Plant Molecular Biology **42:** 667–678, 2000.

The rice *R* gene family: two distinct subfamilies containing several miniature inverted-repeat transposable elements

Jianping Hu[1], Vaka S. Reddy and Susan R. Wessler*
*Departments of Botany and Genetics, University of Georgia, Athens, GA 30602, USA (*author for correspondence); [1]present address: Plant Biology Lab, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA*

Received 17 August 1999; accepted in revised form 13 December 1999

Key words: anthocyanin, gene family, genome evolution, rice, transposable element

Abstract

The *R* and *B* genes of maize regulate the anthocyanin biosynthetic pathway and constitute a small gene family whose evolution has been shaped by polyploidization and transposable element activity. To compare the evolution of regulatory genes in the distinct but related genomes of rice and maize, we previously isolated two *R* homologues from rice (*Oryza sativa*). The *Ra1* gene on chromosome 4 can activate the anthocyanin pathway, whereas the *Rb* gene, of undetermined function, maps to chromosome 1. In this study, rice *R* genes have been further characterized. First, we found that an *Rb* cDNA can induce pigmentation in maize suspension cells. Second, another rice *R* homologue (*Ra2*) was identified that is more closely related to *Ra1* than to *Rb*. Domesticated rice and its wild relatives harbor multiple *Ra*-like and *Rb*-like genes despite the fact that rice is a true diploid with the smallest genome of all the grass species analyzed to date. Finally, several miniature inverted-repeat transposable elements (MITEs) were found in *R* family members. Their possible role in hastening the divergence of *R* genes is discussed.

Introduction

The maize *R* and *B* genes comprise a small gene family of regulatory proteins that act in concert with proteins encoded by the *C1/Pl* gene family to activate most of the structural genes in the anthocyanin biosynthetic pathway (Ludwig and Wessler, 1990; Dooner and Robbins, 1991 for reviews). Genetic analyses found that the *R* genes on chromosome 10 and the *B* gene on chromosome 2 show extensive allelic diversity and are largely responsible for the distinct pigmentation patterns of different maize strains (Styles *et al.*, 1973). Molecular studies demonstrated that *R* and *B* genes encode homologous proteins containing the basic helix-loop-helix (bHLH) motif (Chandler *et al.*, 1989; Ludwig *et al.*, 1989) and nuclear localization signals (Shieh *et al.*, 1993). Furthermore, maize R and B proteins can induce pigmentation in other monocot and dicot plant species, such as wheat, barley, *Arabidopsis* and petunia (Lloyd *et al.*, 1992; Bilang *et al.*, 1993; Casas *et al.*, 1993; Galway *et al.*, 1994). Homologues of the maize *R* and *B* genes from *Antirrhinum* (*Delila*; Goodrich *et al.*, 1992), petunia (*Jaf13*; Quattrocchio *et al.*, 1998), and rice (Hu *et al.*, 1996) also regulate anthocyanin pigmentation.

The evolution of regulatory proteins and their target genes is believed to play a key role in the diversification of organisms (Dickinson, 1991). A phylogenetic analysis suggested that *R* genes have diverged independently and rapidly in different grass genomes (Purugganan and Wessler, 1994). As such, the *R* gene family may serve as a model system for the study of evolution and diversification of regulatory genes in the grasses. Genetic and molecular studies indicate that the organization of the maize *R/B* gene family has been shaped by polyploidization and transposable element activity (Anderson, 1945; Helentjaris *et al.*, 1988; Robbins *et al.*, 1989; Walker *et al.*, 1995). The study of *R* genes from rice was initiated to understand how this gene family has evolved in a distantly related grass species. Like maize, rice strains exhibit diverse patterns of pigmentation that may result from

the proliferation of *R* genes and/or alleles. Although the genomes of these two distant relatives in the grass family display synteny (Helentjaris *et al.*, 1988; Ahn and Tanksley, 1993), the genome of rice is six times smaller than that of maize (Arumuganathan and Earle, 1991). This difference is due in small part to a difference in ploidy (rice is a true diploid whereas maize is an ancient tetraploid; Helentjaris *et al.*, 1988; Oka, 1988), and in large part to much less retrotransposition in rice (Chen and Bennetzen, 1996; San Miguel *et al.*, 1996; Chen *et al.*, 1997).

Two *R* genes were previously identified in domesticated rice, *O. sativa* (Hu *et al.*, 1996). The active *Ra1* gene (formerly named *Ra*) shows homology along its entire length with the maize *R* (*Lc*) gene (Ludwig *et al.*, 1989) and the *Antirrhinum Delila* gene (Goodrich *et al.*, 1992) and is capable of activating the anthocyanin pathway in maize. The location of *Ra1* on chromosome 4 is in synteny with regions of maize chromosomes 2 and 10 that contain the *B* and *R* loci, respectively. A second *R* gene (*Rb*) of unknown function mapped to chromosome 1. Our data suggested that the common ancestor of rice and maize had a single *R* gene and that the *R* gene families of rice and maize evolved after the divergence of these two members of the grass family.

To see if *Ra* and *Rb* play the same pivotal role in controlling pigmentation patterns as the maize *R/B* genes, it was necessary to first determine whether *Rb* encodes a functional transcriptional activator. In addition, we were interested in determining whether there were additional members of this gene family. Here we describe the cloning of a functional *Rb* gene, which can activate the maize anthocyanin biosynthetic pathway. We also report the isolation of a third *R* gene, *Ra2*, that is more closely related to *Ra1* than to *Rb*. Our data show that most rice strains contain multiple copies of *Ra*-like and *Rb*-like genes, and that miniature inverted-repeat transposable elements (MITEs) are in close association with rice *R* genes where they may act to promote the divergence of family members.

Materials and methods

Rice strains

Rice strains used in this study were as follows: *Oryza sativa* strains JP004, T65, and A58, provided by Shigetoshi Sato (University of the Ryukyus, Japan); *O. sativa* strains Purple 522, BS125 and IR36, and the wild species *O. glumeapatula* and *O. longistaminata*, obtained from Gary Kochert (University of Georgia); wild species *O. glaberrima*, *O. barthii* and *O. meridionalis*, from the International Rice Research Institute (IRRI) at the Philippines. Purple 522 has uniform pigmentation of the leaf. JP004 has pigmented leaf mid-rib, collar and auricle. Other strains used in this study are all unpigmented.

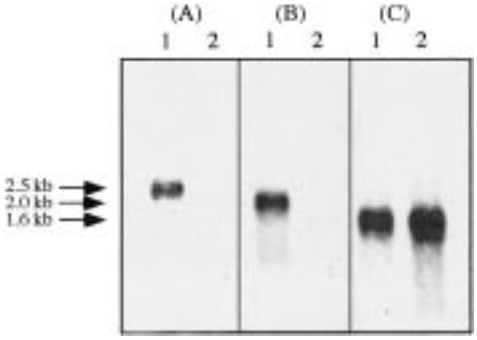


Figure 1. RNA blot analysis of transcripts encoded by *Ra1* and *Rb*. A gel blot containing poly(A)$^+$ RNA from leaf tissue of Purple 522 (lane 1) and JP004 (lane 2) was hybridized with a 450 bp *Ra1* cDNA fragment containing the bHLH region (A). After removing the first probe, the same filter was reprobed with a 550 bp *Rb1* genomic fragment (Hu *et al.*, 1996), and a 500 bp *Hin*dIII/*Sst*I fragment of the maize actin cDNA (C) after the *Rb1* probe was removed.

