Complete Nucleotide Sequence of the Sugarcane (Saccharum Officinarum) Chloroplast Genome: A Comparative Analysis of Four Monocot Chloroplast Genomes

Takayuki ASANO,1 Takahiko TSUDZUKI,2 Sakiko TAKAHASHI,1 Hiroaki SHIMADA,3 and Koh-ichi KADOWAKI1∗

National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan,1 Department of Information and Policy Studies, Aichi-Gakuin University, Araike 12, Nisshin, 470-0195, Japan,2 and Department of Biological Science and Technology, Tokyo University of Science, Yamazaki 2641, Noda, Chiba 278-8510, Japan3

(Received 26 January 2004; revised 9 February 2004)

Abstract

The complete nucleotide sequence of the chloroplast genome of sugarcane (Saccharum officinarum) has been determined. It is a circular double-stranded DNA molecule, 141,182 bp in size, and is composed of a large single copy of 83,048 bp, a small single copy of 12,544 bp, and a pair of inverted repeat regions of 22,795 bp each. A comparative analysis among monocots showed that the sugarcane chloroplast genome was very similar to maize but not to rice or wheat. Between sugarcane and maize at the rps16–trnQ (UUG) region, however, a length polymorphism was identified. With regard to insertions/deletions equal to or longer than 5 bp, a total of 53 insertion and 31 deletion events were identified in the sugarcane chloroplast genome. Of the 84 loci identified, a pair of direct repeat sequences was located side by side in a tandem fashion in 47 loci (56.0%). A recombination event during plant evolution is discussed at two sites between the sugarcane and tobacco chloroplast genomes.

Key words: Poaceae; sugarcane; chloroplast genome; structural changes

1. Introduction

Sugarcane belongs to the grass family (Poaceae) and has the unique and very useful characteristic of high sugar concentration accumulation. Hence, it is an important crop for sugar production. It is mainly cultivated in tropical and subtropical regions. The genus Saccharum comprises six different species: S. officinarum, S. barberi, S. sinense, S. edule, S. robustum and S. spontaneum. Of these, S. officinarum and S. spontaneum are thought to be the ancestors of cultivated sugarcane. S. officinarum was domesticated in Southeast Asia and originally derived from S. robustum, while S. barberi and S. sinense are thought to have been derived by crossing S. officinarum and S. spontaneum.1,2

The chromosome number of different species of the genus Saccharum ranges from 36 to 170 (S. officinarum, 2n = 70 to 140; S. barberi, 2n = 60 to 140; S. sinense, 2n = 104 to 124; S. edule, 2n = 60 to 122; S. robustum, 2n = 66 to 170; S. spontaneum, 2n = 36 to 128).3 The basic chromosome number (x = 10 or 8) of S. officinarum is still controversial.4,5 The polyploid nature and large variation of chromosome number in the genus Saccharum has made its phylogenetic analysis more difficult. Despite its agricultural importance, an understanding of the evolutionary relationship among species at the molecular level is limited. Hence, further analysis is required.

In addition to the nuclear genome, plants have mitochondrial and plastid genomes. Chloroplasts are an important apparatus for photosynthesis and entire chloroplast genomes have been sequenced in a variety of plants that include monocots, dicots, gymnosperms,6 psilotophyta,7,8 bryophytes and algae.9 Among the monocots and dicots, entire sequences are available for Nicotiana tabacum,10 Arabidopsis thaliana,11 Oenothera elata,12 Spinacia oleracea,13 Lotus japonicus,14 Atropa belladonna,15 Oryza sativa,16 Triticum aestivum17 and Zea mays.18 The chloroplast DNA of most land plants is a circular double-stranded molecule that ranges in size from 120 to 160 kb. It is composed of two identical inverted repeats (IRs) that are separated by a large and
small single copy region (LSC and SSC) (see ref. 19 for a review).

To date, complete chloroplast sequences in agriculturally important monocot crops are available only for wheat, rice and maize. As such, chloroplast genome information from sugarcane, which is also agriculturally important, has not been determined. We report here the entire sequence of the *S. officinarum* chloroplast genome and discuss its features in comparison with other monocots. The sugarcane chloroplast genome is the fourth grass plant to be completely sequenced.

