

A Single Base Change Altered the Regulation of the *Waxy* Gene at the Posttranscriptional Level During the Domestication of Rice

Hiro-Yuki Hirano,¹ Mitsugu Eiguchi, and Yoshio Sano²

National Institute of Genetics, Mishima, Japan

The rice *waxy* (*wx*) locus has two functional alleles, *Wx^a* and *Wx^b*, which are defined by a large difference in the amount of the gene product, called *Wx* protein, that accumulates in mature seeds. To elucidate the molecular mechanism underlying this difference and to identify the base change causing the alteration of the regulation of the *Wx* gene during rice evolution, we determined the nucleotide sequences of the regulatory region of *Wx* alleles and analyzed their function in a transient assay system using rice protoplasts. All *Wx^a* alleles from *Oryza sativa* Indica, *O. rufipogon*, and *O. glaberrima* have a normal sequence of GT at the 5' splice junction of the first intron, representing a high expression level of the *Wx* transcripts in the endosperm and a high β -glucuronidase (GUS) activity in protoplasts. On the other hand, *Wx^b* alleles from two strains of *O. sativa* Japonica have TT at the 5' splice junction, representing a low expression level of the mature transcripts and a low GUS activity. Northern blot analysis also indicated that a larger transcript, consisting of the unspliced first intron, is closely correlated with the function of the *Wx^b* allele. These results suggest that a single base change at the 5' splice junction causes inefficient splicing and, as a result, reduces the level of mature transcript and the GUS activity in the *Wx^b* allele. The *Wx^b* allele in *O. sativa* Japonica may have been differentiated from the *Wx^a* allele of *O. rufipogon*, its wild progenitor, by this mutation, and, therefore, a single base change that has altered the regulation of the *Wx* gene at the posttranscriptional level probably occurred during the domestication of rice.

Introduction

Recent studies in molecular evolution have been focused on the importance of gene regulation (Inomata et al. 1995; Britten 1996). Expression of genes is regulated in various ways, such as temporally, spatially, and developmentally. Responses to environmental stimuli or changes are also a key feature of gene regulation. Molecular mechanisms for such regulations at the transcriptional level have been studied in many organisms and an enormous number of cis and trans factors have been identified. Posttranscriptional processes, such as intron splicing and translational processing, have also been shown to play important roles in the regulation of gene expression.

In spite of the accumulating knowledge of gene regulation in individual organisms and its importance in evolution, there have been few studies on the evolution of regulatory mechanisms. Insertion of DNA sequences, such as retrotransposons or *Alu* sequences, during evolution has been shown to affect the transcriptional control of the adjacent genes (for a review, see McDonald 1993 and Britten 1996). For example, the insertion of a retroviral sequence changed the tissue-specific expression of an amylase gene from the pancreas to the salivary glands after the duplication of the gene, which was originally expressed in the pancreas (Samuelson et al.

1990; Ting et al. 1992). However, in addition to such insertions of transposable elements, minor base changes in critical locations should also lead to alterations in the regulation of gene expression. Compared with the protein-coding regions, promoter regions responsible for gene regulation have not always been conserved among organisms, even among closely related species. This often makes it difficult to identify minor mutations (base substitutions, insertions, or deletions) that have led to alterations in gene regulation during evolution.

Domestication of plants and animals is regarded as short-term evolution. Since sequence divergences during such processes should be small, even in rapidly changing regions such as the regulatory regions, the key base changes altering gene regulation may be easily detected by comparing the domesticated rice with its wild progenitor. Therefore, if we focus on characteristic features in domesticated species, we may be able to identify regulatory changes and address evolutionary mechanisms of gene regulation.

The evolution of the genus *Oryza*, rice, has been well studied (Oka 1988; Morishima, Sano, and Oka 1992). *Oryza sativa* is domesticated rice that originates in Asia. *Oryza sativa* is classified into three types, Japonica, Indica, and Javanica, which are derived from an ancestral wild rice, *O. rufipogon*, widely distributed in South Asia. Another type of cultivated rice, *O. glaberrima*, is evolved from wild rice, *O. barthii*, both of which are endemic to west Africa.

The *Waxy* (*Wx*) gene, which controls amylose synthesis in endosperm and pollen, has been well studied in rice (Sano 1984; Hirano and Sano 1991; Sano, Hirano, and Nishimura 1991; Hirano et al. 1994; Sano, Eiguchi, and Hirano 1994; Wang et al. 1995; Hirano, Eiguchi, and Sano 1996; Itoh, Nakajima, and Shimamoto 1997) and in maize (Nelson 1962; Marillonnet and Wessler 1997). The regulation of the rice *Wx* gene is characterized by two functional alleles which are defined on

¹ Present address: Graduate School of Agricultural and Life Sciences, University of Tokyo.

² Present address: Faculty of Agriculture, Hokkaido University.

Abbreviations: dap, days after pollination; GUS, β -glucuronidase; UTR, untranslated region.

Key words: amylose, domestication, posttranscriptional control, Rice (*Oryza*), *Waxy* (*Wx*) gene.

Address for correspondence and reprints: Hiro-Yuki Hirano, Laboratory of Plant Breeding and Genetics, Graduate School of Agricultural and Life Sciences, University of Tokyo, Yayoi, Bunkyo-ku 113-8657, Japan. E-mail: ahirano@hongo.ecc.u-tokyo.ac.jp.

Mol. Biol. Evol. 15(8):978–987, 1998

© 1998 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

the basis of the amount of Wx protein that accumulates in the mature seeds (Sano 1984). The allele that controls the production of lower amounts of Wx protein is Wx^b , which is mainly distributed in *O. sativa* Japonica. The other allele, Wx^a , produces levels of Wx protein about 10-fold higher than Wx^b and is widely distributed in domestic rice, including *O. sativa* Indica, *O. glaberrima*, and their wild progenitors. The evolutionary relationships of the species in the genus *Oryza* indicate that Wx^b in *O. sativa* Japonica has evolved from Wx^a of *O. rufipogon* (Sano 1984; Sano, Hirano, and Nishimura 1991). Previously, posttranscriptional regulation has been suggested as the cause of the quantitative difference in Wx protein between Wx^a and Wx^b alleles of cultivated rice (Okagaki 1992; Wang et al. 1995).