Nucleic acid isolation and blot analysis

DNA isolation and DNA blot analysis were performed as described (Hu *et al.*, 1996). *Hin*dIII was used to digest genomic DNA in all of the DNA blot analyses in this study except in Figure 4D where *Sph*I was used. High-stringency hybridization (65 °C) and washing (0.1× SSC, 0.5% SDS at 65 °C) conditions were used for all of the DNA blots except for lane 1 in Figure 4A, where conditions of intermediate stringency (60 °C hybridization and 1× SSC, 0.5% SDS washes at 60 °C) were used.

Total RNA was isolated from 4 g of leaf tissue from Purple 522 and JP004, following the manufacturer's instructions for Trizol Reagent (Gibco-BRL). Poly(A)$^+$ RNA was extracted from total RNA using the Poly(A) Tract mRNA Isolation System IV (Promega). RNA blot analysis was conducted as described (Hu *et al.*, 1996). Stringencies employed were as follows: high stringency (used in Figure 1A and 1B), 65 °C hybridization and 0.1× SSC, 0.5% SDS washes at 65 °C; low stringency (used in Figure 1C), 55 °C hybridization and 2× SSC, 0.5% SDS washes at 55 °C. The maize actin clone used as a probe in the

RNA blot analysis was obtained from Rich Meagher (University of Georgia).

PCR reaction

The 720 bp *Ra2* fragment was amplified from rice genomic DNA using degenerate primers P1 (5′-GGGAACGGCAARAANCAYGTNATG-3′) and P2 (5′-AGGTGCRTCRAANACNCKNGTCAT-3′) and conditions as described (Purugganan and Wessler, 1994). Primers P3 (5′-ACGCACGACACTCTGGTT-GC-3′) from exon 2 and P4 (5′-TGGGCGGCTGGTT-GAAATGG-3′) from exon 4, and P5 (5′-CTGACAA-ACGCTC AGTCTGC-3′) from exon 6 and P6 (5′-CGATACGGT CGACGAGAGCC-3′) from exon 9 of the *Ra1* gene, were used to amplify additional sequences of *Ra2*. Reactions with P3, P4, P5 and P6 were also carried out as described (Purugganan and Wessler, 1994), with an annealing temperature of 60 °C. Primer positions are depicted in Figure 5A.

Isolation of an Rb1 *genomic fragment containing the* Stow 5 *element*

An *Xba*I fragment containing part of the *Rb1* gene was isolated from a Purple 522 genomic library that was probed with a 550 bp fragment from the *Rb1* gene (Hu *et al.*, 1996). A 2.7 kb subclone, which includes sequence identical to the 550 bp fragment, was found to contain the *Stow 5* element.

Isolation of a 1.9 kb Rb2 *cDNA clone*

Poly(A)$^+$ RNA isolated from Purple 522 leaf tissue was used for cDNA library construction with the ZAP cDNA Synthesis Kit (Stratagene). Screening of the library with the 550 bp *Rb1* fragment (Hu *et al.*, 1996) led to the isolation of a 1.9 kb *Rb2* cDNA clone.

Amplification of an Rb2 *3′ RACE product containing the* Stow 6 *element*

3′ RACE was conducted using the 3′ RACE System for Rapid Amplification of cDNA Ends (Gibco-BRL), with an annealing temperature of 56 °C for the PCR amplification. The *Rb*-specific primer derived from the putative exon 9 of *Rb1* and *Rb2* was 5′-GTCCCGTCCATCGACAAG-3′, shown as P7 in Figure 6A. A ca. 700 bp RACE product was amplified, and found to be identical to the corresponding region in the 1.9 kb *Rb2* cDNA clone, except that the two clones have different polyadenylation sites (depicted in Figure 7D). The 700 bp RACE product contains the *Stow 6* element while the 1.9 kb *Rb2* cDNA clone does not.

Bombardment assay

The expression vector, pRiceRb2, containing the *Rb2* cDNA, was constructed by inserting the 1.9 kb *Eco*RI/*Xho*I *Rb2* fragment between the CaMV 35S promoter and the nopaline synthase terminator in pBI221 (Clontech). Plasmid DNAs were precipitated onto 1.0 μm gold particles (60 mg/ml; Klein *et al.*, 1989) and bombarded into maize suspension cells (Goff *et al.*, 1990) with the Biolistic PDS-1000 (Dupont, Wilmington, DE). Each plate of cells was co-bombarded with 1.0 μg of *R* plasmid DNA (pRice-Rb2, or the pLcΔ control that contains the maize *Lc* gene used in Damiani and Wessler, 1993) and 1.0 μg of p35SC1 plasmid containing the maize *C1* gene (Goff *et al.*, 1990), or p35SC1 alone. The plates were then incubated at 28 °C for 48 h.

DNA cloning, sequencing, alignment and database searches

For cloning purposes, PCR products were gel-purified using the QIAquick Gel Purification Kit (Qiagen), and cloned into the TA vector using the Original TA Cloning Kit (Invitrogen). Plasmids were prepared using the QIAprep Spin Miniprep Kit (Qiagen). DNA was sequenced by the University of Georgia Molecular Genetics Facility. Sequences were aligned with the LINEUP, PILEUP and BOXSHADE programs of the GCG package (Wisconsin package, version 8, 1994). Database searches were conducted using BLAST (National Center for Biotechnology Information, National Institute of Health, Bethesda, MD).

Results

Cloning and characterization of a functional Rb *gene*

As a first step in determining whether the previously identified *Rb* gene located on chromosome 1 encoded a functional protein, it was necessary to ascertain whether it was expressed in pigmented rice tissue. To this end, a 550 bp *Rb* genomic fragment (including parts of exon 9 and 10 and all of intron 9; Hu *et al.*, 1996) was used to probe an RNA blot containing poly(A)$^+$ RNA isolated from the pigmented leaves

of two strains, Purple 522 and JP004. Under high-stringency conditions, an *Ra* cDNA probe detected the 2.5 kb *Ra* mRNA in Purple 522 (Figure 1A). The *Rb* fragment, which is 70% identical to the sequence of the *Ra* probe in the exonic region, hybridized with a ca. 2.0 kb transcript from Purple 522 (Figure 1B). Interestingly, the pigmented leaves of strain JP004 did not contain detectable levels of either transcript (see below).

To obtain a full-length *Rb* cDNA, the 550 bp fragment was used to screen a cDNA library constructed from RNA isolated from the young leaves of Purple 522. A 1.9 kb *Rb* cDNA was isolated and found to be 96% identical to the sequence of the 550 bp genomic fragment in the exonic region (data not shown). We therefore predicted that Purple 522 contains at least two very similar *Rb*-like genes, which we named *Rb1* and *Rb2*. The 550 bp fragment was derived from *Rb1*, while the *Rb2* gene encoded the 1.9 kb cDNA. The amino acid sequence deduced from the *Rb2* cDNA shares extensive similarity with *Ra* and with the maize *R* (*Lc*) gene (Figure 2).

A vector containing a transcriptional fusion of the *Rb2* cDNA and the 35S CaMV promoter was constructed to determine whether *Rb2* encodes a protein capable of activating the anthocyanin pathway. Like the maize *R* (*Lc*) control (Figure 3A), co-delivery of the *Rb2* plasmid and the maize *C1* gene into maize suspension cells resulted in the appearance of numerous pigmented cells within 48 h of bombardment (Figure 3B). In contrast, no red spots were seen in cells bombarded only with maize *C1* (Figure 3C), maize *Lc* (data not shown), or rice *Rb2* alone (data not shown). These data suggest that the RB2 protein can interact with the maize C1 protein to activate the maize structural genes.

