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
The number, gene content and order of the functional genes were deduced. The position of all genes identified (61.5%) and wheat (61.7%). A total of 108 functional genes were identified, which is similar to those of rice (61.0%), maize (61.6%), and wheat (61.7%).

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 trnL (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).

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 trnL (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).
Figure 1. Gene organization of the *S. officinarum* chloroplast genome. Genes shown outside the circle are transcribed clockwise, while those located inside are transcribed counterclockwise. Intron-containing genes are indicated by asterisks and white boxes. Red boxes, ribosomal RNA genes; black boxes, transfer RNA genes; orange boxes, ribosomal protein genes; brown boxes, transcription/translation component genes; green boxes, photosynthetic apparatus genes; yellowish-green boxes, ATP synthase genes; yellow boxes, NADH dehydrogenase genes; light green boxes, deduced ORFs.

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′-TCCATCTCTTTGATAAATAGAGGG-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.
Figure 4. Schematic representation of models of conformational change in the two intergenic regions of the chloroplast genome. (a) Intergenic region of trnL (CAA) and ndhB. (b) Intergenic region of the 4.5S and 5S ribosomal RNA genes. The pair of inverted repeat sequences is shown by boxes. Direct repeat sequences are underlined. Arrowheads show the putative excision points. Bold letters show the putative inserted or substituted sequence in the Sugarcane chloroplast genome. Arrows indicate the possible process of DNA alteration during evolution.

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

in the genus *Saccharum* by physical mapping of ribosomal RNA genes, *Genome*, **41**, 221–225.


