Theor Appl Genet (1994) 88:933–944


S. M. Al-Janabi · M. McClelland · C. Petersen
B. W. S. Sobral

# Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae

Received: 30 June 1993 / Accepted: 21 December 1993

**Abstract** To study the phylogenetics of sugarcane (*Saccharum officinarum* L.) and its relatives we sequenced four loci on cytoplasmic genomes (two chloroplast and two mitochondrial) and analyzed mitochondrial RFLPs generated using probes for COXI, COXII, COXIII, Cob, 18*S*+5*S*, 26*S*, ATPase 6, ATPase 9, and ATPase $\alpha$ (D'Hont et al. 1993). Approximately 650 bp of DNA in the intergenic spacer region between *rbcL* and *atpB* and approximately 150 bp from the chloroplast 16*S* rDNA through the intergenic spacer region $tRNA^{val}$ gene were sequenced. In the mitochondrial genome, part of the 18*S* rRNA gene and approximately 150 bp from the 18*S* gene 3′ end, through an intergenic spacer region, to the 5*S* rRNA gene were sequenced. No polymorphisms were observed between maize, sorghum, and '*Saccharum* complex' members for the mitochondrial 18*S* internal region or for the intergenic $tRNA^{val}$ chloroplast locus. Two polymorphisms (insertion-deletion events, indels) were observed within the 18*S*–5*S* mitochondrial locus, which separated the accessions into three groups: one containing all of the *Erianthus, Eccoilopus, Imperata, Sorghum*, and 1 *Miscanthus* species; a second containing *Saccharum* species, *Narenga porphyrocoma, Sclerostachya fusca*, and 1 presumably hybrid *Miscanthus* sp. from New Guinea; and a third containing maize. Eighteen accessions were sequenced for the intergenic region between *rbcL* and *atpB*, which was the most polymorphic of the regions studied and contained 52 site mutations and 52 indels, across all taxa. Within the *Saccharum* complex, at most 7 site mutations and 16 indels were informative. The maternal lineage of *Erianthus/Eccoilopus* was nearly as divergent from the remaining *Saccharum* complex members as it was from sorghum, in agreement with a previous study. Sequences from the *rbcL-atpB* spacer were aligned with GENBANK sequences for wheat, rice, barley, and maize, which were used as outgroups in phylogenetic analyses. To determine whether limited intra-complex variability was caused by under sampling of taxa, we used seven restriction enzymes to digest the PCR-amplified *rbcL-atpB* spacer of an additional 36 accessions within the *Saccharum* complex. This analysis revealed ten restriction sites (none informative) and eight length variants (four informative). The small amount of variation present in the organellar DNAs of this polyploid complex suggests that either the complex is very young or that rates of evolution between the *Saccharum* complex and outgroup taxa are different. Other phylogenetic information will be required to resolve systematic relationships within the complex. Finally, no variation was observed in commercial sugarcane varieties, implying a world-wide cytoplasmic monoculture for this crop.

**Key words** Cycle sequencing · PCR · Chloroplast Sugarcane · Polyploid · *rbcL* · *atpB* · Mitochondria *Saccharum*

## Introduction

*Saccharum* L. is part of the Saccharinae subtribe of the Andropogoneae (Watson et al. 1985, Clayton and Renvoize 1986). The Andropogoneae are frequently polyploid and many are also of great socio-economic importance (Stebbins 1956; Soltis et al. 1992). Very little is known about the speciation and evolution of polyploids even though about 50–70% of all grass species are polyploid (Stebbins 1956; Soltis et al. 1992). Furthermore, it is difficult to delimit some polyploid plant species as interspecific boundaries are unclear because of the possibility of hybridization and chromosome doubling. These biological phenomena have helped keep the taxonomy of *Saccharum* in a state of flux, especially at lower ranks. Mukherjee (1957) coined the term '*Saccharum* complex' to refer to members of *Narenga* Bor, *Sclerostachya* (Anderss. ex Hackel) A. Camus, *Erianthus* Michx. section *Ripidium*, and *Saccharum* that

Communicated by A. R. Hallauer

S. M. Al-Janabi · M. McClelland · C. Petersen · B. W. S. Sobral (✉)
California Institute of Biological Research,
11099 North Torrey Pines Road,
Suite 300, La Jolla CA 92037, USA

can interbreed and in which species (1) are endemic to the Bengal-Assam-Sikkim area in India, (2) have been classified as *Saccharum* at one time or another, and (3) have been directly implicated in the evolution of *Saccharum* by various authors. More recently, Daniels and Williams (1975) included *Miscanthus* Anderss. section *Diandra* species in a revision of the complex and extended the geographic range to include the Indo-Burma-China border region. A recent taxonomy of the Andropogoneae considers *Erianthus* and *Narenga* to be synonymous with *Saccharum*, *Eccoilopus* synonymous with *Spodiopogon*, and *Miscanthus* synonymous with *Miscanthidium* and *Sclerostachya* (Clayton and Renvoize 1986). It seems that the *Saccharum* complex is a group of plants with their own evolutionary history. However, it remains unclear as to what has been the role of the processes of hybridization and chromosome doubling in the evolution of this complex, though there has been much speculation. In addition, it is unclear what is the main reproductive mode in nature: vegetative or sexual. The taxonomy of *Saccharum* has been and still is complex and controversial. Some of the controversy may stem from the use of morphological traits as such traits may be difficult to score in polyploid species that can interbreed. Furthermore, it is clear that human activities over the past 10,000 years or more have played a crucial role in the dispersal, selection, and vegetative propagation of particular genotypes of sugarcane and its relatives, including interspecific and intergeneric hybrids (Brandes 1928; Artschwager and Brandes 1958; Bellwood 1985). Studies using leaf flavonoids (Williams and Harborne 1974; Daniels et al. 1980) and leaf waxes (Smith and Martin-Smith 1978) suggest that a large amount of variability is present in some species (notably *S. spontaneum* and *Erianthus* species).

