

Allelic Diversification at the *C* (*OsCI*) Locus of Wild and Cultivated Rice: Nucleotide Changes Associated With Phenotypes

Kumi Saitoh, Kazumitsu Onishi, Ichiho Mikami, Khin Thidar and Yoshio Sano¹

Plant Breeding Laboratory, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

Manuscript received June 4, 2004

Accepted for publication June 25, 2004

ABSTRACT

Divergent phenotypes are often detected in domesticated plants despite the existence of invariant phenotypes in their wild forms. One such example in rice is the occurrence of varying degrees of apiculus coloration due to anthocyanin pigmentation, which was previously reported to be caused by a series of alleles at the *C* locus. The present study reveals, on the basis of comparison of its maps, that the *C* gene appears to be the rice homolog (*OsCI*) of maize *CI*, which belongs to the group of R2R3-Myb factors. Two different types of deletions causing a frameshift were detected in the third exon, and both of the deleted nucleotides corresponded to the positions of putative base-contacting residues, suggesting that the Indica and Japonica types carry loss-of-function mutations with independent origins. In addition, replacement substitutions were frequently detected in *OsCI* of strains carrying the previously defined *C* alleles. Molecular population analysis revealed that 17 haplotypes were found in 39 wild and cultivated rices, and the haplotypes of most cultivated forms could be classified into one of three distinct groups, with few shared haplotypes among taxa, including Indica and Japonica types. The genealogy of the *OsCI* gene suggests that allelic diversification causing phenotypic change might have resulted from mutations in the coding region rather than from recombination between preexisting alleles. The McDonald and Kreitman test revealed that the changes in amino acids might be associated with selective forces acting on the lineage of group A whose haplotypes were carried by most Asian cultivated forms. The results regarding a significant implication for genetic diversity in landraces of rice are also discussed.

PHENOTYPIC diversity tends to increase during the domestication process as a general trend in crops (DARWIN 1859) although genetic diversity is expected to decrease in cultivated forms due to genetic bottlenecks during the domestication process (HARLAN 1975; TANKSLEY and MCCOUCH 1997; GOTTLIEB *et al.* 2002). Although there are abundant differences in the DNA sequence among individuals, it is not easy to delineate how this abundant genetic diversity actually contributes to phenotypic diversity (MACKAY 2001; GOTTLIEB *et al.* 2002; LAUTER and DOEBLEY 2002). The importance of regulatory genes in rapid changes of phenotypes has been proposed (DOEBLEY 1993; PURUGGANAN and WESSLER 1994). Thus, the source and maintenance of natural genetic variation reflected by discernible phenotypes is needed to understand the nature of crop gene pools, since crop evolution is not an event but a process at the infraspecific level (HARLAN 1975). Regarding divergent morphology in maize, the wild progenitor teosinte was found to maintain cryptic genetic variation, which can contribute to the phenotypic diversity in maize after reorganization exceeding the threshold necessary for phenotypic changes (LAUTER and DOEBLEY 2002). Another pos-

sibility is that mutations were accumulated during the domestication process, which suggests the presence of agronomically valuable genes in landraces as well as in wild relatives. Naturally occurring allelic diversity could be a source of genetic diversity (HARTL and CLARK 1997). Here, we tested this last possibility by investigating allelic diversity in a regulatory gene to ask how phenotypic diversity is related to nucleotide diversity in wild and cultivated rice and to what extent phenotypic diversity in domesticated forms is explained by mutations that emerged before the origin of rice.

One of the diverged phenotypes in cultivated rice is observed in the patterns of coloration due to anthocyanin pigmentation in spite of invariant pigmentation in the wild forms of rice. Flavonoid derivatives, including anthocyanins, are responsible not only for the pigmentation pattern but also for a wide range of biological functions such as protection against UV radiation, signal molecules in plant-microbe interactions, and plant defense responses (reviewed in DOONER *et al.* 1991; KOES *et al.* 1994). In rice, an extensive genetic study revealed that two genes, *C* and *A*, are basic to forming anthocyanin pigments, since these two genes are required for coloration in all tissues (TAKAHASHI 1957). In addition, a series of multiple alleles at these loci contribute varying degrees of apiculus coloration, resulting in a continuous variation among cultivated forms (TAKAHASHI 1957,

¹Corresponding author: Plant Breeding Laboratory, Department of Applied Biotechnology, Hokkaido University, Sapporo 060-8589, Japan.
E-mail: rysano@abs.agr.hokudai.ac.jp

1982) although little is known of the mechanisms at the molecular level. In maize, the biosynthesis of anthocyanin pigments requires complex interactions between genes with both structural and regulatory roles (DOONER *et al.* 1991). Structural genes such as *a1*, *a2*, *c2*, *chi*, *bz1*, and *bz2* encode the biosynthetic enzymes in the pathway. Expression of the structural genes is controlled at the transcriptional level by the products of regulatory genes belonging to the two gene families, *R/B* and *CI/Pl*. Transcriptional activation of the structural genes in any particular tissue of the plant requires a functional allele of the *R/B* family and a functional allele of the *CI/Pl* family. The *R/B* genes encode related proteins with homology to the basic-helix-loop-helix (bHLH) DNA-binding protein dimerization domain of Myc proteins (CHANDLER *et al.* 1989), while the *CI/Pl* genes encode related proteins with homology to the DNA-binding domain of Myb proteins (PAZ-ARES *et al.* 1987). The rice homologs of the maize genes *R*, *CI*, and *A1* have been characterized (HU *et al.* 1996; NAKAI *et al.* 1998; REDDY *et al.* 1998; SAKAMOTO *et al.* 2001), suggesting that similar genes are also responsible for anthocyanin pigmentation in rice as might be expected. We report here that, on the basis of the syntenic relationship between maize and rice, a candidate for *C* in rice is the rice homolog (*OsCI*) of maize *CI*. Maize *CI* belongs to the group of R2R3-Myb factors that activate the transcription of genes encoding enzymes involved in the biosynthesis of the anthocyanin pigment (MARTIN and PAZ-ARES 1997; REDDY *et al.* 1998). This prompted us to carry out a molecular population study on the *C* (= *OsCI*) locus to examine the association between changes in phenotypes and nucleotides.