Ra2, *another* Ra-*like gene in rice*

To further characterize the rice *R* gene family, DNA blots of genomic DNA digested with *Hin*dIII were used to visualize additional family members. Five bands were detected when DNA from strain Purple 522 was probed under intermediate-stringency conditions with an *Ra1* cDNA fragment containing the conserved bHLH region (Figure 4A, lane 1). Gene-specific probes from *Ra* and *Rb*, derived from sequences corresponding to intron 9, were employed to determine the identity of these bands. Under high-stringency conditions, the *Rb* probe hybridized to bands of 18, 12 and 7.5 kb (Figure 4A, lane 2), while the *Ra* probe detected the 7 kb fragment (Figure 4A, lane 3). These results suggested that Purple 522 contains at least five *R* genes: three *Rb*-like, one *Ra* and *R* homologous sequences that reside on the 3 kb *Hin*dIII fragment.

In order to isolate part of the *R* sequence associated with the 3 kb *Hin*dIII fragment, a PCR assay utilizing degenerate primers capable of amplifying *R* genes from several grass species (Purugganan and Wessler, 1994) was undertaken. The primers, shown as p1 and p2 in Figure 5A, amplified products of ca. 725 bp and ca. 550 bp. Sequencing of clones from the ca. 550 bp PCR product revealed two highly similar sequences that were identical to *Rb1* or *Rb2*, confirming the presence of two closely related *Rb* genes in Purple 522. The ca. 725 bp product was composed of fragments of 725 bp and 720 bp; the former was identical to *Ra*, while the latter was more closely related to *Ra* than to *Rb* (Figure 5B). Thus, the 720 bp fragment seems to be a part of a third *R* gene that is more closely related to *Ra*. For this reason, this putative *R* gene was designated *Ra2*, and the gene previously designated *Ra* was renamed *Ra1*.

The status of *Ra2* as a distantly related *Ra* gene was confirmed in two ways. First a gene-specific probe of *Ra2*, derived from the region corresponding to intron 9 (which shares only 35% sequence identity with *Ra1*), detected only the 3 kb band upon reprobing of the DNA blot (Figure 4A, lane 4) and when additional rice genomic DNAs were probed with either *Ra1*- or *Ra2*-specific probes (data not shown). Analysis of these blots led to the surprising finding that the deeply pigmented strain, JP004, contains *Ra2* but lacks *Ra1* (Figure 4B, lane 1). The availability of JP004 allowed us to rapidly determine the relatedness of *Ra2* to the previously characterized *Ra1* gene along its entire length. To this end, select regions of the *Ra2* gene were amplified with primers from different regions of *Ra1* (p3 and p4, p5 and p6 shown in Figure 5A). The overall sequence identity between the two *Ra*-like genes was found to be ca. 92% in the corresponding exonic regions (data not shown).

Multiple Ra-*like and* Rb-*like genes in* O. sativa *and its wild relatives*

Given that the rice *R* gene family is believed to have evolved recently and independently (Hu *et al.*, 1996) and that at least four *R* genes were found in Purple 522, we were interested in knowing the distribution of *Ra* and *Rb* genes in more rice strains. The availability

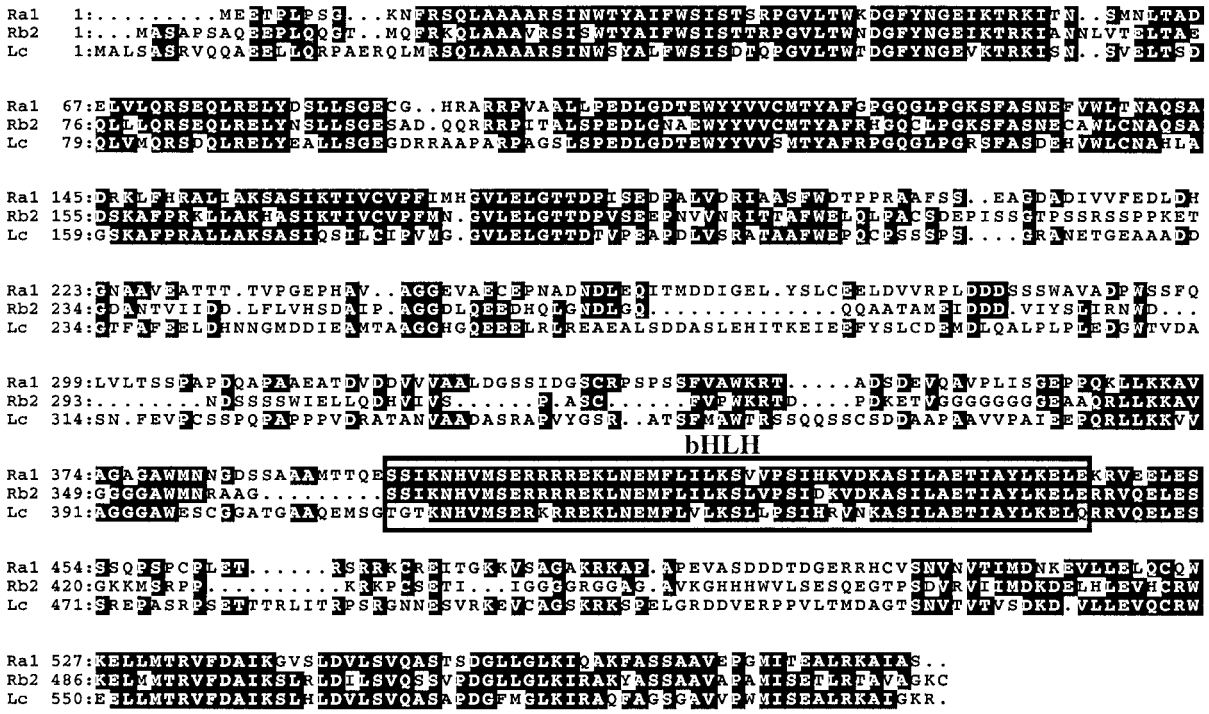


Figure 2. Amino acid sequence alignment of *Ra1*, *Rb2*, and maize *R* (*Lc*). Identical residues are in black boxes.

of several specific probes for members of the *R* gene family permitted a survey of the family in additional *O. sativa* strains and some of their wild relatives by high-stringency genomic blot analysis. The five relatives, of *O. sativa* analyzed were: *O. glaberrima*, another domesticated species, *O. barthii* and *O. glumeapatula*, two close wild relatives, and *O. longistaminata* and *O. meridionalis*, two distant relatives. All strains analyzed share the AA genome.

When probed with the *Ra1* cDNA fragment containing the bHLH region, two bands were detected in most *O. sativa* strains and their close relatives (Figure 4B, lanes 2–7), whereas only a single band of hybridization was found in the two distant relatives (Figure 4B, lanes 8 and 9). The *Rb*-specific sequence hybridized to from one to three bands in different rice strains (Figure 4C). Taken together, our data suggest that domesticated rice and related species have multiple and various copies of *Ra*-like and *Rb*-like genes.

