The Plant Cell, Vol. 18, 283–294, February 2006, www.plantcell.org © 2006 American Society of Plant Biologists

RESEARCH ARTICLES

Caught Red-Handed: *Rc* Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice [W] [OA]

Megan T. Sweeney,[a] Michael J. Thomson,[a,1] Bernard E. Pfeil,[b] and Susan McCouch[a,2]

[a] Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14953-1901
[b] Department of Plant Biology, Cornell University, Ithaca, New York 14853

***Rc* is a domestication-related gene required for red pericarp in rice (*Oryza sativa*). The red grain color is ubiquitous among the wild ancestors of *O. sativa*, in which it is closely associated with seed shattering and dormancy. *Rc* encodes a basic helix-loop-helix (bHLH) protein that was fine-mapped to an 18.5-kb region on rice chromosome 7 using a cross between *Oryza rufipogon* (red pericarp) and *O. sativa* cv Jefferson (white pericarp). Sequencing of the alleles from both mapping parents as well as from two independent genetic stocks of *Rc* revealed that the dominant red allele differed from the recessive white allele by a 14-bp deletion within exon 6 that knocked out the bHLH domain of the protein. A premature stop codon was identified in the second mutant stock that had a light red pericarp. RT-PCR experiments confirmed that the *Rc* gene was expressed in both red- and white-grained rice but that a shortened transcript was present in white varieties. Phylogenetic analysis, supported by comparative mapping in rice and maize (*Zea mays*), showed that *Rc*, a positive regulator of proanthocyanidin, is orthologous with *INTENSIFIER1*, a negative regulator of anthocyanin production in maize, and is not in the same clade as rice bHLH anthocyanin regulators.**

INTRODUCTION

Most rice (*Oryza sativa*) that is grown and consumed throughout the world has white pericarp, but rice can also produce grains with brown, red, and purple pericarp. The color is visible when the grains are dehulled, but it can be removed by polishing to reveal the white endosperm. Red pericarp is ubiquitous among the wild ancestors of cultivated rice (*Oryza rufipogon*), and in some regions of the world red cultivars are preferred for their taste, texture, and ceremonial or medicinal value. Consumer interest in red and purple rices represents a growing specialty market in the United States, but at the same time, the constant presence of weedy red rice in farmers' fields is the most economically important pest and grain-quality problem faced by U.S. rice growers (Gealy et al., 2002). Red rices, which typically show seed shattering and dormancy along with a red pericarp, can belong to either *O. sativa* or *O. rufipogon* (Vaughan et al., 2001), neither of which is native to the United States. They interbreed freely with cultivated, white-grained types, making transgenic herbicide-resistant varieties impractical.

The red pigment in rice grains is proanthocyanidin, also called condensed tannins (Oki et al., 2002). Proanthocyanidins are a branch off the anthocyanin pathway and share many of the same biosynthetic genes (Winkel-Shirley, 2001). Proanthocyanidins have been shown to have important deterrent effects on pathogens and predators, so it is not surprising that spontaneous mutations that inhibit pigment production would be selected against in the wild (Shirley, 1998). On the other hand, white grain appears to be associated with the domestication syndrome and remains under strong selection in most rice breeding programs today.

Regardless of the problems associated with red rice as a weed, the red pigment is of interest for nutritional reasons. It serves as a powerful antioxidant that has been demonstrated to reduce atherosclerotic plaque formation, a risk factor associated with cardiovascular disease (Ling et al., 2001). On the negative side, proanthocyanidin pigments reduce the bioavailability of iron, protein, and carbohydrates (Eggum et al., 1981; Carmona et al., 1996; Glahn et al., 2002), which has important implications for people with low nutritional status. A better understanding of the genetics and molecular biology of red pericarp and the association of this characteristic with other wild/weedy traits will provide important information for the better management of both the negative and positive features associated with red rice.

Two loci have been identified using classical genetic analysis, *Rc* (brown pericarp and seed coat) and *Rd* (red pericarp and seed coat). When present together, these loci produce red seed color (Kato and Ishikawa, 1921). *Rc* in the absence of *Rd* produces

[1] Current address: Department of Plant Breeding, Genetics, and Biotechnology, International Rice Research Institute, Los Baños, Laguna, Philippines.
[2] To whom correspondence should be addressed. E-mail srm4@cornell.edu; fax 607-255-6683.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Susan R. McCouch (srm4@cornell.edu).
[W] Online version contains Web-only data.
[OA] Open Access articles can be viewed online without a subscription.
Article, publication date, and citation information can be found at www.plantcell.org/cgi/doi/10.1105/tpc.105.038430.

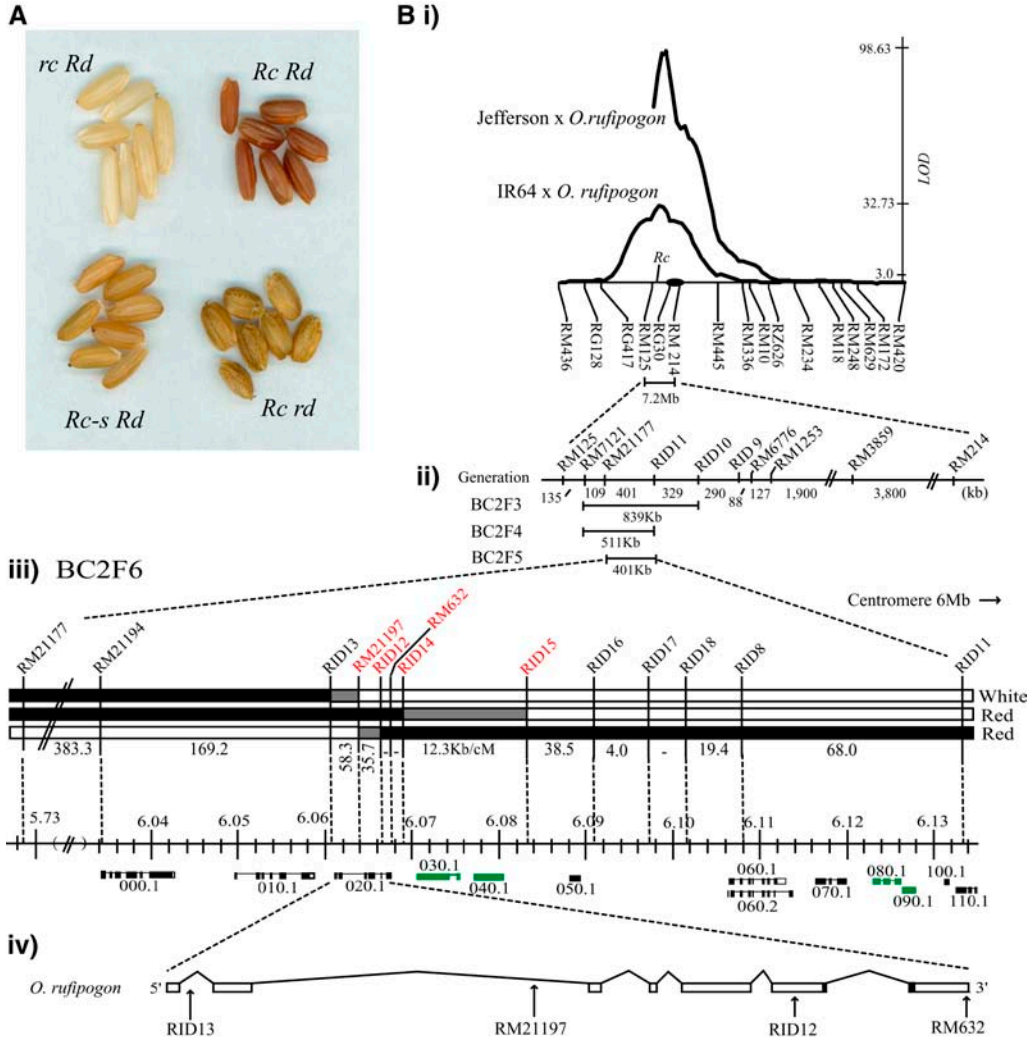


Figure 1. Phenotypes and Fine-Mapping of *Rc*.