MATERIALS AND METHODS

Plant materials: The rice strains used were obtained from the genetic stocks preserved at Hokkaido University, Sapporo, Japan, the National Institute of Genetics, Mishima, Japan, the International Rice Research Institute, Los Banos, Philippines, and the National Institute of Agricultural Sciences, Tsukuba, Japan. The species and strains analyzed were as follows: *Oryza sativa* Japonica (A5, A18, A38, A55, A56, A58, A83, A108, A136: 544 from Japan; T65 from Taiwan; 734 from China); *O. sativa* Japonica (226 from the Philippines; 647 from Indonesia); *O. sativa* Indica (PTB10, ARC6622 from India; C6172 from Laos; IR36 from the Philippines; Acc35618 from Indonesia; 706, N303, 719 from China; 108, 868, 160 from Taiwan; I32, I33, I45, I47 from India); *O. rufipogon* annual type (W107, W2002, W2012 from India; W1865 from Thailand); *O. rufipogon* intermediate type (W1819 from Bangladesh); *O. rufipogon* perennial type (W120 from India; W593 from Malaysia; W1294 from the Philippines; W1943, W1944 from China). Five strains of *O. glaberrima* (W025 from Guinea), *O. barthii* (W1468 from Cameroon; W1647 from Tanzania), and *O. glumaepatula* (W1187 from Brazil) were used as outgroups. Within *O. sativa*, three varietal types, Indica (continental), Japonica (Jv; tropical insular or tropical Japonica), and Japonica (Jp; temperate insular or temperate Japonica) were used in the literature (OKA 1953, 1988; CHANG 1976). In the present study, comparisons were carried out between the Indica type (In) and the

other two types (Jp + Jv) when needed, because the Japonica and Japonica types are genetically similar, as indicated by isoenzymes (GLASZMANN 1987) and restriction fragment length polymorphisms (RFLP; WANG and TANKSLEY 1989). The growth habits of *O. rufipogon* were divided into annual (Ra), intermediate (Rin), and perennial (Rp) forms according to SANO and MORISHIMA (1982).

The *C* locus is located on the short arm of chromosome 6 and the extensive crossing experiments have revealed that multiple alleles are present in cultivated strains (TAKAHASHI 1957, 1982). Strains with the 10 presumed alleles at the *C* locus were examined in this study. The strains and their allelic states are as follows: I33, C^{Bs} ; A58 and T65, C^B ; A38 and A83, C^{Bp} ; A55 and A108, C^{Bt} ; A5 and A56, C^{Br} ; I47, C^{Bd} ; I32, C^{Bk} ; I45, C^{Bc} ; A136, C^{Bm} ; IR36, 868, 108, and 414, C^+ . The C^+ allele was examined by introducing an alien allele into T65 by repeated backcrosses, as mentioned later. The 10 alleles are responsible for varying degrees of apiculus pigmentation in the presence of the *A* and *P* genes. For example, their phenotypes (apiculus color) are blackish purple for C^B , blackish red purple for C^{Bp} , pansy purple for C^{Bt} , tryan rose for C^{Br} , and rose red for C^{Bm} , while C^+ shows no apiculus pigmentation. C^{Bs} causes a deeper pigmentation than C^B , while the three alleles C^{Bd} , C^{Bk} , and C^{Bc} show an intermediate pigmentation between C^{Br} and C^{Bm} . The degree of apiculus pigmentation due to the alleles was estimated to be in the order $C^{Bs} > C^B > C^{Bp} > C^{Bt} > C^{Br} \geq C^{Bd} \geq C^{Bk} \geq C^{Bc} \geq C^{Bm} \geq C^+$ (TAKAHASHI 1957, 1982) although the differences were not always distinct among C^{Bd} , C^{Bk} , C^{Bc} , C^{Bm} , and C^+ . Thus, the multiple alleles adjust the color tone from blackish purple to colorless. Most of the multiple alleles were found as rare mutants not in a collection but in landraces from Hokkaido island (the northernmost area of rice cultivation in Japan). We determined five allelic states of T65, IR36, 868, 108, and 414 in the present study.

As mentioned above, landraces of rice show varying degrees of coloration, in contrast to the wild progenitor, which tends to be uniformly colored. Although phenotypic diversity in anthocyanin pigmentation may be caused by numerous enzymatic and regulatory loci, as seen in maize, a series of alleles found at the *C* and *A* loci play a major role in the continuous variation in anthocyanin pigments observed only in cultivated forms of rice. TAKAHASHI (1957, 1982) conducted his extensive investigation of apiculus pigmentation in Sapporo (43° N), where cool temperatures enhanced anthocyanin pigmentation. His investigation of alleles was largely confined to early heading strains. In addition, it is known that the Indica type frequently carries a nonfunctional allele of *C* (C^+) that was not detected in wild strains except in those of hybrid swarms (OKA 1989). In the present study, genetically defined strains were included in our random sample of rice alleles to survey germplasms over a broad geographic range.

DNA sequencing: Genomic DNA was isolated from 2-month-old plants by the CTAB method according to MURRAY and THOMPSON (1980). The coding region of the *OsCI* gene was determined by direct sequencing of polymerase chain reaction (PCR) products. The region upstream of the first exon was not amplified in the present study. The *OsCI* gene was amplified using specific primers designed according to the published *OsCI* sequence (REDDY *et al.* 1998). After PCR amplification using primers 1 (5'-ATCGCTCAGTCTCACACCGCA CAG-3') and 2 (5'-CGTACGGACGACGAACCTAATGTAC-3'), *OsCI* was amplified in two pieces using the PCR primers 3 (5'-GAGGGA GAATGGGGAGGAGAGC-3') and 7 (5'-ATGG CCGTCTCCTAATTCCTGTC-3'), as well as the primers 4 (5'-TAATGTGATCTGTATGGATGCTG-3') and 2 (Figure 1B). PCR conditions were as follows: 1 min at 94°, for initial denaturation, followed by 30 cycles of denaturation at 94° for

30 sec, primer annealing at 56° for 30 sec, extension at 72° for 90 sec, and termination by 7 min at 72°. Secondary PCR amplification was carried out for 1 min at 94°, followed by 30 cycles of 30 sec at 94°, 30 sec at 53°, 30 sec at 72°, and termination by 7 min at 72°. Sequencing primers were located ~300 bp apart. Each accession was sequenced in both directions using an ABI 377 automatic sequencer (Applied Biosystem, Foster City, CA) using a Big Dye terminator cycle sequencing kit (Applied Biosystem). The sequences outside of the coding region were not determined in the present study. The DNA sequences are available from DDBJ with the accession nos. AB111867–AB111885.

Mapping of the *OsCl* gene: The *C* gene for apiculus pigmentation is located on the short arm of chromosome 6 of rice. To examine the location of the *OsCl* gene, two near isogenic lines (NILs), T65*wx* and T65 C^+ , were used. T65*wx* carries *wx* from Kinoshitamochi and a functional allele of C^B , while T65 C^+ carries C^+ from 868; these lines are derivatives of BC12 and BC8, respectively (DUNG *et al.* 1998). Regarding the synteny map between maize and rice, it was reported that a part of rice chromosome 6 is homologous to a part of chromosome 9 of maize, including the maize *c1* (AHN and TANKSLEY 1993). The *C* gene of rice was reported to be located between RZ588 and G200 (or RZ144) on rice chromosome 6 and to show a tight linkage with RM253 (XIONG *et al.* 1999; LORIEUX *et al.* 2000). To map the *OsCl* gene, 164 F₂ plants of T65*wx* (C^B) × IR36 (C^+) were grown and used for a linkage study. Genomic DNAs from individuals were extracted as described above and RFLP analysis was performed. After digestion with *Bam*HI, *Eco*RV, and *Kpn*LI, the DNAs were subjected to electrophoresis on 1% agarose gels and transferred to BIODYNE B membranes (Pall BioSupport, East Hills, NY). DNA markers were gifts from S. McCouch, Cornell University, and from T. Sasaki, Rice Genome Research Program, National Institute of Agrobiological Resources, Tsukuba, Japan. Southern blotting was performed by using ECL direct nucleic acid labeling and detection systems (Amersham Bioscience, Piscataway, NJ). To detect a deletion (10 bp) in the third exon of *OsCl*, PCR products using specific primers 5 (5'-GATCGATCGTGTATA TATGTTGTCAGGT-3') and 6 (5'-GTTGCTGTGTCGGTGT CGGCG-3') were separated on a 5% acrylamide gel. Recombination values were calculated by the maximum-likelihood method (ALLARD 1956) and converted to centimorgans using the Kosambi function (KOSAMBI 1944).

