Theor Appl Genet (2005) 110: 1523–1529
DOI 10.1007/s00122-005-1990-z

ORIGINAL PAPER

S. Takahashi · T. Furukawa · T. Asano · Y. Terajima
H. Shimada · A. Sugimoto · K. Kadowaki

Very close relationship of the chloroplast genomes among *Saccharum* species

Received: 21 December 2004 / Accepted: 7 March 2005 / Published online: 8 April 2005

Abstract We recently determined the complete sequence of the sugarcane chloroplast genome. Here, we have used the information for a comprehensive phylogenetic analysis of the genus *Saccharum*, using all six species (13 accessions). The polymorphisms between sugarcane and maize in 26 chloroplast genome regions were used for the analysis. In 18 of the 26 regions (a total of 5,381 bp), we found 41 mutations involving 17 substitutions, three inversions, six insertion/deletion mutations, and 15 simple sequence repeat length polymorphisms. Based on these results, we calculated a phylogenetic tree of the genus *Saccharum*, in which all six species are clearly separated. By the analysis, (1) *S. sinense* and *S. barberi*, which have identical sequences, belong to the same clade, whereas the other four species, *S. officinarum*, *S. robustum*, *S. edule*, and *S. spontaneum*, form an independent clade; (2) *S. spontaneum* has a paraphyletic relationship with the other five species; and (3) no or very low intraspecific variation was observed in *S. officinarum*, *S. robustum*, *S. sinense*, *S. barberi*, and *S. edule*, whereas higher intraspecific variation was observed in *S. spontaneum*. Based on the number of nucleotide substitutions, the divergence time between *S. officinarum* and *S. spontaneum*, and between *S. officinarum* and maize were calculated to be about 730–780 thousand years ago and about 5.9 million years ago, respectively. These results suggest that the cytoplasm of *Saccharum* species are very closely related.

Communicated by R. Hagemann

S. Takahashi · T. Asano · K. Kadowaki (✉)
Department of Genetic Diversity,
National Institute of Agrobiological Sciences,
Kannondai 2-1-2, Tsukuba Ibaraki, 305-8602, Japan
E-mail: kadowaki@affrc.go.jp
Tel.: +81-29-8387449
Fax: +81-29-8387408

T. Furukawa · H. Shimada
Department of Biological Science and Technology,
Tokyo University of Science, Yamazaki 2641,
Noda Chiba, 278-8510, Japan

Y. Terajima · A. Sugimoto
Department of Crop and Food Science,
National Agricultural Research Center
for Kyushu Okinawa Region, Anno 1742-1,
Nishino-omote Kagoshima, 891-3021, Japan

Introduction

Sugarcane, belonging to the family Poaceae and the tribe Andropogoneae, is mainly cultivated in tropical and subtropical regions and contributes 60% of the raw sugar production worldwide. The genus *Saccharum* comprises six species: *S. officinarum*, *S. robustum*, *S. spontaneum*, *S. sinense*, *S. barberi*, and *S. edule*. Of these, *S. robustum* and *S. spontaneum* are recognized as wild species (Guimaraes and Sobral 1998). *S. robustum* and *S. officinarum* are distributed in New Guinea and its surrounding islands. *S. officinarum*, which accumulates large amounts of sugar in its stalks, is thought to have been domesticated from *S. robustum*, because *S. officinarum* has features very similar to those of *S. robustum* except for its sugar and fiber content, and it is always found in places associated with humans (Irvine 1999). *S. spontaneum* is broadly distributed from Japan and New Guinea to the Mediterranean and Africa, with a putative center of origin in India. The remaining three species are thought to have interspecific or intergeneric origins. *S. sinense*, cultivated in China, and *S. barberi*, cultivated in India, are believed to have arisen from natural crosses between *S. officinarum* and *S. spontaneum*. *S. edule*, characterized by its abortive flowers, is cultivated as a traditional vegetable in Melanesia, and is thought to have arisen from intergeneric crosses between *S. officinarum* or *S. robustum* and a related genus (e.g., *Miscanthus*), or derived from *S. robustum* (Daniels and Roach 1987; Guimaraes and Sobral 1998; Irvine 1999).

The genomes of the *Saccharum* species are highly polyploid and aneuploid, and have complex structures. The chromosome numbers of *Saccharum* species vary from $2n = 36$–170. Furthermore, *S. officinarum* and

S. robustum have a basic chromosome number of $x = 10$, whereas the basic chromosome number of *S. spontaneum* is $x = 8$ (Irvine 1999).