(A) Rc allele phenotypes. Top row, left to right: seeds from cv Jefferson and *O. rufipogon*; bottom row, left to right: seeds from Surjamkuhi and H75. **(B)** Fine mapping. (i) QTL log of the odds (LOD) plot (*y* axis) and marker order (*x* axis). The black oval indicates the position of the centromere; *Rc* indicates the map position of *Rc*. (ii) Progress made narrowing the QTL by generation. (iii) Scheme of genotypes for three recombinant classes. Black bars represent DNA from the *O. rufipogon* parent, white bars represent DNA from the cv Jefferson parent, and gray bars represent an interval containing a break point between cv Jefferson and *O. rufipogon* DNA. White or red indicates the color of the pericarp conditioned by each class. The ruler shows where the markers and break points are positioned along the psuedomolecule for chromosome 7, and the numbers indicate megabase pairs. Markers that bracket or are included in the 18.5-kb target region are shown in red. The Institute for Genomic Research (TIGR) gene models are shown below the ruler, and numbers indicate the last digits of the gene identifiers, where all begin with LOC_Os07g11_. Transposable elements (TEs) are shown in green. Alternative splice variants are indicated by gene names ending with .1 and .2. The last four genes are staggered for clarity of presentation. (iv) The bHLH gene model, predicted from the *O. rufipogon* sequence, is enlarged to show the location of intragenic markers (indicated with arrows). The bHLH domain is indicated by black boxes at the end of exon 6 and the beginning of exon 7.

brown seeds, whereas *Rd* alone has no phenotype (Figure 1A). There are three known alleles of *Rc*: *Rc*, which exhibits brown spots on a reddish-brown background; *Rc-s*, which produces a light red color; and *rc* (historically, *Rc*+), which is a null allele. Although *Rc* is referred to as a mutant allele because its phenotype differs from that of common rice cultivars, the action of *Rc* is dominant over white pericarp (*rc*). This suggests that the modern cultivated (white) allele might actually be the mutant (nonfunctional) version of the ancestral *O. rufipogon* (red) allele. Both loci have been mapped using standard two-point analysis on the morphological map of rice: *Rc* on chromosome 7 and *Rd* on chromosome 1.

Proanthocyanidin biosynthesis is a branch of the anthocyanin biosynthetic pathway, a well-studied system in multiple species as a result of its visible phenotype and lack of detrimental effects on the plant. These pathways are regulated by homologs with similar functions in different species. The first gene is a R2R3 Myb homolog that contains an acidic activation domain in the C

terminus. Myb domains exhibit DNA binding ability and function in protein–protein interactions. These proteins are encoded by the *PURPLE PLANT1/COLORED ALEURONE1* (*PI1/C1*) genes in maize (*Zea mays*), by the *ANTHOCYANIN2* (*AN2*) and *AN4* genes in petunia (*Petunia hybrida*) (Spelt et al., 2000), and by the *TRANSPARENT TESTA2* (*TT2*) gene in *Arabidopsis thaliana* (Nesi et al., 2001).

The Myb proteins have been shown to interact with a basic helix-loop-helix (bHLH) protein in each of the three model systems. In *Arabidopsis*, *TT8* encodes a bHLH protein (Baudry et al., 2004), whereas petunia has two bHLH proteins involved in anthocyanin regulation, AN1 and JAF13 (Spelt et al., 2000). In maize, several genes belonging to the *RED/BOOSTER* (*R1/B1*) families encode these proteins (Goff et al., 1992). These genes have different tissue specificities and exhibit no activation domains or DNA binding activity alone (Goff et al., 1992). Recent experiments in maize suggest that R functions in part to free C1 from interaction with a repressor protein as well as to recognize R-specific anthocyanin promoter elements (Hernandez et al., 2004). Maize also contains a bHLH protein, *INTENSIFIER1* (*In1*), whose dominant allele acts as a negative regulator of pigmentation (Burr et al., 1996).

In the vegetative tissue of maize, only one member from each of the Myb and bHLH families has been shown to be required for pigmentation. However, in petunia, *Arabidopsis*, and maize seeds, genes encoding a WD40 protein are also required for the expression of anthocyanin biosynthetic genes. These proteins are encoded by the *TRANSPARENT TESTA GLABRA1* gene in *Arabidopsis*, by *AN11* in petunia, and by *PALE ALEURONE COLOR1* in maize (de Vetten et al., 1997; Baudry et al., 2004; Carey et al., 2004). They have been shown to physically interact with the bHLH protein in petunia and *Arabidopsis* (Walker et al., 1997; Sompornpailin et al., 2002). In petunia and *Arabidopsis*, other regulatory factors, such as *TT1*, a zinc finger protein (Sagasser et al., 2002), and *ANTHOCYANINLESS2*, a homeodomain protein (Kubo et al., 1999), have also been described where loss of function results in a complete lack of pigmentation.

In addition to being constitutively expressed, anthocyanin and proanthocyanidins can be induced by stresses, including cold, drought, and UV light. The regulatory elements that control these processes are only beginning to be understood. Recent studies have found that the basic domain/leucine zipper family of transcription factors, together with the Myb genes, play a role in induced expression (Ithal and Reddy, 2004; Hartmann et al., 2005).

We report here the cloning of a bHLH gene underlying a quantitative trait locus (QTL) for rice pericarp color. The QTL colocalizes with the mutant *Rc*. A frame shift deletion before the bHLH domain results in a knockout of proanthocyanidin production, leading to white rice.

RESULTS

Rough Mapping of QTLs and the *Rc* Mutant

Previous QTL mapping in this laboratory identified a single, significant QTL associated with red grain (*rg7.1*) on chromosome 7 (Figure 1B, i). This QTL was identified in two independent BC2 populations derived from crosses between an accession of *O. rufipogon* (IRGC-105491) from Malaysia and, in one case, a U.S. tropical *japonica* cultivar, Jefferson, and in the other case, a widely planted tropical *indica* cultivar, IR64. The log of the odds scores associated with the *rg7.1* QTL peaks in these two populations were 99 and 33, respectively, and the QTL was detected in multiple environments (Septiningsih et al., 2003). The peak of both QTLs corresponded to the previously mapped position of the mutant locus, brown pericarp, *Rc* (Kinoshita, 1998). All of the BC2F1 plants had red seeds, indicating that the *rg7.1* locus is dominant for red color, with the dominant allele donated by the *O. rufipogon* parent. Using the cv Jefferson/*O. rufipogon* population, *rg7.1* encompassed a 5.1-centimorgan (cM) region that represented ~7.2 Mb straddling the border of the centromere on chromosome 7 (Figure 1B, i). The genetic/physical distance in this region averages 1.4 Mb/cM, much above the genome average of 200 to 250 kb/cM, as expected for a pericentromeric region (Zhao et al., 2002; Wu et al., 2003). An investigation of the genotype–phenotype relationship in 285 BC2F2 families demonstrated that all 18 families with red grain contained the *O. rufipogon* allele at either RM125 or the adjacent marker, RG30, suggesting that the gene underlying *rg7.1* lay between these two markers.

Fine-Mapping of *rg7.1*

To fine-map the gene, 1410 BC2F3 plants were genotyped using markers flanking the QTL. The 72 recombinant plants were genotyped using six markers within the QTL region to locate the recombination break points more precisely, and the seed color of each recombinant line was recorded. Nineteen new simple-sequence repeat (SSR) and insertion/deletion (indel) markers were developed to help define break points across the region (see Supplemental Table 1 online). To narrow the region responsible for *rg7.1* in each successive generation, we determined which segments of DNA from the red donor parent, *O. rufipogon*, were shared in all red-seeded progeny and eliminated from further consideration the *O. rufipogon* segments that appeared in white-seeded progeny. Because the trait is dominant, both the heterozygous and the homozygous *O. rufipogon* classes had identical phenotypes and therefore were grouped together during fine-mapping.