What kind of mutation was required for the differentiation of Wx^b of *O. sativa* Japonica from Wx^a of *O. rufipogon* during domestication? Here we report a key sequence responsible for the quantitative difference in gene regulation of Wx^a and Wx^b and a sequence change required for the differentiation of the Wx^b allele during rice evolution.

Materials and Methods

Plant Materials

The plant materials used were the Japonica type of *O. sativa* (T65, Taichung 65 from Taiwan; N8, Norin 8 from Japan) and three congenic lines of T65 carrying different Wx alleles. The congenic lines carry the Wx alleles from the Indica type of *O. sativa* (Patpaku from Taiwan), *O. rufipogon* (W593 from Malaysia), and *O. glaberrima* (W025 from Guinea), respectively. These congenic lines were established by backcrossing more than 10 times.

Cloning, Sequencing, and Plasmid Construction

General recombinant DNA techniques were carried out as described in Maniatis, Fritsch, and Sambrook (1982). The *Bam*HI fragment containing the upstream sequence of the Wx gene isolated from *O. sativa* Japonica (T65) was subcloned into Bluescript⁺SK from the genomic λ clone (λ T65-2) as described previously (Hirano and Sano 1991). The nucleotide sequence was determined by the dye primer method (Perkin-Elmer). The upstream DNAs (about 2.8 kb) of the Wx gene were amplified using an LA amplification kit (Takara) in *O. sativa* Japonica (T65 and Norin8), *O. sativa* Indica (Patpaku), *O. rufipogon* (W593), and *O. glaberrima* (W025). The primers used were as follows: PD11, CTA-AGCTTGCCGAGGGACCTAATCTGC (including an additional *Hind*III site at the 5' end) and PU11, GTCCCGGGTTGTCTAGCTGTTGCTG (including an additional *Sma*I site at the 5' end). Amplified DNAs were cloned into the *Hind*III and *Sma*I sites of pHIR16, which had a promoterless GUS gene derived from pBI221 (after removal of the 35S promoter) in pUC19. The nucleotide sequence of the upstream DNA, except for that from T65, was determined by the dye terminator method (Perkin-Elmer) with various primers in the previously determined Wx upstream sequence of T65. Chi-

meric plasmids were made using restriction enzyme sites. For use of the *Cl*aI site, which was indigestible because of methylation, *Escherichia coli* strain GM2163 (*dam*⁻) was used as a host.

Northern Blot Analysis

RNA was isolated from developing seeds (10 days after planting [dap]) as described in Naito, Dube, and Beachy (1988), and Northern blotting was conducted following the protocols described by Maniatis, Fritsch, and Sambrook (1982). Hybridization was carried out with the rice Wx cDNA labeled with ³²P as a probe. As a control for the blot, the membranes were reprobed with rice rDNA (Sano and Sano 1990) after removal of the Wx cDNA probe.

Transient Assay

Transient assays were carried out using protoplasts isolated from rice suspension cultures (Oc line; Baba, Hasezawa, and Syono 1986) essentially as described by Izawa et al. (1994). Various plasmids (20 μ g) were introduced into protoplasts (5×10^6 /ml, 0.5 ml) by electroporation (180 mV, 30 ms, 500 μ F) with a Transfector 300 (BTX). Each construct was analyzed in triplicate. After culture of the cells for 42 h with Oc nurse cells using a minicell (Falcon), protein extracts were prepared from the protoplasts. β -glucuronidase (GUS) activity was quantitated by measuring the production of 4-methylumbelliferone by the standard protocol of Jefferson, Kavanagh, and Bevan (1987).

Results

Northern Blot Analysis

The alleles at the wx locus, Wx^a and Wx^b , were initially defined on the basis of the amount of Wx protein that is contained in the mature seed (Sano 1984). Seeds with Wx^a contain levels of Wx protein about 10-fold higher than those with Wx^b . To examine whether this quantitative difference is regulated at the mRNA levels, we carried out Northern blot analysis using one allele of Wx^b from *O. sativa* Japonica (T65) and three alleles from *O. sativa* Indica, *O. rufipogon*, and *O. glaberrima*. All rice strains with Wx^a analyzed were congenic lines, in which each wx locus was introduced from the original strains into the genetic background of *O. sativa* Japonica (T65).

Total RNA was isolated from developing seeds (10 dap) and analyzed by Northern blot using the Wx cDNA probe. The result clearly indicated that the steady-state levels of the mature transcript from rice with Wx^a were much higher than those from rice with Wx^b (fig. 1A). In addition to the mature transcript, an extra band was detected in the lane of *O. sativa* Japonica (fig. 1B). This band corresponds to the precursor of Wx transcripts, which contains the unspliced first intron in the region upstream of the ATG codon (fig. 2A), as pointed out previously (Okagaki 1992; Wang et al. 1995). The extra band was not detected in the lanes of rice with Wx^a . These results suggested that the regulation underlying the quantitative difference between Wx^a and Wx^b was at the transcriptional or posttranscriptional level.

