Note

Plant Biotechnology, 15(4), 221-225(1998)

Molecular Characterization of the Gene for Dihydroflavonol 4-Reductase of *Japonica* Rice Varieties

Kiyoshi NAKAI*, Yoshishige INAGAKI**, Hisanori NAGATA*, Chikara MIYKZAKI*,*** and Shigeru IIDA*,***+

Received 24 February 1998; accepted 12 June 1998

Anthocyanins are the most important pigments in higher plants. Genes for anthocyanin biosynthesis can be classified into two groups; structural genes encoding enzymes in the biosynthesis pathway and regulatory genes acting on the structural genes [1-3]. The *DFR* gene encodes the enzyme, dihydroflavonol 4-reductase, which catalyses the stereospecific reduction of dihydroflavonols to the respective leucoanthocyanidines. The *DFR* genes of dicotyledonous plants consist of six exons and five introns whereas those of monocotyledonous plants barley and maize are comprised of four exons and three introns (**Fig. 1**). Recently, the *DFR* gene of a *indica* variety of rice was reported to consist of three exons and two introns [4]. Interestingly, the positions of the introns and exons are well conserved in these plants.

Although most of *japonica* and *indica* rice varieties bear green leaves and stems, there are varieties with purple leaves and stems due to anthocyanin pigmentation [5]. As an initial step to characterize genes for anthocyanin biosynthesis in rice, we have characterized the *DFR* gene of a *japonica* variety *Murasaki-ine* (purple rice) and shown that it carries a single copy of the *DFR* gene consisting of three exons and two introns. We also found that an *indica* variety *Kurogome* (black rice) with purple leaves and stems and two *japonica* cultivars, *Toride 1* and *Koshihikari*, with green leaves and stems carry a single copy of the *DFR* gene. Clear RFLPs were detected at the flanking regions of the *DFR* genes in these four varieties.

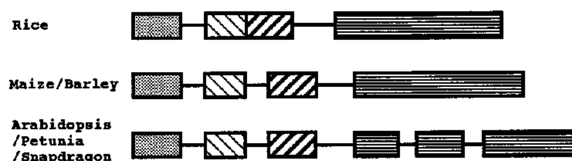


Fig. 1 Organization of the *DFR* genes.
The hatched boxes and the horizontal bars represent exons and introns, respectively. Only four genomic sequences of the *DFR* genes, arabidopsis, barley, maize and *indica* rice have been reported [8, 16, 17]. The *DFR* genes of petunia and snapdragon are reported to consist of six exons and five introns, although their intron sequences were not published [18].

The *japonica* varieties, *Murasaki-ine, Toride 1* and *Koshihikari*, were obtained from Plantech Research Institute, Yokohama, and the *indica* variety *Kurogome* was from Soumaya Co., Iwaki. The plants were grown in a greenhouse. To isolate mRNA, germinated seedlings of *Murasaki-ine* were grown in a growth chamber for three weeks. Total RNA was extracted from UV-induced young leaves of the seedlings [6], and poly$(A)^+$ mRNA was prepared and used for construction of a cDNA library [7]. The cDNA library was prepared by the cDNA Synthesis Kit (Stratagene). To isolate the rice *DFR* cDNA, about 100, 000 recombinant clones in a cDNA library from *Murasaki-ine* seedlings were screened with the maize *a1* cDNA probe [8]. Out of 17 clones isolated, six were chosen to characterize further. The longest clone contains a 1, 465 bp sequence carrying an 1, 119 bp open reading frame (**Fig. 2**), and it agrees in size with the *DFR* cDNA since mRNA of about 1. 4 kb was detected by RNA gel blot analysis (**Fig. 3**). Recently, the *DFR* cDNA sequence from the *indica* rice *Purpleputtu* and the genomic *DFR* sequence from the *indica* variety *Teqing* have been deposited in the nucleotide sequence databases (accession numbers Y07956 and U70541). When the cDNAs from the *japonica* and *indica* varieties were compared, only four nucleotide alterations were detected apart from the 5′ and 3′ ends. Only one of these affects the predicted amino acid sequences of the DFR enzymes at the position 101 (Val in *japonica* and Ile in *indica* varieties).