In the final BC2F6 generation, 4000 plants were genotyped and three classes of informative recombinants were identified, which narrowed the *rg7.1* QTL to an 18.5-kb region (Figure 1B, iii). Class 1 consisted of a single plant with a break point defined by markers RID13 and RM21197, located between the first and second exons of the gene LOC_Os07g11020.1, as illustrated in panel iv of Figure 1B. Upstream of RID13, these plants inherited *O. rufipogon* DNA, and downstream of RM21197, they were homozygous for cv Jefferson alleles, with the break point delimited to the region between the two markers. Because these plants had white seeds, we concluded that *rg7.1* could not be located in the region upstream of RID13, which was heterozygous for *O. rufipogon* DNA. Recombinant class 2 consisted of 22 individuals having recombination break points between RID14 and RID15. This 14-kb region was highly repetitive, precluding the development of additional markers to help resolve the

precise position of each break point. All plants in this class had red pericarp, and because they were all homozygous for cv Jefferson DNA downstream of RID15, we were able to eliminate that region as the location of *rg7.1*. Class 3 consisted of a single red-seeded recombinant plant that contained an intragenic break point between RM21197 and RID12. It had cv Jefferson DNA upstream of RM21197, allowing us to eliminate the region between RID13 and RM21197 from further consideration. We thus defined an 18.5-kb target region for *rg7.1* bracketed by RM21197 and RID15 (Figure 1B, iii).

Physical/Genetic Distance

The evaluation of 4000 plants for recombination allowed us to identify recombinational hot and cold spots within the 401-kb region. The most recombinagenic interval was a 6-kb region that contained no annotated gene models, defined by markers RID16 and RID17 (Figure 2). Within this interval, the physical/genetic distance was 4 kb/cM, a recombination rate significantly higher than the average across the 401-kb region ($P < 0.001$). The 4-kb interval just downstream (RID17 to RID18) also lacked gene models but had no detectable recombination. The second most recombinagenic region (12.3 kb/cM; $P < 0.001$) was the 14-kb interval between RID14 and RID15 that contained two TEs. Again, this region was juxtaposed with a 2.3-kb region upstream in which there was no detectable recombination. The large interval between RM21177 and RM21194 (~300 kb), with 50 annotated gene models only 3 of which are TEs, had a recombination rate significantly lower than the average for the region ($P < 0.001$), even after recombination rates from the two hot spots were excluded from the average (Figure 2).

Positional and Functional Candidate Genes

One non-TE gene was detected within the 18.5-kb target region. LOC_Os07g11020.1 is a single-copy gene 668 amino acids in length and containing a predicted bHLH domain. This domain is common among transcription factors known to regulate pigment

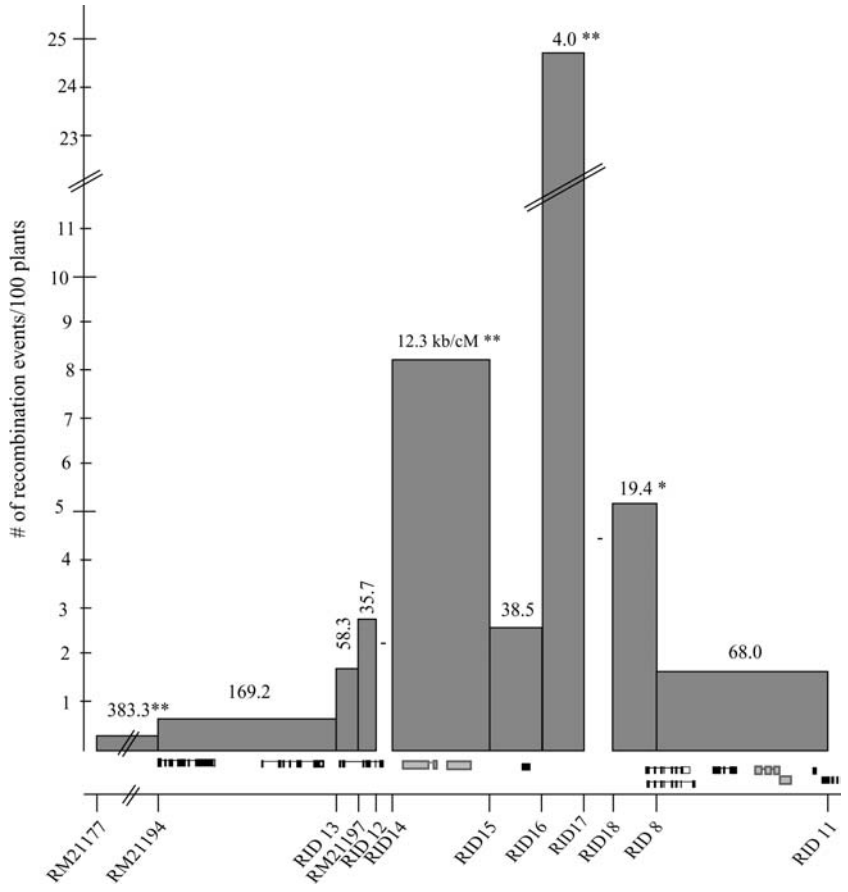


Figure 2. Physical/Genetic Distance around the Rc Locus.

Graph showing recombination rates across the 410-kb region examined in the BC2F6 generation of fine-mapping. The *x* axis is positioned along the psuedomolecule, and the *y* axis shows the number of recombination events per 100 plants. Physical/genetic distance is given in kb/cM above each interval. TIGR gene models are shown above the marker designations, with TEs in gray. Gene models are ordered as in Figure 1. Intervals indicated by asterisks have recombination rates significantly different from the average at the 5% level after Bonferroni correction; those indicated by double asterisks are significant at 0.1%.

synthesis. As illustrated in Figure 1B, panel iv, a recombination break point between RM21197 and RID12 in recombinant class 3 had eliminated the promoter region and the first two exons of the bHLH gene as the source of the functional nucleotide polymorphism, leaving only exons 3 to 7 within the 18.5-kb target region. In addition to the bHLH protein, two putative transposon proteins, LOC_Os07g11040.1 and LOC_Os07g11030.1, both of the CACTA type, En/Spm subclass, were also present within the 18.5-kb target region. These were eliminated from further consideration based on three lines of evidence. First, the sequences of both of these proteins have >100 BLAST hits in the rice genome, with >98% identity over the entire length of the sequences. Mutations in either of these highly repeated genes would have no phenotypic consequences, because many other copies would remain functional in the genome. Second, domain analysis showed that LOC_Os07g11040.1 contains a transposase domain and LOC_Os07g11030.1 contains a proteinase domain, neither of which has been found in any of the regulatory or biosynthetic proteins responsible for anthocyanin/proanthocyanidin pigmentation in plants. Third, although TEs can be responsible for phenotypic changes if they are inserted within functional genes, neither of the two TEs within the *rg7.1* QTL region shows any evidence of having disrupted any other genes. Therefore, based on our positional analysis and in silico functional interpretation, we postulated that the bHLH gene was responsible for red pericarp at the *rg7.1* locus. We proceeded to test this hypothesis using three additional lines of evidence: sequence comparison of parental lines, sequence comparison of an allelic series, and expression analysis of the bHLH gene.

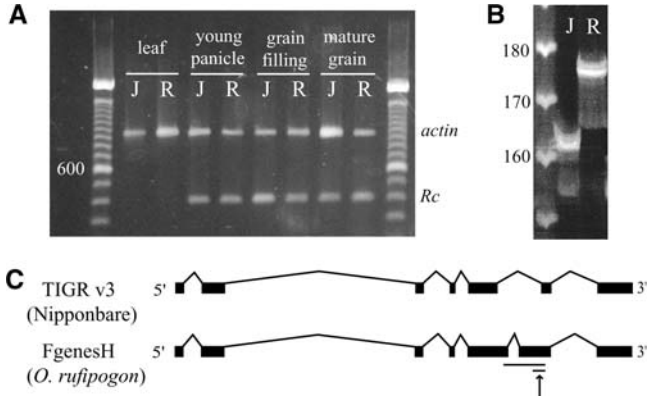


Figure 3. Expression Analysis of *Rc*.

(A) Transcripts of *Rc* and *actin* detected by RT-PCR in leaves, panicles before fertilization, pericarp from grains in the milk or dough stage of filling, and pericarp from mature seeds from both cv Jefferson (J; white seeds) and *O. rufipogon* (R; red seeds).

(B) Short transcripts of *Rc* detected by RT-PCR from the milk and dough stages of filling from cv Jefferson (J) and *O. rufipogon* (R) run out on polyacrylamide.

(C) Gene models from TIGR version 3 using the cv Nipponbare sequence and FgenesH prediction using the *O. rufipogon* sequence. The regions of mRNA amplified are indicated by horizontal lines, the longer one used in **(A)** and the shorter one used in **(B)**. The arrow indicates the location of the 14-bp deletion.