2. Materials and Methods

Sugarcane (*S. officinarum* L. cv. NCo310) was grown in a greenhouse at 28°C. Genomic DNA was extracted from fresh green tissues using the procedure described by Kadowaki et al.20 and was used as a template for PCR amplification. Chloroplast DNA was amplified using one unit of LA-Taq polymerase with appropriate primers (Takara, Japan). The PCR products were purified using the QIAquick PCR purification kit (Qiagen, Germany) and subsequently used for DNA sequencing. The nucleotide sequence of the chloroplasts was determined using the CEQ Dye Terminator Cycle Sequencing Quick Start kit (Beckman Coulter, USA). The sequences were evaluated and assembled using Sequencer 3.0 software (Gene Codes Corporation, USA). A database search was undertaken using the BLAST algorithm at the National Center for Biotechnology Information. A comparative analysis of the chloroplast genome among cereals was carried out according to the program of Tsuzuki et al.21 A Harr plot analysis was carried out using GENETYX software (GENETYX, Japan).

3. Results and Discussion

3.1. Determination of the sugarcane chloroplast genome: structure and gene content

The entire chloroplast genome of sugarcane was a circular double-stranded DNA molecule of 141,182 bp. It was composed of a LSC of 83,048 bp, a SSC of 12,544 bp and a pair of IRs of 22,795 bp each (Fig. 1). The A + T content of the sugarcane chloroplast genome was 61.6%, which is similar to those of rice (61.0%), maize (61.5%) and wheat (61.7%). A total of 108 functional genes were deduced. The position of all genes identified in the sugarcane chloroplast genome is shown in Fig. 1. The number, gene content and order of the functional chloroplast genes are identical to those of rice, maize and wheat.16–18 In addition, 20 open-reading frames with unknown functions were also annotated.

3.2. Comparison of chloroplast genomes among cereals

The overall genomic organization of sugarcane was compared with those of maize, rice and wheat, and the extent of sequence similarity is depicted by white, gray and black colors, respectively, in Fig. 2. Our results indicated that the chloroplast genome of sugarcane was very similar to that of maize but not to rice or wheat. It is worth noting that the gene content and order of genes was very similar among all four species. More polymorphism was observed in the single-copy regions than in the IRs.

The IR region of sugarcane (22,795 bp) and maize (22,748 bp) was larger than that of rice and wheat (20,799 bp and 20,703 bp, respectively). This was because the *trnA* (CAU)–*trnL* (CAA) region of IR from the sugarcane and maize chloroplast genomes had larger insertions, 2,133 bp and 2,130 bp, respectively, than those of rice and wheat. The larger IR region in the sugarcane and maize chloroplast genomes may be attributed to additional sequences (junk DNA).16–18

3.3. Comparative analysis of the region in LSC showing nucleotide diversity

Large nucleotide diversity was observed in two regions of the LSC (Fig. 2): *rps16–*trnQ (UUG) and *trnS* (UGA)–*trnC* (GCA). A Harr plot analysis was conducted to analyze the polymorphism among the four monocots in detail. The results indicate that there are length polymorphisms between sugarcane and other cereals in the intergenic region between *rps16* and *trnQ* (UUG) (Fig. 3a). A number of insertions/deletions, as well as base substitutions, that differed from rice and wheat were observed in the *trnS* (UGA)–*trnC* (GCA) region of sugarcane (Fig. 3b). Furthermore, comparison with the corresponding region in rice and wheat also showed several insertion/deletion events, which suggested that both regions were the potential “hot spots” for genetic diversity.17

An 84-bp sequence that showed a high similarity (87.1% identity) with the first exon of *rpl2* was identified between *trnG* (UCC) and *trnT* (GGU), within the hot spot region. This sequence was truncated, did not encode functional information, and was concluded to be a pseudo *rpl2*. The same sequence was reported to be present in the corresponding region of the maize and wheat chloroplast genomes.22,23 However, the *rpl2*-related sequence was not identified in tobacco or rice chloroplast genomes. Therefore, it is speculated that the *rpl2*-related sequence was introduced into the prototype of the monocot chloroplast genome and, subsequently, the *rpl2*-related sequence was lost in the region of the rice chloroplast genome during evolution.
3.4. Detailed analysis of insertion/deletion events between sugarcane and maize