To characterize the genomic *DFR* sequences of the *japonica* variety *Murasaki-ine*, approximately 100 μg of plant DNA prepared from leaves [9] was digested with *Eco*RI or *Bam*HI and the resulting fragments were cloned into the *Eco*RI site of [λEMBL4] (Stratagene) or the *Bam*HI site of [λGEM11] (Promega). General methods for genomic cloning

* Department of Biological Science and Technology, Science University of Tokyo, 2641 Yamazaki, Noda 278-0022, Japan
** Division of Gene Expression and Regulation I, National Institute for Basic Biology, Myodaiji, Okazaki 444-8585, Japan
*** Nagasaki Laboratory, Japan Science and Technology Cooperation, 2-1303 Ikeda, Ohmura 856-0026, Japan
+ To whom correspondence should be addressed.
National Institute for Basic Biology, Myodaiji, Okazaki 444-8585, Japan

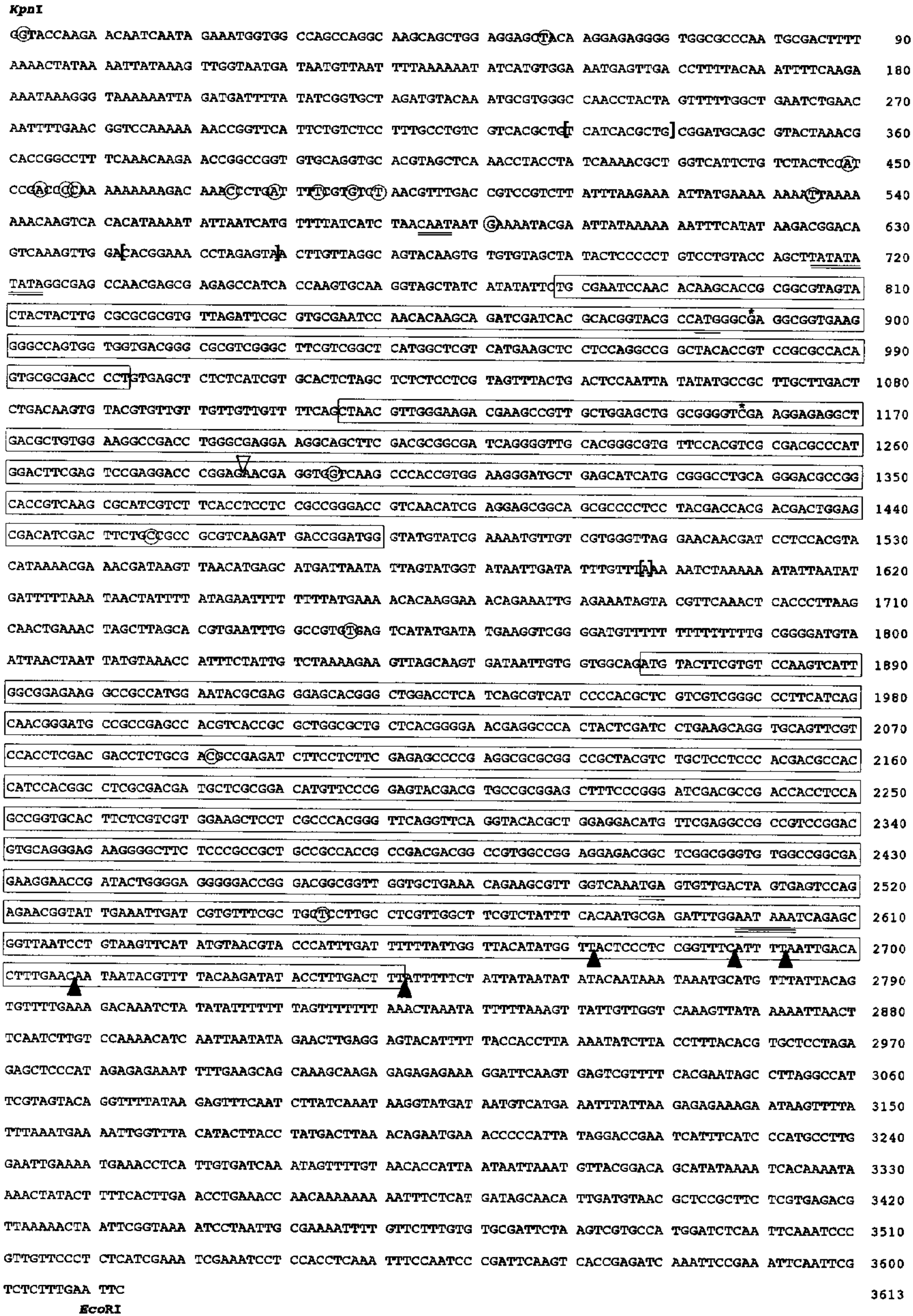


Fig. 2 Nucleotide sequences of the *DFR* gene region in *Murasaki-ine*.
The exon sequences are enclosed in the opened boxes. The ATG initiation codon at the position 883 and the termination codon TGA at the position 2498 are indicated by the underlines. Putative CAAT-box and TATA-box in the promoter region and a possible polyadenylation signal within exon 3 are indicated by the double underlines. The upward filled arrowheads indicate the 3′ ends of the cDNA clones determined. At the position 1286 indicated by the open downward arrowhead, about 0. 1 kb introns were found in barley and maize (see **Fig. 1**). Single base-pair substitutions and insertions found by comparison with the sequence of the *indica* rice *Teqing* are enclosed in circles and brackets, respectively. The sequence alterations leading to the stop codons found in *Koshihikari* and *Toride 1* are indicated by the asterisks. The nucleotide sequences reported here have been deposited to DDBJ under the accession numbers AB003495 (genomic sequence) and AB003496 (cDNA sequence).

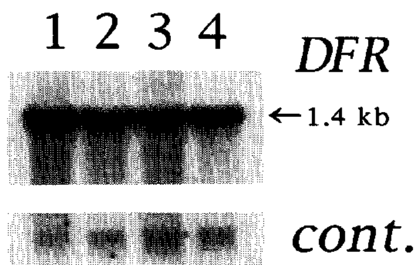


Fig. 3 RNA gel blot analysis.
Total RNAs (10 μg) from young leaves of the seedlings were hybridized with the *DFR* cDNA from *Murasaki-ine* as described before [6, 7]. The cDNA of acyl carrier protein [19] was used as a probe for an internal control, indicated by cont. Lane 1, *Kuro-gome*; lane 2, *Murasaki-ine*; lane 3, *Koshihikari*; lane 4, *Toride 1*.

