosomal crossing over between inverted elements appears to be uncommon.

The presence of two functional *S* copies at *R-r* explains the failure of genetic tests to detect transfer of mutations to the *S* component of *R-r* by intragenic recombination with a simple *R* allele. An *r-g* allele was generated through



**Fig. 8.** *r-r* NCO clones. (A) Positions of deletions. The restriction map of the intact 5 kbp *Hind*III fragment containing the promoter region, first exon and part of the first intron of *S1* and *S2* is shown above. Below, the position and extent of the deletion in each *r-r* NCO derivative is denoted by brackets. (B) Sequence analysis of recombination breakpoints of *r-r* NCO clones. Filler DNA is shown separated by spaces and underlined. The various duplicated sequence motifs are indicated with bold underlines and letters. The positions of each breakpoint relative to the end of the *S* cDNA isolated by Perrot and Cone (1989) are indicated.

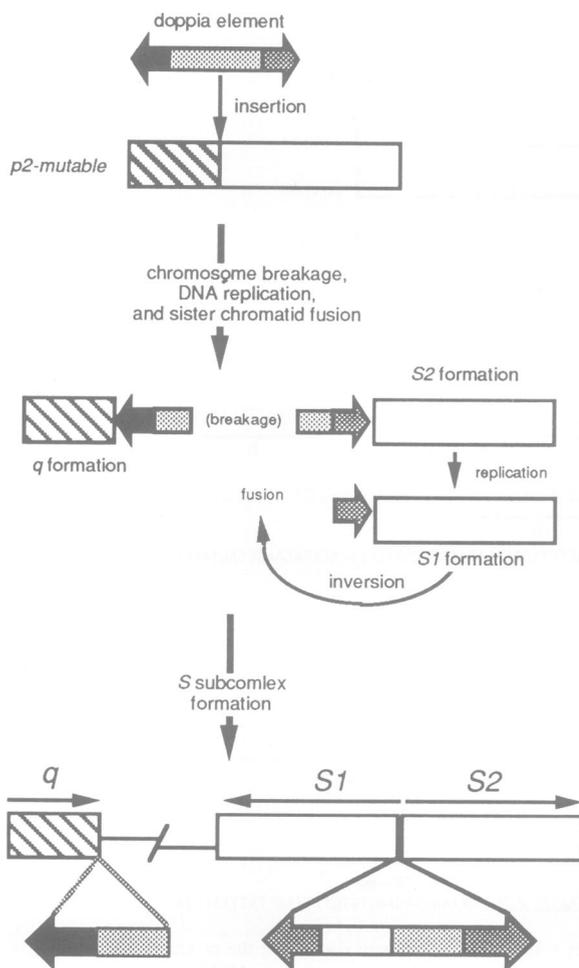
the insertion of a *Ds* element at the *P* component of a *r-r* derivative of *R-r*. The simple duplication model for *R-r* structure predicted that when such *p:Ds* alleles recombine with the *S* component of *R-r*, either reciprocal product of the crossover event should be recoverable because, in one product, *P* function will be restored, and in the reciprocal product, a mutable *S* component (*s:Ds*) should be generated. *P* restitution was readily achieved in these experiments, but in no case was a mutable *S* component identified (Kermicle, 1985, 1988). It is now clear that the failure to detect *s:Ds* chromosomes is due to the presence of two functional *S* genes. Crossing over between *S2* and *P:Ds* can restore *P* function, but a reciprocal exchange would result in a lesion in only one of the two copies of *S*, leaving the inverted *IS* copy intact.

### Evolution of the *R-r* complex

Any model for the evolution of the *R-r* complex from a simplex *R* progenitor gene must take into consideration the following observations. The model must explain the large physical distance that separates *P* from the *S* subcomplex. The model must account for the complementarity of the *q* and *S* components, which appear to have been derived through the fracturing of an intact ancestral component. The model must take into account the presence

of transposable element sequences that are present adjacent to the *q* and *S* breakpoints, as well as the lack of a transposable element footprint at the corresponding site at *P*. Finally, the model should take into account the formation of the inverted duplication which is the *S1* component.