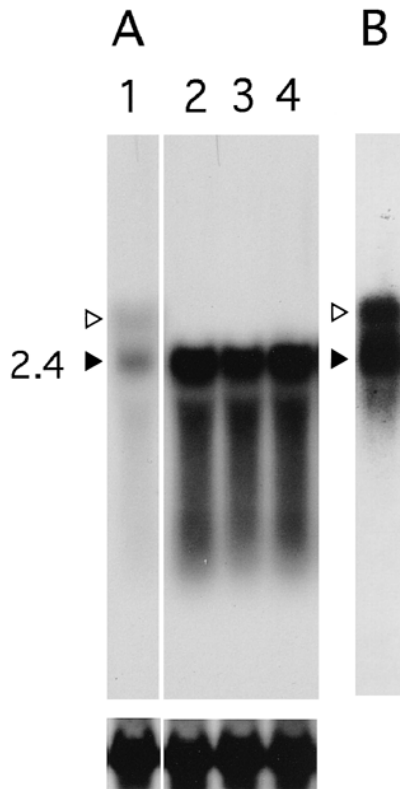


FIG. 1.—Northern blot analysis of the *Wx* gene. RNA was isolated from developing seeds (10 dap) of *Oryza sativa* Japonica (T65, *Wx*^b) and congenic lines (genetic background, *O. sativa* Japonica T65). Total RNA (10 µg) was analyzed using the rice *Wx* cDNA as a probe. Closed triangles indicate the mature *Wx* transcript of 2.4 kb and open triangles indicate the intermediate transcript containing the unspliced first intron. (A) Northern blot of RNA from *O. sativa* Japonica (T65, *Wx*^b) and congenic lines. The origin of *Wx* alleles is as follows: 1, *O. sativa* Japonica (T65, *Wx*^b); 2, *O. sativa* Indica (*Wx*^a); 3, *O. rufipogon* (*Wx*^a); and 4, *O. glaberrima* (*Wx*^a). The lower panel shows the control for the blot using the rice rDNA as a probe. (B) Northern blot of RNA from *O. sativa* Japonica (T65, *Wx*^b) obtained by long exposure.

Sequence Comparison

To find the sequences responsible for the quantitative difference between *Wx*^a and *Wx*^b, we determined the nucleotide sequences of the regulatory region of the *Wx* alleles using *O. sativa* Japonica (T65) and *O. sativa* Indica, because the two strains were closely related and mutations unrelated to the quantitative difference are expected to be small (fig. 3). The sequences determined were mainly composed of two regions, the promoter region (P region) and the intron region (I region) containing the first intron flanked by the 5' untranslated region (UTR). The structure of this region is schematically described in figure 2B.

No large insertions/deletions or inversions were found between *Wx*^a and *Wx*^b. The most evident difference was in the size of microsatellite DNA (TC repeats) located in the 5' UTR. The other differences were substitutions or deletions of one or two bases. The number of base changes, only 27 changes out of 2776 bases compared, was small. There seemed to be no significant difference in the number of changes between the P and I regions. Thus, the sequence comparison suggests that

one (or some) of the base changes between *Wx*^a and *Wx*^b greatly affect the expression level of the two alleles.

Transient Assay

To identify which base change is responsible for the quantitative regulation of *Wx*^a and *Wx*^b alleles, we examined the function of the DNA in the P and I regions (P-I DNA) using a transient assay system on rice protoplasts. P-I DNAs from *Wx*^a of *O. sativa* Indica and *Wx*^b of *O. sativa* Japonica (T65) were fused to the GUS gene as a reporter, and were introduced into protoplasts by electroporation. After 2 days of culture, GUS activity, which was monitoring the function of the two *Wx* alleles, was measured. It should be noted that the P-DNA and I-DNA in the plasmid constructs did not exactly correspond to the P and I regions, respectively. P-DNA was defined as the DNA fragment between the *Pma*CI and *Pst*I sites, including the transcriptional initiation site, and I-DNA was defined as the fragment between the *Pst*I and *Sma*I sites (derived from the linker added at position 2776).

First, WJ40 and WI40, which contained the entire region of P-I DNA from both alleles, were analyzed in this system (fig. 4A). The construct WI40 had much higher GUS activity than did the construct WJ40. This result raised the possibility that base changes responsible for the quantitative regulation among *Wx* alleles can be identified using this transient assay system.

Next, to find the region responsible for the higher activity of the *Wx*^a allele, we made a series of WI constructs (WI40–WI46), in which part of the P-DNA was sequentially deleted from the upstream region (fig. 2C). Although WI46, which lacked the entire portion of the P region, showed very low GUS activity, WI45, which contained only a short segment of P-DNA (120 bp), showed activity that was as high as that of WI40, the original construct (fig. 4B). There was only a slight difference in this short region: one unit of the GA repeat was deleted in *Wx*^a (Indica) (fig. 3). Furthermore, WI50, which had almost the entire length of P-DNA but lacked I-DNA, showed activity that was as low as that of WI46 (fig. 4B).

Identification of a Key Sequence Responsible for the High Activity of the *Wx*^a Allele

The above results indicated that an ordinary analysis that used the clones derived from the simple deletion of a “strong promoter” was not able to identify the element responsible for the high activity of *Wx*^a. Furthermore, the results of the experiment using WI50 also indicated that I-DNA played an important role in the normal high activity of *Wx*^a. Therefore, we constructed the chimera P-I DNA, in which the original I-DNA was replaced by the I-DNA from the other allele (fig. 2C, PJ43P and PI43P). The results of the transient assay showed that PJ43P containing I-DNA from *Wx*^a had an activity that was as high as that of WI43 (fig. 4C). This clearly demonstrated that I-DNA, but not P-DNA, from *Wx*^a (Indica) contains a key sequence for the high activity of *Wx*^a.