Although the overall genomic structures of sugarcane and maize were very similar, a detailed analysis revealed a reasonable number of insertion/deletion events in sugarcane that differed from maize. The insertion/deletion events equal to or longer than 5 bp were taken into consideration in this investigation. A total of 53 insertion and 31 deletion events were identified (Fig. 2). Most of the insertion/deletion loci corresponded to places that showed low similarity with the rice and wheat chloroplast genomes. It was apparent that the insertion/deletion events were not evenly distributed throughout the genomes. Furthermore, most of the insertion/deletion events were located in the intergenic region [45 insertions (84.9%) and 23 deletions (74.2%)]. Regarding the intron region, the intron of the rpl16 (three sites), rps16 (two sites), trnL (UAA) (one site) and ycf3 (two sites) genes showed length polymorphism between the sugarcane and maize chloroplast genomes. Regarding the coding region, it is worth noting that five insertion (rpoC2, rbcL, orf81, orf251; IRa, orf251; IRb) and three deletion (rps18, orf251; IRa, orf251; IRb) deletion events occurred in the protein-coding region. A 6-bp insertion was present in rbcL, whereas another two insertions/deletions of 21 bp were observed in each of rpoC2 and rps18. The reading frames of these genes did not shift despite the insertion/deletion events.
Figure 2. Comparative analysis of the chloroplast genomes of four monocots. Sequences from each plant are aligned in the vertical figure. Each plant name is shown above the figure. Sequence identity is: black boxes, 0–30%; gray boxes, 31–79%; and white boxes, 80–100%. Genes shown on the left side are transcribed sense strands (A-chain), while those on the right side are transcribed antisense strands (B-chain). Intron-containing genes are indicated by white boxes. Insertion/deletion events compared with the maize chloroplast genome are indicated by bars and triangles at the A-chain (short bar, 5 bp to 9 bp; intermediate bars, 10 bp to 19 bp; long bars, >19 bp. Black triangles, deletion events; white triangles, insertion events).
3.5. Mechanism of structural alterations

In order to understand the genetic process of insertion/deletion events during evolution, a computer search was conducted. The involvement of repeat sequences in the insertion/deletion event has been reported in the tobacco and *Atropa* chloroplast genomes. Of the 84 loci identified with insertion/deletion events, a pair of direct repeat sequences was located side by side in a tandem fashion in 47 loci (56.0%). One copy or partial sequence of the direct repeat was found to be deleted or inserted at all sites (data not shown).

Events of genetic recombination were identified between sugarcane and tobacco. A detailed analysis of two recombination events was conducted between the sugarcane and tobacco chloroplast genomes (Fig. 4). In the intergenic region of *trnL* (CAA)—*ndhB*, a pair of inverted repeats of 14 bp was observed in both plants. Furthermore, a 21-base nucleotide between the inverted repeat of sugarcane, 5′-TGATGATCGAGTCGATTCCAT-3′, was inverted to 5′-ATGGAATCGACTCGATCATCA-3′ in tobacco (Fig. 4a). This type of flip-flop recombination is known to occur through inverted repeats. Therefore, the recombination event might have taken place through the pair of inverted repeats.

In the intergenic region of 4.5S and 5S ribosomal RNA genes from the sugarcane chloroplast genome, a pair of inverted repeats of 9 bp was also found (Fig. 4b). It is likely that a sequence between the inverted repeats was itself inverted because a 26-base nucleotide of sugarcane, 5′-TCCATCTCTTGGATAAATAGAGAGGG-3′, was inverted to 5′-CCCTCTCTATCTATCCAAGGGATGGA-3′ in tobacco with a substitution at two positions. In tobacco, two copies of the 32-bp sequence were located in a tandem fashion just upstream of the inverted repeat sequences. Because one copy of the same size (32 bp) is present in the chloroplast genomes of monocots *L. japonicus* (dicot), liverwort, and hornwort, the chloroplast genome with one copy of the 32-bp sequence appears to be an ancestral form. Hence, a duplication event of the 32-bp sequence might have occurred in the tobacco
chloroplast genome. The sequences of sugarcane illustrated in Fig. 4 were highly conserved in monocot chloroplast genomes.

In summary, the complete chloroplast sequence of sugarcane was determined and a genome-wide analysis between sugarcane and maize showed that insertion/deletion events were not evenly distributed throughout the genomes. Most insertion/deletion events between sugarcane and maize were observed at places where direct repeat sequences were located in a tandem fashion. We identified some polymorphic regions, although the chloroplast genome of sugarcane was very similar to that of maize. The information on genetic diversity obtained from this study will be useful for taxonomic analyses of the genus *Saccharum* and its related species, as well as for future chloroplast engineering.

**Acknowledgements:** We express our thanks to Prof. J. Obokata and Drs. M. Shahid, T. Nishikawa, A. Sugimoto and A. Kaga for helpful discussions, comments and advice. Sugarcane was kindly provided by the domed greenhouse of NIAS.