We propose a three-step model for the formation of the *R-r* complex. In the first step, we postulate that a simplex ancestral *R* gene (*P-ancestral*) became duplicated. The exact mechanism of this is not known; a possible mechanism for this duplication would be through crossing over between duplicated segments outside of *R*. We propose this as a first step to account for two observations: the physical distance between *P* and the *S* subcomplex and the lack of a transposable element footprint at the *P* component of *R-r*. Thus, the distance between *P* and the *S* subcomplex would represent the length of this original sequence duplication. As a second step, we propose that the *doppia* element moved into the distal gene copy to produce the intermediate component, *p2-mutable* (see Figure 9). Had the initial event been insertion of a transposable element within the simplex *R* progenitor allele, we would expect to have observed some sort of modification at the target site within all components. Instead we find that the sequence of *P* in this region is



**Fig. 9.** Formation of the *S* subcomplex of *R-r*. We postulate that an ancestral *P*-like gene became duplicated, thus forming two *P*-like components, *P1* and *P2*, with *P2* distal to *P1* (not shown). Insertion of *doppia* into the *P2* component could result in the formation of *p2-mutable*. The *S* subcomplex could have formed from *p2-mutable*. The *doppia* element at *p2-mutable* may have been the site of a chromosome breakage event, possibly due to an aberrant transposition event. This breakage resulted in the separation of the 5' and 3' ends of *p2-mutable* to form the *q* and *S* components. Sister chromatid fusion could lead to the formation of the  $\sigma$  region between the two *S* genes, with subsequent healing of the fused chromosome fragment joining *q* and the duplicated *S* genes at a distance of 10 kb. *Doppia* sequences left behind during these events are indicated by light and dark arrows.

extremely similar to that of the maize *B* gene, and to an *R* gene from teosinte (Figure 3D). Hence we consider it unlikely that the extant *P* component was ever visited by a transposable element at this site.

As a final step we propose that an aberrant transposition event fractured *p2-mutable* to generate the *S* subcomplex (see Figure 9). Although the precise mechanism of the gene fracturing event(s) that gave rise to the *S* subcomplex are not known, it seems likely that the transposable element provided the instability needed for restructuring and the generation of a clustered family of related regulatory genes. Chromosomal restructuring mediated by aberrant transposition events, including inversions, deletions and duplications has been observed in maize (Weil and Wessler, 1993), *Antirrhinum* (Martin and Lister, 1989; Robbins et al., 1989a) and *Drosophila* (Tsubota et al., 1989; Montgomery et al., 1991; Sheen et al., 1993). In

each of these systems, a copy of the transposable element responsible for the restructuring event often remains at the site of rearrangement.

The chromosome breakage, replication and chromatid fusion events depicted in Figure 9 may not have occurred in the order shown. It is possible, for instance, that all three events took place at one time, e.g. during an abortive transposition event of *doppia*. Molecular evidence indicates that the right inverted repeat of *doppia* suffered a deletion of 5 bp prior to the formation of *S1*, since both *S1* and *S2* have this deletion. Therefore, it is possible that the mutation in the inverted repeat at the *doppia* ends found at  $\sigma$  may have caused or resulted from an abortive transposition, responsible for the formation of the *S* subcomplex. If the abortive transposition took place during a period of DNA replication, as has been shown to be the case for *Ac* (Chen et al., 1992), then a replication fork, moving in the distal to proximal direction, having passed through the coding region of *p2-mutable* may have encountered a transpositional intermediate that initiated a chromosomal breakage–fusion event, resulting in the displacement of *q* and duplication and inversion of *S*.

### *S* genes may share promoter elements

Cloned genomic DNA from each of the three functional *R* components was capable of producing anthocyanin pigmentation of appropriate tissues in a microprojectile bombardment assay. This assay was used to confirm that both *S1* and *S2* are functional components. RT-PCR analysis suggests that *S1* mRNA may represent the majority of *S* expression in the aleurone. While it is not possible to assess whether the *S1* and *S2* genes are controlled by separate or overlapping promoters, it is important to note the extremely limited size of the region intervening between these two genes. The unique  $\sigma$  region is only 387 bp in length (see Figure 4) bordered by 160 bp of *R* sequence upstream before reaching transcribed *S* sequences (Perrot and Cone, 1989). This leaves a maximum of 707 bp (387 + 160 + 160) to include the promoters for both genes. The 387 bp  $\sigma$  region is itself critical for the function of both the *S1* and *S2* genes. Evidence for the importance of  $\sigma$  comes from the *r-r:NI-3-1* derivative allele in which there has been a nearly precise deletion of  $\sigma$  which left the rest of *S1* and *S2* intact, and yet rendered both *S* genes non-functional.