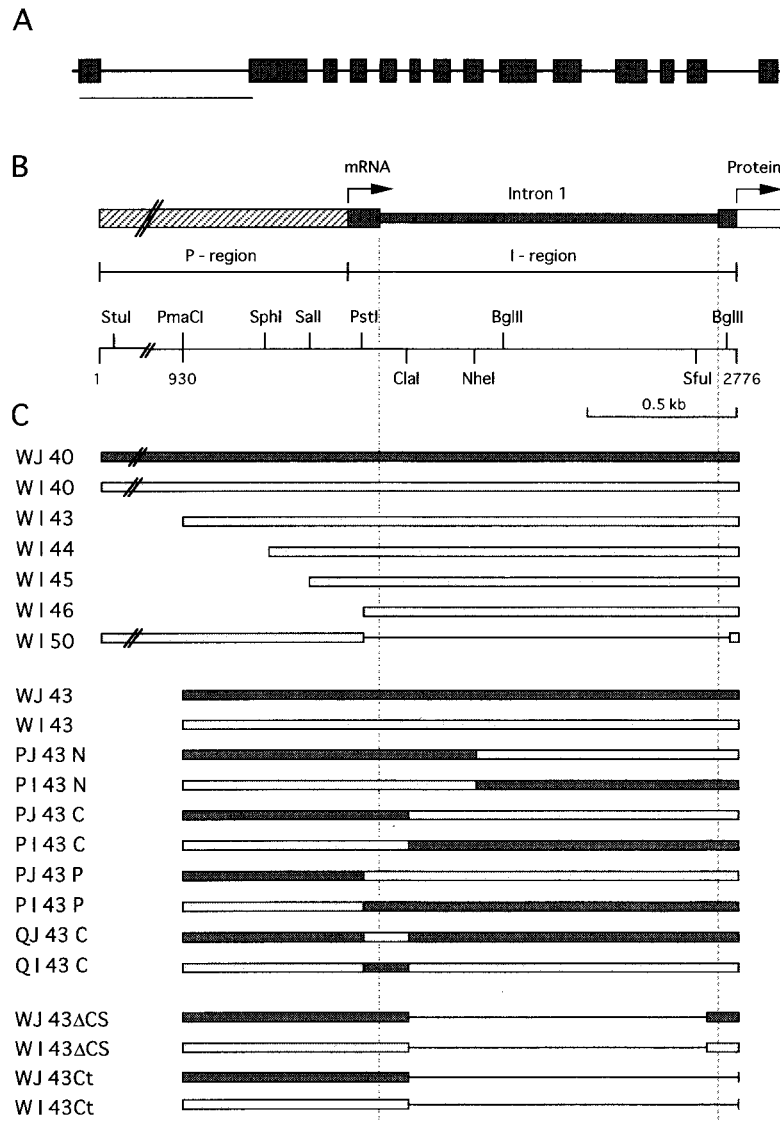


FIG. 2.—Schematic representation of the *Wx* gene structure and DNA constructs used for the transient assay. (A) The exon-intron structure of the rice *Wx* gene. The boxes indicate exons, and the lines between boxes indicate introns. The line under the gene corresponds to the boxes with dark gray color (I region) in fig. 2B. (B) The upstream structure of the *Wx* gene and the restriction enzyme map. The box with slashed lines is the promoter (P) region and the boxes with dark gray color are the intron (I) region, consisting of the first intron (thin box) and the 5' UTR (thick boxes). The three numbers denote the nucleotide numbers as indicated in figure 3. (C) Constructs used for transient assay. The boxes with dark gray indicate the DNA derived from *Wx^b* (Japonica), and the boxes with light gray indicate the DNA from *Wx^a* (Indica). The regions lacking DNA sequences are shown as solid lines. The dashed vertical lines indicate the position of the 5' and 3' splice junctions of the first intron.

To identify the base changes in I-DNA responsible for the quantitative difference between *Wx^a* (Indica) and *Wx^b* (Japonica), we made various chimeric constructs as indicated in figure 2C (PJ43N–QI43C) and analyzed them in the transient assay system. We found that the constructs that showed high GUS activity, such as WI43, included the DNA fragment (145 bp) between the *PstI* and *ClaI* sites derived from *Wx^a* (Indica) (fig. 5). Thus, we concluded that a key element for the high activity of *Wx^a* was present at the 145-bp region between the *PstI* and *ClaI* sites, which was located downstream of the transcriptional initiation site.

Two sequence differences between *Wx^a* (Indica) and *Wx^b* (Japonica) were found in this 145-bp region:

the number of TC repeats (microsatellite DNA), as mentioned earlier, and a single base substitution at the 5' splice junction of the first intron (figs. 3 and 6). Judging from the GT-AG rule for RNA splicing (Breathnach and Chambon 1981; Mount 1982), *Wx^a* has the normal sequence (GT) of the 5' splice junction, but *Wx^b* has a T in place of the G.

Identification of a Base Change Responsible for the Differentiation of the *Wx^b* Allele During Rice Evolution

Since the DNA region responsible for the quantitative difference between *Wx^a* and *Wx^b* was shown to be restricted to the narrow region, we analyzed the regu-

FIG. 3.—Nucleotide sequence of the P-I region of Wx^b from *O. sativa* Japonica (T65; accession number AB008794) and Wx^a derived from *O. sativa* Indica (accession number AB008795). The first intron is shown in lower case. Positions having different bases are marked by asterisks. The double line shows putative TATA-like sequences, and the closed circle shows the transcriptional initiation site (K. Shimamoto, personal communication). The translation initiation codon (ATG) is boxed. The sequences used as primers for DNA amplification are underlined.

the genomic DNA of each rice, and the nucleotide sequences of the 145-bp region were determined (fig. 6). Each of the rice species carrying *Wx^a* (*O. sativa* Indica, *O. rufipogon*, and *O. glaberrima*) had GT at the 5' junction