and DNA gel blot analysis were performed as described before [7, 10]. Using the rice *DFR* cDNA as a probe, we cloned a *Bam*HI-*Bam*HI fragment of about 15 kb genomic DNA into the *Bam*HI site of λGEM11 (**Fig. 4-A**), and the fragments obtained were subcloned into pHSG398 (Takara) or pBluescriptII SK$^-$ (Stratagene) for further analysis. The 3.6 kb *Kpn*I-*Eco*RI fragment containing the *DFR* gene was sequenced by the chain termination method [7], and the result was shown in **Fig. 2**. By comparing the *DFR* cDNA sequence, we have ascertained that the *DFR* gene of the *japonica* rice *Murasaki-ine* is comprised of three exons and two introns (**Fig. 1**).

We also compared the genomic sequences of the 2,790 bp region between the *japonica* variety *Murasaki-ine* and the *indica* rice *Teqing*, since the 3' end of the published sequence of *Teqing* corresponds to the position 2,790 in **Fig. 2** [4]. The sequences of these varieties showed 98.3% identity, and only 18 single base-pair substitutions and 3 tiny insertions (1 bp, 11 bp and 17 bp) were detected in the 2,790 bp region of *Murasaki-ine*. Interestingly, 11 out of 18 base-pair substitutions were localized at the 300 bp segment flanked by the two insertions (11 bp and 17 bp) in the promoter region.

We have also cloned around 10 kb *Eco*RI-*Eco*RI fragments containing the *DFR* gene regions from the genomes of *Toride 1*, *Koshihikari* and *Kuro-gome*, and constructed the restriction cleavage maps of these fragments (**Fig. 4-A**). The restriction map of the *indica* variety *Kuro-gome* is identical to that deduced from the genomic sequence of the *indica* variety *Teqing* [4]. Clear RFLPs were detected in the cloned *DFR* gene regions among these four rice varieties. Some of these RFLPs appear to be due to presence or absence of small insertions. RFLPs due to such insertions were also detected at the *DFR* gene regions in different lines of the common morning glory [11]. Since small insertions are often due to the presence of interspersed repetitive sequences [12], we analyzed whether the regions containing RFLPs due to presence of small insertions carry repetitive DNA elements. As **Fig. 4-B** shows, sequences detected by the probe B are unique in the genomes of *Kuro-gome*, *Murasaki-ine* and *Koshihikari* whereas those by the probe A are present in multiple copies in these genomes. The same conclusions were also drawn from DNA gel blot analysis using several other restriction enzymes (data not shown). **Fig. 4-B** also indicated that the *DFR* regions detected by the probe C are unique in the genomes of *Kuro-gome*, *Murasaki-ine*, *Koshihikari*, and *Toride 1*.