Phylogenetic analysis: The sequence alignment was done by using the CLUSTAL W computer program (THOMPSON *et al.* 1994) with additional minor modifications by visual inspection. The first nucleotide of the translational unit was assigned as coordinate position 1. Program package DnaSP, version 3.14 (ROZAS and ROZAS 1999), was used to analyze intra- and interspecific variation via the estimation of nucleotide diversity (π ; NEI and LI 1979). Sliding-window analysis was conducted to examine changes in the level of variation along the *OsCl* gene. For haplotype diversity, the Shannon information measure ($H = -\sum p_i \ln p_i$, where p_i is frequency of haplotypes) was used.

Phylogenetic trees were constructed using PAUP*, version 4.0 (SWOFFORD 1998). Both the neighbor-joining (NJ) and maximum-parsimony (MP) methods were conducted, and the topologies obtained thereby were compared. The NJ (SAITOU and NEI 1987) method was conducted with KIMURA's (1980) two-parameter distances. The full heuristic maximum-parsimony analyses were also carried out with tree bisection-reconnection branch swapping and random order of taxon addition. To test the robustness of the tree topologies, 1000 bootstrap replicates were performed. Both methods gave the same topology in the present study. To root the constructed trees, reproductively isolated taxa (*O. glaberrima*, *O. barthii*, and *O. glumae-*

patula) from the *O. sativa*-*O. rufipogon* complex were used as outgroups. Recombination produces networks of sequences rather than strictly bifurcating evolutionary trees. A phylogenetic network for haplotypes of the *OsCl* locus was constructed by the procedures of BANDELT (1994) and SAITOU and YAMAMOTO (1997). The network can be considered as a generalization of the discordancy diagram, which appears when two nucleotide positions show incongruent configuration patterns. The minimum number of recombination events (R_M ; HUDSON and KAPLAN 1985) was calculated. The R_M gives the minimum number of recombination events in the history of a sample and it is determined on the four-gamete test, which infers a recombination event between pairs of diallelic loci at which all four possible gametic types are present (HUDSON and KAPLAN 1985). The ratio (K_a/K_s) of the number of replacement (or nonsynonymous) substitutions per replacement site (K_a) to the number of silent substitutions per silent site (K_s) was calculated. In addition, to examine the neutral hypothesis, the tests of TAJIMA (1989), FU and LI (1993), and McDONALD and KREITMAN (1991) were performed using DnaSP, version 3.14. The McDonald and Kreitman test is based on a comparison of silent and replacement variation within and between species. Under neutrality, the ratio of replacement to silent fixed differences between species should be the same as the ratio of replacement to silent polymorphisms within species.

RESULTS

***OsCl* as a candidate for the *C* gene:** The rice homolog (*OsCl*) of the maize *c1* was identified as a possible candidate of the *C* gene in rice on the basis of a comparative mapping between maize and rice (Figure 1A). The genomic sequences of the functional and dysfunctional alleles were compared to examine this possibility. T65 had a colored apiculus due to C^B while the latter two strains (IR36 and 868) had a colorless apiculus due to C^+ . The *OsCl* gene of T65 consisted of three exons and two introns and encoded a 272-amino-acid protein containing a DNA-binding Myb domain at the N terminus as shown in Figure 1B. The positions of the introns were conserved in rice and maize.

The sequence of the *OsCl* gene was compared among T65 (C^B), IR36 (C^+), and 868 (C^+). The results revealed that the two colorless lines (IR36 and 868) had the same sequence, which differed from that of T65 by a 10-bp deletion (at position 795–804 from the translation initiation site) and a replacement substitution (at position 918) both of which were located in the R3 repeat within the third exon (Table 1). The deletion was expected to cause a frameshift, and the deleted residues corresponded to the positions of putative base-contacting residues in the R3 (MARTIN and PAZ-ARES 1997), suggesting that the *OsCl* allele of IR36 and 868 was a loss-of-function mutation. The replacement substitution (Pro to Gln) detected in the two strains was considered to have no effect on the function since it was the same as that reported in Purpleputtu, which has a colored apiculus (REDDY *et al.* 1998).

To examine the location of *OsCl*, a NIL (T65 C^+) was compared with T65 carrying C^B regarding the 10-bp deletion in the third exon of the *OsCl* gene. The

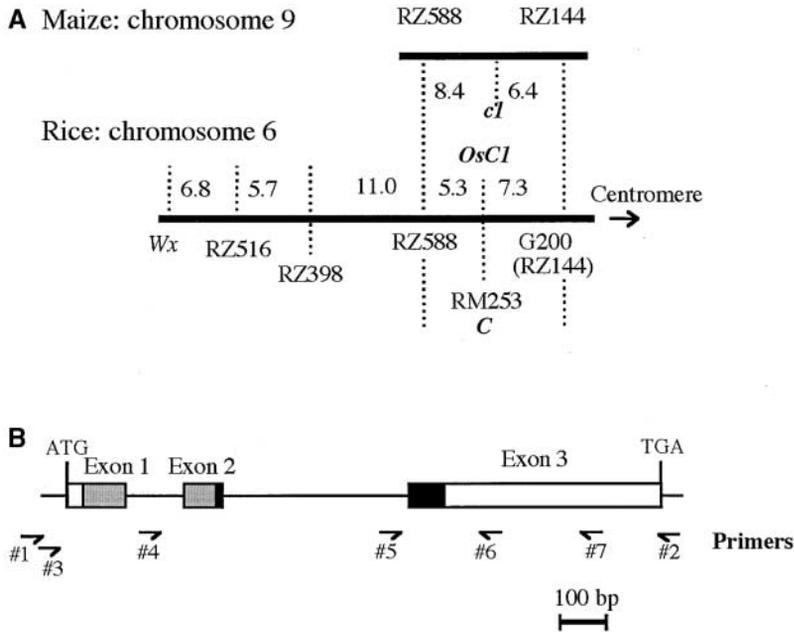


FIGURE 1.—The location of a rice homolog (*OsCI*) of maize *CI* on the short arm of chromosome 6. Microlinearity between maize and rice (A) and the structure of the *OsCI* gene (B) are shown. (B) R2 and R3 in Myb protein are shown by ■ and ■, respectively. The primers used for amplification are indicated by small half-arrows with numerals. Linkage data in maize are based on AHN and TANKSLEY (1993).