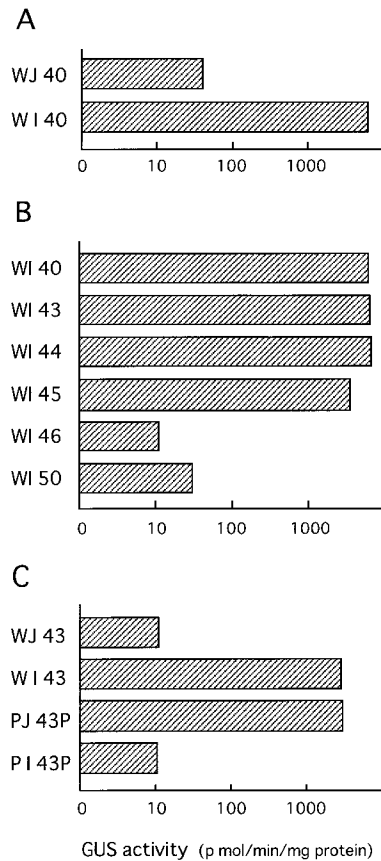


FIG. 4.—GUS activity obtained in the transient assay. Plasmid DNAs (20 μ g) were introduced into protoplasts by electroporation. After culturing them for 2 days, protein was extracted and GUS activity was measured. The bars indicate the average of GUS activity obtained from three independent experiments carried out using the same protoplast preparation.

tion site, while the two strains of *O. sativa* had TT at the same site. On the other hand, the length of the TC repeat of *O. rufipogon* was as long as that of the two strains of *O. sativa* Japonica. This sequence comparison revealed that the key sequence difference between the function of Wx^a and Wx^b was the single base change at the 5' splice junction, not the length of the TC repeat. To test this hypothesis, we analyzed the function of PI-DNA from these rice strains by the transient assay system. All constructs showing high GUS activity were derived from *O. sativa* Indica, *O. rufipogon*, and *O. glaberrima*, whose *Wx* genes have G at the 5' splice junction (fig. 7). Thus, GT at the 5' splice junction of the first intron was coincident with the high GUS activity analyzed by the transient assay (fig. 7), as well as the high expression level of the transcripts found by Northern blotting (fig. 1).

It is possible that the normal sequence of the 5' splice junction (GT) ensured the high expression levels of the mature *Wx* transcript and the GUS activity of fusion genes in protoplasts. A single base change from G to T at this site in *O. sativa* Japonica may cause inefficient splicing and, as a result, the level of mature transcript may be reduced (discussed below).

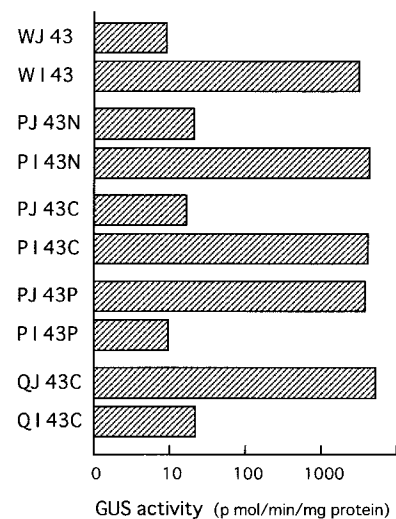


FIG. 5.—GUS activity obtained using the chimeric constructs consisting of the DNA from Wx^a (Indica) and Wx^b (Japonica). Experimental conditions were the same as described in the legend for figure 4. Almost identical results were obtained from another independent experiment using the same constructs (data not shown).

Effect of the First Intron on the Expression of the Wx^a and Wx^b Alleles

To examine the above possibility and the effect of the first intron on gene expression, we made the constructs, in which the first introns were partially deleted (fig. 2C, WJ43 Δ CS–WI43Ct). In WJ43 Δ CS and WI43 Δ CS, the middle part of the first intron was deleted, but the 5' and 3' splice junction sites were intact. WJ43Ct and WI43Ct lacked almost all the region of the first intron and its 3' junction site but contained a region around the 5' junction site. The GUS activities of WI43 Δ CS and WI43Ct were reduced from that of the original construct of WI43, and the extent of reduction was greater in WI43Ct than in WI43 Δ CS (fig. 8). This result suggests that the sequences in the first intron have positive effects on the expression of the *Wx* gene. On the other hand, the GUS activity of WJ43 Δ CS was as low as that of the original construct, WJ43. However, GUS activity was recovered in WJ43Ct and was similar to that of WI43Ct. Since WJ43Ct and WI43Ct lack the 3' splice junction, it is plausible that the remaining sequences of the first intron in both constructs cannot be recognized as an intron but can be recognized as 5' UTR-like sequences by the splicing machinery. As a consequence, the mutation at the 5' junction site of WJ43Ct may not affect the GUS expression analyzed in the transient assay system. Thus, WJ43Ct and WI43Ct showed similar GUS activities. Taking these results together, we concluded that the key alteration responsible for the quantitative difference in the expression level of Wx^a and Wx^b was a single base change at the 5' splice junction of the first intron.

Discussion

We found that a single base change at the 5' splice junction of the first intron is responsible for the quan-

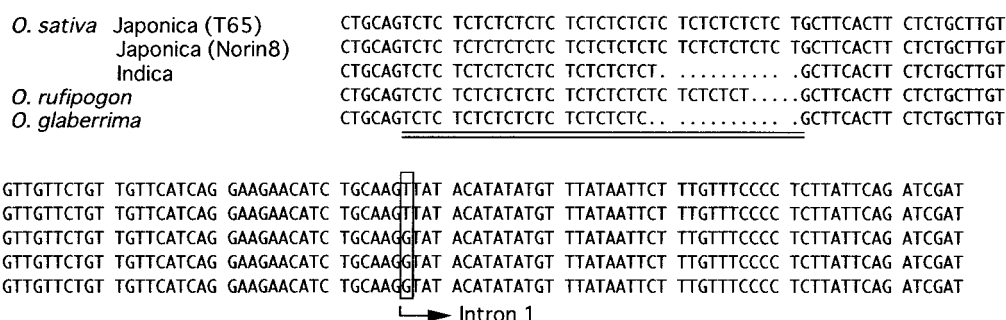


FIG. 6.—Comparison of the 145-bp sequence between the *Pst*I and *Cl*aI sites. Microsatellite DNA (TC repeats) is shown with a double line, and bases at the 5' splice junction site of the first intron are indicated with a box.

titative difference in the *Wx* gene regulation. This change may have caused the differentiation of the *Wx^b* allele from the *Wx^a* allele in the wild progenitor during the domestication of *O. sativa* Japonica.