The  $\sigma$  region containing the promoter(s) of *S1* and *S2* was formed by the fusion of *doppia* sequences directly to *P* coding sequences along with the insertion of rearranged filler DNA. The fusion of *doppia* and filler DNA to the *P* coding region directs the expression of the *R* product, a transcriptional regulator of anthocyanin biosynthetic genes, in the aleurone of the seed. Thus, aleurone expression could have been conferred by *doppia* sequences themselves, by the rearranged region, or by juxtaposition of the two. It is intriguing to note that the *S* genes of *R-r* lack the first six amino acids of an upstream open reading frame (ORF) that has been shown to repress expression of the related *Lc* gene in aleurone (Damiani and Wessler, 1993). The absence of aleurone repression mediated by this  $\mu$ ORF may be partially responsible for allowing the expression of the *S* genes in aleurone. Elucidation of the exact nature of the aleurone enhancer element will be important to distinguish whether the *doppia* sequences

themselves are directing aleurone-specific expression or whether aleurone expression is the result of a fortuitous new combination of sequences at the promoter region of the *S* genes.

### Meiotic instability of *R-r* and formation of the *r-r* NCO alleles

Approximately 3/10 of the *r-r* derivatives that have been examined are apparently not the result of crossing over between *P* and *S2* (Stadler and Emmerling, 1956; Dooner, 1971; Kermicle and Axtell, 1981) and behave genetically as *P s*, i.e. they retain the duplicated structure but have a non-functional *S* component. Every *P s* derivative that was examined at a molecular level has suffered a deletion encompassing the  $\sigma$  region. Thus, retention of the genetic duplication in *r-r* NCO derivative formation is correlated with loss of the  $\sigma$  region between *S1* and *S2*. Moreover, sequence analysis of the *r-r:NI-3-1* allele, which has suffered a deletion that almost precisely deletes  $\sigma$ , implicates the  $\sigma$  region as the promoter for both the *S1* and *S2* genes.

Each of the three *r-r* NCO alleles that were sequenced contained 'filler' DNA at the deletion endpoints. In the case of the *r-r:n156* allele, the origin of the filler DNA could not be established. In the *r-r:NI-3-1* and *r-r:NI-3-2* alleles, the filler DNA was derived from nearby sequences but was complex in the sense that the filler DNA was derived from two or more discontinuous stretches of DNA which often partially overlapped. Several spontaneous deletions of the *waxy* and *bronze1* loci were similarly found to contain filler DNA from nearby sequences (Ralston *et al.*, 1988; Wessler *et al.*, 1990). However, in these cases, the filler DNA was simple, i.e. it was derived from a single intact stretch of nearby sequence. In addition, the nearby sequence which gave rise to the filler DNA of the *wx* and *bz1* deletions was flanked by short sequences homologous to the deletion endpoints (Wessler *et al.*, 1990). This is not the case for *r-r:NI-3-1* and *r-r:NI-3-2*, which do not have homologous flanking sequences in the donor DNA.

Filler DNA is also often found at rearrangement junctions in mammalian cells. In these cases it is important to distinguish between junctions formed by immune rearrangements (i.e. addition of N-regions) in which addition of extra nucleotides is caused by terminal deoxynucleotidyl transferase activity (Alt and Baltimore, 1982; Landau *et al.*, 1987) and non-immune rearrangements which appear to result from other mechanisms (Roth *et al.*, 1989). Various mechanisms that have been suggested to account for the presence of filler DNA in non-immune rearrangement junctions include addition of oligonucleotides from a cellular pool (Roth *et al.*, 1989), and slipped mispairing and repair synthesis (Roth *et al.*, 1985, 1989). A modified version of the slipped mispairing model has also been invoked to account for the filler DNA present at *wx* and *bz1* deletion alleles (Wessler *et al.*, 1990). The presence of two or four discontinuous stretches of nearby sequences at the deletion junction of *r-r:NI-3-1* and *r-r:NI-3-2* is not readily explained by either of these models. It is difficult to imagine how the use of a cellular pool of oligonucleotides would lead to addition of sequences from near the rearrangement endpoints; but it is also difficult to account for the presence of more

than one continuous stretch of DNA using the slipped mispairing model. To do so would necessitate two (or more) slipped mispairing events to account for the fact that the motifs overlap and are present in filler DNA in different orders than in the original donor DNA.