Miniature inverted-repeat transposable elements (MITEs) and rice R *genes*

Sequence comparison of *Ra1* and *Ra2* revealed ca. 90% sequence identity in the putative exonic regions and in sequences corresponding to introns 2, 3, 4 and 6 (data not shown). However, regions corresponding to introns 7 and 8 differ in length, and intron 9 sequences only share ca. 30% identity (see below). Our previous study (Hu *et al.*, 1996) found two miniature inverted-repeat transposable elements (MITEs) of the *Stowaway* element family (Bureau and Wessler, 1994) in intron 7 (*Stow 1*) and intron 9 (*Stow 2*) of *Ra1*. For this reason we were interested in determining if MITEs accounted for the difference in intron size between the two *Ra* genes and if additional MITEs could be identified in the *R* genes. To this end, database searches were undertaken using the available sequences of *Ra1*, *Ra2*, *Rb1* and *Rb2* as queries. These searches led to the identification of four additional *Stowaway* elements in the non-coding regions of rice *R* genes (Figure 6A).

Members of the *Stowaway* family share the following features: (1) conserved 11 bp terminal inverted repeats (TIRs) with a consensus sequence of CTCCCTCCGTT, (2) AT-richness, (3) small size (80–323 bp), (4) target site preference (TA), and (5) potential to form stable DNA secondary structure (Bureau and Wessler, 1994). The *Stowaway* elements found in rice *R* genes fit the above criteria and share an overall sequence identity of ca. 58% (Figure 6B).

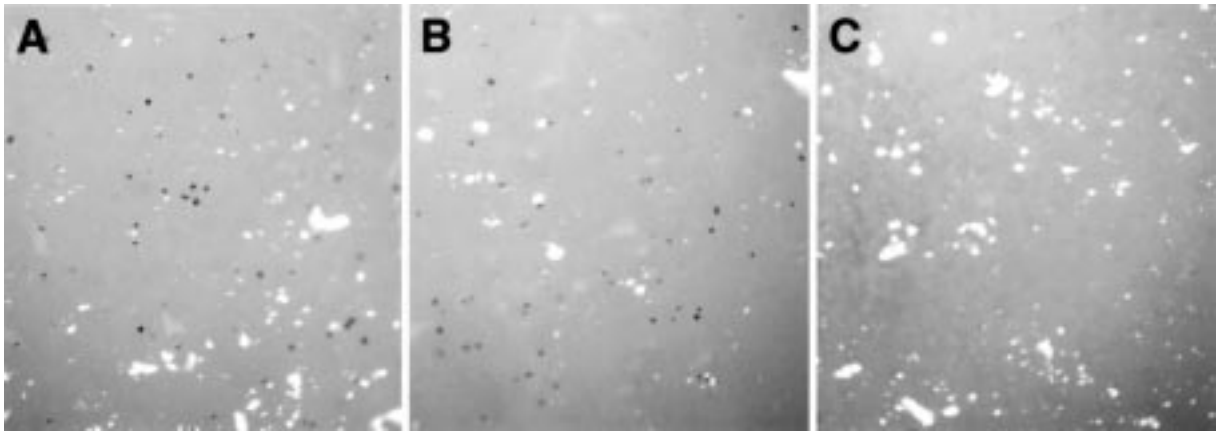


Figure 3. Expression of maize *R* (*Lc*) and rice *Rb2* in maize suspension cells. Maize cells were bombarded with gold particles coated with pLCΔ and p35SC1 (A), pRiceRb2 and p35SC1 (B), or p35SC1 alone (C).

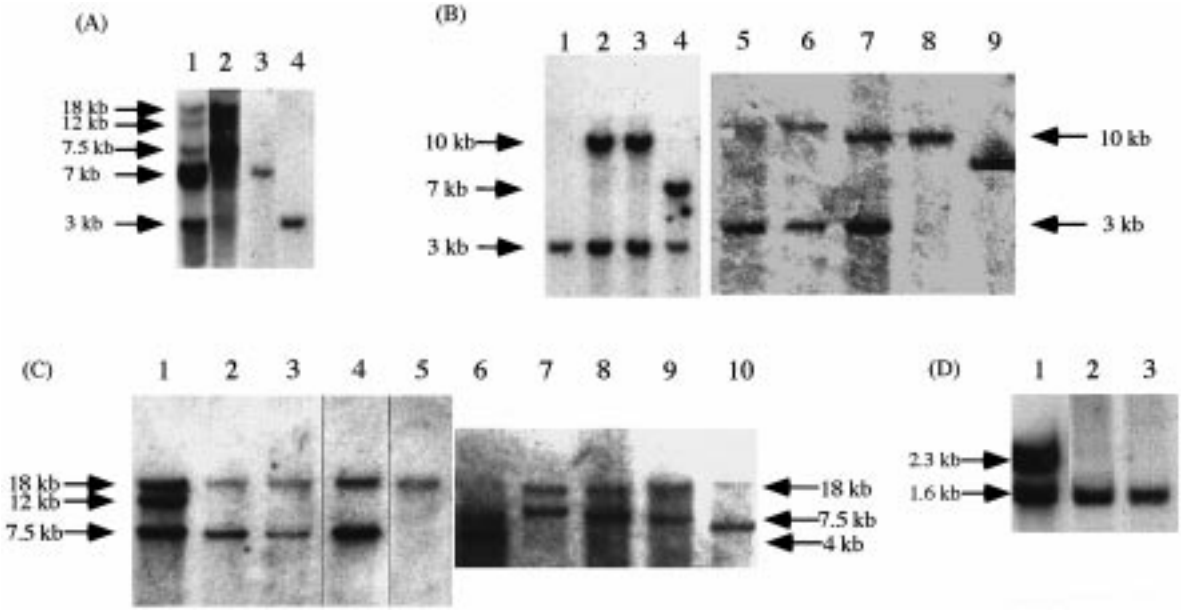


Figure 4. Genomic DNA blot analyses of *Ra* and *Rb* genes in *O. sativa* and other rice species. Intermediate-stringency conditions were used in lane 1 of (A), whereas high-stringency conditions were used in the other blots. A. *R* genes in Purple 522. Blots containing DNA from Purple 522 were hybridized with the following probes: lane 1, an *Ra1* cDNA fragment containing the highly conserved bHLH region; lane 2, *Rb*-specific (from *Rb1*); lane 3, *Ra1*-specific; or lane 4, *Ra2*-specific. All the gene-specific probes were derived from intron 9. B. *Ra*-like genes in *O. sativa* and other AA-genome species. Blots containing DNA from different rice strains were probed with an *Ra1* cDNA fragment containing the bHLH region. Lanes 1–4: *O. sativa* strains JP004, T65, IR36, and Purple 522; lane 5: *O. glaberrima*; lane 6: *O. glumeapatula*; lane 7: *O. barthii*; lane 8: *O. longistaminata*; lane 9: *O. meridionalis*. C. *Rb*-like genes in *O. sativa* and its AA-genome relatives. Blots containing DNA from different rice strains were probed with an *Rb*-specific sequence containing intron 9. Lanes 1–5: *O. sativa* strains Purple 522, JP004, BS125, T65 and A58 respectively; lane 6: *O. glaberrima*; lane 7: *O. glumeapatula*; lane 8: *O. barthii*; lane 9: *O. longistaminata*; lane 10: *O. meridionalis*. D. Both *Ra2* and *Rb* contain *Sph*I fragments of 1.6 kb. A blot containing DNA from Purple 522 that was digested with *Sph*I was hybridized with the following probes: an *Ra1* fragment containing the bHLH region (lane 1), *Rb*-specific (lane 2) and *Ra2*-specific (lane 3).