The taxonomy of *Saccharum* has been adjusted many times and is still complex and controversial (Irvine 1999). Conventionally, *Saccharum* has been classified on morphological characteristics, chromosome number, and sugar content. Some of the controversy may stem from the use of these traits, which can be difficult to score in polyploid species that can interbreed.

Modern sugarcane cultivars are complex interspecific hybrids involving *S. officinarum*, which contributes high sugar content, and *S. spontaneum* for other desirable traits (Berding and Roach 1987). For sugarcane breeding, it is particularly important to access the diversity existing among species, because major breakthroughs in sugarcane improvement have been achieved by distant hybridizations. Recently, phylogenetic analyses of the genus *Saccharum* have been performed using biochemical and molecular markers, and the genetic relationships among the genus *Saccharum* determined by conventional analyses have been verified. Based on the analysis of several isozymes, Glaszmann et al. (1989) reported that *S. spontaneum* was distinguished from *S. officinarum* and *S. robustum*, whereas the latter two were not differentiated from one another. Using 195 nuclear random amplification of polymorphic DNA (RAPD) markers, Nair et al. (1999) analyzed the phylogeny of all *Saccharum* species (except *S. edule*) and related genera. In their cluster analysis, *S. officinarum*, *S. robustum*, and *S. spontaneum* formed discrete groups, whereas *S. sinense* and *S. barberi* formed a single cluster, and *S. officinarum* was positioned close to *S. robustum* and distant from *S. spontaneum*. Selvi et al. (2003) used 34 maize nuclear simple sequence repeat (SSR) markers for their phylogenetic analysis of all *Saccharum* species (except *S. edule*), commercial sugarcane cultivars, and related genera, and obtained results very similar to those of Nair et al. (1999). Based on data from genomic in situ hybridization (GISH) analysis using *S. officinarum*-specific and *S. spontaneum*-specific probes, D'Hont et al. (2002) showed that *S. sinense* and *S. barberi* were derived from interspecific crosses between *S. officinarum* and *S. spontaneum*. These molecular data roughly support the traditionally accepted genetic relationships among *Saccharum* species. Precise analysis of the relationships of *Saccharum* species using nuclear markers is difficult in general because of their highly polyploid nature.

The chloroplast genomes of most land plants have a simple, unicircular structure and are uniparentally inherited. Understanding the cytoplasmic lineage of *Saccharum* species based on chloroplast genome diversity should facilitate an understanding of the genetic relationships of *Saccharum* species, which have complex nuclear genome structures. So far, there have been two reports of large-scale phylogenetic analysis of the genus *Saccharum*, using chloroplast genome sequences. Sobral et al. (1994) carried out restriction fragment length polymorphism analysis of all six *Saccharum* species and related genera, using 15 restriction enzymes and 12 chloroplast DNA probes that span the complete rice chloroplast genome, and demonstrated that the genus *Narenga*, *Miscanthus*, *Sclerostachya*, and *Saccharum* form a monophyletic group, whereas the genera *Erianthus* and *Eccoilopus* form a separate group. In their analysis, the genus *Saccharum* was separated into *S. spontaneum* and all the other *Saccharum* species by a single-site mutation. Al-Janabi et al. (1994) analyzed the phylogeny of all *Saccharum* species (except *S. barberi*) and related genera by direct comparison of 664 bp sequences from the *rbcL–atpB* region of the chloroplast DNA. Although seven site mutations and 16 insertion/deletion mutations (indels) were informative when comparing the genus *Saccharum* and its related genera, no reliable phylogenetic tree was produced. They found very few informative mutations within the genus *Saccharum*, but did not describe them in detail.

Previous reports suggest that very little diversity exists among the chloroplast genomes of *Saccharum* species. Very recently, we have determined the complete sequence of the sugarcane chloroplast genome. Comparative analysis of sugarcane and three other monocots showed that the chloroplast genome of sugarcane is very similar to those of other monocots, especially maize. On the other hand, there was sufficient diversity between sugarcane and maize for successful phylogenetic analysis (Asano et al. 2004). Here, we selected 26 chloroplast genome regions displaying intensive polymorphisms between sugarcane and maize. Using these regions, we performed phylogenetic analysis of all six *Saccharum* species, and calculated a phylogenetic tree on which all six species are clearly separated. Our results suggest that all *Saccharum* species, except *S. spontaneum*, share *S. officinarum–S. robustum*-type cytoplasm, and that the cytoplasm of all six *Saccharum* species are very closely related. We also discuss the cause of the diversity among the *Saccharum* species.