deletion was readily detected from PCR products produced using primers 5 and 6 (Figure 1) and separated on a 5% acrylamide gel (Figure 2), which shows that T65 C^+ had the 10-bp deletion and indicates that *OsCI* as well as C^+ is located on the short arm of chromosome 6. To map *OsCI* precisely, Southern blotting was carried out in 164 F_2 plants of the T65 (C^B) \times IR36 (C^+) cross using molecular markers on chromosome 6 as probes. The linkage between *C* and *OsCI* was not examined in the F_2 population since a monogenic segregation was not observed due to the presence of other interacting genes. *OsCI* was linked to RZ588 and G200 markers with map distances of 5.3 and 7.3 cM, respectively (Figure 1), suggesting it as a candidate for *C*.

Naturally occurring nucleotide variation in the *OsCI* gene: A total of 29 strains of *O. sativa* and 10 of *O. rufipogon* were sequenced to examine the naturally occurring variation within the *O. sativa-O. rufipogon* complex (Table 1). In these strains, a total of seven indels and 44 polymorphic sites were detected, of which 27 sites were phylogenetically informative and 17 sites were singletons. To investigate the distribution of the polymorphic sites at the *OsCI* locus, a sliding-window analysis was performed within the *O. sativa-O. rufipogon* complex (Figure 3). This analysis of the 39 strains revealed that polymorphic sites were frequently observed in the second intron in the *O. rufipogon* and Indica strains, in contrast to frequent polymorphic sites in the third exon of Japonica and Javanica types. A total of 21 mutation sites were found in the coding region, and 14 were replacement substitutions while 7 were synonymous. Of the seven indels, five were present in the third exon and three resulted in a frameshift. The 43 sequences, including outgroups, were divided into 19 distinct haplotypes (Table 1). A total of 17 haplotypes were found

in the *O. sativa-O. rufipogon* complex; however, only haplotype SA9 was found in both *O. sativa* and *O. rufipogon*, suggesting that the haplotypes of *OsCI* differ markedly between *O. rufipogon* and *O. sativa*. Each strain of *O. rufipogon* had a unique haplotype except for W2012, reflecting a higher value of haplotype diversity in *O. rufipogon* ($H = 2.164$) than in *O. sativa* ($H = 1.980$). We used the NJ and MP methods to determine the *OsCI* gene phylogeny on the basis of the 19 haplotypes found in the 43 strains examined. Both the methods produced the same topology. Phylogenetic analysis divided the 17 haplotypes in the *O. sativa-O. rufipogon* complex into three different groups (Figure 4). Groups A, B, and C included 10, 4, and 3 haplotypes, respectively, showing that the primary gene pool consisted of *OsCI* genes from different lineages. Among 27 strains of *O. sativa*, 23 carried the 8 haplotypes in group A while only 4 strains of the Indica type carried haplotype SA7 in group B. The two haplotypes, RU1 and RU2, which were carried by *O. rufipogon* from China, were included in group A.

Association of nucleotide mutations with phenotypes: The presence of the candidate gene in the *C* locus gave us an opportunity to examine the association of the nucleotide mutations with the phenotypic changes in anthocyanin pigmentation and to gain insight into the molecular changes that occurred during the process of rice domestication. The *OsCI* alleles were compared in 14 cultivated strains of *O. sativa*, including representatives of the 10 alleles examined by TAKAHASHI (1982). Among the 17 strains (including T65, IR36, and 868), eight haplotypes were detected and the deduced amino acid sequences were compared with the corresponding phenotypic changes (Table 2). The two additional C^+ strains (108 and 414, Indica type) showed the same sequence as that of IR36 and 868, exhibiting the 10-bp

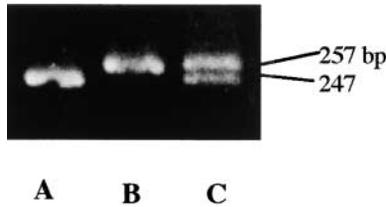


FIGURE 2.—Detection of a 10-bp deletion in the third exon by PCR amplification in IR36 (Indica) carrying C^+ . The primers (5 and 6) used are presented in Figure 1. (A) C^+C^+ (IR36); (B) $C^B C^B$ (T65); (C) the heterozygote ($C^B C^+$).

were associated in part with the phenotypic changes observed in rice with different C alleles.

Origin of multiple alleles in *OsCI*: The number of replacement substitutions detected was five among 29 sequences of *O. sativa* and six among 10 sequences of *O. rufipogon* while the number of silent mutations was much lower in *O. sativa* (Table 3). The nucleotide diversity (π) in silent positions was lower in *O. sativa* than in *O. rufipogon*, suggesting a population bottleneck due to domestication or the recent origin of the lineage. Neither Tajima's D nor Fu and Li's D^* statistic differs significantly from zero for either species examined, although the K_a/K_s ratio was higher in *O. sativa* (1.13) than in *O. rufipogon* (0.26).

A phylogenetic network for haplotypes of the *OsCI* gene sequences revealed networked evolution, indicative of recombination (Figure 5). An interconnected network rather than strictly bifurcating evolutionary trees was detected between the haplotype groups; however, no such network was found within group A and group B, into which most of alleles in *O. sativa* were classified. The minimum number of R_M was estimated to be four in the 43 sequences and all the positions were within the second intron, as shown in Table 1, although R_M was estimated to be 2 (between positions

492–516 and 516–528) in the 39 sequences of *O. rufipogon* and *O. sativa* (Table 1). The results showed that the allelic diversity in cultivated forms mostly resulted from changes in nucleotides rather than from recombination events.

Figure 5 indicated the existence of mutually incompatible sites (positions 571, 516, 428, 844, 481, 491, 528, and 1114), all of which were found between the haplotype groups. Further, changes of nucleotides frequently occurred in the coding region within group A (10/12). The McDonald and Kreitman test revealed that the ratio of replacement to silent fixed differences between group A and *O. glumaepatula* significantly differed from the ratio of replacement to silent polymorphisms within group A, suggesting the signature of positive selection in the *OsCI* locus of group A (Table 4).