Studies using the tools of molecular systematics in the *Saccharum* complex were pioneered by Waldron et al. (1974). They used $\beta$-amylase isoenzymes as genetic markers in *Saccharum* and its relatives and found that collection site was correlated with banding pattern, which was proposed to indicate widespread occurrence of hybridization. The limited existing body of molecular data support a recent common ancestry for sorghum and sugarcane. Hamby and Zimmer (1988) used nuclear rRNA sequences of taxa within Poales to infer phylogeny. Their phylogenetic hypothesis showed that *Sorghum* and *Saccharum* were very closely related; in fact these had the smallest pairwise genetic distance ($d$=0.00611) of any two genera sampled. This is interesting because *Sorghum* is thought to have radiated from Africa whereas *Saccharum* has been suggested to have radiated from Southeastern Asia. Current taxonomy suggests that the genetic distance from *Saccharum* to any member of the *Saccharum* complex should be smaller than the *Saccharum–Sorghum* distance. Springer et al. (1989) analyzed the genomic organization of nuclear rDNA in *Sorghum* and its close relatives using restriction fragment length polymorphisms (RFLPs). They found that sorghum, sugarcane, and maize had very similar rDNA monomer sizes and restriction maps and that sorghum and sugarcane are more closely related to each other than either is to maize. Both studies used only one genotype of *Saccharum*, and it was a commercial interspecific hybrid (*S. officinarum* × *S. spontaneum* hybrids). Glaszmann et al. (1990) determined RFLP variation of nuclear rDNA in some *Saccharum* species, commercial hybrids, one *Erianthus arundinaceus* accession, and one *Miscanthus* species. They reported 15 rDNA length variants, some of which appeared to be phylogenetically informative, though no phylogenies were inferred, possibly because most accessions studied had more than one rDNA RFLP phenotype and hybridizing fragments varied in intensity. D'Hont et al. (1993) studied organellar RFLPs in 57 accessions of *Saccharum* and some *Saccharum* complex members; they found no variation using 2 chloroplast probes that covered approximately 20% of the wheat chloroplast genome, although a single fragment was missing for the single *Erianthus* and *Miscanthus* species studied. In addition, nine mitochondrial probes yielded ten different pattern types, although no phylogenetic hypothesis was generated from their data. Sobral et al. (1994) studied chloroplast RFLPs using 32 accessions representing eight genera and 19 species with 15 restriction enzymes and 12 rice chloroplast probes that covered the rice chloroplast genome entirely. Sixty-two mapped restriction site mutations (18 informative) placed the accessions into nine clades: seven from within the Saccharinae, maize, and sorghum. Maternal lineages of *Saccharum, Narenga, Sclerostachya,* and *Miscanthus* were shown to form a monophyletic group displaying little variation, whereas *Erianthus* and *Eccoilopus* were shown to be different from other *Saccharum* complex members. Because the level of polymorphism detected by previous studies was very low, we sequenced potentially polymorphic regions in the chloroplast genome to allow a finer level of phylogenetic resolution within the *Saccharum* complex and herein report the results.

## Materials and methods

### Plant materials

Plant accessions and their origins are listed in Table 1. These accessions were chosen to represent a wide range of species, and they differ in their origin and chromosome numbers; some have been studied previously (Sobral et al. 1993).

### Primer design

To study cytoplasmic DNA sequence variation we chose regions in the genome that were highly conserved though interspersed with polymorphic regions. We constructed primers to the conserved regions that would amplify across regions that were expected to be polymorphic based on previous work (Zurawski et al. 1984; Salts et al. 1984; Bowman et al. 1983; Birky 1988; Zurawski and Clegg 1987) and GENBANK sequences. To study chloroplast sequence variation, primers were designed to target two loci: the intergenic spacer region between the highly conserved $tRNA^{val}$ gene and the 16*S* rRNA gene, and the intergenic region between *rbcL* and *atpB*. In the mitochondrial genome the intergenic spacer region between the 18*S* and 5*S* rRNA genes was amplified as well as an internal region of the 18*S* gene.

Primers were designed such that a pair of external primers was flanked by a pair of internal, or nested primers. The external set was used to amplify the primary product from total DNA, and the inter-