Materials and methods

Plant materials

S. robustum (two accessions), *S. edule* (two accessions), and *S. sinense* (one accession) were selected from the collections at Okinawa Prefectural Agricultural Experimental Station (OPAES) or the Okinawa Subtropical Station of the Japan International Research Center for Agricultural Sciences (JIRCAS). Another eight *Saccharum* accessions were selected from the collection at the National Agricultural Research Center for the Kyushu Okinawa Region. These 13 *Saccharum* accessions were used for sequence analysis. The whole chloroplast genome sequence of *S. officinarum* cv. NCo310, available in the GenBank/EMBL/DDBJ database (accession number, AP006714), was used as the standard. The whole chloroplast genome sequence of *Zea mays* (accession number, X86563) was used as the outgroup. Species

Table 1 *Saccharum* accessions used in this study

Species	Accession names	Abbreviations in this study	Origin
Saccharum officinarum	Badila[b]	off-Ba	New Guinea
	Fiji40[b]	off-Fi	Fiji
S. robustum	6	rob-6	New Guinea
	16	rob-16	New Guinea
S. sinense	Ooshima	sin-Oo	Japan
	Tekcha	sin-Te	China
S. barberi	Chunnee	bar-Ch	India
	Kewali14	bar-Ke	India
S. edule	IN95-009	edu-IN	New Guinea
	COL/PAPUA N.G/TARC/E12	edu-PN	New Guinea
S. spontaneum	Glagah[b]	spo-Gl	Indonesia
	JW385[b]	spo-Jw	Japan
	SES205A	spo-SE	India
Sugarcane (commercial hybrid)	cv. NCo310	NCo	Breeding
Zea mays[a]	–	ZM	Breeding

[a]Used as an outgroup for the phylogenetic analysis
[b]Used for evaluation of 26 selected chloroplast regions

names, accession names, and the abbreviations used in this study and the origins of the collection are shown in Table 1.

DNA extraction, PCR amplification, and DNA sequencing

Genomic DNA was extracted from fresh green tissues by the method of Murray and Thompson (1980) and used as the template for PCR amplification. Chloroplast DNA fragments of 3.2–12 kb were amplified using one unit of LA *Taq* DNA polymerase (TaKaRa, Japan). The PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Calif., USA), and subsequently sequenced with the appropriate primers (shown in Table 2). The DNA sequence of the amplified product was determined with the BigDye Terminator Cycle Sequencing Kit, version 3.1 (Applied Biosystems, Calif, USA), on an ABI PRISM 310 automated sequencer (Applied Biosystems).

Table 2 Twenty-six chloroplast regions analyzed in this study

Region no.	Position in NCo310	P or N[a]	Accession no.[b]	Sequences of primers (5′–3′)
1	3952–4286	P	AP006879–AP006891[C]	CTGGACGAATTTGTTGCTTC
				GTGCCAATCCAACAAAAGTC
2	6087–6140	P	AP006892–AP006904	ACTGTATAGAAAATGAGACC
3	9459–9792	N	AP006905–AP006908	ACAGACCGAGCAAGTTCA
4	10819–11229	N	AP006909–AP006912	TCCTTACTTTGTATCCGGAG
				ACCAGACAAAACAACAAGCG
5	12999–13392	P	AP006913–AP006925	GCAAGCGTAGTTCAATGTAG
6	19102–19455	P	AP006926–AP006938	GCATCCACTTAATTTCAA
				TGGACTCTAGGGATACTACC
7	20948–21128	P	AP006939–AP006951	CAGTTTCGCTTATTCTCCTC
8	29163–29670	N	AP006952–AP006955	GTCCCATATTCCTTCTCTGA
9	31988–32480	N	AP006956–AP006959	CTGCATCAAAAACTAAATCA
10	35640–36020	N	AP006960–AP006963	TTTCTTCTTTGTTCGTCCTA
				CGATATGAGTGTTCTATATC
11	38642–39208	P	AP006964–AP006976	TGAAGGAAGCTATTCAGGAA
				ACTTTGGGCTATCCGGACAC
12	49126–49176	P	AP006977–AP006989	GACACCCCGCTCGCTTATTG
13	50351–50806	N	AP006990–AP006993	GGCAAGGAATGTCGATTA
14	53110–53247	P	AP006994–AP007006	CTTCACTTTGTCTCACTTTC
15	56696–56908	P	AP007007–AP007019	ACATTAATTTGCTTATCGGC
				CAGCAATCTATGCTTCACAG
16	58970–59114	P	AP007020–AP007032	CATATGCCAGCTCTGACC
17	60884–61362	N	AP007033–AP007036	TGGCCTATTTCTTGCGTGTA
18	63003–63309	P	AP007037–AP007049	AATGCAACGTCAACAGGT
19	65488–66202	P	AP007050–AP007062	AATCCTTGTCTTGTTTGTTG
				ATTAGGCCTAAGACGATTCC
20	67953–68303	P	AP007063–AP007075	AAAGAAATAGGAGCATCGTG
				CCTTGTTCACTAATAAATCG
21	78460–78521	P	AP007076–AP007088	TAAAACGGGCATTCCTACGC
				AAGGATTCGAAGCGTACC
22	81008–81672	P	AP007089–AP007101	CGTGGTAAAGTATTCTAATC
				AACCATGTCTTCCCATTCCG
				ATAACCAACCTATTGCTTCG
23	99004–99316	N	AP007102–AP007105	CTACAGGAGAACCAGGAACG
24	108144–108742	P	AP007106–AP007118	TGATATGTATGTTCCATAAG
				CATCTATTGCCGCAAATC
25	110230–110306	P	AP007119–AP007131	TAACTCCAAGAATCCAAACG
26	110683–110855	P	AP007132–AP007144	TACGAATTCCGCACTTGTAG