DISCUSSION

Importance of regulatory loci in plant domestication: Recent interest in crop evolution studies has been focused on identification of genes that control phenotypes of biological and agronomic importance. On the basis of an increase in the nonsynonymous relative to the synonymous nucleotide substitution rate, it was proposed that morphological evolution proceeds via diversification in regulatory loci and that phenotypic evolution may correlate better with regulatory than with structural gene divergence (DOEBLEY 1993; PURUGGANAN and WESSLER 1994). This prediction suggests that a few mutations in regulatory loci have the potential to affect a number of interacting genes, causing distinct changes in phenotype. However, the molecular population genetics of developmental loci that control morphological traits are poorly understood, particularly within crop species groups that exhibit marked morphological divergence as a result of domestication. In maize, the *teosinte branched 1* (*tb1*) gene belongs to the TCP gene family,

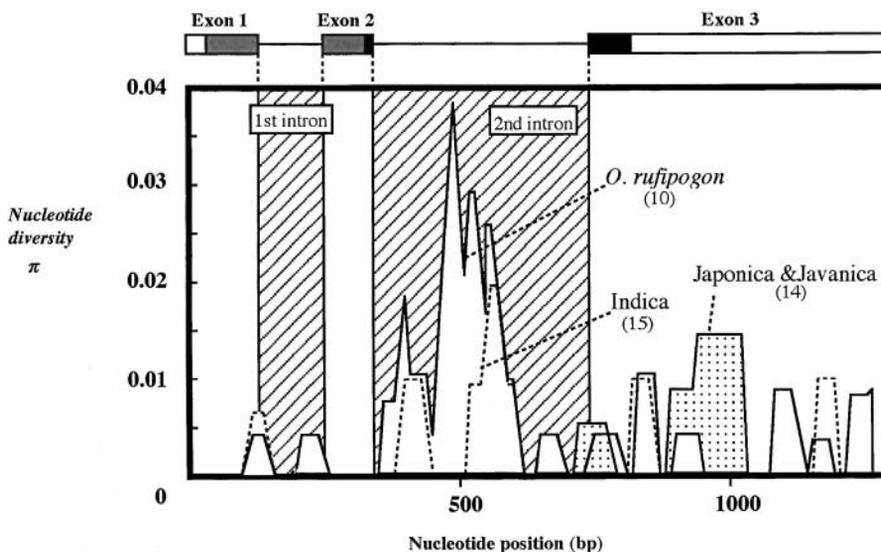


FIGURE 3.—Sliding-window analysis of nucleotide diversity for the *OsCI*. R2 and R3 in Myb are shown by ■ and ■.

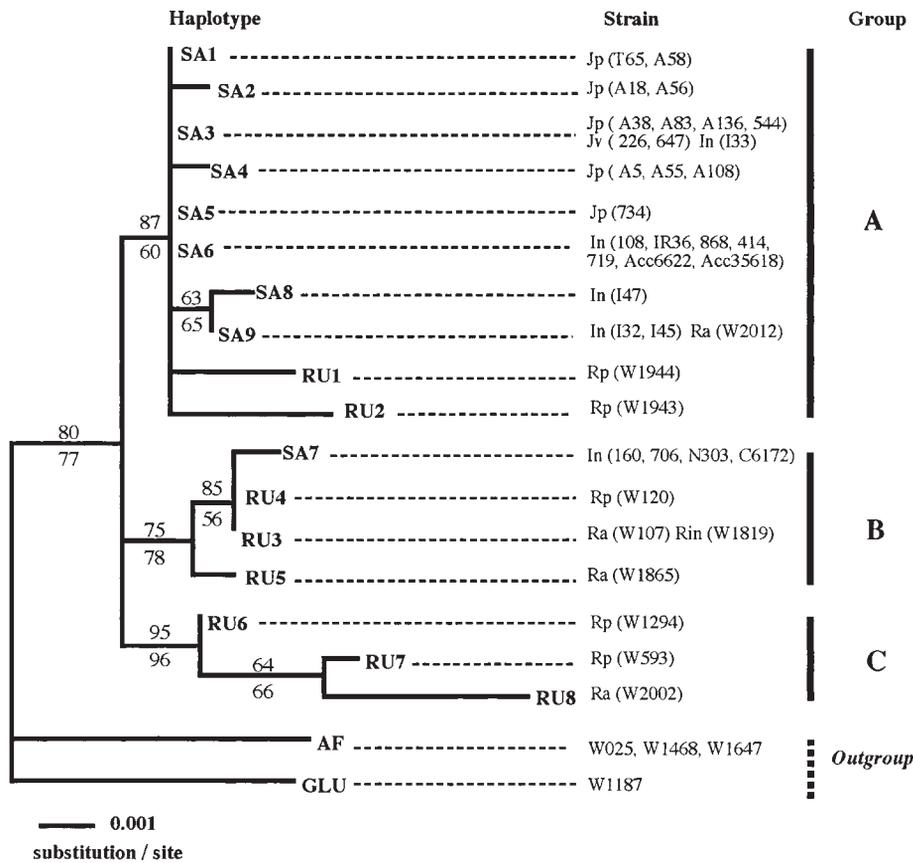


FIGURE 4.—NJ reconstruction of the genealogical relationships among *OsCI* haplotypes of 43 strains in wild and cultivated rice. *O. glaberrima*, *O. barthii* (AF), and *O. glumaepatula* (GLU) were used as outgroups. The MP method gave the same topology. Bootstrap values for nodes supported in >50% of 1000 bootstrap replicates are shown above (for NJ) and below (for MP) the branches. Jp, Jv, and In represent Japonica, Javanica and Indica types of *O. sativa*, respectively. Ra, Rin, and Rp represent annual, intermediate, and perennial types of *O. rufipogon*, respectively. AF includes *O. glaberrima* (W025) and *O. barthii* (W1468, W1647).

whose members encode putative bHLH DNA-binding proteins. *tb1* affects plant architecture that distinguishes maize from its wild ancestor teosinte, showing evidence of a selective sweep during maize domestication (WANG *et al.* 1999). Although a K_a/K_s ratio >1.0 is considered as an indication for positive selection, the K_a/K_s ratio provided no evidence for positive selection in *tb1* (LUKENS and DOEBLEY 2001). On the other hand, HANSON *et al.* (1996) analyzed the evolution of anthocyanin-pigmented kernels in maize from colorless kernels of its wild progenitor, teosinte. Unexpectedly, molecular evolutionary analysis of *c1* showed an exceptionally low

level of nucleotide polymorphism, suggesting that the evolution of the purple kernels resulted from changes in *cis* regulatory elements at regulatory loci and not from changes in either regulatory protein function or the enzymatic loci. In tomato, fruit weight and size distinctly differ between the domesticated and wild tomato species. One of the QTL (*fw2.2*) with the largest effect was cloned and postulated to act by causing changes in the timing of gene expression, suggesting the importance of regulatory functions in crop evolution (CONG *et al.* 2002). These results suggest the importance of regulatory functions in crop evolution; however, it re-

TABLE 2
Nucleotide changes among the eight haplotypes in the *OsCI* gene observed within *O. sativa*

Haplotype	Allele (strain)	Position	Mutation ^a	Effect on protein
SA1	C^B (T65, A58)	918	C → A	Replacement (Pro → Gln) in third exon
SA2	C^{Br} (A56)	744	C → T	Replacement (Ser → Phe) in R3
		918	C → A	Replacement (Pro → Gln) in third exon
SA3	C^{Bs} (I33), C^{Bp} (A38, A83), C^{Bm} (A136)	—	—	—
SA4	C^{Bt} (A55, A108), C^{Br} (A5)	968–970	ACG → GCA	Replacement (Thr → Ala) in third exon
		973	34-bp deletion	Frameshift in third exon
SA5	C^+ (734)	786	2-bp deletion	Frameshift in R3
SA6	C^+ (IR36, 868, 108, PTB10)	796	10-bp deletion	Frameshift in R3
SA8	C^{Bd} (I47)	122	C → G	Replacement (Pro → Arg) in R2
SA9	C^{Bk} (I32), C^{Bc} (I45)	122	C → G	Replacement (Pro → Arg) in R2

^a Compared with the SA3 sequence, which is the same as that of Purpleputtu (REDDY *et al.* 1998).