**Table 1** Plant accessions and their origin (*U* undetermined)

| Species | Group[a] | Genotype | Origin[b] | Cytol[c] | Source[d] |
|---|---|---|---|---|---|
| Accessions used only in MRSP study: | | | | | |
| *Coix gigantea* | | IS 76-199 | Indonesia | U | WC |
| *Erianthus arundinaceus* | *Ripidium* | SES 288 | India | U | Houma |
| | *Ripidium* | Mardan | U | 54 | Houma |
| *E. bengalense* | *Ripidium* | IMP 2886 | U | U | Houma |
| *E. procerus* | *Ripidium* | Kalimpong | U | 40 | Houma |
| *E. ravennae* | *Ripidium* | SES 372 | U | 20 | Houma |
| *Miscanthus sinensis* | *Diandra* | NG 77-22 | PNG | 38 | TAES |
| *Saccharum barberi* | Saretha | Katha | India | 90 | Houma |
| | Nargori | Nargori | India | 124 | TAES |
| *S. officinarum* | | NG 57-72 | PNG | 80 | TAES |
| | Kassoer | U | 80 | TAES | |
| *S. robustum* | Teboe Salak | Teboe Salak Toewa | Celebes | 90 | CTC |
| | Wau-Bulolo | NG 57-11 | PNG | 60 | CTC |
| | Goroka | NG 57-208 | PNG | 80 | HSPA |
| | Port Moresby | NH 1 | NH | 80 | HSPA |
| *S. sinense* | Pansahi | Chuk Chee | U | U | TAES |
| | Pansahi | Uba Nanquim | U | U | TAES |
| *S. spontaneum* | | NG 51-2 | PNG | 80 | HSPA |
| | | SES 370 | Nepal | 40 | HSPA |
| | | SES 208 | India | 64 | HSPA |
| | | SES 113A | India | 48 | WC |
| | *Aegyptiacum* | Aegyptiacum | Egypt | 112 | WC |
| | | SES 517 (Canton 1) | China | 112 | WC |
| | | Okinawa | Japan | 112 | WC |
| *Saccharum* sp. | Comm. hybrid | EK 28 | Breeding | U | TAES |
| *Saccharum* sp. | Comm. hybrid | POJ 100 | Breeding | U | TAES |
| *Saccharum* sp. | Comm. hybrid | POJ 2878 | Breeding | U | TAES |
| *Saccharum* sp. | Comm. hybrid | Co 206 | Breeding | U | TAES |
| *Saccharum* sp. | Comm. hybrid | SP 70-1143 | Breeding | U | TAES |
| *Saccharum* sp. | Comm. hybrid | CP 65-357 | Breeding | U | TAES |
| *Saccharum* sp. | Comm. hybrid | CP 70-321 | Breeding | U | TAES |
| *Sclerostachya fusca* | | US-58-5-2 | U | 30 | Houma |
| *Sorghum alum* | | | U | U | ASCL |
| *S. bicolor* | | Wray | U | 20 | TAES |
| *S. halepense* | | | U | 20 | ASCL |
| *S. plumosum* | | US 71-17 | U | U | WC |
| Accessions sequenced for entire *rbcL-atpB* spacer: | | | | | |
| *Eccoilopus longisetosus* | | US 57-11-2 | India | 60 | Houma |
| *Erianthus kanashiori* | | MIA 33247 | U | U | WC |
| *E. trinii* | New World | US 65-14 | Argentina | U | Houma |
| *Miscanthus sininses* | | Zebrinus | U | 38 | TAES |
| *Miscanthus* sp. | | NG 77-193 | U | 192 | Houma |
| *Narenga porphyrocoma* | | US 58-4-1 | U | 30 | TAES |
| *Saccharum barberi* | Saretha | Chunnee | India | U | TAES |
| *S. edule* | New Guinea | NH 70-23 | New Hebrides | U | TAES |
| *S. officinarum* | | NG 51-131 | PNG | 80 | TAES |
| | | Black cheribon | Java | 80 | TAES |
| *S. robustum* | Sanguineum | NG 28-218 | PNG | 70 | TAES |
| *S. sinense* | Pansahi | China | U | 112 | WC |
| *S. spontaneum* | | Coimbatorre | India | 64 | TAES |
| | | SES 561 | Sudan | 128 | HSPA |
| *Sorghum bicolor* | | Sweetchew | U | 20 | TAES |
| *Vetiveria* sp. | | ND 81-53 (IMP9755)U | 20 | WC | |
| *Zea mays* | | Dk77w Dekalb | U | 20 | TAES |
| | | Conlee 202 | U | 20 | TAES |

[a] Group=non-taxonomic grouping used by sugarcane biologists; generally related to cytological or geographic groups
[b] Origin=original place of collection (Artschwager 1954; Brandes et al. 1939; Moriya 1940; Price 1968; Panje and Babu 1960)
[c] Cytol=2n chromosome number (Moriya 1940; Panje and Babu 1960; Price 1957; Burner 1991; Mohan and Sreenivasan 1983)
[d] Source refers to the place from which we obtained a sample. (*TAES* Texas Agricultural Experiment Station, Weslaco Texas; *Houma* USDA Sugercane Laboratory at Houma, LA; *WC* World collection Miami FL; *HSPA* Hawaiian Sugarcane Planters Association; *CTC* Copersucar Sugarcane Technology Center collection, Piracicaba, Brazil)

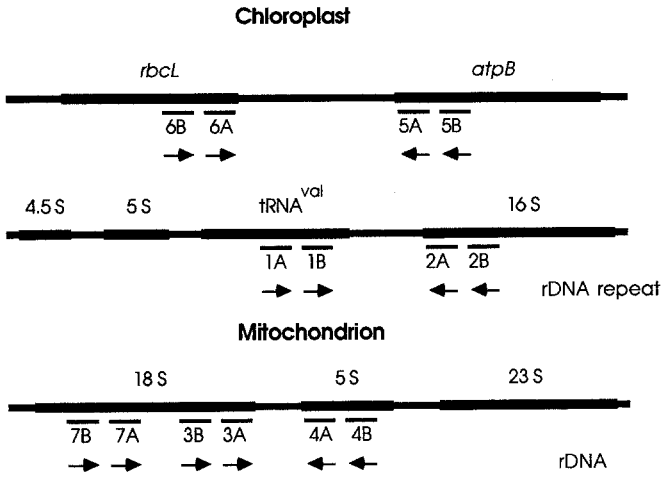


**Fig. 1** Oligonucleotide primers used to amplify loci for sequencing. Primer designations are as described in Materials and methods. Maps are not drawn to scale

nal or nested set of primers was used for cycle-sequencing, giving cleaner results than if the external primers were used to prime sequencing reactions. Primer sequences were checked against GENBANK to ensure that they did not match well at the 3′ end with any structural RNAs except those in targeted regions. These primers should work with all angiosperms because we selected regions that were conserved across monocot-dicot divergence, based on current GENBANK sequences. Primer details and map positions are shown in Fig. 1, along with information about the amplified regions. The sequences of the primers are as follows (designations in parentheses correspond to those in Fig. 1):

| | |
|---|---|
| *atpB*P1 (5A) | 5′-GGAAAAGTGATATCCAGCAC-3′ |
| *atpB*P2 (5B) | 5′-GGCAACTTGCCCGGGGGAAA-3′ |
| *rbcL*P1 (6A) | 5′-TTGAGTTGTTGTTATGGTAA-3′ |
| *rbcL*P2 (6B) | 5′-TATACACCCTGTGTACGTTC-3′ |
| cpval3P2 (1A) | 5′-AGTTCGAGCCTGATTATCCC-3′ |
| cpval3P1 (1B) | 5′-AAGTCATCAGTTCGAGCCTG-3′ |
| cp16S5P1 (2A) | 5′-GCATGCCGCCAGCGTTCATC-3′ |
| cp16S5P2 (2B) | 5′-TTGCATGTGTTAAGCATGCC-3′ |
| mt18S1180 (7A) | 5′-GACATGCGCCTAAGGAGAAA-3′ |
| mt18S1170 (7B) | 5′-GTGTTGCTGAGACATGCGCC-3′ |
| mt18S3P1 (3A) | 5′-TGCCACAAAGGCCTTTGGTG-3′ |
| mt18S3P2 (3B) | 5′-GTAACAAGGTAGCCGTAGGG-3′ |
| mt5S5P1 (4A) | 5′-AAACACGTCTCACCGTAGTG-3′ |
| mt5S5P2 (4B) | 5′-ATATGGCGCAAGACGATTCC-3′ |