**References**

1. Daniels, J. and Roach, B. T. 1987, In: Heinz, D. J. (ed) Taxonomy and evolution, Elsevier, Amsterdam, pp. 7–84.
2. Berding, N. and Roach, B. T. 1987, In: Heinz, D. J. (ed) Germplasm collection, maintenance, and use, Elsevier, Amsterdam, pp. 143–210.
3. Irvine, J. E. 1999, *Saccharum* species as horticultural classes, *Theor. Appl. Genet.*, 98, 186–194.
4. D’Hont, A., Ison, D., Alix, K., Roux, C., and Glaszmann, J. C. 1998, Determination of basic chromosome numbers
in the genus *Saccharum* by physical mapping of ribosomal RNA genes, *Genome*, 41, 221–225.

5. Ha, S., Moore, P. H., Heinz, D., Kato, S., Ohmido, N., and Fukui, K. 1999, Quantitative chromosome map of the polyploid *Saccharum spontaneum* by multicolor fluorescence in situ hybridization and imaging methods, *Plant Mol. Biol.*, 39, 1165–1173.

6. Wakasugi, T., Tsuzuki, J., Ito, S., Nakashima, K., Tsuzuki, T., and Sugimura, M. 1994, Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*, *Proc. Natl. Acad. Sci. U.S.A.*, 91, 9794–9798.

7. Ohyama, K., Fukuzawa, T., Kohchi, T. et al. 1986, Chloroplast Gene Organization Deduced from Complete Sequence of Liverwort *Marchantia polymorpha* Chloroplast DNA, *Nature (London)*, 322, 572–574.

8. Kugita, M., Kaneko, A., Yamamoto, Y., Takeya, Y., Matsumoto, T., and Yoshinaga, K. 2003, The complete nucleotide sequence of the hornwort (*Anthoceros foemosae*) chloroplast genome: insight into the earliest land plants, *Nucleic Acids Res.*, 31, 716–721.

9. De Las Rivas, J., Lozano, J. J., and Ortiz, A. R. 2002, Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns, *Genome Res.*, 12, 567–583.

10. Shinozaki, K., Ohme, M., Tanaka, M. et al. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression, *EMBO J.*, 5, 2043–2049.

11. Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. 1999, Complete structure of the chloroplast genome of *Arabidopsis thaliana*, *DNA Res.*, 6, 283–290.

12. Hupfer, H., Swiatek, M., Hornung, S., Herrmann, R. G., Maier, R. M., Chiu, W. L., and Sears, B. 2000, Complete nucleotide sequence of the eukaryotic plastid chromosome, representing plastome I of the five distinguishable eu-nuclear plastomes, *Mol. Gen. Genet.*, 263, 581–585.

13. Schmitz-Linneweber C., Maier, R. M., Alcaraz J. P., Cottet A., Herrmann R. G., and Mache R. 2001, The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization, *Plant Mol. Biol.*, 45, 307–315.

14. Kato, T., Kaneko, T., Sato, S., Nakamura, Y., and Tabata, S. 2000, Complete structure of the chloroplast genome of a legume, *Lotus japonicus*, *DNA Res.*, 7, 323–330.

15. Schmitz-Linneweber, C., Regel, R., Du, T. G., Hupfer, H., Herrmann, R. G., and Maier, R. M. 2002, The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation, *Mol. Biol. Evol.*, 19, 1602–1612.

16. Hiratsuka, J., Shimada, H., Whittier, R. et al. 1989, The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals, *Mol. Gen. Genet.*, 217, 185–194.

17. Ogihara, Y., Isono, K., Koijima, T. et al. 2002, Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA, *Mol. Genet. Genomics*, 266, 740–746.

18. Maier, R. M., Neckermann, K., Igloi, G. L., and Kossel, H. 1995, Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing, *J. Mol. Biol.*, 251, 614–628.

19. Sugiura, M. 1992, The chloroplast genome, *Plant Mol. Biol.*, 19, 149–168.

20. Kadowai, K., Kubo, N., Ozawa, K., and Hirai, A. 1996, Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals, *EMBO J.*, 15, 6652–6661.

21. Tsuzuki, J., Tsuzuki, T., Wakasugi, T., Kinoshita, K., Konodo, T., Ito, Y., Sugiura, M. (2003), In: Sugiura, H. (ed) Comparative analysis of the whole chloroplast genomes from rice, maize and wheat. In ICES8 Proceedings, Logos Verlag, Berlin.

22. Shimada, H. and Sugiura, M. 1989, Pseudogenes and short repeated sequences in the rice chloroplast genome, *Curr. Genet.*, 16, 293–301.

23. Bowman, C. M., Barker, R. F., and Dyer, T. A. 1988, In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination, *Curr. Genet.*, 14, 127–136.

24. Linne von Berg, K. H. and Kowallik, K. V. 1992, Structural organization of the chloroplast genome of the chromophytic alga *Vaucheria bursata*, *Plant Mol. Biol.*, 18, 83–95.