A more attractive model accounting for the complex structure of the filler DNA in the *r-r* NCO derivatives is that it results from a gap repair mechanism needed to join two broken chromosome ends. Repair would occur off templates located nearby. Short stretches of sequence would be added and the novel DNA would be used continually to search for short homologies that would allow joining of the two broken ends. This repair-and-search type of mechanism would result in the series of slightly overlapping motifs found in the filler DNA of *r-r:NI-3-1* and *r-r:NI-3-2*, and would account for the slight overlap of the last filler DNA motif with the endpoint of the deletion.

Two complex *niv<sup>rec</sup>* alleles of *Antirrhinum majus*, *niv<sup>rec</sup>:554* and *niv<sup>rec</sup>:557* (Martin and Lister, 1989), share some striking similarities to the *IS*  $\sigma$  *S2* region of *R-r* and may shed light on the mechanism of *S* instability. Both the *niv<sup>rec</sup>:554* and *niv<sup>rec</sup>:557* alleles resulted from a *Tam3*-mediated rearrangement of the *niv* locus in which the *Tam3* element remained at its original position in the *niv* promoter but an inverted duplication of *niv* sequences was generated just downstream of *Tam3*. These structures are similar to the *S* components of *R-r* in that in each case there is a long inverted duplication in close association with a transposable element. The *niv<sup>rec</sup>:554* and *niv<sup>rec</sup>:557* alleles are unstable, yet simple excision does not restore *niv* function; at least part of the inverted duplication must also be deleted. The *S* components of *R-r* are also unstable due to the formation of deletions. The deletion derivatives resulting from both the *niv<sup>rec</sup>:554* and *niv<sup>rec</sup>:557* alleles and the *S* components of *R-r* have filler DNA present at the deletion junctions. The origin of the filler DNA in the *Niv* derivative alleles is unknown, but does not appear to come from nearby sequences.

Instability of the associated transposon and inverted duplication in the *niv<sup>rec</sup>:554* and *niv<sup>rec</sup>:557* alleles is clearly due to transposition-related functions, since conditions which suppress *Tam3* transposition reduce the instability of *niv<sup>rec</sup>:554* and *niv<sup>rec</sup>:557*. Furthermore, following simple excision of the *Tam3* element, the inverted duplication becomes stable. Thus the close association of an active transposable element with an inverted duplication causes the instability of the whole structure. It is interesting to speculate that the *r-r* NCO alleles are the result of transposition activity involving *doppia* sequences which, while probably incapable of excision, might still be recognized by transposase supplied in *trans* from an active *doppia* element elsewhere in the genome. This would suggest that the transposable element responsible for the formation of the *R-r* complex is still active in causing instability of *R-r*.

Large differences in the frequency of NCO derivative formation from hemizygous versus homozygous or heterozygous *R-r* stocks suggested that hemizygoty itself might increase the lability of *S* (Kermicle and Axtell, 1981). It is intriguing, therefore that of the *r-r* NCO alleles that were sequenced in the present study, the two derived from hemizygotes (*r-r:NI-3-1* and *r-r:NI-3-2*) have a deletion

breakpoint located within or very close to the the right inverted repeat sequence of *doppia*. No indication of involvement of the *doppia* sequences was found for the *r-r:n156* derivative, which was derived from a homozygous *R-r* stock. If our speculation that the *r-r* NCO derivatives result from latent transposition activity of  $\sigma$  is correct, then the high frequency of NCO derivative formation from hemizygous *R-r* stocks noted by Kermicle and Axtell (1981) could be explained by the presence of an active *doppia* element in this stock. Hemizygosity might itself play a role in allowing transposition activity at  $\sigma$  by promoting association of the left inverted repeat at *q* with the right inverted repeat at  $\sigma$ ; a configuration which might be necessary for transposase activity on the  $\sigma$  end. Alternatively, hemizygosity could promote *S* instability by lack of normal synapsis which could promote the formation of secondary structures, e.g. hairpins, that would be substrates for cellular enzymes. In this case it would not be necessary to postulate that *doppia* is still active in promoting instability of the *R* region.