[a]*P* and *N* denote polymorphism and no-polymorphism among two *S. officinarum* (acc. Badila and Fiji40) and two *S. spontaneum* (acc. Glagah and JW385) accessions, respectively

[b]Accession numbers in the GenBank/EMBL/DDBJ database
[c]AP006879–AP006891 denotes 13 accession numbers from AP006879 to AP006891

Table 3 Character matrix of mutations. *Indels*, insertions/deletions; *SSLPs*, simple sequence length polymorphisms

Accessions[b]	Types of mutations and their locations[a]																													
	Substitutions								Inversions			Indels						SSLPs													
	2	5	7	14	16	19	24	26	6	18	25	5	11	19	20	21	22	1	6	7	11	12	14	15	19	20	21	22	24	
NCo	000	000	00	000	0	00	00	0	0	0	0	0	0	0	0	0	0	0	0	0	0 0	0	0	0	0 0	0	0	0 0	0	
off-Ba (= off-Fi)	000	000	00	000	0	00	00	0	0	0	0	0	0	0	0	0	0	0	0	0	0–1	0	0	0	0 0	0	0	0 0	0	
rob-6	000	000	01	001	1	00	00	0	0	0	1	0	0	0	0	0	0	0	0	0	0–1	0	0	0	0 0	−1	0	+1 0	0	
rob-16	000	000	01	001	1	00	00	0	0	0	1	0	0	0	0	0	0	0	0	0	0 0	0	0	0	0 0	−1	0	+1 0	0	
sin-Oo (= sin-Te)	100	000	01	001	1	00	00	0	0	0	0	0	0	0	0	0	0	0	0	0	0–3	0	0	0	+1 0	0	0	+1 0	0	
bar-Ch (= bar-Ke)	100	000	01	001	1	00	00	0	0	0	0	0	0	0	0	0	0	0	0	0	0–3	0	0	0	+1 0	0	0	+1 0	0	
edu-IN (= edu-PN)	000	000	01	001	1	00	00	0	0	0	1	0	0	0	0	0	0	0	0	0	0+1	0	0	0	0 0	0	0	+1 0	0	
spo-Gl	011	111	01	111	1	11	11	1	1	1	1	1	1	1	1	1	1	+1	+1	+2	−1–5	+1	−1	+3	0+3	−2	−1	+1 0	−1	
spo-Jw	011	111	11	111	1	11	11	1	0	1	0	1	1	0	1	0	1	+1	0	+2	−1–5	+1	−1	+2	+1+2	0	0	+1+1	−1	
spo-SE	011	111	01	111	1	11	11	1	1	1	0	1	1	1	1	1	1	+1	+1	+2	−1–5	+1	−1	+2	0+3	−2	0	+1 0	−1	
ZM	...[b]	111	01	011	1	X[c]	1	10	0	1	1	1	1	1	0	0	0	0	+1	+3	+2	−1–5	−1	0	−1	0 0	0	−7	−4.[b]	−1

For substitutions, inversions, and indels, the sequence of NCo310 was used as the standard ("0"), and mutations are designated "1." For SSLPs, nucleotide length difference (+/− for longer/shorter) from the sequence in NCo310 are shown

[a]The number of analyzed regions corresponds to Table 2

[b]*Periods* indicate missing nucleotide at corresponding position

[c]*X* indicates nucleotide different from those designated as "0" or "1" (in this column, 0, 1, and X correspond to C, G, and T, respectively)