### DNA manipulations

Total DNA was extracted by the method of Honeycutt et al. (1992) and stored in aliquots at –20 °C. Working stocks were generated for polymerase chain reaction (PCR): these contained 100, 10, or 1 ng $\mu l^{-1}$ DNA in TE (Maniatis et al. 1982). PCR was used to amplify target regions: reactions used 10–30 ng of template DNA, 0.2 m*M* of each dNTP, 4 m*M* $MgCl_2$, 10 m*M* KCl, 20 m*M* TRIS-HCl pH 8.8, 10 m*M* $(NH_4)SO_4$, 0.45 μ*M* of each primer, and 2.5 U Vent DNA polymerase (New England Biolabs, Beverly, Mass.) in a 50 μl reaction volume. Thermal cycling was done in a System 9600 (Perkin-Elmer, Norwalk, Conn.), using the following thermal profile: 94 °C for 1 min, 60 °C for 1 min, and 72 °C for 2 min, for a total of 30 cycles. For each reaction 3 μl were resolved on a 1.7% LE agarose gel to confirm amplification of products.

### DNA sequencing

Products positively identified on agarose gels were subjected to PCR cycle-sequencing (Ruanto and Kidd 1991) with the internal ("nested") primers (Fig. 1). For cycle-sequencing reactions, a 50 μl cocktail was prepared that contained 6 m*M* each dNTP, 3 m*M* $Mg^{++}$, 1×*Taq* buffer (10 m*M* TRIS-HCl, pH 8.3, 50 m*M* KCl), 1 μ Ci [$^{32}$P]-dCTP (3,000 mCi mmol, NEN), 4 μ*M* of 1 internal primer, 5 U *Taq* polymerase and 2 μl template from primary amplification (approximately 40 ng template). This cocktail was divided into four tubes to which ddNTPs (Pharmacia, Piscataway, N. J.) were added separately at the following final concentrations: 667 μ*M* ddATP, 667 μ*M* ddTTP, 400 μ*M* ddCTP, 80 μ*M* ddGTP. The ddNTP solutions were prepared in 1×*Taq* buffer supplemented with 2.5 m*M* $Mg^{++}$. These sequencing reactions were then cycled 30 times in the Gene Amp 9600 using the following temperature profile: denaturation at 95 °C for 45 s, annealing at 55 °C for 35 s, and extension at 72 °C for 75 s with a ramp of 30 s between each change of temperature. Six microliters of the sequencing reaction products were denatured by diluting 1:1 in formamide dye solution (final formamide concentration was 50%) at 85 °C for 5 min, and then 8 μl of the denatured mix was loaded on a 0.35-mm 5% polyacrylamide sequencing gel. The gel was run in 1×TBE (Maniatis et al. 1982) at a constant power of 50 W for approximately 1.5 h, then submitted to autoradiography for 24–48 h. Approximately 150–200 bp of sequence could be read using direct cycle-sequencing.

Because of its length and polymorphic nature, the amplified *rbcL-atpB* intergenic region was cloned into PCR-Script SK(+) vector (Stratagene, La Jolla Calif.). Two hundred nanograms of the PCR products were incubated with 50 ng of PCR-Script SK(+) DNA, 0.5 m*M* rATP, 15 U *Srf*I (Stratagene), 4 U T4 DNA ligase, 100 μ*M* KOAc, 25 m*M* TRIS-OAc, pH 7.6, 10 m*M* MgOAc, and 10 μg $ml^{-1}$ BSA in 11 μl of total reaction volume. Single-stranded and double-stranded DNAs were used for DNA sequencing of the cloned *rbcL-atpB* spacer region. The chain-termination method was used (Sanger et al. 1975). Sequenase v 2.0 sequencing kit (United States Biochemical, Cleveland, Ohio) was used with primer 6B (Fig. 1) for single-stranded DNA sequencing and primer 5B (Fig. 1) for double-stranded DNA sequencing, allowing the resolution of up to 500 bases in each direction. In each reaction 5 μCi of α-[$^{35}$S]-dATP (Amersham, Arlington, Ill.) was used. Sequencing reactions were resolved on 50% urea-6% polyacrylamide denaturing gels for 3–5 h at 50 W constant power. Gels were dried for 1 h at 80 °C, and autoradiography was done overnight at -70°C using X-omat-AR film (Kodak) and an intensifying screen.

### Mapped restriction site polymorphism (MRSP) analysis

Site and length mutations within the *rbcL-atpB* spacer were determined in all genotypes by restriction analysis of the target region using the map of the region that was established from the DNA sequences (Ralph et al. 1993; Liston 1992). Seven restriction enzymes with 4-bp target sites [*Hinf*I, *Hph*I, *Ple*I, *Hha*I, *Mbo*I, (NEB) *Sau*I, *Eco*O109I (Stratagene)] were used according to the manufacturer's instructions. The *rbcL-atpB* region was first amplified by PCR following the same conditions mentioned in DNA manipulation (above) except that we included 0.2 μCi α–[$^{32}$P] dCTP to label the resulting fragments. This allowed visualization of the restriction fragments on high resolution polyacrylamide sequencing gels. A 2-μl aliquot of the PCR reaction was mixed with 10 μl of formamide dye and 1.5 μl was used for electrophoresis for 2.5–3 h on 6% polyacrylamide gels. The gels were then dried for 1 h at 80 °C and autoradiography was done overnight at –70 °C using X-omat-AR film (Kodak). The presence or absence of a restriction site or of a particular length mutation was scored (Table 3).