TABLE 3
Level of DNA variation in the *OsCI* region of *O. sativa* and *O. rufipogon*

Species	No. of strains	No. of mutations ^a			Nucleotide diversity per nucleotide (π) ^b							
		Total	Synonymous	Replacement	Silent	Total	Synonymous	Replacement	Silent	K_a/K_s ^c	Tajima's <i>D</i>	Fu and Li's <i>D</i> *
<i>O. sativa</i>	29	10	1	5	5	0.0016 (0.0010)	0.0010 (0.0007)	0.0017 (0.0011)	0.0015 (0.0009)	1.13	-0.6307 ^{NS}	0.9431 ^{NS}
<i>O. rufipogon</i>	10	21	3	6	15	0.0052 (0.0032)	0.0047 (0.0031)	0.0021 (0.0019)	0.0080 (0.0043)	0.26	-0.4857 ^{NS}	-0.6092 ^{NS}

NS, nonsignificance.

^a Silent substitutions include synonymous substitutions and changes in the noncoding positions. Indels were ignored for computation.

^b Numbers of synonymous, replacement, and silent positions are 182, 6, 578, 4, and 659, 6, respectively. Standard deviations are shown in parentheses.

^c K_a and K_s are nucleotide diversities for replacement and silent mutations per nucleotide, respectively.

mains unclear whether evolutionary changes in regulatory proteins themselves can cause the diversification of phenotypes.

Phenotypic alterations caused by changes of nucleotides in the *OsCI*: The present results strongly suggest that the *C* locus in rice is the homolog (*OsCI*) of the maize *CI*, which belongs to the group of R2R3-Myb factors. In the case of a multigene family like Myb-related genes, one problem in genealogical analysis results from the possibility that PCR with conserved primers might amplify nonorthologous gene family members. Linkage analysis and the expected difference among NILs suggested that the orthologous members were probably compared in the present study, although strict orthology cannot be assured. Comparisons of nucleotides in *OsCI* revealed that the series of alleles at *OsCI* is partly explained by nucleotide changes in the coding region (Table 2), which strongly supports the idea proposed by TAKAHASHI (1957) that a series of *C* alleles contributes to produce the varying degree of apiculus pigmentation observed in cultivated forms in spite of uniform pigmentation in the wild progenitor. Of the presumed 10 alleles, C^B , C^{Br} , and C^+ had unique mutations causing changes in the gene product in comparison with the sequence of the functional allele in Purpleputtu. Two different deletions in R3 of C^+ showed that mutations with loss of function occurred independently in Indica and Japonica types. Two alleles, C^{Bd} and C^{Bk} , which cause a slightly pigmented apiculus, had the same replacement substitution at position 122 although they were distinguished by a silent mutation at position 432. It is not certain whether the slight difference in pigmentation is caused by an additional mutation(s) out of the sequenced region or by other factors such as temperature and the genetic background, since only the coding regions were compared in the present study. Furthermore, three alleles, C^{Bs} , C^{Bp} , and C^{Bm} , also had haplotype SA3, which gives the same deduced amino acid sequence as that in Purpleputtu. As their phenotypes are distinct, it is considered that an additional change(s) outside of the sequenced region might explain their phenotypic difference. Thus, the present results revealed that altered proteins in *OsCI* play a role in determining the diversified phenotypes observed in cultivated forms, although an additional role of the regulatory regions cannot be excluded.

Haplotype diversity in the primary gene pool: Haplotype diversity is affected by introgression among taxa within the primary gene pool, and introgression has been reported to take a place among various rice taxa when they coexist within populations even across isolating barriers (CHU and OKA 1970). Therefore, wild rice often absorbs genes from cultivars through hybridization since the wild progenitor tends to be cross-pollinated with its surrounding cultivars (OKA 1988). Asian wild and cultivated rice are widely distributed in tropical and subtropical areas while Japonica extends its distribu-

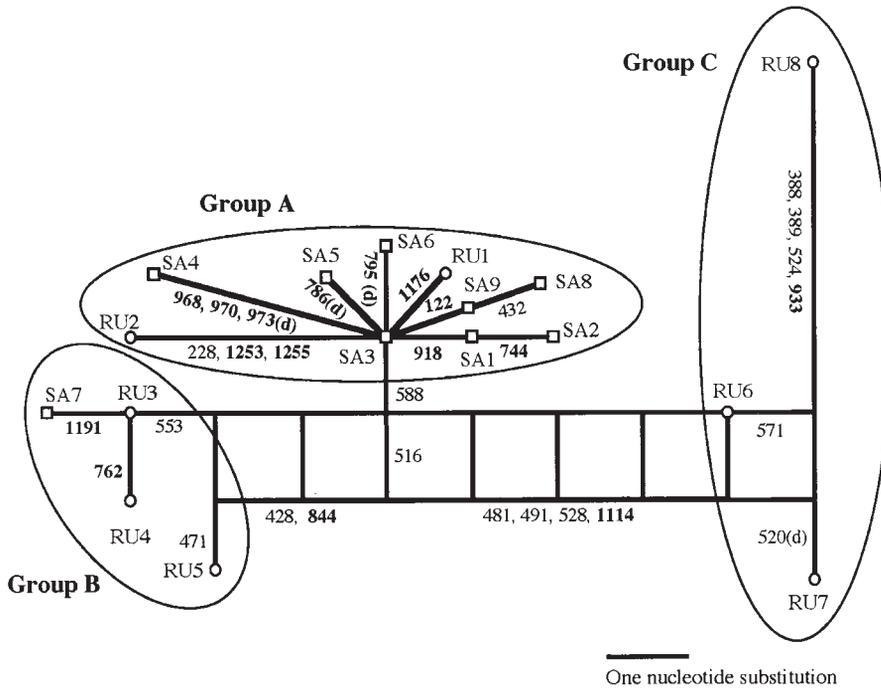


FIGURE 5.—The phylogenetic network for alleles of the *OsCl* locus found in wild and cultivated rice species based on the procedures of SAITOU and YAMAMOTO (1997). Haplotypes observed in wild (○) and cultivated (□) rice strains correspond to those in Figure 4. Numbers on branches denote the nucleotide positions in which substitutions or deletions (d) were found to occur, and those in boldface type denote changes in the coding region.

tion to temperate areas with overlapping. The distribution of Japonica and Indica types is also affected by altitude (TING 1949; OKA 1988). Recently, it was reported that glutinous landraces carry intragenic recombinants at the *wx* gene (OLSEN and PURUGGANAN 2002), showing evidence for introgression within *O. sativa* since glutinous endosperms are restricted to cultivated rice. The haplotypes found in cultivated forms were mostly included in group A (25/29) together with those of *O. rufipogon* from China (RU1 and RU2). Shared haplotypes were found between Japonica and Indica types (SA7) and between Indica type and *O. rufipogon* (SA9), which is indicative of introgression between taxa or of retention of an ancient haplotype. A higher level of haplotype diversity in *O. rufipogon* than in *O. sativa* is expected from the population bottleneck due to domestication as observed in the present study. A significant loss of diversity in cultivated forms is also caused not only by selective forces but also by the “hitchhiking” effect (MAYNARD SMITH 1998).

