Gene-associated CpG Islands and the Expression Pattern of Genes in Rice

Ikuo Ashikawa*

National Agricultural Research Center, Inada, Joetsu, Niigata 943-0193, Japan

(Received 30 May 2002; revised 18 July 2002)

Abstract

In an attempt to understand the role of gene-associated CpG islands in the expression of plant genes, I determined the position of CpG islands within their associated genes and the expression of the genes in rice tissues. I examined the expression patterns of 75 rice genes by Northern hybridization analysis using RNAs isolated from four rice tissues: leaf, root, callus, and panicle at flowering stage. From the results of this analysis, I classified most of the genes into one of two groups: expression in a single tissue and expression in two or more tissues. There was a marked correlation between the expression of a gene in two or more tissues and the presence of a CpG island in its 5'-end (class 1 CpG island). Among the genes expressed in a single tissue, the genes expressed in callus were distinct from those expressed in other tissues in that a large proportion contained a class 1 CpG island. These results suggest that plant CpG islands may be useful for deducing the expression pattern of uncharacterized genes.

Key words: CpG island; genome; tissue-specific gene expression

CpG islands are discrete DNA regions in the genome in which an unmethylated CpG dinucleotide frequently occurs. In contrast, in other DNA regions the dinucleotide CpG occurs at a very low frequency. Until quite recently, knowledge of CpG islands has been obtained mostly from analyses of the human genome.1–4 Thanks to recent developments in plant genome research, a large amount of plant DNA sequence information, indispensable for studying CpG islands, has been accumulated.5 Analysis of these DNA sequences has revealed that plants also contain unmethylated CpG-rich clusters in their genomes.6 One of the characteristics of human and plant CpG islands is their close association with genes; about 60% of human5,7,8 and Arabidopsis thaliana and rice CpG islands were associated with genes.6 The position of a CpG island within its associated gene varies. In some genes, an island occurs around the 5'-end of the gene; in other genes, it covers the whole gene region or occurs at a considerable distance downstream of the 5'-end. In humans, the position of a CpG island within its associated gene may be related to the tissue distribution of gene expression. Several researchers reported that almost all genes classified as “housekeeping” enzymes contained a CpG island near the 5'-end, whereas only about 25% of genes with a limited expression pattern did.2,3,9 This kind of information could be useful for deducing the expression pattern of uncharacterized genes in the genome. Gardiner-Garden and Frommer10 and Ashikawa6 first reported such a relationship in plants. To evaluate the tissue specificity of gene expression, I previously used in silico expression analysis6 in which a protein-coding sequence in the rice genome was compared to a database of ESTs (Expressed Sequence Tags) representing four rice tissues: shoot, root, callus and panicles.11 Where a matched EST was identified, the rice tissues the gene was expressed in were known because the name of each catalogued rice EST carries information about its expressed tissue. One of the limitations of the previous study6 is that the accuracy of information obtained in this way is highly dependent on the quality of the cDNA libraries used for EST cataloguing and on the number of genes sequenced in each library. In the present study, using Northern hybridization analysis I obtained direct information about the tissue-specific expression pattern of rice genes, and tested whether the position of a plant CpG island within its associated gene is really related to the tissue distribution of gene expression.

I analyzed the tissue-specific expression and the associated CpG island of each of 75 rice genes that are present in rice chromosome 1. In this analysis, I selected only genes for which partial sequences were identified in the rice EST catalogue. Thus, all of these 75 genes can be regarded as expressed genes. The genome sequences were screened for CpG islands, and the positions...
of CpG islands in their associated genes were examined as described. For purposes of this study, I defined a CpG island as a CpG-rich region larger than 100 bp in length with a moving average CpG frequency of > 0.09. Information on the location of the exons and introns for each gene in the genomic sequences came from public databases. Tissue-specific expression of each gene was examined using RNAs isolated from 4 tissues of the rice cultivar Nipponbare, whose genomic sequence was used for screening for CpG islands.

Among these 75 genes, I obtained at least one discernible Northern fragment on the rice RNA blots for 51 genes (http://cse.naro.affrc.go.jp/ashikawa/Expression.html). These 51 genes could be divided into two groups on the basis of their tissue-specificity of gene expression: 25 genes were expressed in only one of the four tissues (tissue-specific genes), while the remaining 26 genes were expressed in two or more tissues (widely expressed genes). Among the 26 widely expressed genes, 11 were detected in two tissues, 7 in three tissues and 8 in all four tissues. In most of these widely expressed genes (92%), one of the expressed tissues was callus. Although a large proportion of the genes examined exhibited the same expression patterns as those inferred from my previous in silico analysis, the remaining 19 genes (37%) showed different patterns: in the in silico analysis, 15 genes (C62458, S10696, C52296, S20978, E50136, E10840, R844, C10106, S20025, E50230, C1560, S5588, C62003, E11049, C668) were regarded as tissue-specific, whereas Northern hybridization analysis showed that they were expressed in two or more tissues. Conversely, four genes (S20908, E20660, R3292, C30430), previously classified as widely expressed, were found to be expressed mainly in single tissues and thus were grouped with the tissue-specific genes. This result indicates that information obtained from my previous in silico analysis was not so accurate and that the relationship between the tissue distribution of gene expression and CpG islands must be re-examined using information from this Northern analysis.

