ORIGINAL PAPER

A natural mutation in *rc* reverts white-rice-pericarp to red and results in a new, dominant, wild-type allele: *Rc-g*

Steven A. Brooks · Wengui Yan · Aaron K. Jackson · Christopher W. Deren

Abstract The *Rc* locus regulates pigmentation of the rice bran layer, and selection for the *rc* allele (white pericarp) occurred during domestication of the crop. White bran is now ubiquitous among cultivated varieties throughout rice growing regions of the world. We identified a new allele that arose by natural mutation within the *rc* pseudogene of the cultivar 'Wells'. The mutation restored the reading frame of the gene, and reverted the bran layer pigmentation to red (wild-type). By sequencing the *Rc* locus in plants derived from red seeds, and linkage analysis in a segregating population, we were able to demonstrate that mutation within *rc* resulted in the new, dominant, wild-type allele *Rc-g*.

Introduction

Since the origins of agriculture humans have selected traits in plants that were useful in cropping systems. Propagation of specific variants, with repeated selection over time, imposed genetic bottlenecks that increased the uniformity of crops and fixed desirable traits in cultivated germplasm. Such traits in agricultural crops are often conferred by so-called 'domestication genes', which distinguish cultivated crops from their wild and frequently weedy progenitors. Cultivated rice (*Oryza sativa* L.) is a crop of worldwide importance. The vast majority of modern varieties have a white pericarp, which is conferred by the *rc* domestication pseudogene (Furukawa et al. 2007; Sweeney et al. 2006). White pericarp distinguishes cultivated rice from wild relative species, and also from 'red rice' which is a common weed of the same *O. sativa* species. These wild-types have the dominant red pericarp allele at the *Rc* locus.

Recently two groups independently cloned the *Rc* gene (Furukawa et al. 2007; Sweeney et al. 2006); and sequenced the wild-type allele (*Rc*), the domestication allele (*rc*), and a mutant allele (*Rc-s*; Sweeney et al. 2006). Both groups demonstrated that *rc* was a null-allele, produced by a 14 bp deletion (in exon seven using the mRNA-based gene model of Furukawa et al. 2007), which caused a frame shift mutation and a premature stop codon. The *rc* mutation results in a loss of proanthocyanidin synthesis and corresponding pigmentation of the pericarp (Furukawa et al. 2007).

It is speculated that selection for white rice grains was historically for sanitary purposes, where the trait permitted easy identification of impurities in stored grain. Post-harvest insect and rodent damage being more easily detectable against a white rice background. The trait is also important to date for producing polished (milled) white rice. However, red pericarp can be desirable in specialty rices, and there are many good-quality, locally-adapted, red rice varieties grown throughout Asia. Unfortunately, these varieties are not adapted to the cultivation systems in the United States, and introgressing the *Rc* allele from non-adapted germplasm would require years of backcrossing and selection to produce useful varieties. Therefore, a source of red pericarp in adapted germplasm would be very valuable for specialty rice breeding in the United States.

Communicated by T. Sasaki.

S. A. Brooks (✉) · W. Yan · A. K. Jackson
USDA ARS, Dale Bumpers National Rice Research Center,
2890 Hwy. 130 E., Stuttgart, Arkansas 72160, USA
e-mail: ricegenes@mac.com

C. W. Deren
University of Arkansas, Rice Research and Extension Center,
Stuttgart, Arkansas 72160, USA

Arkansas is the leading producer of rice in the United States, and all varieties in commercial production have a white pericarp (*rc*). Red rice is a common weed of Arkansas rice culture. Although it has the niche-desirable *Rc* allele, it does not possess cultivated plant characteristics, and is a noxious weed with no agronomic value. Red rice has medium grain length, is taller than conventional cultivars, shatters easily, and its seed can remain dormant in the soil between growing seasons. Red rice also outcrosses to cultivated rice at low frequency (Gealy et al. 2002), causing great concern over genetic purity of cultivars and transfer of herbicide resistance to weed populations. Therefore, strong selection against red rice has been enforced in breeding programs, foundation seed, and on-farm operations, which limits the spread of weedy red rice and the monetary losses resulting from mixed-seed crops.