Phylogenetic analysis

A sequence matrix of 18 chloroplast regions was obtained by multiple alignment using Clustal X, version 1.81 (Thompson et al. 1997). Based on the matrix, pairwise comparisons were performed using the three-parameter method (Kimura 1981) to clarify the differences between sequences. The information on inversions, indels, and simple sequence repeat length polymorphisms (SSLPs) (Table 3) was converted into binary matrix data and added to the sequence matrix. Based on the resulting matrix of 5,744 characters, phylogenetic analysis was performed as a maximum parsimony estimation using PAUP*, version 4.0b8 (Swofford 1998). The most parsimonious tree was calculated using the heuristic search option involving 100 replications of random addition sequences and tree bisection reconnection (TBR) branch swapping. All characters were specified as unweighted. The strict consensus tree was computed from all trees obtained. Bootstrap analysis (Felsenstein 1985) was conducted to assess the reliability of the tree. One thousand replications were calculated using the heuristic search option with TBR branch swapping and random sequence addition.

Results and discussion

Selection of diagnostic regions of the chloroplast genome for phylogenetic analysis of the genus *Saccharum*

Comparative chloroplast genome analysis of sugarcane and three monocots revealed that the four chloroplast genomes have the same gene order, and that the chloroplast genome of sugarcane is more similar to that of maize than to that of rice or wheat (Asano et al. 2004). In this study, we assumed that regions conserved between sugarcane and maize will also be conserved among *Saccharum* species, but that the polymorphic regions between sugarcane and maize might be variable among *Saccharum* species, and selected 26 such regions at random (Table 2). Previous reports have indicated that *S. officinarum* and *S. spontaneum* are the most distantly related by analysis using nuclear RAPD and SSR markers (Nair et al. 1999; Selvi et al. 2003). Therefore, we chose those two species to examine the utility of these 26 regions. The selected 26 regions were assessed for their diversity, and 18 regions of the 26 were polymorphic across the two species (Table 2). Therefore, these 18 regions were used in further analyses.

Characteristics of mutations among the genus *Saccharum*

The DNA sequences of the 18 chloroplast genome regions (total of 5,381 bp in NCo310 sequences) from the six *Saccharum* species (13 accessions of Table 1) were determined and compared. We found 41 polymorphic sites among the *Saccharum* species (Fig. 1; Table 3). In the 18 regions, polymorphisms were identified at 194 sites between sugarcane (cv. NCo310) and maize; 25 sites (12.9%) were polymorphic both in the genus *Saccharum* and between *Saccharum* and maize (Fig. 1). This result indicates that the selection of diagnostic regions based on the variation between *S. officinarum* and *S. spontaneum* was effective in surveying the intrageneric poly-

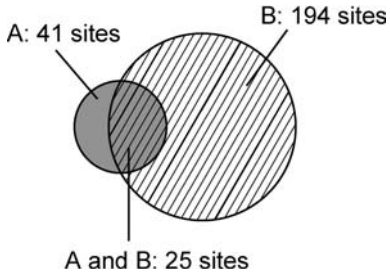

Fig. 1 Polymorphic sites identified within the genus *Saccharum* or between sugarcane and maize in 18 chloroplast regions. The *gray circle* (A) represents 41 polymorphic sites identified within the genus *Saccharum*. The *striped circle* (B)represents 194 polymorphic sites identified between sugarcane and maize. Area overlapped by the two circles represents 25 sites those were polymorphic both in the genus *Saccharum* and between *Saccharum* and maize

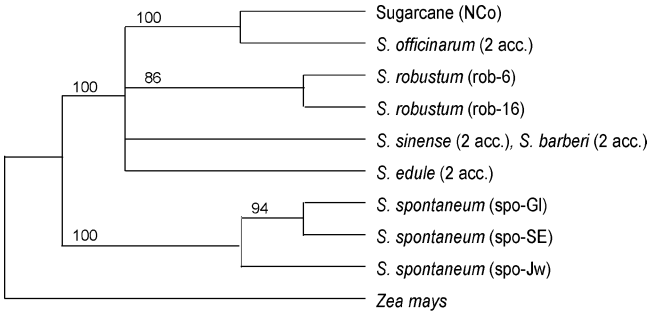

Fig. 2 Phylogenetic relationships among *Saccharum* species. A strict consensus tree was calculated using the maximum parsimony method based on the sequence data of 18 chloroplast DNA regions. Tree length = 264, confidence interval = 0.8659, retention index = 0.9299, rescaled consistency index = 0.8912. Species names and abbreviations of each accession (in *parentheses*) are shown. If two accessions of the same species have identical sequences, they are shown as "2 acc." The *numbers above the nodes* represent bootstrap values expressed as the percentage of 1,000 bootstrap replications. Characters were calculated under default conditions

morphisms within *Saccharum*. On the other hand, four other intrageneric polymorphisms were identified within *Saccharum*, the corresponding sequences of which were deleted in maize. The remaining 12 intrageneric polymorphisms were identified within sequences conserved between sugarcane and maize (Fig. 1; Table 3). Regarding of the interspecific polymorphisms of *Saccharum*, almost all polymorphic sites within the genus *Saccharum* (40 of 41 sites) were polymorphic between *S. officinarum* and *S. spontaneum*, whereas only one site showing another type of polymorphism was identified (Table 3).