Wx gene expression is affected by other loci, such as the *du* locus (Okuno, Fuwa, and Yano 1983; unpublished data). Moreover, since the expression level of the *Wx* transcript varies temporally through a wide range in developing seeds (Hirano and Sano 1991), and since the rate of seed maturation differs among species, this variation should be taken into consideration in comparison of the transcript level. For these reasons, we carried out Northern blot analysis using congenic lines to compare the expression levels of the *Wx* transcript in developing seeds. The results clearly indicated that lines with *Wx^a* from *O. sativa* Indica, *O. rufipogon*, and *O. glaberrima* showed much higher levels of mature transcripts than did *O. sativa* Japonica with *Wx^b*. Thus, the quantitative difference in the amount of *Wx* protein in mature seeds, which was the original criterion for allele definition (Sano 1984), was shown to depend on the steady-state level of mature transcripts. The presence of a larger extra band in *Wx^b*, which contains the unspliced first intron, has been pointed out in the studies using rice cultivars with different genetic backgrounds (Okagaki 1992; Wang et al. 1995). Therefore, there remained the possibility that the extra band was produced not only by the function of *Wx^b* but also depending on the genetic background of *O. sativa* Japonica. Here, we were able to demonstrate clearly that the extra band found in *O. sativa* Japonica was closely related to the function of the *Wx^b* allele per se. Thus, we concluded that the splic-

ing mechanism may not efficiently function in rice with the *Wx^b* allele.

We have found that the sequence responsible for the high activity of *Wx^a* is located within the 145-bp region downstream of the transcriptional initiation site. This region contained two differences between *Wx^a* and *Wx^b*, that is, the sequences at the 5' splice junction of the first intron and the length of the TC repeat. The result of the transient assay using constructs with partial deletions of the first intron (fig. 8) strongly suggests that GT at the 5' splice junction is essential for the high expression level in *Wx^a*. All *Wx* alleles with GT at the 5' splice junction were associated with a high expression level of the *Wx* transcript in the Northern blot analysis and with a high GUS activity in the transient assay. On the other hand, the size of the TC repeats of *Wx^a* (*O. rufipogon*) is more similar to that of two *Wx^b* alleles (Japonica T65 and Japonica Norin8) than to that of *Wx^a* (Indica). Akagi et al. (1996) examined many rice cultivars and found that the sizes of the TC repeats were highly polymorphic. Furthermore, the size distribution was indistinguishable between *O. sativa* Japonica and *O. sativa* Indica (H. Akagi, personal communication). Moreover, our recent study indicated that two alleles of *Wx^a* (Indica) had the same number of TC repeats (17 or 18) as two alleles of *Wx^b* (17 or 18), while the sequences at the 5' junction were coincident with the type of allele (G for *Wx^a* and T for *Wx^b*). Taken together, these results lead us to conclude that the base change at the 5' splice

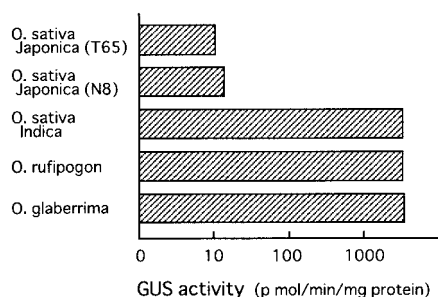


FIG. 7.—GUS activity obtained using the P-I DNA derived from two cultivars of *O. sativa* Japonica, *O. sativa* Indica, *O. rufipogon*, and *O. glaberrima*. Experimental conditions were the same as described in the legend for figure 4.

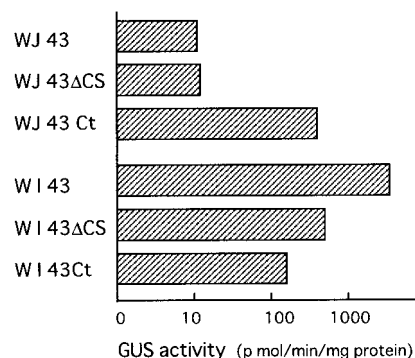


FIG. 8.—GUS activity using constructs made to analyze the effect of the first intron. Experimental conditions were as described in the legend for figure 4.

junction of the first intron might play a key role in the quantitative regulation of Wx^a and Wx^b .

In the GT-AG rule for RNA splicing, the 5'-junction sequence of TT in Wx^b seems to function poorly. However, several GT sequences are present near the original junction site, and, therefore, some of these sequences may provide cryptic sites for splicing (Treisman, Orkin, and Maniatis 1983). Thus, the splicing of the Wx transcript seems to be inefficient in Wx^b . The presence of the larger extra transcripts detected in the Northern blot analysis agrees with this conclusion.

The inefficient splicing of the first intron may be responsible for the reduction in the level of mature transcript. If the abnormal transcript containing the unspliced intron is unstable and is degraded quickly, the level of the mature transcript may be reduced. Alternatively, the accumulation of the abnormal transcript, which was an intermediate in the process of RNA splicing, may suppress the transcription of its own gene. The mutation at the splice junction may not affect the transcriptional level, because the transient assay shows that the promoter activity of Wx^b is as high as that of Wx^a (fig. 4C). The posttranscriptional regulation of the Wx gene has been described in maize (Marillonnet and Wessler 1997). In this case, the insertion of a retrotransposon into an intron altered the tissue-specific quantitative regulation in the endosperm and pollen, causing abnormal splicing, the efficiency of which differed in a tissue-specific manner.