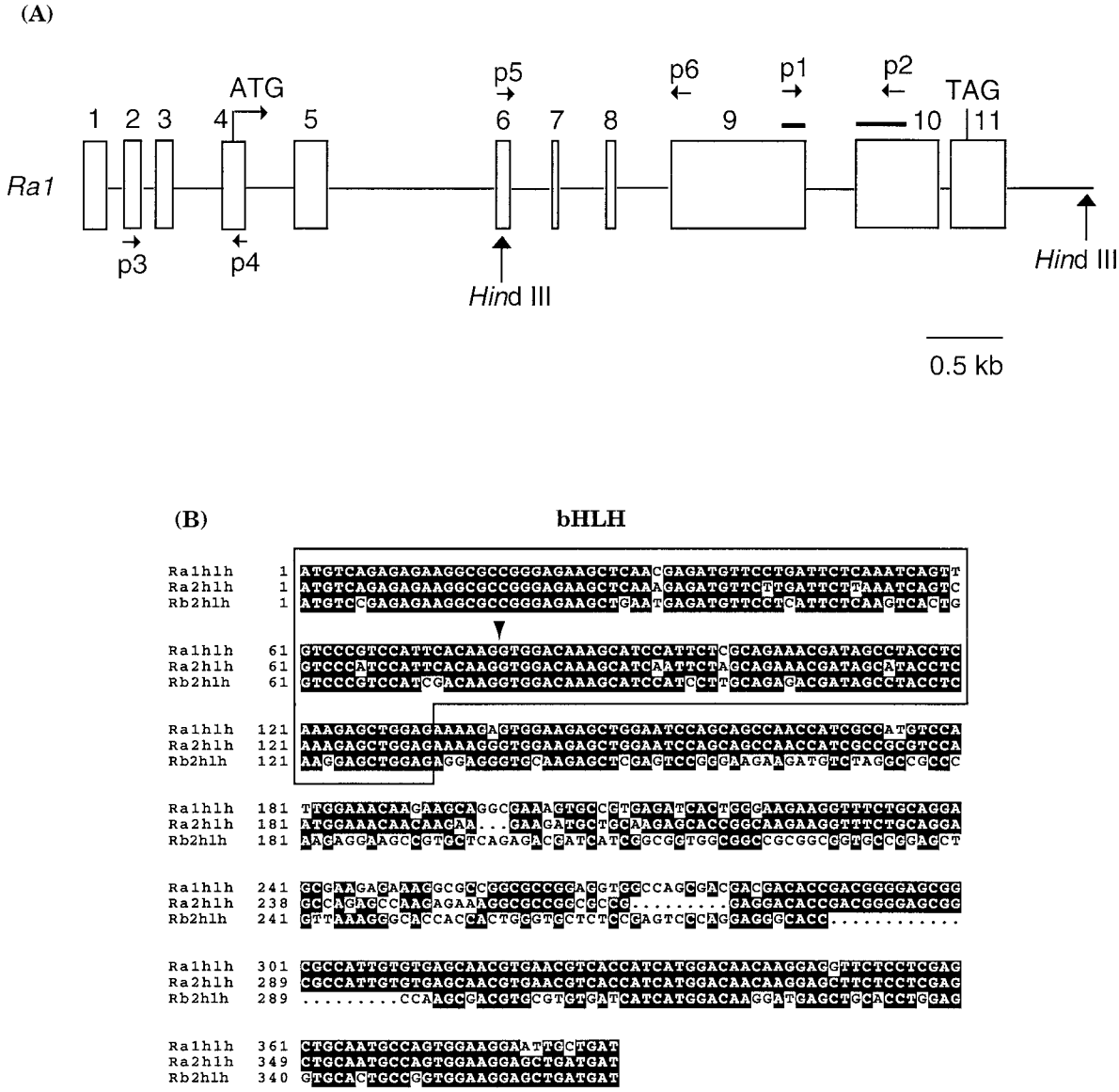
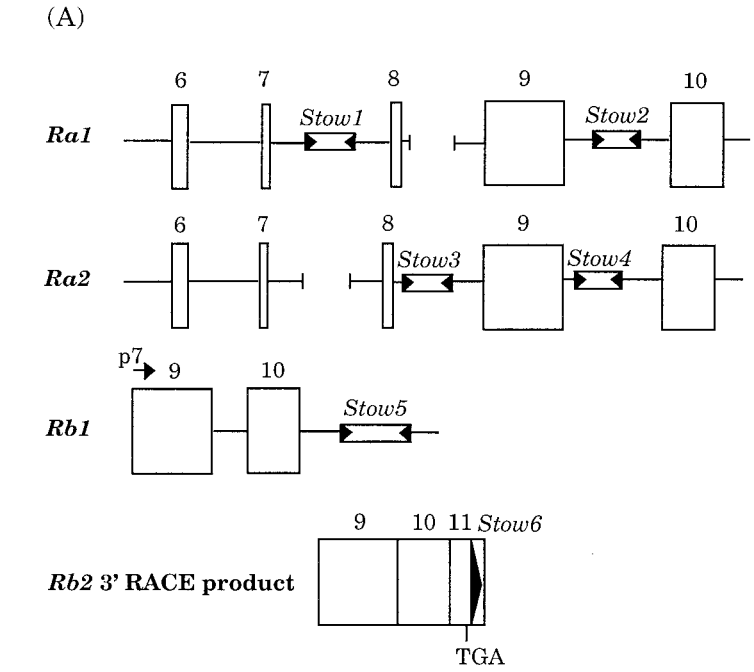


Figure 5. The structure of *Ra1* and the comparison of rice *R* sequences. A. The structure of *Ra1* with the positions of primers described in the text. Exons are shown as open boxes and the connecting lines represent introns. The black bar at the end of exon 9 and the beginning of exon 10 indicates the conserved *Ra1* cDNA probe used in DNA blot analyses. P1 and P2 show the positions of degenerate primers used to amplify *R* sequence from rice and other grasses. P3 and P4, P5 and P6 were used to amplify additional segments of *Ra2*. B. Alignment of partial sequences from *Ra1*, *Ra2* and *Rb2*. The arrowhead indicates the junction of exons 9 and 10 in the *Ra1* gene.

Four *Stowaway* elements were found in *Ra1* and *Ra2*. Intron 7 of *Ra1* but not *Ra2* contains *Stow 1* while intron 8 of *Ra2* but not *Ra1* contains *Stow 3*. A comparison of these intron sequences and the position of the *Stowaway* insertions are shown in Figure 7A and 7B. Although both *Ra1* and *Ra2* have *Stowaway* elements in intron 9, two factors indicate that these are independent insertions. First, the insertion sites differ, albeit by only a few base pairs (Figure 7C). Second, the elements themselves are distantly related (ca. 54% identity) (Figure 6B). Thus, the presence of all four *Stowaway* elements in *Ra1* and *Ra2* appear to have resulted from independent insertion events.