These 41 mutations involved 17 substitutions, three inversions, six indels, and 15 SSLPs; 39 of them were phylogenetically informative (Table 3). Most of the mutations (36 of 41) were found within intergenic regions, but three were identified within introns and two were identified within an exon. The two mutations within the exon were found in the same *ccsA* gene, and resulted in one silent mutation and one substitution in the encoding amino acid sequence. Three inversions (6, 6, and 4 bp) were identified. Because inverted repeats (14, 14, and 20 bp) were present adjacent to the inversions, this type of mutation probably occurred via a flip-flop mechanism (Linne von Berg and Kowallik 1992). All six indels (2–21 bp) probably arose from the duplication or deletion of direct repeats. All SSLPs were found at single-nucleotide repeats of $(A/T)_n$ ($n = 8$–12 in NCo310 sequences). The A/T content of the chloroplast genome is higher than that of the nuclear genome, and most chloroplast microsatellites are single-nucleotide repeats of $(A/T)_n$ (Weising and Gardner 1999). Many of the mutations (31 of 41) were found within or adjacent to various types of repeated sequences, suggesting that mutations have preferentially arisen among sequences surrounded by repeated sequences.

Phylogenetic analysis of *Saccharum* species

Based on the results discussed above, a strict consensus tree was calculated using the maximum parsimony method (Fig. 2). On the tree, *S. sinense* and *S. barberi* formed a clade because their sequences are identical, whereas the other species formed another independent distinct clades. In analyses using nuclear RAPD and SSR markers, *S. sinense* and *S. barberi* also formed a single cluster (Nair et al. 1999; Selvi et al. 2003). Although these two species are distinguishable by morphological characteristics (Daniels and Roach 1987) and chromosome number (Price 1968), our results together with those of previous nuclear analyses suggest that these two species have an extremely close relationship.

It is noteworthy that *S. spontaneum* and the other five species (*S. officinarum*, *S. robustum*, *S. sinense*, *S. barberi*, and *S. edule*) show paraphyletic relationships. The relationships of all *Saccharum* species (except *S. edule*) were previously analyzed based on nuclear RAPD or SSR markers (Nair et al. 1999; Selvi et al. 2003). A close relationship between *S. officinarum* and *S. robustum* and a distant relationship between *S. officinarum* and *S. spontaneum* were also shown in previous reports based on nuclear DNA analysis and isozyme analysis (Glaszmann et al. 1989). On the other hand, the clustering patterns of *S. sinense* and *S. barberi* in this study are different from those of previous analyses, in which *S. sinense* and *S. barberi* clustered between *S. officinarum* and *S. spontaneum* (Nair et al. 1999; Selvi et al. 2003). *S. sinense* and *S. barberi* are thought to be derived from interspecific hybrids of *S. officinarum* and *S. spontaneum* (Daniels and Roach 1987). GISH analysis showed that these two species both have chromosomes derived from *S. officinarum* and *S. spontaneum* (D'Hont et al. 2002). Our results suggest that the cytoplasmic parents of *S. sinense* and *S. barberi* are genetically closely related to *S. officinarum* or *S. robustum*. Very few molecular analyses of *S. edule* have been done, and none of these have explained the origin of this species, although it has been suggested, based on chromosome numbers and mor-

Table 4 Pairwise comparisons of total mutations and genetic distances between taxa

Accession	1	2	3	4	5	6	7	8	9	10	11
1. NCo	–	1	7	6	7	7	6	37	33	35	194
2. *S. officinarum* (2 acc.)	0.000	–	6	7	7	7	6	37	33	35	194
3. *S. robustum* (rob-6)	0.056	0.056	–	1	5	5	1	33	30	33	191
4. *S. robustum* (rob-16)	0.056	0.056	0.000	–	5	5	2	32	31	32	191
5. *S. sinense* (2 acc.)	0.075	0.075	0.019	0.019	–	0	4	35	29	33	192
6. *S. barberi* (2 acc.)	0.075	0.075	0.019	0.019	0.000	–	4	35	29	33	192
7. *S. edule* (2 acc.)	0.056	0.056	0.000	0.000	0.019	0.019	–	32	30	32	190
8. *S. spontaneum* (spo-Gl)	0.282	0.282	0.225	0.225	0.244	0.244	0.225	–	12	3	181
9. *S. spontaneum* (spo-Jw)	0.300	0.300	0.244	0.244	0.263	0.263	0.244	0.019	–	9	180
10. *S. spontaneum* (spo-SE)	0.282	0.282	0.225	0.225	0.244	0.244	0.225	0.000	0.019	–	182
11. ZM	2.278	2.278	2.220	2.220	2.220	2.220	2.220	2.068	2.087	2.068	–