All alleles from *O. sativa* Indica, *O. rufipogon*, and *O. glaberrima* had GT at the 5' splice junction and exhibited high levels of expression of the Wx transcripts in the endosperm and high GUS activity in protoplasts. On the other hand, the two alleles from the two strains of *O. sativa* Japonica had TT at the 5' splice junction and exhibited low levels of expression of the mature transcripts and low GUS activity. The evolutionary relationships within the genus *Oryza* indicate that TT at the 5' splice junction of Wx^b was produced by a single base change from GT in Wx^a of *O. rufipogon*. This mutation had spread to the population during domestication of *O. sativa* Japonica (fig. 9). In *O. sativa* Indica and other species, the mutation may not have been fixed in the population. Due to the function of the Wx gene (discussed below), the mutation producing Wx^b may have been spread quickly by an artificial selection. It is of great interest whether Wx^b originated from a single mutation event or from independent multiple events. To address this question, the sequences independent of the Wx gene from a number of strains of *O. sativa* Japonica, *O. sativa* Indica, and *O. rufipogon* may need to be compared.

Since Wx protein catalyzes amylose synthesis, the expression level of the Wx gene affects the production of amylose. Consistent with the low levels of the Wx transcript and Wx protein in Wx^b , the amylose content in *O. sativa* Japonica is lower than in other rice species carrying Wx^a (Sano 1984; Sano, Katsumata, and Amano 1985; Sano, Hirano, and Nishimura 1991). The amylose content is a major factor that affects the quality of the rice grains, in particular their stickiness after

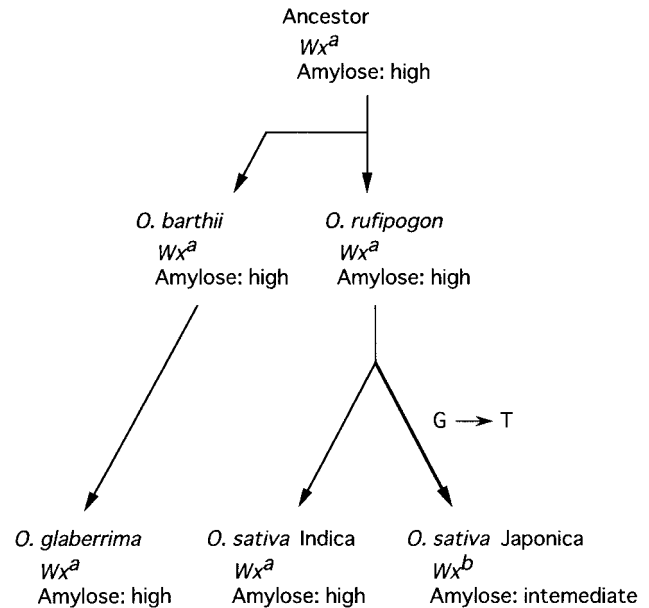


FIG. 9.—Evolutionary process of domesticated rice and their wild progenitors. The type of allele and the amylose content of a putative "ancestor" were deduced from those of the other five rice strains. In the process indicated with a thick line, the change from G to T at the 5' splice junction may have occurred, differentiating Wx^b in *O. sativa* Japonica from Wx^a in *O. rufipogon*.

cooking. Cooked grains from rice with Wx^a , such as *O. sativa* Indica, are not sticky. The rice carrying Wx^b is the intermediate type with respect to amylose content, and its cooked grains show moderate stickiness. Therefore, the Wx^b allele may have been selected during domestication by the people in East Asia who preferred cooked rice with moderate stickiness. In other words, a point mutation, G to T, was selected for during the domestication of *O. sativa* Japonica. In this sense, a strong artificial selection may have occurred on this mutation.

Recently, Ayres et al. (1997) showed the relation between the polymorphism at the 5' splice junction and the amylose content by analyzing 92 U.S. rice cultivars. This result strongly supports our above conclusion.

The Wx gene is not essential for the growth of rice, because null mutants of this locus (wx) grow normally in the fields. However, neither Wx^b or wx have yet been found in wild rice populations (Sano, Katsumata, and Okuno 1986). Therefore, the function of both alleles may be slightly deleterious in wild populations. In fact, the wx alleles that introgressed into the wild rice from wx cultivars grown around them are gradually eliminated from wild populations with successive generations (Oka 1988). The Wx gene controls amylose synthesis in the pollen as well as in the endosperm, and amylose may be used for a carbon source for the pollen tube growth. It is plausible that the pollen with Wx^b or wx may have a disadvantage, on an evolutionary timescale, in the fertilization race. Thus, the differentiation of Wx^b is closely associated with the domestication of rice.

Some studies have reported that the insertion of DNA, such as retrotransposons, has changed gene regulation during evolution (Samuelson et al. 1990; Ting et al. 1992; Britten 1996). We demonstrated here for the first time that a single mutation altered the gene regulation at the posttranscriptional level during the evolution of domestic rice. The molecular mechanisms that lead to the domestication of crops from wild relatives is of great interest. One clear example is the origin of cauliflower in the genus *Brassica*. Here, the key mutation appears to be in the *CAL* gene, which regulates flower development. In *Arabidopsis thaliana*, a defect in this gene has been shown to produce a cluster of undifferentiated inflorescence meristems, which give rise to a cauliflowerlike structure. In the cultivated garden variety of cauliflower, *Brassica oleracea* var. botrytis, the *CAL* gene has a non-sense mutation causing the loss of its function (Kempin, Savidge, and Yanofsky 1995).

Acknowledgments

We are grateful to Dr. K. Shimamoto for communicating to us the position of the transcription initiation site of the *Wx* gene before publication and to Dr. T. Hattori for providing *Oc* suspension cells. We thank Dr. F. Tajima for a critical reading of the manuscript. We thank Miss M. Matsumoto for technical assistance. This work was supported in part by grants-in-aid for scientific research from the Japanese Ministry of Education, Science, Culture, and Sports.