### Phylogenetic analyses

DNA sequence were scored independently at least twice. Sequences were analyzed for restriction sites using DNA Strider v 1.0.1. Manual sequence alignment was done to establish the database for the phylogenetic analyses. Nucleotides were treated as individual characters and analyzed by cladistic parsimony and maximum likelihood using PAUP v 3.1.1 (Swofford 1991) and PHYLIP v 3.4 (Felsenstein 1989). Insertion/deletion events (indels) were obtained by aligning sequences in the most parsimonious way. The presence or absence of a nucleotide or group of nucleotides was scored as a unordered binary character. The resulting matrix of 719 characters, containing

**Table 2** Classification of DNA sequences in cytoplasmic genomes of sugarcane and related grasses (*ND* not determined)

| Accession | Mitochondrion | | Chloroplast | | |
|---|---|---|---|---|---|
| | 3A to 4A[a] | 7A[b] | 6A[c] | 5A[d] | 1B to 2A[e] |
| *Erianthus arundinaceus* | A | ND | A | ND | ND |
| *E. bengalense* | A | ND | A | ND | ND |
| *E. longisetosus* | A | ND | A | ND | ND |
| *E. procerus* | A | ND | A | ND | ND |
| *E. ravennae* | A | ND | A | ND | ND |
| *E. trinii* | A | ND | A | D | ND |
| *Imperata cylindrica* | A | ND | A | ND | ND |
| *Miscanthus sinensis* | A | A | C | C | ND |
| *Miscanthus* sp. (New Guinea) | B | ND | A | D | ND |
| *Narenga porphyrocoma* | B | ND | A | ND | ND |
| *Saccharum edule* | B | ND | A | A | A |
| *S. barberi* | B | A | A | A | ND |
| *S. officinarum* | B | A | A | A | A |
| *S. robustum* | B | ND | A | ND | A |
| *S. sinense* | B | A | A | A | A |
| *S. spontaneum* | B | A | A | A | A |
| *Sclerostachya fusca* | B | A | A | ND | ND |
| *Sorghum bicolor* | A | A | A | B | A |
| *Zea mays* | C | A | B | D | A |
| Commercial hybrids[f] | ND | ND | A | A | A |

[a] All primer numbers refer to Fig. 1. Species with the same letter (A, B, or C) had the same sequence. Data is based on scoring of 1 polymorphism (a deletion/insertion event) that separated types A from B and 2 polymorphisms that separated both A and B from C. Both strands have been sequenced

[b] No polymorphisms were observed between sugarcane species and outgroup species in the first 200 bases using 7A as a sequencing primer

[c] Based on analysis of the first 150 bases using 6A as a primer. The two polymorphisms that differentiated the types were insertion/deletion events

[d] Same as note (c) above, except 3 polymorphisms were scored

[e] No polymorphisms observed between outgroups and *Saccharum* complex members. *Arabidopsis thaliana* sequence has also been determined

[f] Based on evaluation of 7 accessions (Table 1), some of which are in the lineage of almost all commercial clones used by modern sugarcane growers

the nucleotide sequence and 55 indels, was analyzed with settings to ignore invariant characters. In some analyses, transitions were weighted 1.5×transversions, according to results of others from analysis of this sequence (Zurawski et al. 1984).

The analysis of mitochondrial data from D'Hont et al. (1993) was done by re-coding their data (which had been scored qualitatively) into binary characters based on the presence or absence of a restriction fragment of a particular size. As these data represent restriction enzyme sites, we weighted the costs of transformation for each character state using stepmatrices in PAUP, as described (Albert et al. 1992; Sobral et al. 1993). For these data, our method of scoring is called fragment "direct" analysis (FDA) by Bremer (1991). We are confident of homology of fragments because D'Hont et al. (1993) generated their data using Southern hybridization with nine conserved mitochondrial gene probes from wheat. However, independence of characters is violated by scoring in this manner without knowing the restriction maps of the probes (not reported by D'Hont et al. 1993).

## Results

### Primer performance

Initially, we tested primers on *Zea mays* and *Arabidopsis thaliana* to compare the sequences amplified with those reported in GENBANK. Amplified sequences matched reported sequences. We then tested the spectrum of plant taxa in which the primers would be useful. We included two *Pinus* species (pinyon pines, kindly provided by Paul Keim, Northern Arizona University, Flagstaff), *P. edulis* and *P. californiarum* in our amplifications, in addition to the plants listed in Table 1. Generally, the primers worked well in angiosperms and poorly or not at all in pinyon pine, at least under our high stringency conditions. Primer pairs that did yield products in pinyon pine were from the mitochondrial genome and more than one fragment was amplified (not shown).

### Sequence variation

The sequenced regions from representatives of the *Saccharum* complex were aligned with GENBANK sequences. In Table 2, these data are shown and genomes having the same scores are grouped. Our strategy was to first sequence and compare maize and sorghum with representative *Saccharum* genotypes (Table 2). If polymorphisms were not observed between maize and sugarcane, as in the region of the mitochondrial 18*S* primed by oligonucleotide 7A or the intergenic region of the chloroplast genome primed by oligonucleotides 1B and 2A, then sequencing of accessions within the *Saccharum* complex was not done. Cycle-sequencing of the ends of the intergenic region between *rbcL* and *atpB*, using oligonucleotides 6A and 5A, revealed some chloroplast-borne polymorphism (Table 2), so this region was cloned and sequenced in its entirety.