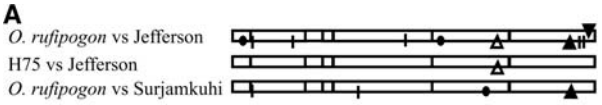
Sequence Comparison of the bHLH Gene

We sequenced the bHLH locus in both mapping parents, *O. sativa* (cv Jefferson) and *O. rufipogon*, to search for sequence changes that could explain the observed change in pericarp color. Having eliminated the promoter of the bHLH gene through recombination, we focused on changes that could affect the protein sequence. Six indels and 22 single-nucleotide polymorphisms were detected across the genomic sequence. We also compared the sequences of the mapping parents with the publicly available cv Nipponbare sequence. The cv Jefferson allele was identical to the Nipponbare sequence, both of which are *japonica* cultivars having white seeds.

Keeping in mind the possibility that annotation from the Nipponbare sequence might provide a gene model that differed from the dominant allele, we annotated the allele obtained from *O. rufipogon* and compared it with the gene model available at TIGR (http://www.tigr.org/tdb/e2k1/osa1). In the Nipponbare annotation, the 3′ end of exon 5 and the 5′ end of exon 6 were truncated relative to the *O. rufipogon* gene model, predicting an mRNA in Nipponbare that was 513 bp shorter than the mRNA predicted from the *O. rufipogon* annotation (Figure 3C). To confirm the accuracy of the different gene models, we amplified a segment of cDNA from both *O. rufipogon* and cv Jefferson within the only region in which the gene models differed. The cDNA amplicons were both 400 bp, the size expected from the *O. rufipogon* annotation (Figure 3A). Sequencing of this amplicon from *O. rufipogon* confirmed the splice sites predicted from the *O. rufipogon* annotation. When the polymorphisms between cv Jefferson and *O. rufipogon* were aligned with the new gene model, 10 of the sequence polymorphisms fell within the coding sequence, and 5 of those are expected to affect the protein sequence (Figure 4).

To help identify which of the sequence polymorphisms between the parents was responsible for the altered function of the gene, we also sequenced the bHLH locus in H75, an *Rc* mutant stock belonging to the *japonica* subspecies (Figure 1A). H75, like *O. rufipogon*, carries a functional allele, but it is much more closely related to cv Jefferson than to *O. rufipogon*. Thus, a sequence comparison between H75 and cv Jefferson was expected to help eliminate some of the nonfunctional polymorphisms detected between the parents in the bHLH gene. We found that the coding sequence of the bHLH allele in H75 was identical to the cv Jefferson sequence except for a 14-bp indel in exon 6 (Figure 4). This 14-bp sequence was present in the H75 stock as well as in *O. rufipogon*, but it was deleted in cv Jefferson and cv Nipponbare. The deletion induces a frame shift in the sequence, resulting in two premature stop codons before the end of exon 6. The stop codons truncate the protein before the bHLH domain. Given that this deletion was the only difference between the alleles of LOC_Os07g11020.1 in the H75 mutant stock (pigmented seeds) and the *japonica* cultivars cv Jefferson and cv Nipponbare (white seeds), that its location in exon 6 is consistent with the recombinational data, and that it would have a clear and important impact on gene function, we conclude that the 14-bp deletion is the only apparent reason for

A



B

nucleotide position in mRNA	brown	white	red	light red	phenotype
	japonica	*japonica*	*O. rufipogon*	*indica*	sub population
	H75	Jefferson	*O. rufipogon*	Sujamkuhi	name
	Rc	*rc*	*Rc*	*Rc-s*	allele
21	T	T	A	A	
96	G	G	T	G	
324	T	T	G	G	
660	A	A	A	G	
897	A	A	G	G	
1191	G	G	A	A	
1353	C	C	C	A	
1408-1421	ACGCGAA AAGTCGG	deleted	ACGCGAA AAGTCGG	ACGCGAA AAGTCGG	
1833-1844	deleted	deleted	(cgg)x4	(cgg)x2	
1923	T	T	C	C	
1962	T	T	C	C	
1982-1987	AAATGC	AAATGC	deleted	deleted	

Figure 4. Coding Sequence Differences between *Rc* Alleles.

(A) Graphic representation of coding sequence differences in LOC_Os07g11020.1 between several pairs of genotypes. The mRNA is represented by rectangles, and the beginning and end of the exons are indicated by vertical lines. Sequence changes are annotated as follows: closed circles, nonsynonymous substitution; lines, synonymous substitution; closed triangles, in-frame indel; open triangles, frame-shift indel; point-up triangles, deletion from *O. rufipogon* or H75; point-down triangle, insertion into *O. rufipogon*.

(B) Table showing polymorphic sites within the coding region of LOC_Os07g11020.1 for several different alleles of *Rc*. Functional nucleotide polymorphisms are highlighted in gray.

the lack of pigment in the pericarp of cv Jefferson and cv Nipponbare seeds.

To confirm that the bHLH protein underlying *rg7.1* is also *Rc*, we analyzed a second mutant stock for sequence variation in the bHLH gene. The stock, Surjamkuhi, is an *indica* line that carries a third allele, *Rc-s*, conditioning light red seed pigmentation. This genetic stock offered independent confirmation of the identity of the *Rc* gene because different sequence polymorphisms in the same gene would be expected to distinguish the *Rc-s*, *Rc*, and *rc* alleles. The sequence of the bHLH gene in Surjamkuhi differed from the sequence of the *japonica* cultivars at many sites (as expected for varieties from different subspecies) but differed from the *O. rufipogon* allele at only four sites (positions 96, 660, 1353, and 1833 to 1844) (Figure 4). The first two changes proved to be synonymous substitutions. The change at position 1353 consisted of a C-to-A change in exon 6. This single-nucleotide polymorphism was independent of any change seen in previous comparisons and represented a premature stop codon before the bHLH domain, truncating the protein and rendering the effect of the remaining indel immaterial. The fact that the different alleles of *Rc* show sequence polymorphisms that clearly account for the observed phenotypic differences is consistent with the conclusion that the bHLH protein is the *Rc* gene.

Expression Profiles of *Rc* and Biosynthetic Genes in White and Red Rice

To examine the timing and localization of the *Rc* transcript, we used RT-PCR to amplify mRNA from leaf, young panicle (before fertilization), pericarp of young seeds (at the milk or dough stage of grain filling), and pericarp from mature seeds. The mRNA was collected from both cv Jefferson (white seeds) and *O. rufipogon* (red seeds) plants. RT-PCR showed no expression of *Rc* in leaf tissue, as expected for a gene associated with a seed phenotype; however, expression was seen in several stages of panicle development (Figure 3A). Because the promoter of the bHLH gene had been eliminated as the source of polymorphism based on the recombination data, we anticipated that similar expression levels of *Rc* would be detected in red and white seeds. Our results confirmed this expectation and further demonstrated that the RNA transcript from cv Jefferson contained the 14-bp deletion predicted from the sequence information (Figure 3B).

Phylogenetic Comparison

To explore the evolutionary origin of *Rc* in rice and to identify putative orthologs in other species, we compared the sequence of *Rc* with other, previously identified bHLH transcription factors involved in anthocyanin and proanthocyanidin regulation as well as proteins with unknown effects that were recovered from BLAST using *Rc* as the query. The alignable portion of these sequences extended well beyond the bHLH domain (see Supplemental Figure 1 online), indicating that homology was not restricted to a single conserved functional domain. Analyzed using maximum parsimony or Bayesian analyses, these sequences fell into several clades (Figure 5).

The divergence between sequences of different clades is substantial, making outgroup selection and the position of the root uncertain. Among clades 1 and 2, clades 4 and 5, and within clade 3, further alignment was possible, strengthening our findings that these groups of sequences are more closely related to each other. Therefore, it is likely that the root lies on one of the branches separating the three main groupings (clades 1 and 2, clade 3, and clades 4 and 5) from one another.

Several copies of this type of transcription factor appear to have been present in the ancestor of the monocots and eudicots, as clades 1 and 2 contain both monocot and eudicot sequences (Figure 5). A third copy, present in the common ancestor of maize and rice, gave rise to clade 3, which shows gene duplication within each species. The paralogs within maize are known to confer tissue specificity of the anthocyanin pigmentation (Goff et al., 1992). Clades 4 and 5 contain only eudicot sequences (Figure 5).