Total numbers of mutations are shown *above the diagonal*. Base substitutions per 100 sites (adjusted for missing data) are shown *below the diagonal*. Characters are shown under default conditions

phological characteristics, that *S. edule* arose from intergeneric crosses between *S. officinarum* or *S. robustum* and related genera (e.g., *Miscanthus*) or was derived from *S. robustum* (Irvine 1999). Our results suggest that the cytoplasm of *S. edule* is similar to that of *S. officinarum* and *S. robustum*.

Commercial hybrid NCo310, used as the standard, occurs in the same clade as *S. officinarum*. This result confirms the fact that modern sugarcane cultivars are mainly derived from *S. officinarum*.

Different extent of intraspecific variation among *Saccharum* species

Pairwise comparisons of total mutations (base substitutions, inversions, indels, and SSLPs) among the 13 accessions are shown in Table 4. No intraspecific variation was observed within the two accessions of *S. officinarum*, the two accessions of *S. sinense*, the two accessions of *S. barberi*, or the two accessions of *S. edule*. One SSLP was found in the two *S. robustum* accessions (Tables 3 and 4). Therefore, these five *Saccharum* species show no or very little intraspecific variation. On the other hand, *S. spontaneum* displays higher intraspecific variation. For example, 12 mutations were found between *S. spontaneum* acc. Glagah and *S. spontaneum* acc. Jw385 (Table 4). These results show greatly contrasting levels of intraspecific variation. *S. spontaneum* is thought to be the ancestral species of the genus *Saccharum* because of its large intraspecific diversity in terms of morphology, species distribution, and chromosome number (Guimaraes and Sobral 1998). Our results provide supporting evidence for this view.

Phylogenetic relationships among *Saccharum* species

High levels of interspecific variation (30–37 total mutations) were observed between *S. spontaneum* and the other five species (*S. officinarum*, *S. robustum*, *S. sinense*, *S. barberi*, and *S. edule*) (Table 4). However, low levels of interspecific variation were observed among these five species (0–7 total mutations), which are even lower than the intraspecific variation of *S. spontaneum* (3–12 total mutations, Table 4). These results suggest that the cytoplasm of these five species are genetically closely related. The significant divergence between *S. spontaneum* and the other *Saccharum* species (except *S. edule*, which was not analyzed) was also identified in previous analyses using nuclear RAPD or SSR markers (Nair et al. 1999; Selvi et al. 2003). *S. robustum* is known to be very closely related to *S. officinarum*, and is thought to be the progenitor of *S. officinarum* (Guimaraes and Sobral 1998). On the other hand, 60–70% of *S. sinense* and *S. barberi* chromosomes are derived from *S. officinarum* (D'Hont et al. 2002). Therefore, the close relationships between the nuclear genomes of these three species and that of *S. officinarum* are reasonable. Based on morphological, cytological, and molecular analyses, Irvine (1999) proposed that *Saccharum* should be classified into *S. spontaneum* and *S. officinarum* including the other four species. Our results support this view in principle.

No interspecific variation was observed between *S. sinense* and *S. barberi* (Table 4), suggesting a significantly close relationship between these species. Furthermore, very low interspecific variation was observed between *S. edule* and *S. robustum* (one and two mutations between *S. edule* and *S. robustum* acc. rob-6 and rob-16, respectively), which is much lower than the interspecific variation between *S. edule* and *S. officinarum* (six mutations) (Table 4). This further suggests that *S. robustum* is more closely related to *S. edule* than to *S. officinarum*.

Estimation of divergence times among *Saccharum* species

Based on the base substitution rate in the 18-chloroplast regions between maize and wheat, we estimated the divergence times among the *Saccharum* species and maize. We compared the 18 chloroplast regions of wheat (accession number for the whole chloroplast genome,

AB0422409) with those of maize and found that the region showed 0.1924 substitutions/site overall. Wolfe et al. (1989) estimated the divergence time of wheat and maize at 50 million years ago. Therefore, the substitution rate of this region was calculated as 3.85 nucleotide substitutions/site per one billion years. Pairwise comparisons of the substitutions per 100 sites between the *Saccharum* accessions were performed and are shown in Table 4. Based on the substitution rate calculated here, the divergence times were estimated as follows: between *Saccharum* and maize, about 5–6 million years ago; between *S. spontaneum* and the other five *Saccharum* species, about 580–780 thousand years ago; and within the five *Saccharum* species, 0–220 thousand years ago, suggesting that the *Saccharum* species diverged very recently.