Recently typical long grains of the cultivar 'Wells', the most extensively grown cultivar in Arkansas, were identified with red pericarp. Wells is a high yielding, long-grain cultivar developed and patented by the University of Arkansas (Moldenhauer et al. 2007). Plants grown from red seed were of cultivated idiotype, which raised immediate concern regarding the genetic purity of Wells. Red rice is the only source of *Rc* in Arkansas rice production, and an outcross of Wells to red rice would have dramatic consequences. In this study, we sought to identify the source of red pericarp in the off-type plants using molecular markers, linkage and DNA sequence analysis. We have shown that Wells is a genetically pure cultivar, and red pericarp arose by natural mutation within the *rc* pseudogene that created the new *Rc-g* allele. Furthermore, the germplasm is an ideal source of red pericarp for specialty rices, as it has all the properties of an elite cultivar with no weedy or un-adapted traits.

Materials and methods

Plant material

The plant materials described herein were identified by the University of Arkansas, Rice Research and Extension Center, foundation seed program, located in Stuttgart Arkansas. Rice kernels of typical long-grain-cultivated type were identified with red pericarp in 2005-produced seed of the cultivar Wells. These seeds were clearly not of weedy red rice type, and were saved to determine if outcrossing was the source of red pericarp. Plants grown from the red-pericarp-rice were phenotypically identical to Wells, except for pericarp color, and the single plants selections were named Red Wells.

Molecular marker analysis

Microsatellite markers (SSRs, Table 1) were chosen based on in-house availability, robust amplification, high polymorphism, and previous utility for 'fingerprinting' rice cultivars (M. Jia, personal communication). Polymerase chain reaction (PCR) for SSRs was performed according to Gealy et al. (2002). The RID12 marker developed by Sweeney et al. (2006) was used to detect the *Rc/rc* functional nucleotide polymorphism (FNP). PCR reactions for RID12 were 20 μl in volume, used 50 ng template DNA, 1 pmol of each primer, 1 U of Taq DNA polymerase (Promega, Madison, WI), 1.6 μl of 25 mM $MgCl_2$, 2 μl of 10× buffer, and 0.2 μl of 10 mM dNTPs. PCR was performed in an Eppendorf (Hamburg, Germany) Mastercycler, where reaction conditions were: 3 min at 94°C, followed by 35 cycles each of 1 min at 92°C, 1 min at 55°C, 1 min at 72°C, and a final cycle of 72°C for 5 min.

DNA sequence analysis

A PCR-based approach was used to amplify the *Rc* locus, in overlapping fragments, for DNA sequencing. PCR prim-

Table 1 SSR marker comparisons of Wells and Red Wells

SSR marker[a]	Chromosome location	Allele size (bp)		
		Wells	Red Wells	STG-S
RM 5	1	114	114	109
RM 22	3	191	191	194
RM 124	4	271	271	267
RM 125	7	126	126	123
RM 154	2	179	179	165
RM 174	2	221	221	206
RM 190	6	113	113	107
RM 210	8	151	151	153
RM 214	7	149	149	112
RM 219	9	190	190	220
RM 224	11	139	139	128/154
RM 225	6	142	142	130/142
RM 231	3	180	180	184
RM 232	3	158	158	148
RM 234	7	135	135	153
RM 248	7	84	84	91
RM 345	6	163	163	169
RM 408	8	127	127	125
RM 420	7	186	186	186
RM 481	7	217	217	176
RM 484	10	298	298	293
RM 490	1	97	97	93

[a] All information for 'RM' SSR markers, including primer sequences and physical genome position, is available on the Gramene (2007) web site (http://www.gramene.org)

ers were designed using PrimerSelect (DNASTAR®, Madison, WI) based on the *rc* allele sequence of the rice cultivar 'Jefferson' (DQ204736). PCR conditions were nearly identical to those used for RID12, with the only difference being optimization of specific primer pair annealing temperatures. Direct sequencing of PCR products was performed by the Genomics Core Facility at the Dale Bumpers National Rice Research Center. DNA sequence assembly and alignment was performed with SeqMan II (DNASTAR®, Madison, WI).

RCG-FNP marker development

Two single nucleotide polymorphism (SNP) based PCR primer pairs were designed to selectively amplify the *Rc-g* and *rc* alleles. Forward primers were designed to terminate on a single selective base, at position 1388 of the *O. rufipogon Rc* coding sequence (DQ204737). pSNP-red (5′-AGA-AACACCTGAATCAATGGC-3′) and pRedReverse (5′-GAGCTCTTGTATGCGGTTCCTTAG-3′) produce a 889 bp fragment from the *Rc-g* allele. pSNP-white (5′-AGA-AACACCTGAATCAAGTGG-3′) and pWhiteReverse (5′-GGATACGGGTAGGATTCACTTCTG-3′) produce a 668 bp fragment from the *rc* allele. Since individual primer pairs were dominant markers, both sets were used individually (on all samples) and pooled after amplification, to produce an essentially co-dominant FNP fingerprint. Different reverse primers were used to produce amplicons of significantly different size that could be resolved on agarose gels. Note: the pSNP-white primer will also amplify the *Rc* allele (red), but the fragment will be 14 bp larger than the *rc* fragment, as these primers also span the *Rc/rc* FNP. Therefore, the RCG-FNP marker, a combination of two dominant PCR products, can be used to score all three (*Rc*, *Rc-g* and *rc*) alleles simultaneously.