It is clear from this analysis that *Rc* is not closely related to the rice bHLH proteins regulating anthocyanin, because they fall in different clades (Hu et al., 1996; Sakamoto et al., 2001). These rice anthocyanin regulators are sister to the maize anthocyanin regulators, and they map to homologous locations on rice chromosome 4 and maize chromosomes 2 and 10, respectively. The phylogenetic analysis clusters *Rc* with *In1* from maize. *Rc* and *In1* are located in homologous chromosomal regions on chromosome 7 of both genomes. *In1* is within maize bin 7.02, and of the 29 markers within 7.02 that map to the rice genome, 14 of them hit a

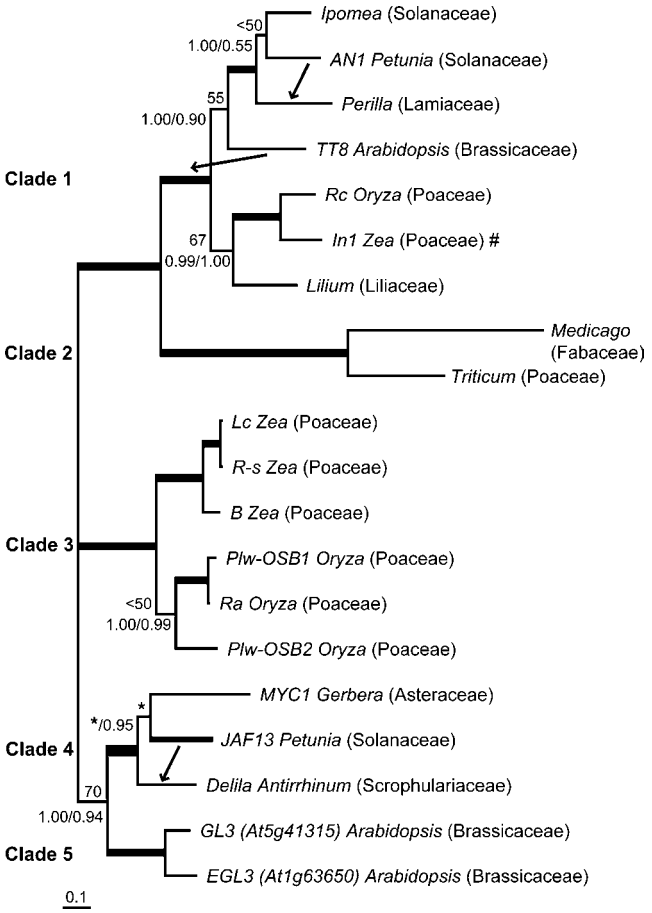


Figure 5. Phylogenetic Analysis of *Rc* and Other bHLH Proteins.

Topology derived from Bayesian analysis using the GTR+G model after 5 million generations (saving 5001 trees), discarding the first 100 trees as burnin. It was compared with a single most parsimonious tree of 3629 steps found after a heuristic search using 100 random addition sequence replicates holding a maximum of 100,000 trees (maximum not reached). The parsimony tree differed from the Bayesian analysis. The parsimony-preferred positions of taxa are shown by arrows where they connect to other branches; for example, in the parsimony tree, JAF13 and Delila are grouped together to the exclusion of MYC1. Parsimony bootstrap percentages (from 1000 pseudoreplicates) are shown above the branches, whereas posterior probabilities from the HKY model and the GTR+G model are shown, left and right, below the branches (after 100 burnin trees of 5001 trees were discarded). Thicker branches have >80% bootstrap and >0.95 posterior probability in each analysis. The topology should be considered unrooted. Genes that have been shown to function as regulators of the anthocyanin or proanthocyanidin pathways are named; additional sequences were retrieved from BLAST searches using *Rc* as the query and align across more than just the bHLH domain of ~60 amino acids. The pound sign indicates a known negative regulator.

3-Mb region on rice chromosome 7, surrounding *Rc*. The intron–exon structures of *In1* and *Rc* are similar, but not identical. *Rc* has seven exons, whereas *In1* has nine, but in both genes the bHLH domain spans an intron–exon boundary. Although the phylogeny of *Rc* and *In1* and the positions of the 14 homologous genes support orthology, *In1* and *Rc* have different functions. *In1* is a negative regulator of anthocyanin synthesis, which gives an intense purple color when mutated, whereas *Rc* is a positive regulator of proanthocyanidin synthesis, which has no color when the function is lost (Burr et al., 1996). All of the other genes on the tree whose functions have been established are positive regulators, suggesting that the ancestral function is positive regulation (Goff et al., 1992; Nesi et al., 2000; Spelt et al., 2000; Sakamoto et al., 2001; Bernhardt et al., 2003; Elomaa et al., 2003).

It is possible that two orthologous genes, *Rc* and *In1*, have evolved separate functions since the divergence of the common ancestor of rice and maize. To further test the hypothesis of orthology, we compared the nucleotide divergence (Ks) values between these genes and two other pairs of genes from bin 7.02 and the region around *Rc* (see Supplemental Figure 2 online). The Ks values for the *Rc–In1* comparison are consistent with the expected values of orthologous gene pairs from these taxa. The surrounding genes had Ks values above and below those from the *Rc–In1* comparison, showing that *Rc* and *In1* were not changing faster or slower than other genes in the homologous region. Given the homologous positions of these loci and Ks values that are in agreement with expected values for two of the three genes analyzed, there is no evidence that contradicts the orthologous relationship between *Rc* and *In1*.

Thus, it appears that the rice and maize bHLH genes associated with pigment production in plants have evolved separate functions over time. The rice genes found in different clades have specialized, becoming part of either the proanthocyanidin or anthocyanin pathway, whereas in maize, the duplicated genes have become positive and negative regulators of the anthocyanin pathway.

DISCUSSION

QTL analysis was used to identify the location of *rg7.1*, a locus for red grain, near the centromere on rice chromosome 7. A combination of fine-mapping, mutant analysis, and sequence comparisons demonstrated that a bHLH protein corresponding to the gene LOC_Os07g11020.1 was responsible for *rg7.1* as well as for the classically defined mutant alleles *Rc* and *Rc-s*.

Allelic Variation at *Rc*

The functional nucleotide polymorphism, a 14-bp deletion that knocked out gene function, was sufficient to explain the change in seed color between *O. rufipogon* and cv Jefferson. It remains to be seen whether independent mutations in this gene or other genes also give rise to the white phenotype in other lineages. We are currently undertaking an association study to determine the predictive power of these mutations for pericarp color in rice.

When comparing the *rc* and *Rc-s* alleles, it is not immediately apparent why the 14-bp deletion that frame shifts the bHLH domain should result in no pigment production and why a premature stop codon before the bHLH domain would give an allele conditioning light red color. In petunia and maize, *an1* and *b1* mutants lacking the bHLH domain are able to promote anthocyanin synthesis, much like the *Rc-s* allele (Liu et al., 1998; Spelt et al., 2002). Insertions that cause a frame shift within

the bHLH domain of *an1* produce null alleles, a parallel to *rc* in rice (Spelt et al., 2002).

Functional Parallels

In maize, petunia, and *Arabidopsis*, the bHLH proteins *R*/*B*, *AN1*, and *TT8* interact with a Myb transcriptional activator (*C1*, *AN2*, and *TT2*, respectively) to promote transcription of the anthocyanin and proanthocyanidin structural genes. In rice, only *Rc* has been reported as necessary for the production of colored seeds, although *Rd* is needed for the seed color to be red. The action of *Rd* is not consistent with that of the Myb transcription, as a functional copy of *Rd* is not required for the presence of any color in the pericarp. The *Rd* locus may be a biosynthetic gene that, when nonfunctional, results in the accumulation of a brown proanthocyanidin precursor. The rice gene *OsC1*, with similarity to the maize Myb gene *C1*, has been cloned on rice chromosome 6 (Saitoh et al., 2004). This locus is acknowledged to have a role in purple (anthocyanin) rice pigmentation but has not been shown to play a role in red (proanthocyanidin) rice pigmentation.