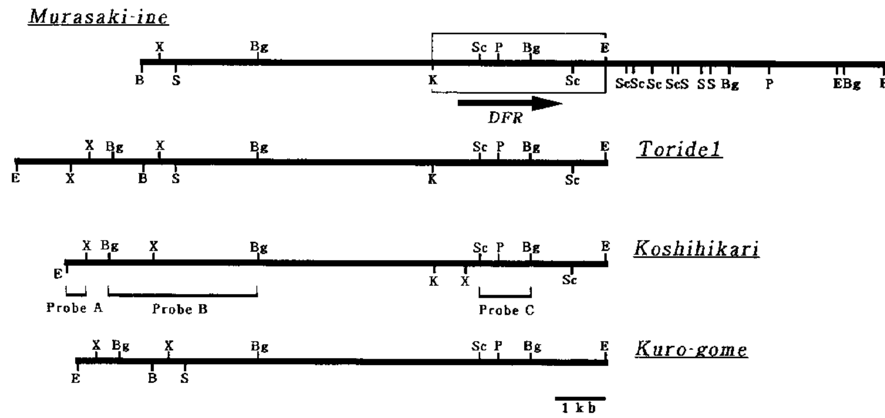
We have also sequenced an approximately 1.2 kb region containing exon 1, intron 1 and 5′ part of exon 2 in *Koshihikari* (accession number AB010744) and *Toride 1* (accession number AB010745). In these cultivars with green leaves and stems, either the sequences GAG at the exon 1 in *Koshihikari* or TCG at the exon 2 in *Toride 1* has changed into the stop codon TGA (**Fig. 2**). We do not know whether these mutations of the two varieties are solely responsible for deficiency in anthocyanin pigmentation. However, it is clear that the *DFR* mRNA was produced in these cultivars with green leaves and stems as well as the *indica* cultivar, *Kuro-gome*, with pigmented stems and leaves (**Fig. 3**).

The *A1* gene encoding the DFR enzyme has been mapped at 198 cM (the near the terminus of the long arm) on the maize chromosome III in 1997 [13]. Comparative linkage maps of the rice and maize indicated that the 140-170 cM region on the maize chromosome III corresponds to the 80-170 cM area of the rice chromosome I [14]. Among the several loci affecting anthocyanin pigmentation and mapped on the rice chromosomes, the *A* locus of rice is assigned at approximately 93 cM on the rice chromosome I [15]. Whether the *A* locus has any association with the *DFR* gene described here remains to be investigated.

Acknowledgments

We thank Udo Wienand for providing the maize *a1* cDNA, Takuji Sasaki, Masahiro Yano, Ayahiko Shomura, Kiichi Fukui, Nobuko Ohmido and Fumiyoshi Myouga for valuable discussions and sharing unpublished data, Tomoko Endo for helping to prepare cDNA library partly, and Rie Terada for reading the manuscript. This work was supported in part by grants from the Ministry of Education, Science, and Culture of Japan.

(A)



(B)

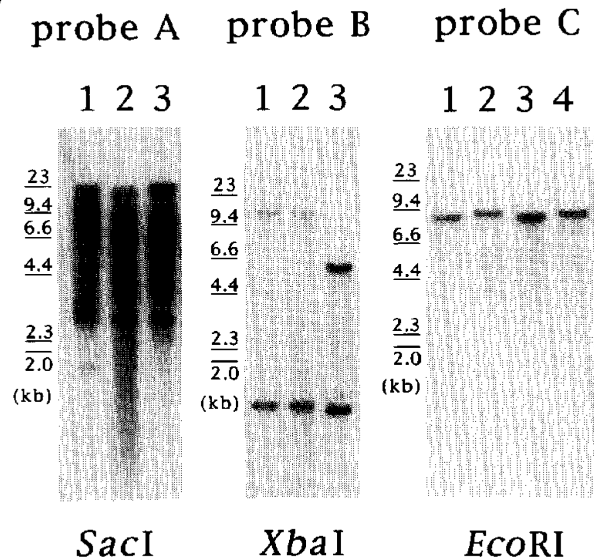


Fig. 4 Genomic structures of the *DFR* gene regions in rice varieties.