Polymerase chain reactions for the RCG-FNP were 20 μl in volume, used 100 ng template DNA, 0.5 pmol of each primer, 1 U of Taq DNA polymerase (Promega, Madison, WI), 1.4 μl of 25 mM MgCl₂ 2 μl of 10× buffer, and 0.2 μl of 10 mM dNTPs. PCR was performed in an Eppendorf (Hamburg, Germany) Mastercycler and reaction conditions were: 94°C for 3 min, followed by 35 cycles of 92°C for 2 min, 45 s at 60°C (pSNP-red) or 62°C (pSNP-white), 72°C for 1 min 30 s, and a final cycle of 72°C for 10 min. All samples were tested in triplicate and analyzed on 1.5% agarose (Bio-Rad, Hercules, CA) gels.

F₂ mapping population

Reciprocal crosses from single plant selections of Wells and Red Wells were made in the spring of 2006, and F₁ plants grown in the summer of the same year. F₂ seed was collected from a single F₁ (Wells × Red Wells), and 92 F₂s

were grown in the greenhouse during the winter of 2006–2007. Leaf tissue was harvested for DNA extraction and molecular marker analysis at the seedling 4 leaf stage. A single panicle from each F₂ was collected at maturity and de-hulled (palea and lemma removed) to score pericarp color.

Results

Identification and frequency of Red Wells

Using standard procedures required for certification of foundation seed (Arkansas State Plant Board 2002), approximately 150 red seeds were found in a 56 ton seed lot of the rice cultivar Wells. Single plant selections from white and red seeds were phenotypically identical for plant type, panicle morphology and grain type (Fig. 1). Furthermore, 24 F₁s of reciprocal crosses made from the selected plants, and 96 derived F₂s (not shown), were all uniform for plant type. Quantitative measurements were also made for plant height, tiller number, days to heading, glabrous leaves (plus vs. minus), shattering and kernel dimensions (not shown). The only differences observed between Wells (*n* = 18) and Red Wells (*n* = 18) were in mean plant height (92 ± 3 vs. 96 ± 4 cm respectively) and kernel dimensions (7.6 ± 0.2 vs. 7.1 ± 0.2 mm for kernel length; and 2.0 ± 0.1 vs. 1.9 ± 0.2 mm for kernel width; Wells vs. Red Wells). Although the differences were statistically significant (*P* = 0.05, *df* = 34) for the single plant selections, they were negligible in degree and the values fall within the known standards for Wells (Slaton et al. 2000).

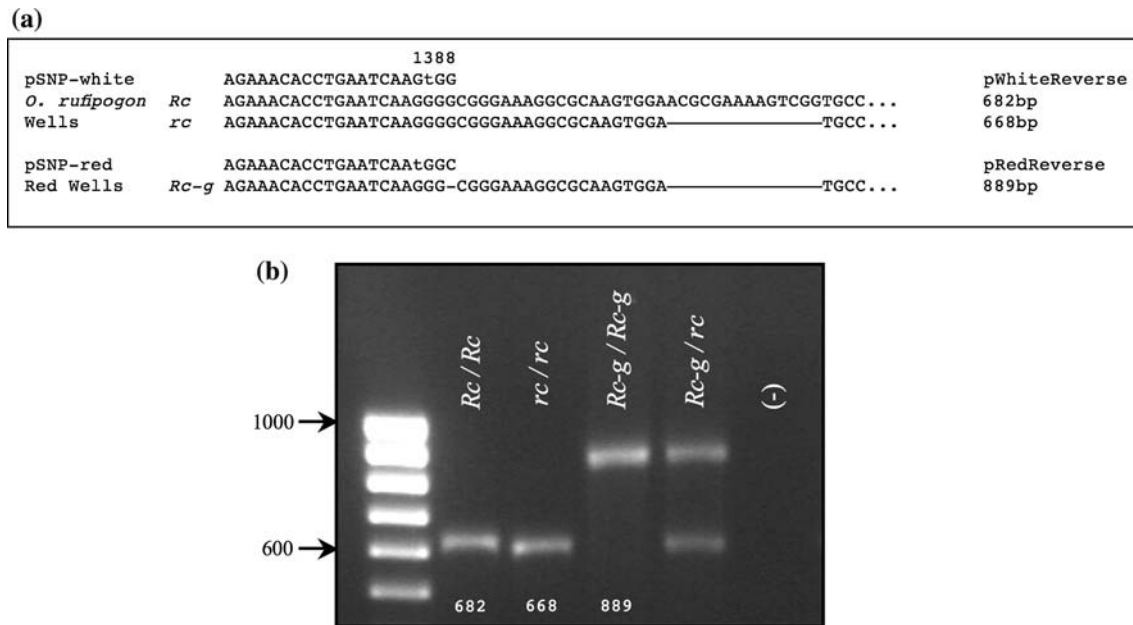### Red Wells did *not* arise from an outcross to red rice