## Materials and methods

The *R-r:standard* allele used in this study is typical of the A group of *R* alleles described by Stadler (1948). In the W22 genetic background used in this study this allele confers strong pigmentation of the aleurone of the seed, the coleoptile, the roots and leaf tip of seedlings, and to the roots and anthers of mature plants. The isolation and characterization of many of the derivative alleles used in this study have been described (Dooner, 1971; Dooner and Kermicle, 1974; Kermicle and Axtell, 1981; Demopoulos, 1985), and summarized (Robbins *et al.*, 1991) previously.

The terminology used to describe *R-r* and its derivatives is as follows: the first part of the symbol describes the color of the seed, *R* = colored aleurone, *r* = colorless aleurone. The second part of the symbol describes the color of plant parts, with *-r* = pigmented plant parts and *-g* = green plant parts with no anthocyanin pigment. Hence, *R-r* pigments plant and seed, *R-g* pigments only the seed, *r-r* pigments only the plant and *r-g* produces no anthocyanin.

### Cloning and sequencing

*R-r* genomic libraries were prepared with partial *Sau3A*-digested W22 *R-r* DNA size-fractionated on glycerol gradients (Sambrook *et al.*, 1989). Partially digested DNA of the appropriate molecular weight (15–20 kbp) was ligated into prepared  $\lambda$ DashII vector arms (Stratagene) according to the manufacturers instructions, packaged using Gigapack Gold (Stratagene), and plated on *Escherichia coli* strain ER1647 (New England Biolabs). Libraries were screened with a mixture of pR-nj:1, R-sc:3231, 4.7H2.5 and 4.7E1.8 probes, to isolate all *R*-homologous sequences represented in the library. Size-selected genomic libraries were constructed from homozygous *r-r:NI-3-1*, *r-r:NI-3-2*, and *r-r:n156* DNA digested to completion with *HindIII* and fractionated in 1% agarose gels. The DNA from the appropriate size fraction was excised from the gel, purified using GeneClean (Bio101) according to the manufacturers instructions, ligated to prepared arms of  $\lambda$ NM1149, as previously described (Robbins *et al.*, 1989b), and packaged using Gigapack Gold (Stratagene). These libraries were screened with the probe S2.0AH. The 5' region of each *R* component was subcloned as a *HindIII* fragment into the plasmid pTZ18U (US Biochemical). Sequencing reactions were performed on both strands of double-stranded plasmid DNA using Sequenase II (US Biochemical) according to the manufacturer's instructions.

A small (100 bp) portion of an *R* gene from teosinte was amplified by the polymerase chain reaction (PCR) using the primers oR1 (5' TCTACCCTTTCGCATGGAAGTTCTT 3') and oR3 (5' TCTACTGAT-CATCAGACGATGCTT 3'). The product was cloned into the vector pCRII (Clontech). Sequencing reactions were performed on both strands of double-stranded plasmid DNA using Sequenase II (US Biochemical) according to the manufacturer's instructions.

### DNA isolation and genomic blot analysis

Genomic DNA was isolated as described previously (Chen *et al.*, 1987) and was transferred to Zeta-probe membranes (Biorad) according to the

manufacturer's instructions. Membranes were hybridized and washed as described previously (Dellaporta *et al.*, 1988).

**Preparation of DNA probes.** The pR-nj:1, R-sc:3231, 4.7H2.5 and 4.7E1.8 probes have been described previously (Dellaporta *et al.*, 1988; Robbins *et al.*, 1991). The pR-nj:1, R-sc:3231, 4.7H2.5 and 4.7E1.8 probes contain sequences from the 5' end of the *R* transcription unit, the coding region, the 3' transcribed, untranslated region (E.Walker, unpublished data) and a 3' flanking region from *P*, respectively. The S2.0AH probe is two *AvaI*–*HindIII* fragments derived from the *S2* genomic clone,  $\lambda$ 3D (Robbins *et al.*, 1991), as shown in Figure 2. These fragments correspond to a region of each *S* component from the translation start into the large first intron (E.Walker, unpublished results). The  $\sigma$ 1011 probe was generated by PCR, as follows. The primers oR10 (5' TTTGCCACC-ACATACTTCATGAGCA 3') and oR11 (5' TCGTGGGCCAAAACA-CGAAAATA 3'), which prime DNA synthesis at the borders of the  $\sigma$  region were used to generate a DNA fragment from  $\lambda$ R3D. The appropriately sized fragment was subcloned into the plasmid, pTZ18 (US Biochemical). The probe fragment, s1011, was prepared by digestion with *KpnI*, which removes a small (116 bp) highly repetitive region from one end of the cloned PCR fragment, and releasing the 230 bp fragment,  $\sigma$ 1011, shown in Figure 2. DNA fragments were labeled in low melting temperature agarose by the random priming method (Feinberg and Vogelstein, 1984).