**Table 3** Summary of the mapped restriction site polymorphism (MRSP) study. Each group of accessions separated by a blank from the next group has the same scores as the first representative of that group. Numbers in parentheses indicate number of accessions studied

| Accessions | S1[a] | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | LV1[b] | LV2 | LV3 | LV4 | LV5 | LV6 | LV7 | LV8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *E. kanashiori*<br>*Sorghum plumosum* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| *S. officinarum* (2)<br>*Coix gigantea*<br>*E. arundinanceus*<br>*E. bengalense*<br>*Eccoilopus longisetosus*<br>*E. procerus*<br>*E. ravennae*<br>*S. barberi* (2)<br>*S. edule*<br>*S. robustum* (4)<br>*S. sinense* (2)<br>*Vetiveria sp*<br>*Saccharum sp.* (7) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| *S. spontaneum* (8)<br>*Miscanthus sinensis*<br>*E. trinii*<br>*Narenga porphyrocoma*<br>*Sclerostachya fusca*<br>*Imperata cylindrica* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| *Sorghum bicolor*<br>*Sorghum alum*<br>*Sorghum halepense* | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| *Zea mays* | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |

[a] S1-S10: restriction sites: S1, *Ple*I, S2, *Sau96*I; S3, *Hph*I; S4, *Mbo*I; S5, *EcoO109*I; S6, *Hha*I; S7-S10, *Hinf*I
[b] LV1-LV8: length variants (position in bp based on map in Fig. 3): LV1, at position 172; LV2, at position 324; LV3, at position 346; LV4, at position 391; LV5, at position 427; LV6, 504; LV7, at position 525; and LV8, at position 544

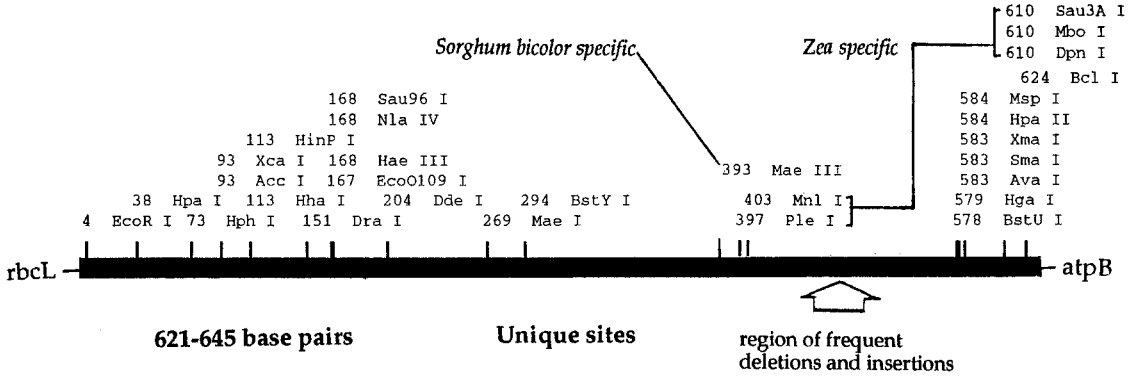


**Fig. 2** Unique restriction sites in the *rbcL-atpB* spacer. A restriction map of the spacer was created by analyzing the sequence data collected on 18 accessions. Unique sites occurred only once in the sequence

Mitochondrial polymorphisms were observed after sequencing with oligonucleotides 3A and 4A (Table 2), but the presence or absence of 2 indels were all that separated the accessions into three groups. *Erianthus/Eccoilopus* accessions showed the same mitochondrial sequence as sorghum, and a presumably hybrid, high chromosome number (2n=192) *Miscanthus* sp. from New Guinea showed the same sequence as *Saccharum/Narenga*, in agreement with previous results obtained from an analysis of chloroplast restriction site mutations (Sobral et al. 1993). Modern cane varieties, which are interspecific hybrids between *S. officinarum* and *S. spontaneum*, in which *S. officinarum* generally serves as the female, recurrent parent, did not show variability for loci sequenced in this preliminary evaluation (Table 2). This was not surprising because no variation was observed within the *Saccharum* genus.

To determine whether the limited variability observed within the *Saccharum* complex was due to an undersampling of taxa, we conducted a study using the restriction map (Fig. 3) generated by the complete sequence of the intergeneric *rbcL-atpB* region for the taxa listed in Table 1. Seven restriction enzymes were identified that revealed a