Recombination can be a dominant force in shaping genomes and associated phenotypes (HARTL and CLARK 1997; POSADA *et al.* 2002). The non-treelike network suggested that intragenic recombination in the *OsCl* gene might contribute to the haplotype diversity in rice. The estimated number of R_M supported the notion that intragenic recombination contributed to the three distinct groups of haplotypes observed in the Asian wild and cultivated rice (Figure 5), although this does not imply that recombination rarely occurred after the origin of rice since it becomes hard to detect recombination between sequences with few polymorphic sites. In contrast to the presence of recombination events in *OsCl*, no re-

combination events were detected in maize *cl* sequences (HANSON *et al.* 1996) in spite of the fact that maize is an outbreeder.

The genealogy of the *OsCl* gene suggests that the allelic diversification observed in domesticated forms resulted from replacement substitutions or mutations rather than from recombination between preexisting alleles. Furthermore, within *O. sativa*, naturally occurring alleles of *OsCl* possess an excess of intraspecific replacement substitutions, and this variation seems to be associated with altered patterns in anthocyanin pigmentation. Regarding nucleotide substitutions within *O. sativa*, five of six found were replacement mutations. In addition, of the five colorless cultivated strains, two different types of deletions causing frameshifts were detected, but none were found in the wild forms. Regarding the distribution

TABLE 4

McDonald-Kreitman test for the *OsCl* locus between the haplotype groups and *O. glumaepatula*

	Group A		Group B	
	Fixed ^b	Polymorphic ^c	Fixed	Polymorphic
Substitutions ^a				
Silent	10	4	10	3
Replacement	1	6	1	2

Group A: $G = 6.571$, $P = 0.0104$; group B: $G = 2.010$, $P = 0.1562$.

^a Silent and replacement show silent and replacement substitutions, respectively. Numbers of silent and replacement positions are 659.6 and 578.4, respectively.

^b Fixed differences in comparison with *O. glumaepatula*.

^c Polymorphic sites within group A or B.

of C^+ , it was reported that Indica type carried C^+ more frequently (70.3% in 64 strains) than did Japonica type (34.8% in 66) or the wild progenitor (4.1% in 49), and furthermore that two wild rice strains with C^+ probably resulted from introgression from cultivars since they were obtained from hybrid swarms (OKA 1989). A simple scenario for rice history is that C^+ arose as a mutation and became widespread in cultivated forms during rice domestication. This is not the case, however, since the present results showed that C^+ seems to have originated independently in Indica and Japonica types.

An implication of positive selection: Replacement substitutions and deletions causing frameshifts in the coding region of *OsCI* were suggested as possible candidates for mutations leading to affected anthocyanin pigmentation in rice. In *Arabidopsis*, the *GLABROUS1* (*GLI*) gene belongs to the large family of Myb transcription factors and is known to play a role in trichome initiation. Replacement substitutions or frameshifts downstream of the conserved domains R2 and R3 in *GLI* have been reported to cause leaky mutants (HAUSER *et al.* 2001). Within *O. sativa*, especially Japonica type, the small number of silent substitutions in the second intron strongly suggests their recent emergence, since the rate of silent mutation correlates with the divergence time. This suggests that changes in amino acids might be associated with selective forces that acted on the lineage of group A (Table 4). Since the *CI* protein of maize activates a subset of the genes in anthocyanin biosynthesis (CONE *et al.* 1986), it is possible that *OsCI* evolution could involve changes in protein function such that the Myb protein would interact with new protein partners or recognize new target sequences. It might be interesting to inquire if distinct haplotypes in Japonica type are related to unfavorable growth conditions in the northern area of rice cultivation. In rice, the *A* locus also contributes to continuous pigmentation via a series of alleles (TAKAHASHI 1957) and the locus has been suggested to be the rice homolog of the maize *al* (CHEN and BENNETZEN 1996; NAKAI *et al.* 1998), which encodes the enzyme dihydroflavonol 4-reductase (*DFR*). The genealogy of the *DFR* gene is under investigation to compare nucleotide divergence in structural *vs.* regulatory genes during rice domestication.

As described above, the present results strongly support the previous reports by TAKAHASHI (1957, 1982) showing that allelic diversity at *C* plays a significant role in the patterns of anthocyanin pigmentation in rice. Such allelic diversity may be related to agricultural merits; otherwise, positive selection or frequent replacement substitution would be unexpected. Therefore, the present results have the significant implication that landraces might maintain more agronomically valuable genes accumulated since the origin of rice than previously believed, even if a reduction in nucleotide diversity was caused by the population bottleneck due to domestication.

We thank S. R. McCouch and T. Sasaki for molecular markers and N. Saitou for the procedures of the phylogenetical network. We also thank Y. Kishima, H. Nagano, and A. Takahashi for their comments and assistance.