Rb1 and *Rb2* each contain at least one *Stowaway* element. A genomic fragment of *Rb1*, isolated from Purple 522 (see Materials and methods), contains



(B)

```
Ra1stow1      1 TACTCCCTCCGTCACATAATA.TAAG.GCATTTTA......AGTTTTTGGTTACACTCTT
Rb2 stow6     1 TACTCCCTTCATCTTATAATA.TAA....ATTATA......AGGAATT..ATCC......
Ra2stow4      1 TACTCCATCCGTATTTAATG.TATGACGATGTTG......ACTTTTT..AACAAACGTT
Ra1stow2      1 TACTTCATCCGTTTCACAATG.TAAGTC.ATTCTA......GCATTTTCCATATTCATAT
Ra2stow3      1 TACTCCCTCCGTC...CAAAA.AAAGACAAACCT.......GGTTTCCGTGTCTTATGTT
Rb1stow5      1 TACTCTCTCCATTCATACGTGTTTTGCATATTTTATTTTCTAGTTTTTTCTTAATGAGAT

Ra1stow1     53 TAAC.CACTCATCTTATTCAAAAA.......ATT.TATGCAAATA.........TAAAAAAA
Rb2 stow6       ..............................................................
Ra2stow4     52 TGATT.TTTTCGTCTTATAAAAAAA.....TTTATAATCAAATG.........TTCTTTAA
Ra1stow2     53 TGAG..........GGTAATGAA.......TCTATATTCAAAT..............TAA
Ra2stow3     50 TGAA.CGTCCGTCTTATTTAAAAA.....AAT.TATGAAAAAA.........ATTAAAAA
Rb1stow5     61 AGACATATTTGTAGTATCCATGAATCCAAGACTTCAATGAAGTAGTTGTCTGTTTGTTTT

Ra1stow1     97 GAAAAAATGTGCTTAAACTACT..............TTGG..........AATAATAAAG
Rb2 stow6       ..............................................................
Ra2stow4     97 GAA.......TGACATAAGTATT.............TTTG..........TAT.......T
Ra1stow2     82 ..........TC..TTAATGAAT.............TTAG..........A..........
Ra2stow3     94 GACAAGTGGGCTATAAAATATTAATCATGTTTTATCATCT..........AACAACAATG
Rb1stow5    121 TACAATGCTAGACTAAACTGTTGGCCAGAATATATCTAAGCAGTCATCTTAATACTTATT

Ra1stow1    133 ..........TAAGTAAAAAAATCTATTTTTTT.AA.ATAAGACGAG..TGGTCAAAT
Rb2 stow6       ..............................................................
Ra2stow4    121 ..........TGCATAAAAAA..........TTT.GA.ATAAGACGAA..TGATCAAAC
Ra1stow2     98 ..............................TT.CA.TTACCACCAA..TATGAATAT
Ra2stow3    144 ..........AAAATACGAATTATATAAAAATTT.CATATAAGACGGA..CCGTTAAA.
Rb1stow5    181 GGTTCTATGCATGCGTATAATCCTTGGTATTTGTTACATGTAGGATGAGTTTATTTCTT

Ra1stow1    178 AGTGCAAACCA.AAAACTCCAAAATCTTATATTATGGGCCGCAGTGAGTA
Rb2 stow6       ..................................................
Ra2stow4    156 GTTGCTTA.AA.AAGTCAACGGCCTCATACATTAAAATACCGAGGGAGTA
Ra1stow2    121 GG.......GA.AATGCTAGAATCACTTATATTATTAAACCGAGGGAGTA
Ra2stow3    189 GTTAGACACGG.AAACCCATAGTTTGTCTTTTTTGGGACGGACCGAGTA
Rb1stow5    241 GCTCTTGATGACAAAAGAATACACTC.TAAACTTTTGGATGGAGGGAGTA
```

Figure 6. Stowaway elements in rice *R* genes. A. Positions of the *Stowaway* elements in *Ra* and *Rb*. Vertical lines in intron 8 of *Ra1* and the putative intron 7 of *Ra2* indicate the absence of elements at these sites. Solid arrowheads in each element indicate TIRs. P7 is the *Rb*-specific primer used in the 3′ RACE experiment described in Materials and methods. B. Sequence comparison of *Stowaway* elements found in rice *R* genes. Black boxes indicate identical nucleotides, and gaps are shown as dots. Open boxes show the putative target site sequence of *Stowaway* elements. Arrows represent the conserved TIR sequences.

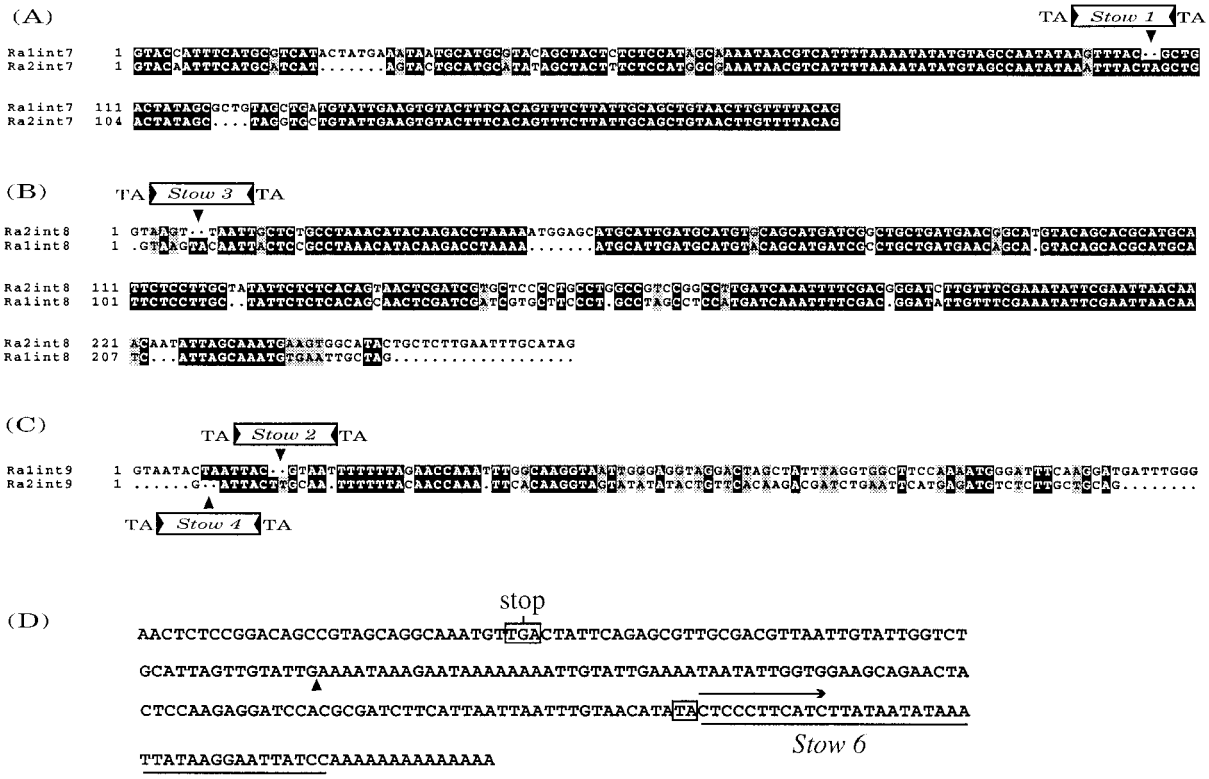


Figure 7. Alignment of *Ra1* and *Ra2* in regions corresponding to introns 7–9 and the position of *Stow 6* in *Rb2*. A–C. Comparison of regions corresponding to intron 7 (A), 8 (B), 9 (C) of *Ra1* and *Ra2* in the absence of the *Stowaway* sequences. Identical nucleotides are shown in black boxes, and similar residues are in gray boxes. Dots are deletions. The sequences of the *Stowaway* elements, whose positions are indicated, have been omitted from the alignment to show maximal sequence similarity. The last 27 bp of *Ra1* intron 9 is not shown in C. D. Sequence of the 3′ end of an *Rb2* 3′ RACE product showing the putative truncated *Stow 6* element. TA is the putative target site for *Stowaway* elements. The horizontal arrow represents the conserved TIR sequence. The vertical arrow indicates the polyadenylation site for the 1.9 kb *Rb2* cDNA clone. The partial *Stow 6* element is underlined.