It is possible that, unlike the anthocyanin regulatory systems in maize, petunia, and *Arabidopsis*, the bHLH gene encoded by the *Rc* locus is sufficient to activate the transcription of the rice structural anthocyanin genes alone. The bHLH genes do function as transcription factors in homodimers or heterodimers in animal systems, so there is no inherent reason that proteins containing this domain could not activate transcription on their own (Heim et al., 2003). In maize, the *PERICARP COLOR1* Myb transcription factor has evolved so that it does not require interaction with a bHLH protein to activate the transcription of genes involved in phlobaphene production (Grotewold et al., 1994). An analogous change may have occurred in rice to allow the bHLH protein to function without an interacting partner. It is more likely, however, that rice also requires an interaction with a Myb transcription factor, but if both the red and white rices used in the genetic studies contain a functional copy of the Myb locus, it would not segregate and hence would not be detected as a contributor to pericarp color. If this is the case, a plant carrying a null allele at the Myb locus and a functional *Rc* should have white seeds.

Despite the differences in function between *Rc* and *In1* from maize, the phylogeny of bHLH genes and Ks values support an orthologous relationship. It should be noted that the expected Ks for orthologous sequences (0.65) is also within the range expected of ancient polyploid events in these taxa (Schlueter et al., 2004). Polyploidy could also have produced duplicated regions in the common ancestor of these taxa. If a paralogous region was lost in both the rice and maize lineages after divergence, then *Rc* rice and *In1* maize could be ancient paralogs but still have the Ks values we found. Regardless of the possibility of ancient paralogy, *In1* would still have diverged in function from the other positive regulators of anthocyanin and proanthocyanidin found in clade 1.

Phenotypic Associations

Red pericarp has long been used as marker for the cluster of domestication traits associated with weedy rice, including dormancy and shattering (Gu et al., 2005). Several studies have placed QTLs for dormancy and shattering in the pericentromeric region of chromosome 7, encompassing the *Rc* locus. With the cloning of *Rc*, it is now possible to ask whether this association is the result of linkage or pleiotrophy. Given the reduced rate of recombination within the *rg7.1* QTL, it is logical that genes for shattering, dormancy, and pericarp color have simply hitchhiked together in a linkage block. Indeed, fine-mapping of a rice shattering gene in this region has recently shown that this gene is tightly linked to *Rc*, although it occupies a different position on chromosome 7 (H. Ji, personal communication). It is also possible that *Rc* acts pleiotropically, as do many of the other bHLH proteins presented in the phylogeny (Payne et al., 2000; Spelt et al., 2002; Bernhardt et al., 2003; Zhang et al., 2003). Using the recombinant lines generated in this work, we will be able to test these different hypotheses.

Rice and wheat (*Triticum aestivum*) are similar in that red pericarp in both species can be eliminated by one locus. Comparative mapping shows no homology between the position of the *Rc* gene in rice and the *R* gene controlling red pericarp in wheat. A reverse genetics approach also failed to locate any ESTs from wheat that map to the *Rc* locus, although this is not surprising, because no rice or maize ESTs have been found for this locus either. Our work confirms that the *Rc* transcript only amplifies with a high sensitivity *Taq* polymerase, and this suggests that low transcript abundance may also explain the lack of EST hits in wheat. The *R* locus in wheat may be orthologous to the *Rd* gene in rice, given their homologous positions on wheat chromosome 3 and rice chromosome 1. Although the systems look similar phenotypically, molecular genetics analysis suggests that the mutations leading to white pericarp occurred at different points in the pathway.

Recombinational Analysis

The *rg7.1* locus was originally mapped to a 7.2-Mb region that included the centromere on rice chromosome 7. Given the low frequency of recombination across this region, it was not clear whether positional cloning would be feasible. This study demonstrates that even in regions that are recombinationally repressed, map-based gene isolation offers a viable approach.

The *Rc*/*rg7.1* locus is an area rich in gypsy retrotransposons and other repetitive elements (www.gramene.org). Several previous studies have noted significant repression of recombination associated with the abundance of TEs and other repetitive sequences (Fu et al., 2002; Wu et al., 2003; Shah and Hassan, 2005). Therefore, we expected that the TEs themselves might contribute to the recombinational repression. However, fine-mapping demonstrated that within the 401-kb region analyzed in the BC2F6 generation, one of the regions with a significantly increased recombination rate was TE-rich.

It has been shown that recombinational break points occur most often in regions with the greatest sequence similarity and that indels (including TEs) decreased the rate of crossovers more than single-nucleotide polymorphisms (Puget et al., 2002). The TE components of varieties may be different, and the lack of similarity would be one explanation for the low recombination in TE-rich regions. A recent study in humans showed that the retrotransposons THE1A and THEA1B are overrepresented in recombinational

hot spots and that the motif CCTCCCT, found associated with the retrotransposons, has a role in hot spot determination (Myers et al., 2005). In our study, one of the TEs in a region with enhanced recombination near *Rc* contains a CCTCCCT motif. This offers an alternative explanation for the significant increase in recombination in this particular TE-rich region.

Implications

The cloning of *Rc* will make possible new methods of fighting weedy rice infestations in rice paddies. Red rice is a noxious weed that is currently responsible for losses of as much as $50 million per year in the United States (Gealy et al., 2002). It is a perfect mimic of elite varieties, as the red pericarp is not visible until after harvest, when the grains are dehulled. Furthermore, the close association between red pericarp, seed shattering, and dormancy allows it to persist in fields for years despite vigorous attempts to remove it. The fact that the pericarp is maternal tissue, so that its color is dependent on the maternal genotype, means that seeds pollinated by red rice can be white (if the maternal plant carries the *rc* allele), but plants grown from these seeds will produce red seeds.

An immediate application of the work presented here involves the use of perfect markers that specifically target the 14-bp functional nucleotide polymorphism within the bHLH gene to screen for red rice contamination within certified seed lots. This will also facilitate the use of genes derived from crosses with wild relatives by allowing breeders to conclusively select against progeny carrying *Rc*, and to do so before the plants set seed.

METHODS

QTL Population Development and QTL Detection

A BC2F2 population was constructed for QTL mapping using *Oryza sativa* subsp tropical *japonica* cv Jefferson as the recurrent parent and a wild accession of *Oryza rufipogon* (IRGC-105491 from Malaysia) as the donor parent (Thomson et al., 2003). QTL detection was followed as described by Thomson et al. (2003).

Fine-Mapping of *rg7.1*

Eighteen BC2F2 families with red grain but with low levels of dormancy and shattering were used to fine-map the QTL. These families represent a subset of the red-grained phenotypes in the BC2F2 population. These segregating families were grown in 50-mm-wide × 178-mm-deep plastic pots in the Guterman Greenhouse at Cornell University. DNA was extracted using the Matrix Mill method (Paris and Carter, 2000). The microsatellite and indel markers were amplified using standard PCR protocols, run on 4% polyacrylamide electrophoresis gels, and silver-stained as described (Panaud et al., 1995). Seeds were harvested, and 10 to 15 seeds/plants were dehulled to determine seed color. Color was scored as either present or absent. In each generation, we determined which segments of DNA from the red donor parent, *O. rufipogon*, were present in all red-seeded progeny and discounted the *O. rufipogon* segments that appeared in white-seeded progeny as possible gene positions. Only plants heterozygous for the region containing *rg7.1* were planted for the next generation of screening. A total of 5922 individuals from BC2F3 to BC2F6 were genotyped.

Molecular Marker Development

The required density of molecular markers in the target region was achieved using previously published SSRs (McCouch et al., 2002) as well as SSR and indel markers developed as part of this study. New markers were designed from the publicly available rice genome sequence (http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice; http://143.48.220.116/genome_browser/index.html) using the SSRIT tool (http://143.48.220.116/db/searches/ssrtool). Indel markers were designed by sequencing areas from both cv Jefferson and *O. rufipogon* and aligning the sequences using the SeqMan program of DNAstar (GeneCodes) to identify indels. Primer sequences, map positions, and amplified lengths of the newly developed SSRs and indel markers used in this study are listed in Supplemental Table 1 online.

Recombination Rate Statistics

The recombination rate for each interval was compared with the average using likelihood ratio test statistics and P values for the test that compares the two-parameter model in which the fragment has its own unique recombination rate (while everything else is uniform) with the model in which everything is uniform. The likelihood of each recombination rate $= e^{\lambda}\lambda^{xi}/x_i!$ where $x_i =$ the number of recombinants in interval i, $\lambda = r$(length of sequence)(number of plants), and r is the per nucleotide recombination rate and is estimated as (total number of recombinants)/(total length of region)/(number of plants). The P values were corrected for multiple tests using a Bonferroni correction. The RM21177 to RM21194 interval was significantly lower than the average, even after the recombination rates from the two hot spots with $P < 0.001$ were excluded from the average.