LITERATURE CITED

- AHN, S., and S. D. TANKSLEY, 1993 Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. USA* **90**: 7980–7984.
- ALLARD, R. W., 1956 Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* **24**: 235–278.
- BANDELT, H. J., 1994 Phylogenetic networks. *Verh. Naturewiss. Ver. Hamburg* **34**: 51–71.
- CHANDLER, V. L., J. P. RADICELLA, T. P. ROBBINS, J. CHEN and D. TURKS, 1989 Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of *B* utilizing *R* genomic sequences. *Plant Cell* **1**: 1175–1183.
- CHANG, T. T., 1976 The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica* **25**: 435–441.
- CHEN, M., and J. L. BENNETZEN, 1996 Sequence composition and organization in the *Sh2/Al*-homologous region of rice. *Plant Mol. Biol.* **32**: 999–1001.
- CHU, Y. E., and H. I. OKA, 1970 Introgression across isolating barriers in wild and cultivated rice species *Oryza* species. *Evolution* **24**: 344–355.
- CONE, K. C., F. A. BURR and B. BURR, 1986 Molecular analysis of the maize anthocyanin regulatory locus *CI*. *Proc. Natl. Acad. Sci. USA* **83**: 9631–9635.
- CONG, B., J. LIU and S. D. TANKSLEY, 2002 Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc. Natl. Acad. Sci. USA* **99**: 13606–13611.
- DARWIN, C., 1859 *The Origin of Species by Means of Natural Selection*. Murray, London.
- DOEBLEY, J., 1993 Genetics, development and plant evolution. *Curr. Opin. Gen. Dev.* **3**: 865–872.
- DOONER, H. K., T. P. ROBBINS and R. A. JORGENSEN, 1991 Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* **25**: 173–199.
- DUNG, L. V., T. INUKAI and Y. SANO, 1998 Dissection of a major QTL for photoperiod sensitivity in rice: its association with a gene expressed in an age-dependent manner. *Theor. Appl. Genet.* **97**: 714–720.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GLASZMANN, J. C., 1987 Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **74**: 21–30.
- GOTTLIEB, T. M., M. J. WADE and S. L. RUTHERFORD, 2002 Potential genetic variance and the domestication of maize. *BioEssays* **24**: 685–689.
- HANSON, M. A., B. S. GAUT, A. O. STEC, S. I. FUERSTENBERG, M. M. GOODMAN *et al.*, 1996 Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* **143**: 1395–1407.
- HARLAN, J. R., 1975 *Crops and Man*. American Society of Agronomy, Madison, WI.
- HARTL, D. L., and A. G. CLARK, 1997 *Principles of Population Genetics*, Ed. 3. Sinauer Associates, Sunderland, MA.
- HAUSER, M.-T., B. HARR and C. SCHLOTTERER, 2001 Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene *GLABROUS1*. *Mol. Biol. Evol.* **18**: 1754–1763.
- HU, J., B. ANDERSON and S. R. WESSLER, 1996 Isolation and characterization of rice *R* genes: evidence for distinct evolutionary paths in rice and maize. *Genetics* **142**: 1021–1031.
- HUDSON, R. R., and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–164.
- KIMURA, M., 1980 Simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- KOES, R. E., F. QUATTROCCHIO and J. N. M. MOL, 1994 The flavonoid

- biosynthetic pathway in plants: function and evolution. *BioEssays* **16**: 123–132.
- KOSAMBI, D. D., 1944 The estimation of map distance from recombination values. *Ann. Eug.* **12**: 172–175.
- LAUTER, N., and J. DOEBLEY, 2002 Genetic variation for phenotypically invariant traits detected in teosinte: implications for the evolution of novel forms. *Genetics* **160**: 333–342.
- LORIEUX, M., M.-N. NDJIONDJIP and A. GESQUIERE, 2000 A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite-based genetic linkage map. *Theor. Appl. Genet.* **100**: 593–601.
- LUKENS, L., and J. DOEBLEY, 2001 Molecular evolution of the teosinte branched gene among maize and related grasses. *Mol. Biol. Evol.* **18**: 627–638.
- MACKAY, T. F. C., 2001 The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**: 303–339.
- MARTIN, C., and J. PAZ-ARES, 1997 MYB transcription factors in plants. *Trends Genet.* **13**: 67–73.
- MAYNARD SMITH, J., 1998 *Evolutionary Genetics*, Ed. 2. Oxford University Press, Oxford.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MURRAY, M. G., and W. F. THOMPSON, 1980 Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**: 4321–4325.
- NAKAI, K., Y. INAGAKI, H. NAGATA, C. MIYAZAKI and S. IIDA, 1998 Molecular characterization of the gene for dihydroflavonol 4-reductase of *japonica* rice varieties. *Plant Biotech.* **15**: 221–225.
- NEI, M., and W.-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**: 5269–5273.
- OKA, H. I., 1953 Variations in various characters and character combinations among rice varieties. *Jpn. J. Breed.* **3**: 33–43.
- OKA, H. I., 1988 *Origin of Cultivated Rice*. JSSP, Tokyo/Elsevier, Amsterdam.
- OKA, H. I., 1989 Distribution of gene diversity in Indica and Japonica rice varieties and their wild progenitors. *Rice Genet. Newslett.* **6**: 70–71.
- OLSEN, K. M., and M. D. PURUGGANAN, 2002 Molecular evidence on the origin and evolution of glutinous rice. *Genetics* **162**: 941–950.
- PAZ-ARES, J., D. GHOSAL, U. WIENAND, P. A. PETERSON and H. SAEDLER, 1987 The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* **6**: 3553–3558.
- POSADA, D., K. A. CRANDALL and E. C. HOLMES, 2002 Recombination in evolutionary genomics. *Annu. Rev. Genet.* **36**: 75–97.
- PURUGGANAN, M. D., and S. R. WESSLER, 1994 Molecular evolution of the plant *R* regulatory gene family. *Genetics* **138**: 849–854.
- REDDY, V. S., B. E. SCHEFFLER, U. WIENAND, S. R. WESSLER and A. R. REDDY, 1998 Cloning and characterization of the rice homologue of the maize *C1* anthocyanin regulatory gene. *Plant Mol. Biol.* **36**: 497–498.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SAITOU, N., and M. NEI, 1987 The neighbor joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SAITOU, N., and F. YAMAMOTO, 1997 Evolution of primate ABO blood group genes and their homologous genes. *Mol. Biol. Evol.* **14**: 399–411.
- SAKAMOTO, W., T. OHMORI, K. KAGEYAMA, C. MIYAZAKI, A. SAITO *et al.*, 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.
- SANO, Y., and H. MORISHIMA, 1982 Variation in resource allocation and adaptive strategy of a wild rice, *Oryza perennis* Moench. *Bot. Gaz.* **143**: 518–523.
- SWOFFORD, D. L., 1998 *PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Sinauer Associates, Sunderland, MA.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAKAHASHI, M.-E., 1957 Analysis on apiculus color genes essential to anthocyanin coloration in rice. *J. Fac. Agr. Hokkaido Univ.* **50**: 266–362.
- TAKAHASHI, M.-E., 1982 Gene analysis and its related problems. *J. Fac. Agr. Hokkaido Univ.* **61**: 91–142.
- TANKSLEY, S. D., and S. R. MCCOUCH, 1997 Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**: 1063–1066.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 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.
- TING, Y., 1949 A preliminary report on the cultivation and the distribution of hsien and keng rices in ancient China and the classification of current cultivars. *Memoir Coll. Agr. Sun. Yatsen Univ.* **6**: 1–32 (in Chinese).
- WANG, R.-L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. *Nature* **398**: 236–239.
- WANG, Z. Y., and S. D. TANKSLEY, 1989 Restriction fragment length polymorphism in *Oryza sativa* L. *Genome* **32**: 1113–1118.
- XIONG, L. Z., K. D. LIN, X. K. DAI, C. G. XU and Q. ZHANG, 1999 Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between *Oryza sativa* and *O. rufipogon*. *Theor. Appl. Genet.* **98**: 243–251.

Communicating editor: S. R. MCCOUCH