A total of 22 simple sequence repeat (SSR) markers, located on 10 of the 12 rice chromosomes, were used to search for non-Wells alleles in Red Wells (Table 1). All

Fig. 1 Close-up images of Wells (*left*) and Red Wells (*right*), de-hulled seed (*top*) and 'paddy rice' (*bottom*)

Fig. 2 PCR amplification with the RID12 (Sweeney et al. 2006) marker. Rice cultivar Wells (*rc*; *lanes 2* and *3*), Red Wells mutant (*Rc-g*; *lanes 4* and *5*), red rice accession STG-S (*Rc*; *lane 6*), red rice accession RR8 (*Rc*; *lane 7*), and red rice accession LA3 (*Rc*; *lane 8*). Negative (no template DNA) control in *lane 9*. Size standard HyperLadder IV (Bioline, Randolph, MA), with 100 and 200 bp markers indicated at left (*lanes 1* and *10*). RID12 target fragment indicated by *white arrows*. Fragments resolved on 2.3% Metaphor® agarose (Cambrex, Rockland, ME)

SSR markers produced identical allele sizes in Wells and Red Wells, and were different from Stuttgart straw hull (STG-S, except monomorphic marker RM420), the most common red rice type on the experiment station where Red Wells was identified. To assure that low marker resolution did not miss a localized introgression at the *Rc* locus, the RID12 marker, that distinguishes *Rc*/*rc* FNP was used to screen Wells, Red Wells, and three accessions of red rice (Fig. 2). Wells (Fig. 2; lanes 2 and 3) and Red Wells (Fig. 2; lanes 4 and 5) had the *rc* FNP, which is 14 bp smaller than the fragment amplified in *Rc*-allele-containing common red rice accessions (STG-S, RR8, and LA3; Fig. 2; lanes 6, 7, and 8 respectively). The presence of the *rc* FNP in Red Wells provides conclusive evidence that the dominant *Rc* phenotype was not the result of an outcross to an *Rc*-type red rice.

Rc-g is a novel, dominant allele that arose by natural mutation in Red Wells

DNA sequence alignments of Wells and Red Wells within the exon seven region of the *O. rufipogon* (DQ204737) *Rc* gene sequence (Fig. 3), revealed a 1 bp deletion (guanine) in Red Wells at position 1388. This feature is 20 bp upstream of the *rc* FNP and lies within the RID12 amplicon, but was undetected in Fig. 2 due to limited resolution of the agarose gel. Therefore, both Wells and Red Wells have the *rc* FNP (14 bp deletion), but Red Wells has an additional 1 bp deletion resulting in a RID12 fragment that is 15 bp shorter than obtained from a *Rc* allele. The 1 bp deletion was the only sequence polymorphism observed between Wells and Red Wells. The *rc* coding sequence in Wells was also identical to the published sequence for the rice cultivar Jefferson (DQ204736). Therefore, additional GenBank submissions were not made.

The close proximity of two deletions within exon seven, upstream of the functional domain, restored the reading frame of the *rc* pseudogene to a functional allele (*Rc-g*) in Red Wells. Putative amino acid (aa) sequence alignments based on the *O. rufipogon* gene model (Sweeney et al. 2006) show the premature stop codon in Wells (*rc*) at aa position 474 (Fig. 3) that results from the 14 bp FNP. The *Rc-g* allele encodes a putative full-length polypeptide, with a five aa deletion (positions 466–470) and four aa substitutions (positions 463, 471, 472, and 474) relative to the *Rc* protein sequence. Perfect alignment exists between *Rc* and *Rc-g* aa sequences before aa-463 and following aa-474 to the end of the protein (668 aa total for *Rc*; 663 aa for *Rc-g*).