LITERATURE CITED

- AKAGI, H., Y. YOKOZEKI, A. INAGAKI, and T. FUJIMURA. 1996. Microsatellite DNA markers for rice chromosomes. *Theor. Appl. Genet.* **93**:1071–1077.
- AYRES, N. M., A. M. MCCLUNG, P. D. LARKIN, H. F. J. BLIGH, C. A. JONES, and W. D. PARK. 1997. Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. *Theor. Appl. Genet.* **94**:773–781.
- BABA, A., S. HASEZAWA, and K. SYONO. 1986. Cultivation of rice protoplasts and their transformation mediated by *Agrobacterium* spheroplasts. *Plant Cell Physiol.* **27**:463–472.
- BREATHNACH, R., and P. CHAMBON. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**:349–383.
- BRITTEN, R. J. 1996. DNA sequence insertion and evolutionary variation in gene regulation. *Proc. Natl. Acad. Sci. USA* **93**:9374–9377.
- HIRANO, H.-Y., M. EIGUCHI, and Y. SANO. 1996. A point mutation, G to T, causes the differentiation of the *Wx^b* allele, which is specific to Japonica rice. *Rice Genet. Newsl.* **13**:148–149.
- HIRANO, H.-Y., and Y. SANO. 1991. Molecular characterization of the waxy locus of rice (*Oryza sativa*). *Plant Cell Physiol.* **32**:989–997.
- HIRANO, H.-Y., N. TABAYASHI, T. MATSUMURA, M. TANIDA, Y. KOMEDA, and Y. SANO. 1994. Tissue-dependent expression of the rice *wx⁺* gene promoter in transgenic rice and petunia. *Plant Cell Physiol.* **36**:37–44.
- INOMATA, N., K. KANDA, M. L. CARIU, H. TACHIDA, and T. YAMAZAKI. 1995. Evolution of the response patterns to dietary carbohydrates and the developmental differentiation of gene expression of α -amylase in *Drosophila*. *J. Mol. Evol.* **41**:1076–1084.
- ITO, K., M. NAKAJIMA, and K. SHIMAMOTO. 1997. Silencing of waxy genes in rice containing *Wx* transgenes. *Mol. Gen. Genet.* **255**:351.
- IZAWA, T., R. FOSTER, M. NAKAJIMA, K. SHIMAMOTO, and N. H. CHUA. 1994. The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. *Plant Cell* **6**:1277–1287.
- JEFFERSON, R. A., T. A. KAVANAGH, and M. W. BEVAN. 1987. GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**:3901–3907.
- KEMPIN, S. A., B. SAVIDGE, and M. F. YANOFSKY. 1995. Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* **267**:522–525.
- MCDONALD, J. F. 1993. Evolution and consequence of transposable elements. *Curr. Opin. Genet. Dev.* **3**:855–864.
- MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- MARILLONNET, S., and S. R. WESSLER. 1997. Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing. *Plant Cell* **9**:967–978.
- MORISHIMA, H., Y. SANO, and H.-I. OKA. 1992. Evolutionary studies in cultivated rice and its wild relatives. *Oxf. Surv. Evol. Biol.* **8**:135–184.
- MOUNT, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**:459–472.
- NAITO, S., P. H. DUBE, and R. N. BEACHY. 1988. Differential expression of α' -conglycinin and β subunit genes in transgenic plants. *Plant Mol. Biol.* **11**:109–123.
- NELSON, O. E. 1962. The waxy locus in maize. I. Interlocus recombination frequency estimates by pollen and by conventional analyses. *Genetics* **47**:737–742.
- OKA, H. I. 1988. Origin of cultivated rice. Japan Scientific Societies Press and Elsevier Science Publishers, Tokyo, Amsterdam.
- OKAGAKI, R. J. 1992. Nucleotide sequence of a long cDNA from the rice waxy gene. *Plant Mol. Biol.* **19**:513–516.
- OKUNO, K., H. FUWA, and M. YANO. 1983. A new mutant gene lowering amylose content in endosperm starch of rice, *Oryza sativa* L. *Jpn. J. Breed.* **33**:387–394.
- SAMUELSON, L. C., K. WIEBAUER, C. M. SNOW, and M. H. MEISLER. 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol. Cell. Biol.* **10**:2513–2520.
- SANO, Y. 1984. Differential regulation of waxy gene expression in rice endosperm. *Theor. Appl. Genet.* **68**:467–473.
- SANO, Y., M. EIGUCHI, and H.-Y. HIRANO. 1994. Gamete eliminator adjacent to the *wx* locus as revealed by pollen analysis in rice. *J. Hered.* **85**:310–312.
- SANO, Y., H.-Y. HIRANO, and M. NISHIMURA. 1991. Evolutionary significance of differential regulation at the *wx* locus of rice. Pp. 11–20 in IRRI, ed. *Rice Genetics II*. IRRI, Manila, The Philippines.
- SANO, Y., M. KATSUMATA, and E. AMANO. 1985. Correlations between the amounts of amylose and *Wx* protein in rice endosperm. *SABRAO J.* **17**:121–127.
- SANO, Y., M. KATSUMATA, and K. OKUNO. 1986. Genetic studies of speciation in cultivated rice. 5. Inter- and intra-specific differentiation in the waxy gene expression of rice. *Euphytica* **35**:1–9.

- SANO, Y., and R. SANO. 1990. Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. *Genome* **33**:209–218.
- TING, C. N., M. P. ROSENBERG, C. M. SNOW, L. C. SAMUELSON, and M. H. MEISLER. 1992. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* **6**:1457–1465.
- TREISMAN, R., S. H. ORKIN, and T. MANIATIS. 1983. Specific transcription and RNA splicing defects in five cloned β -thalassaemia genes. *Nature* **302**:591–596.
- WANG, Z.-Y., F.-Q. ZHENG, G.-Z. SHEN, J.-P. GAO, D. P. SNUSTAD, M.-G. LI, J.-L. ZHANG, and M.-M. HONG. 1995. The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. *Plant J.* **6**:613–622.

NAOYUKI TAKAHATA, reviewing editor

Accepted April 14, 1998