Sequence Analysis of Red and White Cultivars

Overlapping primers between 700 and 900 bp were designed to cover the HLH gene and were used to amplify DNA with Pfu polymerase (Invitrogen). Amplified products were sequenced at the Cornell Biotechnology Resource Center.

RT-PCR

mRNA was collected from cv Jefferson and *O. rufipogon* panicles from the following stages and tissues: mature leaves, spikelets before pollination, dehulled seeds from the milk/dough stage of seed filling minus the starchy endosperm, and pericarp and seed coat scraped from the hardened endosperm of mature dried seeds. Total RNA was extracted with the RNAeasy miniplant RNA extraction kit (Qiagen). Total RNA (1 μg) was converted into cDNA with reverse transcriptase (Invitrogen), according to the manufacturer's instructions. As a control, water was used in place of reverse transcriptase for the reaction. The reaction was diluted threefold, and 2 μL of cDNA was used for amplification. A total of 10 μL of PCR products was loaded on 1% agarose gels. Primers for RT-PCR were designed to span introns. BLAST searches were used to ensure that primers were specific to the candidate gene of interest. The primers used were as follows: actin F, 5′-CGTCCTCTCTCTGTATGCCAG-3′; actin R, 5′-CTGGTACCCTCATCAGGCAT-3′; *Rc* F, 5′-GCCTTGTCACTCTTGGCATT-3′; *Rc* R, 5′-GGTTGGCACTGAAATCACCT-3′; *Rc short* F, 5′-CAGGCACCACACAGAGAATG-3′; *Rc short* R, 5′-CTCCTCTCTTTCAGCACATGG-3′.

Sequence Alignment and Tree Building

DNA sequences were aligned in BioEdit (Hall, 1999) with the assistance of ClustalW (Thompson et al., 1994) and bl2seq (Tatusova and Madden,

1999) to identify similar regions, with extensive manual adjustment. Alignment was undertaken to initially preserve the reading frame of each sequence. Some gaps were later introduced that disrupted small sections of the reading frame to minimize the total number of events (substitutions, insertions, or deletions) required to transform one sequence to another. Only some regions of the sequences could be confidently aligned across all taxa, including the bHLH domain (~1 kb). Further alignment was accomplished among sequences within the three main clades found in our analyses, with gaps for the remaining sequences (~500 bp). Gaps were treated as missing in all analyses. Indels were not coded as additional characters.

Analyses

The aligned matrix was examined for the presence of secondary signal(s) using reverse successive weighting, implemented in RSW1.1a (Trueman, 1998). Using a parsimony bootstrap cutoff of 70% and 1000 fast bootstrap replicates, no evidence of secondary signal was found. Further analyses were conducted with equal character weighting using parsimony implemented in PAUP* (Swofford, 1998) as well as Bayesian analysis using HKY and GTR+G models implemented in Mr Bayes (Huelsenbeck and Ronquist, 2001) version 3.1.1 to examine the sensitivity of the resulting phylogenetic inference to some model and method variation. Of 1449 aligned positions, 810 (55%) were parsimony-informative. A single most parsimonious tree of 3629 steps was found after a heuristic search using 100 random addition sequence replicates holding a maximum of 100,000 trees (maximum not reached); Bayesian analyses for both models used 10 chains and were run for 5 million generations, saving a tree every 1000 generations. Two replicate runs were compared with check convergence using the likelihood versus generation plot. Full mixing was evident from almost the first few generations, so an arbitrary 100 burnin trees were discarded before producing the consensus topologies and clade posterior probabilities. The parsimony tree differed slightly from the Bayesian majority rule tree (Figure 5). The parsimony-preferred positions of taxa are shown by arrows where they connect to other branches.

Synonymous changes per synonymous site (Ks) were calculated for *Rc* rice and *In1* maize (*Zea mays*) sequences along with sequences from two flanking genes from both species, No Apical Meristem putative protein and putative cellulose synthase sequences. We used the method of Yang and Nielsen (2000) implemented in PAML (Yang, 2000; Yang and Nielsen, 2000) to test whether these values were consistent with that expected for orthologous sequences, around Ks = 0.65, assuming a synonymous substitution rate of 0.65×10^{-9} (Gaut et al., 1996) and the age of divergence of rice and maize at 50 million years ago (Gaut, 2002). Two approaches were considered when deciding how different Ks values could be from the expected values and not be significantly different. In the first approach, we compared two standard errors above and below the estimated Ks for each gene with the expected value of 0.65. This represents an approximation of the 95% confidence interval around each estimate. The other approach we used was to treat the expected value as the mean of a distribution of Ks values, and therefore an estimate for any one pair of genes. Extending a similar approach that has been used in comparisons of genes duplicated by polyploidy (Zhang et al., 2003), we might expect the variation for 90% of genes duplicated by the same event (here speciation) to show 2.6-fold variation in Ks. Therefore, a range of 0.361 to 0.939 around 0.65 can be expected for most genes.

Accession Numbers

Sequence data for the genomic sequence of *Rc* can be found in the GenBank/EMBL data libraries under the following accession numbers: H75, DQ204735; cv Jefferson, DQ204736; *O. rufipogon*, DQ204737; Surjamkuhi, DQ204738. The cDNA fragment amplified from *O. rufipogon* has accession number DQ315482.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Table 1. Primers Developed for Fine-Mapping.

Supplemental Figure 1. Sequence Alignments Used in Phylogenetic Analysis.

Supplemental Figure 2. Ks Ranges for *Rc–In1* Surrounding Gene Pairs.

ACKNOWLEDGMENTS

We thank W. De Jong and J.J. Doyle for critical reading of the manuscript and L. Swales for administrative assistance. The *Rc* mutant stocks were provided by H.J. Koh. We are grateful to Scott Williamson for statistical consultation about hot spot analysis and to Dave Gealey for helpful discussions and for providing information about field problems associated with weedy red rice. This material is based upon work supported by National Science Foundation Grant 0110004. B.E.P. acknowledges support from National Science Foundation Grant 0321664.

Received October 6, 2005; revised December 13, 2005; accepted December 19, 2005; published January 6, 2006.

REFERENCES

Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana.* Plant J. **39,** 366–380.

Bernhardt, C., Lee, M.M., Gonzalez, A., Zhang, F., Lloyd, A., and Schiefelbein, J. (2003). The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the *Arabidopsis* root. Development **130,** 6431–6439.

Burr, F.A., Burr, B., Scheffler, B.E., Blewitt, M., Wienand, U., and Matz, E.C. (1996). The maize Repressor-like Gene intensifier1 shares homology with the r1/b1 multigene family of transcription factors and exhibits missplicing. Plant Cell **8,** 1249–1259.

Carey, C.C., Strahle, J.T., Selinger, D.A., and Chandler, V.L. (2004). Mutations in the *pale aleurone color1* regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar *TRANSPARENT TESTA GLABRA1* gene in *Arabidopsis thaliana.* Plant Cell **16,** 450–464.

Carmona, A., Borgudd, L., Borges, G., and Levy-Benshimol, A. (1996). Effect of black bean tannins on in vitro carbohydrate digestion and absorption. J. Nutr. Biochem. **7,** 445–450.

de Vetten, N., Quattrocchio, F., Mol, J., and Koes, R. (1997). The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes Dev. **11,** 1422–1434.

Eggum, B., Alabata, E., and Juliano, B. (1981). Protein utilization of pigmented and nonpigmented brown and milled rice by rats. Qual. Plant Plant Foods Hum. Nutr. **31,** 175–179.

Elomaa, P., Uimari, A., Mehto, M., Albert, V.A., Laitinen, R.A.E., and Teeri, T.H. (2003). Activation of anthocyanin biosynthesis in *Gerbera hybrida* (Asteraceae) suggests conserved protein-protein and protein-promoter interactions between the anciently diverged monocots and eudicots. Plant Physiol. **133,** 1831–1842.

Fu, H., Zheng, Z., and Dooner, H.K. (2002). Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. Proc. Natl. Acad. Sci. USA **99,** 1082–1087.