RCG-FNP distinguished the *Rc-g* and *rc* alleles

A SNP-based PCR marker was developed to resolve the *Rc-g*/*rc* FNPs on standard agarose gels. Figure 4a is a diagrammatic representation of primer binding positions, and Fig. 4b shows the results of amplification from homozy-

Fig. 3 DNA and putative protein sequence alignments of *Rc*, *rc*, and *Rc-g* alleles. *Rc* (*O. rufipogon* cv. IRGC105491; DQ204737; Sweeney et al. 2006), *rc* (*O. sativa* cv. Wells), and *Rc-g* (*O. sativa* mutant Red Wells). The *Rc-g* 1 bp-FNP at position 1388 and the *rc* 14 bp-FNP (positions 1408–1421) are indicated by *arrows*. The region of DNA sequence translation is delineated be dashed lines. *Asterisks* indicate a premature stop codon in *rc*. Deleted nucleotide and amino acid positions are filled with *dashes* (–) to justify sequence alignments. Nucleotide and amino acid positions (*O. rufipogon*; *Rc*) are indicated above the alignments

Fig. 4 a SNP primer alignments to corresponding DNA target sequences of *Rc*, *rc*, and *Rc-g* alleles. *Rc* (*O. rufipogon* cv. IRGC105491; DQ204737; Sweeney et al. 2006), *rc* (*O. sativa* cv. Wells), and *Rc-g* (*O. sativa* mutant Red Wells). The *Rc-g* 1 bp-FNP at position 1388 and the *rc* 14 bp-FNP (positions 1408–1421) are indicated by *dashes*. Amplicon sizes for primer pairs are indicated at *right*. **b** PCR amplification with the RCG-FNP marker. Size standard HyperLadder IV (Bi-oline, Randolph, MA), with 600 and 1,000 bp markers indicated at left (*lane 1*), and band sizes in bp indicated below *lanes 2–4*. Red rice accession STG-S (*lane 2*), rice cultivar Wells (*lane 3*), Red Wells mutant (*lane 4*), and a F$_1$ from the Red Wells/Wells cross (*lane 5*). Negative (no template DNA) control in lane 6. Fragments resolved on 1.5% agarose (Bio-Rad, Hercules, CA)

gous *Rc*, *rc*, *Rc-g*, and a *Rc-g/rc* heterozygote. The RCG-FNP was then used to score 92 segregating F$_2$s derived from a cross between Wells and Red Wells, which had also been scored for pericarp color (F$_3$ seed, not shown). The critical values for chi-square are 3.84 for trait data (*df* = 1, alpha = 0.05), and 5.99 for marker data (*df* = 2, alpha = 0.05). The calculated chi-squares were 2.84 and 4.46, respectively, and failed to reject the null hypothesis for a single dominant gene. In all F$_2$ progeny perfect marker-trait co-segregation was observed, where *Rc-g/Rc-g* and *Rc-g/rc* genotypes had red seed and *rc/rc* genotypes were white seeded.

Discussion

Red rice contamination of cultivated seed lots is a common phenomenon in the southern United States, where red rice is a prevalent and problematic weed species. Red rice grain type is different from cultivated rice and can be excluded based upon grain shape and color. The identification of typical long grains with red pericarp and cultivated idiotype in derived plants was alarming and posed a serious problem for certifying weed-free foundation seed. Since the probability of a mutation resulting in a dominant phenotype is low, we first considered an outcross as cause of red pericarp in Red Wells. However, identical plant and grain types without the

presence of red rice traits caused immediate doubt for the outcross hypothesis. Using the recently published DNA sequence for *Rc* (Sweeney et al. 2006) and markers for the gene, we were able to demonstrate that mutation restored function to the *rc* pseudogene in Red Wells.

The presence of the *rc* FNP in Red Wells eliminated an outcross from consideration as the source of red pericarp, since all wild-type red rices lack this 14 bp deletion. Therefore, mutation either within *rc* or another gene in the same biochemical pathway were considered as the basis for the reversion phenotype. DNA sequence analysis revealed a 1 bp deletion within the *rc* pseudogene in Red Wells. This feature was considerable as it could cause a reversion from domesticated-type to wild-type. The close proximity of two deletions within exon seven could restore the reading frame of the gene, and result in an allele with little deviation from wild-type. Interpretation of the DNA sequences was convincing, but the perfect co-segregation of FNP and trait data provided sound evidence that a reversion to wild-type occurred via this mutation. The new allele was designated as *Rc-g*, symbolizing the guanine deletion as the allele's distinctive feature.