Although the *Saccharum* species show large phenotypic diversity, phylogenetic analysis of the chloroplast genomes shows very close relationships among these species, suggesting recent and rapid evolution of the genus *Saccharum*. The huge and complex nuclear genomes of *Saccharum* species could be one reason that the species achieved such great diversity within a short period. Ming et al. (1998) reported threefold higher recombination in both *S. officinarum* and *S. spontaneum* (chromosome numbers $2n = 52$–140) than in sorghum (chromosome number $2n = 20$), a genus closely related to *Saccharum* with a compact genome, based on the genetic distances between two corresponding genetic markers within conserved chromosomal regions.

Furthermore, *S. sinense* and *S. barberi* are thought to have gained some diversity from the *S. officinarum* and the *S. spontaneum* interspecific crosses. Based on GISH analysis, the ratio of chromosome numbers originating from *S. officinarum* or *S. spontaneum* are different in the $2n = 82$ and 91 clones of *S. barberi* (D'Hont et al. 2002).

In conclusion, information from the complete chloroplast genomes of sugarcane and maize enabled us to perform comprehensive phylogenetic analysis of the genus *Saccharum* in hyperpolymorphic regions. We found 41 polymorphisms among the six *Saccharum* species. Based on these data, we calculated a phylogenetic tree with sufficient resolution. This is the first phylogenetic tree showing the overall genetic relationships of the genus *Saccharum*. The cytoplasmic lineage of the genus *Saccharum* shown here is very useful not only for comparison with the phylogeny based on nuclear genomic diversity, but also, of itself, in providing important data for sugarcane breeding using related wild species.

Acknowledgements We thank Drs. M. Matsuoka and M. Sato from JIRCAS and K. Miyagi from OPAES for providing the plant materials, Dr. T. Nishikawa for his useful suggestions, and Dr. K.K. Wu of Hawaii Agriculture Research Center for information exchange regarding plant materials. We also thank Ms. K. Miyashita and N. Nohara for their technical assistance.

References

Al-Janabi SM, McClelland M, Petersen C, Sobral BWS (1994) Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. Theor Appl Genet 88:933–944

Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K (2004) Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. DNA Res 11:93–99

Berding N, Roach BT (1987) Germplasm collection, maintenance, and use. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 143–210

Daniels J, Roach BT (1987) Taxonomy and evolution. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 7–84

D'Hont A, Paulet F, Glaszmann JC (2002) Oligoclonal interspecific origin of 'North Indian' and 'Chinese' sugarcanes. Chromosome Res 10:253–262

Felsenstein J (1985) Confidence limits on phylogenies an approach using the bootstrap. Evolution 39:783–791

Glaszmann JC, Fautret A, Noyer JL, Feldmann P, Lanaud C (1989) Biochemical genetic markers in sugarcane. Theor Appl Genet 78:537–543

Guimaraes CT, Sobral BWS (1998) The Saccharum complex: relation to other Andropogoneae. Plant Breed Rev 16:269–288

Irvine JE (1999) Saccharum species as horticultural classes. Theor Appl Genet 98:186–194

Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci USA 78:454–458

Linne von Berg KH, Kowallik KV (1992) Structural organization of the chloroplast genome of the chromophytic alga *Vaucheria bursata*. Plant Mol Biol 18:83–95

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Nair NV, Nair S, Sreenivasan TV, Mohan M (1999) Analysis of genetic diversity and phylogeny in *Saccharum* and related genera using RAPD markers. Genet Resour Crop Evol 46:73–79

Price S (1968) Cytology of Chinese and North Indian sugarcane. Econ Bot 22:155–164

Selvi A, Nair NV, Balasundaram N, Mohapatra T (2003) Evaluation of maize microsatellite markers for genetic diversity analysis and fingerprinting in sugarcane. Genome 46:394–403

Sobral BWS, Braga DPV, Lahood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the *Saccharinae* Griseb. subtribe of the *Andropogoneae* Dumort. tribe. Theor Appl Genet 87:843–853

Swofford DL (1998) PAUP*. Phylogenetic analysis using parsimony (*and other methods). 4 4. Sinauer, Sunderland

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201–6205