(**A**) Restriction maps.

The open box represents the *Kpn*I-*Eco*RI segment shown in **Fig. 2**. The thick horizontal arrow indicates the location and orientation of the *DFR* gene. Restriction sites present in all the cloned fragments are represented by the vertical lines above the thick horizontal lines and restriction sites unique to each fragment are below the thick lines. Restriction sites are: B, *Bam*HI ; Bg, *Bgl*II ; E, *Eco*RI; K, *Kpn*I ; P, *Pst*I ; S, *Sal*I ; Sc, *Sac*I ; X, *Xba*I . The segments used for probes in DNA gel blot analysis are indicated under the maps.

(**B**) DNA gel blot analysis.

Genomic DNAs (10μg) from young leaves were digested with the restriction enzymes indicated and hybridized with the probes A, B and C indicated above [7, 10]. Lane 1, *Kuro-gome*; lane 2, *Murasaki-ine*; lane 3, *Koshihikari*; lane 4, *Toride 1*.

References

[1] Forkmann, G., 1993. In "The Flavonoids, Genetics of flavonoids", (ed. Harborne, J. B.), p. 536-564. Chapman & Hall, London.

[2] Martin, C., Gerats, T, 1993. In "The molecular biology of flowering, The control of flower coloration.", (ed. by Jordan, B. R.), p. 219-255, CABI, Wallingford, UK.

[3] Holton, T. A., Cornish, E. C., 1995. Plant Cell, **7**: 1071-1083.

[4] Chen, M., Bennetzen J. L., 1996. Plant Mol. Biol. **32**: 999-1001.

[5] Nagao, S., Takahashi, M., Kinoshita, T., 1962. J. Facul. Agr., Hokkaido Univ., Sapporo, **52**: 20-50.

[6] van Tunen, A. J., Koes, R. K., Spelt, C. E., van der Krol, A. R., Stuitje, A. R., Mol, J. N. M., 1988. EMBO J. **7**, 1257-1263.

[7] Sambrook, J., Fritsch, E. F., Maniatis, T., 1989. "Molecular Cloning: A Laboratory Manual, 2nd ed.", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

[8] Schwarz-Sommer, Z., Shepherd, N., Tacke, E., Gierl, A., Rohde, W., Heclercq, L., Mattes, M., Berndtgen, R., Peterson, P. A., Saedler, H., 1987. EMBO J. **6**: 287-294.

[9] Murray, M. G., Thompson, W. F., 1980. Nucl. Acids Res., 8: 4321-4325.

[10] Inagaki, Y., Hisatomi, Y., Suzuki, T., Kasahara, K., Iida, S., 1994. Plant Cell **6**: 375-383.

[11] Hisatomi, Y., Yoneda, Y., Kasahara, K., Inagaki, Y., Iida, S., 1997. Theor, Appl. Genet., **95**: 509-515.

[12] Weising, K., Nybom, H., Wolff, K., Meyer, W., 1995. "DNA Fingerprinting in Plants and Fungi", CRC Press, Boca Raton.

[13] Neuffer, M. G., Coe, E. H., Wessler, S., 1997. In "Mutants of Maize", p. 38-39, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

[14] Ahn, S., Tanksley, S. D., 1993. Proc. Natl. Acad. Sci. USA, **90**: 7980-7984.

[15] Kinoshita, T., 1995. Rice Genet. Newslett., **12**: 9-153.

[16] Shirley, B. W., Hanley, S., Goodman, H. M., 1992. Plant Cell, **4**: 333-347.

[17] Kristiansen, K. N., Rohde, W., 1991. Mol. Gen. Genet., **230**: 49-59.

[18] Beld, M., Martin, C., Huits, H., Stuitje, A. S., Gerats, A. G. M., 1989. Plant Mol. Biol., **13**: 491-502.

[19] Uchimiya, H., Kidou, S., Shimazaki, T., Aotsuka, S., Takamatsu, S., Nishi, R., Hashimoto, H., Matsubayashi, Y., Kidou, N., Umeda, M., Kato, A., 1992. Plant J., **2**: 1005-1009.