Rc-g is a naturally occurring, low frequency mutation. At present the precise frequency of *Rc-g* occurrence in Wells is unknown, though the observed occurrence was equivalent to a single plant in 56 tons of foundation seed. Low frequency of *Rc-g* is expected as a zero tolerance for

red bran exists in the University of Arkansas foundation seed program, and is selected against to avoid red rice contamination of seed lots. In addition, Moldenhauer et al. (2007) reported off-type occurrence in Wells at a frequency of less than 1 in 5,000 plants, and did not report red pericarp as an off type for the cultivar.

Significance

Red rice is the most important weed problem of rice culture in the southern United States, causing significant economic losses in infested areas (summarized in Gealy et al. 2002). Since outcrossing of cultivated rice with red rice is known (Gealy et al. 2002), the ability to determine the source of red pericarp in Red Wells was of paramount importance. Wells has been the leading inbred rice cultivar in Arkansas since 2002 (Wilson 2002), and genetic contamination of the cultivar would have had enormous economic impact. Identification of the *Rc-g* allele removes the assumption that cultivated rice-grain-types with red pericarp are the result of outcrosses to red rice [Similar phenotypic observations have been made in other cultivars, where typical long-grain-rice with red pericarp has been found (A. M. McClung, personal communication)]. Use of the RID12 (Sweeney et al. 2006) and the RCG-FNP markers will prove useful to identify the *Rc*, *rc*, and *Rc-g* alleles and to certify foundation seed lots.

The Red Wells mutant represents a novel germplasm resource for specialty rice breeding. *Rc* encodes a regulatory protein that enhances the accumulation of proanthocyanidins in the rice pericarp (Furukawa et al. 2007). The health beneficial properties of proanthocyanidin and related secondary metabolites are well known. Ling et al. (2001) demonstrated that red rice consumption reduced progression of atherosclerotic plaque development induced by dietary cholesterol (the area of aortic atherosclerotic plaques was 50% lower in rabbits fed red rice diets than those fed a white rice diet). In future research, we intend to evaluate Red Wells for beneficial properties (i.e. antioxidant capacity)

versus its isogenic parent cultivar Wells. The exploitation of Red Wells circumvents linkage drag from weedy traits (seed shattering and dormancy are tightly linked to *Rc* on chromosome 7; Ji et al. 2006), that occur when introgressing alleles like *Rc* from wild species. Red Wells has immediate utility as a specialty rice variety, and will prove useful as a source of the *Rc-g* allele (selectable with the RCG-FNP perfect marker) in an elite genetic background.

References

Arkansas State Plant Board (2002) Official standards for seed certification in Arkansas. Circular 15

Furukawa T, Maekawa M, Oki T, Suda I, Iida S, Shimada H, Takamure I, Kadowaki K-i (2007) The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. Plant J 49:91–102

Gealy DR, Tai TH, Sneller CH (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Sci 50:333–339

Gramene (2007) Cold Spring Harbor Laboratory and Cornell University, USA. http://www.gramene.org. Cited 26 December 2007

Ji H-S, Chu S-H, Jiang W, Cho Y-I, Hahn J-H, Eun M-Y, McCouch SR, Koh H-J (2006) Characterization and mapping of a shattering mutant in rice that corresponds to a block of domestication genes. Genetics 173:995–1005

Ling WH, Cheng QX, Ma J, Wang T (2001) Red and black rice decrease atherosclerotic plaque formation and increase antioxidant status in rabbits. J Nutr 131:1421–1426

Moldenhauer KAK, Lee FN, Bernhardt JL, Norman RJ, Slaton NA, Wilson CE, Anders MM, Cartwright RD, Blocker MM (2007) Registration of 'Wells' rice. Crop Sci 47:442–443

Slaton N, Moldenhauer K, Gibbons J, Blocker M, Wilson C, Dilday R, Robinson J, Koen B (2000) Grain characteristics of rice varieties. University of Arkansas, Cooperative Extension Service, Rice Information Bulletin, No. 146

Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283–294

Wilson C (2002) Arkansas rice acreage by variety survey. Proceedings of the 30th Rice Technical Working Group, New Orleans, LA