Gaut, B.S. (2002). Evolutionary dynamics of grass genomes. New Phytol. **154,** 15–28.

Gaut, B.S., Morton, B.R., McCaig, B.M., and Clegg, M.T. (1996). Substitution rate comparisons between grasses and palms: Synonymous rate differences in the nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proc. Natl. Acad. Sci. USA **93,** 10274–10279.

Gealy, D.R., Tai, T.H., and Sneller, C.H. (2002). Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Sci. **50,** 333–339.

Glahn, R., Cheng, Z., and Welch, R. (2002). Comparison of iron bioavailability from 15 rice genotypes: Studies using an in vitro digestion/caco-2 cell culture model. J. Agric. Food Chem. **50,** 3586–3591.

Goff, S., Cone, K., and Chandler, V. (1992). Functional analysis of the transcriptional activator encoded by the maize B gene: Evidence for a direct functional interaction between two classes of regulatory proteins. Genes Dev. **6,** 864–875.

Grotewold, E., Drummond, B., Bowen, B., and Peterson, T. (1994). The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell **76,** 543–553.

Gu, X.-Y., Kianian, S.F., Hareland, G.A., Hoffer, B.L., and Foley, M.E. (2005). Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). Theor. Appl. Genet. **110,** 1108–1118.

Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41,** 95–98.

Hartmann, U., Sagasser, M., Mehrtens, F., Stracke, R., and Weisshaar, B. (2005). Differential combinatorial interactions of *cis*-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol. Biol. **57,** 155–171.

Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. **20,** 735–747.

Hernandez, J.M., Heine, G.F., Irani, N.G., Feller, A., Kim, M.-G., Matulnik, T., Chandler, V.L., and Grotewold, E. (2004). Different mechanisms participate in the R-dependent activity of the R2R3 MYB transcription factor C1. J. Biol. Chem. **279,** 48205–48213.

Hu, J., Beth, A., 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.

Huelsenbeck, J.P., and Ronquist, F. (2001). Mr Bayes: Bayesian inference of phylogeny. Bioinformatics **17,** 754–755.

Ithal, N., and Reddy, A.R. (2004). Rice flavonoid pathway genes, OsDfr and OsAns, are induced by dehydration, high salt and ABA, and contain stress responsive promoter elements that interact with the transcription activator, OsC1-MYB. Plant Sci. **166,** 1505–1513.

Kato, S., and Ishikawa, J. (1921). On the inheritance of the pigment of red rice. Jpn. J. Genet. **1,** 1–7.

Kinoshita, T. (1998). Linkage mapping using mutant genes in rice. Rice Genet. Newsl. **15,** 13–74.

Kubo, H., Peeters, A.J.M., Aarts, M.G.M., Pereira, A., and Koornneef, M. (1999). ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis.* Plant Cell **11,** 1217–1226.

Ling, W.H., Cheng, Q.X., Ma, J., and Wang, T. (2001). Red and black rice decrease atherosclerotic plaque formation and increase antioxidant status in rabbits. J. Nutr. **131,** 1421–1426.

Liu, Y., Wang, L., Kermicle, J., and Wessler, S.R. (1998). Molecular consequences of *Ds* insertion into and excision from the helix-loop-helix domain of the maize R gene. Genetics **150,** 1639–1648.

McCouch, S.R., et al. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res. **9,** 199–207.

Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. Science **310,** 321–324.

Nesi, N., Clarisse, J., Debeaujon, I., Caboche, M., and Lepinec, L. (2001). The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell **13,** 2099–2114.

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and Lepiniec, L. (2000). The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. Plant Cell **12,** 1863–1878.

Oki, T., Masuda, M., Kobayashi, M., Nishiba, Y., Furuta, S., Suda, I., and Sato, T. (2002). Polymeric procyanidins as radical-scavenging components in red-hulled rice. J. Agric. Food Chem. **50,** 7524–7529.

Panaud, O., Chen, X., and McCouch, S.R. (1995). Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome **38,** 1170–1176.

Paris, M., and Carter, M. (2000). Cereal DNA: A rapid high-throughput extraction method for marker assisted selection. Plant Mol. Biol. Rep. **18,** 357–360.

Payne, C.T., Zhang, F., and Lloyd, A.M. (2000). GL3 encodes a bHLH protein that regulates trichome development in Arabidopsis through interaction with GL1 and TTG1. Genetics **156,** 1349–1362.

Puget, N., Gad, S., Perrin-Vidoz, L., Sinilnikova, O.M., Stoppa-Lyonnet, D., Lenoir, G.M., and Mazoyer, S. (2002). Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot. Am. J. Hum. Genet. **70,** 858–865.

Sagasser, M., Lu, G.-H., Hahlbrock, K., and Weisshaar, B. (2002). *A. thaliana* TRANSPARENT TESTA1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. Genes Dev. **16,** 138–149.

Saitoh, K., Onishi, K., Mikami, I., Thidar, K., and Sano, Y. (2004). Allelic diversification at the C (OsC1) locus of wild and cultivated rice: Nucleotide changes associated with phenotypes. Genetics **168,** 997–1007.

Sakamoto, W., Ohmori, T., Kageyama, K., Miyzaki, C., Saito, A., Murata, M., Noda, K., and Maekawa, M. (2001). The *purple leaf* (*pl*) locus of rice: The Pl^{w} allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. Plant Cell Physiol. **42,** 982–991.

Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J.J., and Shoemaker, R.C. (2004). Mining the EST databases to resolve evolutionary events in major crop species. Genome **47,** 868–876.

Septiningsih, E.M., Trijatmiko, K.R., Moeljopawiro, S., and McCouch, S.R. (2003). Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon.* Theor. Appl. Genet. **107,** 1433–1441.

Shah, M.M., and Hassan, A. (2005). Distribution of genes and recombination on wheat homoeologous group 6 chromosomes: A synthesis of available information. Mol. Breed. **15,** 45–53.

Shirley, B. (1998). Flavonoids in seeds and grains: Physiological function, agronomic importance and the genetics of biosynthesis. Seed Sci. Res. **8,** 415–422.

Sompornpailin, K., Makita, Y., Yamazaki, M., and Saito, K. (2002). A WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis in *Perilla frutescens.* Plant Mol. Biol. **50,** 485–495.

Spelt, C., Quattrocchio, F., Mol, J., and Koes, R. (2002). ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. Plant Cell **14,** 2121–2135.

Spelt, C., Quattrocchio, F., Mol, J.N.M., and Koes, R. (2000). anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell **12,** 1619–1632.

Swofford, D.L. (1998). PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0b10. (Sunderland, MA: Sinauer Associates).

Tatusova, T.A., and Madden, T.L. (1999). BLAST 2 sequences—A new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. **174,** 247–250.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22,** 4673–4680.

Thomson, M.J., Tai, T.H., McClung, A.M., Lai, X.-H., Hinga, M.E., Lobos, K.B., Xu, Y., Martinez, C.P., and McCouch, S.R. (2003). Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor. Appl. Genet. **107,** 479–493.

Trueman, J.W.H. (1998). Reverse successive weighting. Syst. Biol. **47,** 733–737.

Vaughan, L.K., Ottis, B.V., Prazak-Havey, A.M., Bormans, C.A., Sneller, C., Chandler, J.M., and Park, W.D. (2001). Is all red rice found in commercial rice really *Oryza sativa*? Weed Sci. **49,** 468–476.

Walker, E., Eggleston, W.B., Demopulos, D., Kermicle, J., and Dellaporta, S.L. (1997). Insertions of a novel class of transposable elements with a strong target site preference at the *r* locus of maize. Genetics **146,** 681–693.

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. **126,** 485–493.

Wu, J., et al. (2003). Physical maps and recombination frequency of six rice chromosomes. Plant J. **36,** 720–730.

Yang, Z. (2000). Phylogenetic Analysis by Maximum Likelihood (PAML), Version 3.0. (London: University College).

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. **17,** 32–43.

Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development **130,** 4859–4869.

Zhao, Q., et al. (2002). A fine physical map of the rice chromosome 4. Genome Res. **12,** 817–823.

Caught Red-Handed: *Rc* Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice

Megan T. Sweeney, Michael J. Thomson, Bernard E. Pfeil and Susan McCouch

Plant Cell 2006;18;283-294; originally published online January 6, 2006;
DOI 10.1105/tpc.105.038430