### PCR analysis

cDNA was prepared by reverse transcription of RNA from 21-day post-fertilization aleurones as described (Consonni *et al.*, 1992). Primers oR31A (5' ATGGCTTCATGGGGCTTAAGATAC 3') and oR32 (5' GAATGCAACCAACACCTTATGCC 3'), which correspond to position 1893 t–1916 and 2273–2297 of the *S* cDNA clone of Perrot and Cone (1989), were used in the amplification reactions. This primer set amplifies across the last intron of *R* and allows the products of (intronless) cDNA amplification to be distinguished from those of (intron-containing) genomic DNA. PCR amplification was performed in a 50  $\mu$ l reaction containing: 1 ng plasmid or ~10–50 ng cDNA; 1 $\times$  polymerase buffer (Promega); 2 mM MgCl<sub>2</sub>; 2% formamide; 4  $\mu$ M each dATP, dTTP, dGTP and dCTP; 1 U Taq DNA polymerase (Promega). Reactions were carried out for 30 cycles of 45 min 95°C denaturation, 1 h 55°C annealing, 2 h 72°C polymerization. Restriction enzyme digestions were performed by making a 4-fold dilution of the PCR reaction into the appropriate restriction buffer with 5–10 U of restriction enzyme for 16 h at 37°C. Control plasmids for the PCR reactions were pP2.5H, which contains the 2.5 kbp *HindIII* fragment of  $\lambda$ R4.7 (Robbins *et al.*, 1991); pS110.0H, which contains the 10 kbp *HindIII* fragment of IR68; and pS22.5, which contains the 2.5 kbp *HindIII* fragment of  $\lambda$ R3D (Robbins *et al.*, 1991). All subclones were made in the vector pBluescript SK+ (Stratagene).

### Microprojectile bombardment of maize tissues

Seedlings and intact aleurones from mature, dry seeds were prepared according to Ludwig *et al.* (1990). Gold particles were coated with DNA inserts of  $\lambda$  clones  $\lambda$ 3D,  $\lambda$ 4.7, and  $\lambda$ 55 by precipitation as described (Ludwig *et al.*, 1990). Tissue was bombarded and incubated at 25°C under constant illumination. Photographs were taken 2 days after bombardment.

## Acknowledgements

We thank Tom Brutnell and Alison Delong for critical reading and helpful comments on the manuscript, Lisa Shuster for technical assistance, and Suzy Pafka for photography. This work was supported by Department of Energy, Office of Energy Research (DE-FG02–86ER13627) to S.L.D.

## References

- Alt,F.W. and Baltimore,D. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 4118–4122.
- Chandler,V.L., Radicella,J.P., Robbins,T.P., Chen,J. and Turks.D. (1989) *Plant Cell*, **1**, 1175–1183.
- Chen,J., Greenblatt,I.M. and Dellaporta,S.L. (1987) *Genetics*, **117**, 109–116.
- Chen,J., Greenblatt,I.M. and Dellaporta,S.L. (1992) *Genetics*, **130**, 665–676.
- Coccolone,S.M. and Cone,K.C. (1993) *Genetics*, **135**, 575–588.
- Cone,K.C., Coccolone,S.M., Moehlenkamp,C.A., Weber,T., Drummond,