Stow 5 (Figure 6A). A 700 bp *Rb2* 3′ RACE product, which was amplified from pigmented leaves of Purple 522 (see Materials and methods), harbors a partial element (*Stow 6*) in the 3′-untranslated region (Figure 7D). Only one copy of the dinucleotide TA (the putative target sequence for *Stowaway*) and a single TIR sequence were found in this transcript, indicating that the full-length *Stow 6* element contained the polyadenylation site for the corresponding *Rb2* mRNA. Interestingly, the 1.9 kb *Rb2* cDNA isolated from the same tissue as the 3′ RACE product lacks the *Stow 6* element since it utilizes a different polyadenylation site located ca. 100 bp upstream of *Stow 6* (Figure 7D).

Discussion

Evidence is presented that rice *Rb2*, like *Ra1*, encodes a functional *R* homologue. The amino acid sequence of *Rb2*, as derived from a full-length cDNA, is similar to *Ra1* and the maize *Lc* gene along its entire length. In addition, *Rb2* cDNA activated anthocyanin biosynthesis in maize suspension cells, but only when co-bombarded with the maize *C1* gene. These data indicate that RB2 is competent to interact with C1 to activate the maize structural genes in the anthocyanin pathway.

Experimental results also suggest that *O. sativa* and its wild relatives contain multiple *R* genes that have been classified into two groups: *Ra*-like and *Rb*-like. This classification was facilitated by the identification of *Ra*- and *Rb*-specific probes that detect two distinct sets of rice genes. Under high-stringency conditions, an *Ra* probe (extending downstream from

the bHLH domain) only detected *Ra1*, *Ra2* and other *Ra*-like genes, while the *Rb* probe (containing intron 9) only hybridized with sequences that are highly homologous to *Rb*.

Digestion of genomic DNAs with *Hin*dIII permitted unambiguous resolution of the two *R* gene classes and allowed us to correct a misinterpretation of prior DNA blot results (Hu *et al.*, 1996). Specifically, two bands of 1.6 kb and either 4 or 2.3 kb were detected in all rice strains when DNA was digested with *Sph*I and probed with the 3′ half of the *Ra1* cDNA. The *Rb*-specific probe hybridized with the 1.6 kb band; however, curiously, this band appeared as a doublet in a subset of the strains. At that time we concluded that the larger band (4 or 2.3 kb) was *Ra1* while the 1.6 kb band was from *Rb* and that the 1.6 kb doublet in some strains derived from two sequences that are closely related to *Rb* intron 9. We now know that the two bands detected by the *Ra1* cDNA were in fact *Ra1* (4 and 2.3 kb) and *Ra2* (1.6 kb) and not *Rb*. That is, by an unfortunate coincidence both *Ra2* and *Rb* contain *Sph*I fragments of 1.6 kb. This was confirmed by demonstrating that the *Ra2*-specific probe hybridized with a 1.6 kb *Sph*I fragment (Figure 4D). Furthermore, this explains why the *Ra1* probe detected only a single band of 1.6 kb in all rice strains while a doublet was detected in a subset of strains by the *Rb*-specific probe.

The rice genus (*Oryza*) has at least seven distinct genome types, including AA, BB, CC, EE, FF, BBCC and CCDD. The cultivated rice *O. sativa* and *O. glaberrima*, and six wild species share the AA genome (Vaughan, 1989). Our previous study suggested that CC and EE genomes contained a single copy of *Ra*-like gene while *O. sativa* and its wild progenitor *O. rufipogon* both had at least two *Ra*-like genes (Hu *et al.*, 1996), indicating that *Ra* genes amplified later in rice evolution. Analysis of several AA-genome-containing strains in this study provided additional evidence that *Ra* genes were amplified during the evolution of AA genome species. First, almost all *O. sativa* strains analyzed (both previously and in this study) and their close relatives harbor at least two *Ra*-like genes, yet only a single *Ra*-like gene was found in species distantly related to *O. sativa*. Further, that *Ra1* and *Ra2* are highly homologous argues that they are the result of a recent duplication event.

The finding that rice has evolved a fairly complex family of the *R* genes was unexpected, given that rice has the smallest genome of all the grasses analyzed to date (Arumuganathan and Earle, 1991). As with maize, human domestication may have played a role in the selection of different *R* genes and *R* alleles due to the conspicuous and dispensable nature of the *R* phenotype. In the discussion below we hypothesize that, as in maize, unequal crossing over and transposable element activity may have been responsible for generating the diversity of alleles available for human selection.

Genetic analysis of the loci responsible for pigmentation patterns in rice identified a group of genes that map to either chromosome 1 or 4. These genes include *Pl* (purple leaf), *Pn* (purple node), *Pr* (purple hull), *Ps* (purple stigma), *Pb* (purple pericarp), and *Pin* (purple internode) (Takahashi, 1982). Interestingly, *Ra1* and *Rb1* also map to chromosomes 4 and 1 respectively (Hu *et al.*, 1996), suggesting that the genes identified genetically may be alleles of *Ra* and *Rb* and/or clusters of *Ra* and *Rb* genes. The clustering of *R* genes has already been documented extensively in maize where the *R-r* complex on chromosome 10 is composed of four closely linked *R* components. Unequal crossing over between *R* complexes has resulted in varied numbers of *R* genes in different maize strains (Dellaporta *et al.*, 1988; Robbins *et al.*, 1991; Walker *et al.*, 1995). In this study, the *Rb* probe detected one to three bands in the *O. sativa* strains assayed (Figure 4C), indicating that the copy number of *Rb*-like genes also varies in rice. These data lead us to hypothesize that the *Rb* genes are linked on chromosome 1, and that unequal crossing over may serve to expand or reduce the number of *Rb* genes in *O. sativa* varieties. Similarly, the *Ra*-like genes may also be linked on chromosome 4. If so, the deletion of *Ra1* from JP004 could be due to unequal crossing over between *Ra* genes. This question could be easily addressed by determining the map positions of individual *Ra* and *Rb* genes described in this study. Unfortunately, the mapping of *Ra2* and the other *Rb* genes was not possible because the wild parent used to generate our mapping population only contained a single *Ra*-like gene and a single *Rb*-like gene. Different mapping populations, where both parents contain multiple *Ra* or *Rb* genes, will be required to determine if the *Ra*-like and *Rb*-like genes described in this study are linked on chromosomes 4 and 1, respectively.

Recent comparative studies of the *adh* and *a1-sh2* regions of the maize and rice genomes revealed massive amplifications of retrotransposons in the intergenic regions of maize, but not in rice (Chen and Bennetzen, 1996; SanMiguel *et al.*, 1996; Chen *et al.*, 1997). In fact, the amplification of a few retrotransposon families in maize accounts for most of the 6-fold

difference in size between the two genomes. In contrast, MITEs appear to be thriving in both maize and rice perhaps, in part, because they may contribute to the evolution of gene regulation and expression (Wessler *et al.*, 1995; Bureau *et al.*, 1996 for reviews). In addition to identifying at least six MITE elements of the *Stowaway* family in rice *R* genes, the data, as discussed in Results, also reveals the independence of each insertion and their occurrence after the duplication of *Ra1* and *Ra2* and possibly after the duplication of *Rb1* and *Rb2*.

Perhaps *Stowaway* insertions into *Ra*-like and *Rb*-like genes serve to promote diversification by reducing intergenic interactions that could lead to gene silencing (Flavell, 1991; Vaucheret *et al.*, 1998 for reviews) or unequal crossing over. Among the *Stowaway* elements found in rice *R* genes, *Stow 1–4* are primarily responsible for the variation in intron size and sequence observed in the two otherwise similar *Ra* genes (Figure 7A–C). Alternatively, *Stowaway*-derived sequences may have directly altered gene expression by inserting new regulatory motifs. The polyadenylation site provided by the insertion of *Stow 6* into the *Rb2* gene might be relevant in this regard (Figure 7D).