As mentioned above, the position of a CpG island within its associated gene is variable. In the previous paper, I divided genes into 5 classes on the basis of position. In class 1 genes, a CpG island is located near the 5′-end of the associated gene. In class 2 genes, a CpG island covers the whole gene region. In class 3 genes, a CpG island occurs considerably downstream of the 5′-end. Class 4 genes contain a CpG island near the 5′-end of the associated gene and another island downstream of the 5′-end. Class 5 genes lack CpG islands. In this study, I examined the distribution of the 75 rice genes among the 5 classes. About half were grouped into class 1, and most of the rest belonged to class 2 or 5 (Fig. 1A). Among the 75 genes, I selected 51 that were expressed in at least one of the four tissues. In this selection, the proportions of genes belonging to class 2 or 5 were slightly lower (Fig. 1B). For each of these 51 genes, I compared the position of its gene-associated CpG island with the tissue distribution of the gene. Figure 2 shows two examples...
Figure 2. Expression patterns of 2 rice genes and positions of their associated CpG islands. Upper: Northern hybridization patterns obtained with (A) probe E10840 cDNA and (B) probe C113. RNAs for these analyses were isolated from leaf, root, callus and panicle at flowering stage by the method of Chang et al.\textsuperscript{12} RNA from each tissue was separated on a 1.2% agarose gel containing 5% formaldehyde, blotted onto a positively charged nylon membrane, and hybridized with cDNA probes for each gene sequences around which were used for characterizing the associated CpG islands. To detect a Northern hybridization signal, I used a North2South labelling and detection kit (Pierce, Rockford, IL, USA). Rice callus was induced from Nipponbare seeds by the method of Toki.\textsuperscript{13} Lower: CpG frequency plots in the gene regions of E10840 and C113. The horizontal arrows under the CpG frequency plots indicate the position of each gene. (C) RNA loading was monitored by ethidium bromide staining.

of these comparisons. Northern hybridization analysis showed that gene E10840 was expressed mainly in two tissues: callus and panicle (Fig. 2A). The distribution pattern of CpG dinucleotides in the genome region around the gene showed a CpG-rich cluster over the 5'-end of the gene; hence, the gene was classified into class 1. On the other hand, gene C113 was expressed mainly in callus (Fig. 2B). This gene had no CpG-rich cluster around it; hence, it was classified into class 5. The results of these comparisons are shown in Fig. 1C, D, which shows
the distribution of the positions of CpG islands in the tissue-specific and widely expressed genes. These figures indicate a strong positive correlation between the presence of a CpG island in a class 1 gene and tissue distribution of gene expression: from 24% in the tissue-specific genes to 81% in the widely expressed genes.

In my previous study, I did not find any apparent relationship between the class to which a tissue-specific gene belonged and the tissue in which it was expressed. In contrast, in the present study, I found a relationship between the expressed tissues and the location of their associated CpG islands for the tissue-specific genes; among the genes expressed in callus, about half belonged to class 1 (Fig. 1E), which suggests that the genes expressed in callus are predominantly class 1 genes. Conversely, most of the genes expressed in shoot, root or panicle were grouped into class 2 or 5 (Fig. 1F). A similar situation might exist in humans: the genes expressed in early embryos were different from the genes expressed in differentiated tissues, in that most embryonic genes had a CpG island over their transcription start site.9

The results from this study suggest that genes functionally involved in the maintenance of a single cell, for example genes used for cell growth or cell division, rather than those involved in differentiation and development of tissues, might tend to have CpG islands over their 5′-ends. In my previous report, I suggested that CpG islands were useful landmarks in genome sequences for identifying unknown genes. The present results suggest that analysis of CpG islands would also be useful for evaluating the tissue distribution or tissue-specific expression of uncharacterized genes.

Acknowledgements: I thank Dr. T. Hirose at the National Agricultural Research Center in Japan for his help in the isolation of rice RNAs.

References

1. Bird, A. P. 1987, CpG islands as gene markers in vertebrate nucleus, Trends Genet., 3, 342–347.
2. Gardiner-Garden, M. and Frommer, M. 1987, CpG islands in vertebrate genomes, J. Mol. Biol., 196, 261–282.
3. Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. 1992, CpG islands as gene markers in the human genome, Genomics, 13, 1095–1107.
4. Cross, S. H. and Bird, A. P. 1995, CpG islands and genes, Curr. Opin. Genet. Dev., 5, 309–314.
5. Arabidopsis Genome Initiative 2000, Analysis of the genome sequence of the flowering plant Arabidopsis thaliana, Nature, 408, 796–815.
6. Ashikawa, I. 2001, Gene-associated CpG islands in plants as revealed by analyses of genomic sequences, Plant J., 26, 617–625.
7. Dunham, I., Shimizu, N., Roe, B. A. et al. 1999, The DNA sequence of human chromosome 22, Nature, 402, 489–495.
8. Ioshikhes, I. P. and Zhang, M. Q. 2000, Large-scale human promoter mapping using CpG islands, Nature Genet., 26, 61–63.
9. Ponger, L., Duret, L., and Mouchiroud, D. 2001, Determinants of CpG islands: expression in early embryo and isochore structure, Genome Res., 11, 1854–1860.
10. Gardiner-Garden, M. and Frommer, M. 1992, Significant CpG-rich regions in angiosperm genes, J. Mol. Evol., 34, 231–245.
11. Yamamoto, K. and Sasaki, T. 1997, Large-scale EST sequencing in rice, Plant Mol. Biol., 35, 135–144.
12. Chang, S., Puryear, J., and Cairney, J. 1993, A simple and efficient method for isolating RNA from pine trees, Plant Mol. Biol. Rep., 11, 113–116.
13. Toki, S. 1997, Rapid and efficient agrobacterium-mediated transformation in rice, Plant Mol. Biol. Rep., 15, 159–164.