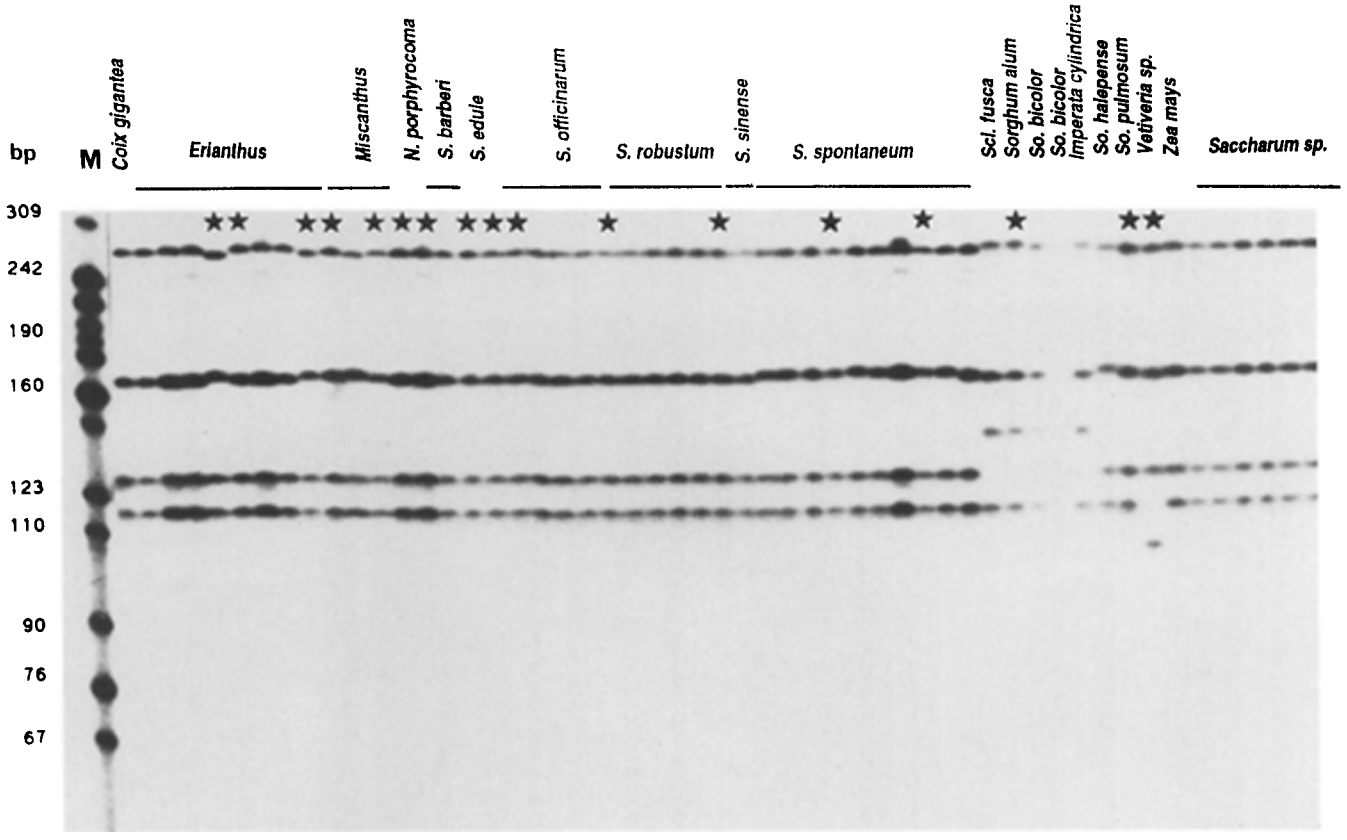


**Fig. 3** Example of MRSP analysis using *Hinf*I restriction digest resolved on 6% polyacrylamide gel. * indicates accessions for which complete sequences of the *rbcL-atpB* spacer were determined, *M* pBR322 /*Hpa*II molecular weight marker

total of ten target sites within the region and eight length variants. None of the site mutations and only four of the indels were informative (Table 3). An additional 36 accessions (besides those sequenced, see Table 1) were analyzed by MRSP including 7 *S. spontaneum* accessions chosen to represent cyto/geographical groups and 7 commercial varieties that are in the maternal lineage of almost all sugarcane grown today. Furthermore, 4 *S. robustum* accessions from New Guinea, some of which have been implicated in the origin of sugarcane (Daniels and Roach 1987), were included. An example of MRSP analysis resolved on a sequencing gel is shown in Fig. 3. At this level of resolution, we obtained five groups of genotypes that each had the same score for all restriction sites and length mutations (Table 3). Although MRSP analysis cannot yield the same level of phylogenetic resolution as sequencing, it revealed that the limited variability observed within the *Saccharum* complex and close relatives was likely not due to undersampling of taxa, at least for this locus. This is a reasonable conclusion because the indels were more frequently polymorphic than point mutations within the ingroup taxa, and MRSP data showed products of the same length as a sequenced accession of the same species for the additional accessions analyzed.

## Phylogenetic analysis

Manual alignment of 664 nucleotides from the *rbcL-atpB* spacer region yielded 52 site mutations and 55 insertion/deletion events (shown in Table 4) when all taxa were considered. Of these, only 7 site mutations and 16 insertion/deletion events were informative within the proposed *Saccharum* complex accessions (Mukherjee 1957; Daniels and Williams 1975). The greatest number of site mutations between any two *Saccharum* complex accessions was between *S. officinarum* Black Cheribon and *Narenga porphyrocoma*, which differed by 7 site mutations and had an average genetic distance difference of 0.302 (Table 5). This was the same number of site mutations observed between these 2 taxa and *Sorghum bicolor*, for which an average genetic distance of less than 0.3 was calculated.

Informative polymorphisms allowed placement of the 18 sequences into 12 terminal taxa, plus wheat, rice, and barley sequences obtained from GENBANK. DNA sequences for *S. barberi, S. edule, S. officinarum* NG 51-131, and *S. robustum* were identical, and the terminal taxon representing this group of accessions is *S. robustum*. Likewise, the two *S. spontaneum* accessions sequenced were identical, as were *S. officinarum* 'Black Cheribon' and *S. sinense*, the latter pair forming a terminal taxon represented by 'Black Cheribon'.

Pairwise genetic distances computed in PAUP from analysis of the character matrix are shown in Table 5. Branch-and-bound searches, sometimes weighting transitions 1.5×transversions (Zurawski et al. 1984), yielded 79

**Table 4** Input data matrix for PAUP. Numbers in header refer to character number in the data matrix. Node refers to taxon designation in trees (Fig. 4)