Acknowledgements

We thank Liangjiang Wang and Ernie Dodd for technical assistance, and Rob Larkin, Bridey Maxwell and Karin Schumacher for critical reading of the manuscript. This work was supported by a grant from the Department of Energy to S.R.W.

References

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

Anderson, E. 1945. What is *Zea mays*? A report of progress. Chron. Bot. 9: 88–92.

Arumuganathan, K. and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9: 208–218.

Bilang, R., Zhang, S., Leduc, N., Iglesias, V.A., Gisel, A., Simmonds, J., Potrykus, I. and Sautter, C. 1993. Transient gene expression in vegetative shoot apical meristems of wheat after biolistic microtargeting. Plant J. 4: 735–744.

Bureau, T.E. and Wessler, S.R. 1994. *Stowaway*: a family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6: 907–916.

Bureau, T.E., Ronald, P.C. and Wessler, S.R. 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc. Natl. Acad. Sci. USA 93: 8524–8529.

Casas, A.M., Kononowicz, A.K., Zehr, U.B., Tomes, D.T., Axtell, J.D., Butler, L.G., Bressan, R.A. and Hasegawa, P.M. 1993. Transgenic sorghum plants via microprojectile bombardment. Proc. Natl. Acad. Sci. USA 90: 11212–11216.

Chandler, V.L., Radicella, J.P., Robbins, T.P., Chen, J. and Turks, D. 1989. Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of *B* utilizing *R* genomic sequences. Plant Cell 1: 1175–1183.

Chen, M. and Bennetzen, J.L. 1996. Sequence composition and organization in the Sh2/A1-homologous regions or rice. Plant Mol. Biol. 32: 999–1001.

Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.S., Zhang, H., Wing, R.A. and Bennetzen, J.L. 1997. Microcollinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes. Proc. Natl. Acad. Sci. USA 94: 3431–3455.

Damiani, R.D. Jr. and Wessler, S.R. 1993. An upstream open reading frame represses expression of *Lc*, a member of the *R/B* family of maize transcriptional activators. Proc. Natl. Acad. Sci. USA 90: 8244–8248.

Dellaporta, S.L., Greenblatt, I., Kermicle, J.L., Hicks, J.B. and Wessler, S.R. 1988. Molecular cloning of the maize *R-nj* allele by transposon tagging with *Ac*. In: J.P. Gustafson and R. Apples (Eds.), Chromosome Structure and Function: Impact of New Concepts, 18th Stadler Genetics Symposium, Plenum Press, New York, pp. 263–282.

Dickinson, W.J., 1991. The evolution of regulatory genes and patterns in *Drosophila*. In: M. Hecht, B. Wallace and R. MacIntyre (Eds.), Evolutionary Biology, vol 25, Plenum Press, New York, pp. 127–173.

Dooner, H.K. and Robbins, T.P. 1991. Genetic and developmental control of anthocyanin biosynthesis. Annu. Rev. Genet. 25: 173–199.

Flavell, R.B. 1994. Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proc. Natl. Acad. Sci. USA 91: 3490–3496.

Galway, M.E., Masucci, J.D., Lloyd, A.M., Walbot, V., Davis, R.W. and Schiefelbein, J.W. 1994. The *TTG* gene is required to specify epidermal cell fate and cell patterning in the *Arabidopsis* root. Dev. Biol. 166: 740–754.

Goff, S.A., Klein, T.M., Roth, B.A., Fromm, M.E., Cone, K.C., Radicella, J.P. and Chandler, V.L. 1990. Transactivation of anthocyanin biosynthetic genes following transfer of *B* regulatory genes into maize tissues. EMBO J. 9: 2517–2522.

Goodrich, J., Carpenter, R. and Coen, E.S. 1992. A common gene regulates pigmentation pattern in diverse plant species. Cell 98: 955–964.

Helentjaris, T., Weber, D. and Wright, S. 1988. Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics 118: 353–363.

Hu, J., Anderson, B. and Wessler, S.R. 1996. Isolation and characterization of rice *R* genes: evidence for distinct evolutionary paths in rice and maize. Genetics 142: 1021–1031.

Klein, T.M., Roth, B.A. and Fromm, M.E. 1989. Regulation of anthocyanin biosynthetic genes introduced into intact maize tissues by microprojectiles. Proc. Natl. Acad. Sci. USA 86: 6681–6685.

Lloyd, A.M., Walbot, V. and Davis, R.W. 1992. *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1*. Science 258: 1773–1775.

Ludwig, S.R. and Wessler, S.R. 1990. Maize *R* gene family: tissue specific helix-loop-helix proteins. Cell 62: 849–851.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L. and Wessler, S.R. 1989. *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein simi-

lar to transcriptional activators and contains the *myc*-homology region. Proc. Natl. Acad. Sci. USA 86: 7092–7096.

Oka, H.I. 1988. The genus *Oryza*. In: Origin of Cultivated Rice, Japanese Science Society Press/Elsevier, Tokyo, pp. 1–14.

Purugganan, M.D. and Wessler, S.R. 1994. Molecular evolution of the plant *R* regulatory gene family. Genetics 138: 849–854.

Quattrocchio, F., Wing, J.F., van der Woude, K., Mol, J.N.M. and Koes, R. 1998. Analysis of bHLH and MYB-domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. Plant J. 13: 475–488.

Robbins, T.P., Chen, J., Norell, M.A. and Dellaporta, S.L. 1989. Molecular and genetic analysis of the *R* and *B* loci in maize. In: D.E. Styles, G.A. Gavazzi and M.L. Racchi (Eds.), The Genetics of Flavanoids. Proceedings of a Post Congress Meeting of the XVI International Congress of Genetics, Edizioni Unicopli, Milan, pp. 105–114.

Robbins, T.P., Walker, E.L., Kermicle, J.L., Alleman, M. and Dellaporta, S.L. 1991. Meiotic instability of the *R-r* complex arising from displaced intragenic exchange and intrachromosomal rearrangement. Genetics 129: 271–283.

SanMiguel, P., Tikhonov, A., Jin, Y., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z. and Bennetzen, J.L. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768.

Shieh, M.W., Wessler, S.R. and Raikhel, N.V. 1993. Nuclear targeting of the maize *R* protein requires two nuclear localization sequences. Plant Physiol. 101: 353–361.

Styles, E.D., Ceska, O. and Seah, K.T. 1973. Developmental differences in action of *R* and *B* alleles in maize. Can. J. Genet. Cytol. 15: 59–72.

Takahashi, M. 1982. Gene analysis and its related problems. J. Fac. Agric. Hokkaido Univ. 61: 93–99.

Vaucheret, H., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J., Mourrain, P., Palauqui, J. and Vernhettes, S. 1998. Transgene-induced gene silencing in plants. Plant J. 16: 651–659.

Vaughan, D.A. 1989. The genus *Oryza* L.: current status of taxonomy. Int. Rice Res. Inst. Res. Pap. Ser. 138: 3–21.

Walker, E.L., Robbins, T.P., Bureau, T.E., Kermicle, J. and Dellaporta, S.L. 1995. Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize *R-r* complex. EMBO J. 14: 2350–2363.

Wessler, S.R., Bureau, T.E. and White, S.E. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5: 814–821.