- B.J., Tagliani, L.A., Bowen, B.A. and Perrot, G.H. (1993) *Plant Cell*, **5**, 1807–1816.
- Consonni, G., Viotti, A., Dellaporta, S.L. and Tonelli, C. (1992) *Nucleic Acids Res.*, **20**, 373.
- Damiani, R.D. Jr and Wessler, S.R. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8244–8248.
- Dellaporta, S.L., Greenblatt, I.M., Kermicle, J.L., Hicks, J.B. and Wessler, S.R. (1988) In Gustafson, J.P. and Appels, R. (eds), *Molecular Cloning of the R-*rj* Gene by Transposon Tagging with Ac*. Plenum Press, New York, Vol. 11, pp. 263–281.
- Demopoulos, J.T. (1985) *Characterization and Fine Structure Analysis of Mutants at the R Locus of Maize*, Ph.D. Thesis, University of Wisconsin.
- Dooner, H.K. (1971) *Recombinational Analysis of the R-r Duplication in Maize*, Ph.D. Thesis, University of Wisconsin.
- Dooner, H.K. and Kermicle, J.L. (1971) *Genetics*, **67**, 427–436.
- Dooner, H.K. and Kermicle, J.L. (1974) *Genetics*, **78**, 691–701.
- Dooner, H.K., Jorgenson, R. and Robbins, T.P. (1991) *Annu. Rev. Genet.*, **25**, 173–199.
- Feinberg, A.P. and Vogelstein, B. (1984) *Anal. Biochem.*, **137**, 266–267.
- Goff, S.A., Klein, T.M., Roth, B.A., Fromm, M.E., Cone, K.C., Radicella, J.P. and Chandler, V.L. (1990) *EMBO J.*, **9**, 2517–2522.
- Kermicle, J.L. (1985) In Freeling, M. (ed.), *Alternative Tests of Allelism*. Alan R. Liss, New York, Vol. 35, pp. 491–507.
- Kermicle, J.L. (1988) In Nelson, O.E. (ed.), *Recombinant Mutable Alleles of the Maize R Gene*. Plenum Press, New York, pp. 81–89.
- Kermicle, J.L. and Axtell, J.D. (1981) *Maydica*, **26**, 185–197.
- Landau, N.R., Schatz, D.G., Rosa, M. and Baltimore, D. (1987) *Mol. Cell Biol.*, **7**, 3237–3243.
- Ludwig, S.R., Habera, L.F., Dellaporta, S.L. and Wessler, S.R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7092–7096.
- Ludwig, S.R., Bowen, B., Beach, L. and Wessler, S.R. (1990) *Science*, **247**, 449–450.
- Martin, C. and Lister, C. (1989) *Devel. Genet.*, **10**, 438–451.
- Montgomery, E.A., Huang, S.-M., Langley, C.H. and Judd, B.H. (1991) *Genetics*, **129**, 1085–1098.
- Perrot, G.H. and Cone, K.C. (1989) *Nucleic Acid Res.*, **17**, 8003.
- Ralston, E.J., English, J.J. and Dooner, H.K. (1988) *Genetics*, **119**, 185–197.
- Robbins, T.P., Carpenter, R. and Coen, E.S. (1989a) *EMBO J.*, **8**, 5–13.
- Robbins, T.P., Chen, J., Norell, M.A. and Dellaporta, S.L. (1989b) In Styles, D.E., Gavazzi, G.A. and Racchi, M.L. (eds), *Molecular and Genetic Analysis of the R and B Loci in Maize*. Edizioni Unicopli, Milan, pp. 105–114.
- Robbins, T.P., Walker, E.L., Kermicle, J.L., Alleman, M. and Dellaporta, S.L. (1991) *Genetics*, **129**, 271–283.
- Roth, D.B., Porter, T.N. and Wilson, J.H. (1985) *Mol. Cell Biol.*, **5**, 2599–2607.
- Roth, D.B., Chang, X.-B. and Wilson, J.H. (1989) *Mol Cell Biol.*, **9**, 3049–3057.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Stadler, L.J. (1942) *Some Observations on Gene Variability and Spontaneous Mutation*. Michigan State College, East Lansing, MI, pp. 3–15.
- Stadler, L.J. (1948) *Am. Nat.*, **82**, 289–314.
- Stadler, L.J. and Emmerling, M. (1956) *Genetics*, **41**, 124–137.
- Stadler, L.J. and Nueffer, M. (1953) *Science*, **117**, 471–472 (abstr.).
- Tonelli, C., Consonni, G., Dolfini, S.F., Dellaporta, S.L., Viotti, A. and Gavazzi, G. (1991) *Mol. Gen. Genet.*, **225**, 401–410.
- Tsubota, S.I., Rosenberg, D., Szostak, H., Rubin, D. and Schedl, P. (1989) *Genetics*, **122**, 881–890.
- Weil, C.F. and Wessler, S.R. (1993) *Plant Cell*, **5**, 515–522.
- Wessler, S., Tarpley, A., Purugganan, M., Spell, M. and Okagaki, R. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8731–8735.

Received on November 8, 1994; revised on February 2, 1995