| Node | 1111111111111111111122223333333333333333344444444444455555555556666666666666666666666666666666666666666666666666777777777777777777 |
|---|---|
| | 1134446788012333345577788899990228001112344566778880112224478990022233335111222334445566666777777777788888888899999999900000000001111111111 |
| | 1473057357850804563894571562578909908245615984902078519567244903344560389213712617567175678901234567890123456789012345678901234567890123456789 |
| *O. sativa* | TACTCTCACCTAAT?CTTC??TACTCTTGGGAGCATAGTTAGGTCGGCCGCGTCTGCTCCGGGTGGCAAGATCTTGCGTTTATACGC10111111110100001110100111010101100111111101110100111110 |
| *H. vulgare* | T??GTCGATCAGG?TTCCC??GTTTCTGAGCCTTGTGCCCAGGTCAACCGCGCCTCCTCCTTTG??CAATAGACTGCGTCGATACTC0000010110011110011011101110101011010000001011101000011O |
| *T. aestivum* | T??GTCGATAAGA?TTTCC??TTTTCTTAGCCTTGTGCCCAGGTCAACCGCGCCTCCTCCTTTG??CAATAGACTCCGCCTATACTC11100011001110101101110111010101101000000101110100001l0 |
| Zm DeKalbW77 | AGAGTCGGCCAGACCTC?TTT?CCCGATACTAGCGTGCTC?T?GCAACTACACCGGC?TTGGGT??CAAGAA?CGGCGTCTACGTGT11111011011011110000011000010101010110000111100110111l0 |
| *E. kanashiori* | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTCTTATCAACCACACCTCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011111110100011110000101010110111l0 |
| *E. longisetosus* | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTCTTATTAACCACACCTCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011110010100010110000101010100111l0 |
| *E. trinii* | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTCTTATCAACCACACCTCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000010111101100010110000101010110111l0 |
| *S. barberi* | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011010010100010110000101010110111l0 |
| *Sorghum bicolor* | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACGCTTC??TTGGGT??ATTCGAATGGCGTCTGCGTGT11111010010111111000011010010101010110000000011011001111 |
| *S. edule* | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGTT????CGAACGGCGTCTACGTGT11111010010111111000011010010100010110000101010110111l0 |
| *So.* NG51-131 | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGTT????CGAACGGCGTCTACGTGT11111010010111111000011010010100010110000101010110111l0 |
| *S. robustum* | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGTT????CGAACGGCGTCTACGTGT11111010010111111000011010010100010110000101010110111l0 |
| *S. spontaneum* C | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011011110100010110000101010110111l0 |
| *S. spontaneum* S | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011011110100010110000101010110111l0 |
| *Vetiveri* spp. | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGTGA???CGAACGGCGTCTACGTGT11111010010111111000011010010100010111101101011100111l0 |
| *So.* B Cheribon | AGAGTCGGCCAGAC?TCCTTAGCCCGCTACTAGCGTGCTC?TATCAACTACACCGCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011010010001010110000111010110111l0 |
| *S. sinense* | AGAGTCGGCCAGAC?TCCTTAGCCCGCTACTAGCGTGCTC?TATCAACTACACCGCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011010010001010110000111010110111l0 |
| *Narenga* Bor | AGAGTCGGCCAGAC?TCCTTAGCCCGATACTAGCGAGCTC?TATCAACCAACCCTCG?TTGGGT?????CGAACGGG?TCTACGTGT11111010010111111000011011110011010110000111010110110l0 |
| *Miscanthus* sp. | AGAGTCGGCCAGAC?TCCTTTGCCCGATACTAGCGTGCTC?TATCAACCACACCTCGGTTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011011110001110110000101010110111l0 |
| *M. sinensis* Z | AGAGTCGGCCAGAC?TCCTAAGCCCGATACTAGCGTGCTC?TATCAACCACACCTCG?TTGGGT?????CGAACGGCGTCTACGTGT11111010010111111000011011110001010110000101010110111l0 |

**Table 5** Pairwise genetic distances[a] between taxa

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 *O. sativa* | – | 0.455 | 0.440 | 0.542 | 0.556 | 0.549 | 0.556 | 0.553 | 0.575 | 0.556 | 0.556 | 0.556 | 0.553 | 0.553 | 0.530 | 0.576 | 0.576 | 0.603 | 0.549 | 0.561 |
| 2 *H. vulgare* | 61 | – | 0.104 | 0.512 | 0.477 | 0.485 | 0.492 | 0.488 | 0.511 | 0.488 | 0.488 | 0.488 | 0.488 | 0.488 | 0.512 | 0.512 | 0.512 | 0.539 | 0.485 | 0.496 |
| 3 *T. aestivum* | 59 | 14 | – | 0.504 | 0.477 | 0.485 | 0.492 | 0.488 | 0.511 | 0.488 | 0.488 | 0.488 | 0.488 | 0.488 | 0.512 | 0.512 | 0.512 | 0.539 | 0.485 | 0.496 |
| 4 *Z. mays* | 71 | 66 | 65 | – | 0.162 | 0.154 | 0.177 | 0.131 | 0.189 | 0.131 | 0.131 | 0.131 | 0.146 | 0.146 | 0.169 | 0.115 | 0.115 | 0.186 | 0.146 | 0.162 |
| 5 *E. kanashiori* | 74 | 62 | 62 | 21 | – | 0.037 | 0.030 | 0.030 | 0.105 | 0.030 | 0.030 | 0.030 | 0.015 | 0.015 | 0.067 | 0.075 | 0.075 | 0.083 | 0.045 | 0.045 |
| 6 *E. longisetosus* | 73 | 63 | 63 | 20 | 5 | – | 0.052 | 0.022 | 0.098 | 0.022 | 0.022 | 0.022 | 0.037 | 0.037 | 0.045 | 0.067 | 0.067 | 0.105 | 0.067 | 0.067 |
| 7 *E. trinii* | 74 | 64 | 64 | 23 | 4 | 7 | – | 0.045 | 0.120 | 0.045 | 0.045 | 0.045 | 0.030 | 0.030 | 0.082 | 0.090 | 0.090 | 0.098 | 0.060 | 0.060 |
| 8 *S. barberi* | 73 | 63 | 63 | 17 | 4 | 3 | 6 | – | 0.075 | 0.000 | 0.000 | 0.000 | 0.015 | 0.015 | 0.037 | 0.045 | 0.045 | 0.083 | 0.045 | 0.045 |
| 9 *Sorghum bicolor* | 77 | 67 | 67 | 25 | 14 | 13 | 16 | 10 | – | 0.075 | 0.075 | 0.075 | 0.090 | 0.090 | 0.113 | 0.105 | 0.105 | 0.136 | 0.105 | 0.105 |
| 10 *S. edule* | 74 | 63 | 63 | 17 | 4 | 3 | 6 | 0 | 10 | – | 0.000 | 0.000 | 0.015 | 0.015 | 0.044 | 0.045 | 0.045 | 0.083 | 0.045 | 0.045 |
| 11 *So* NG51-131 | 74 | 63 | 63 | 17 | 4 | 3 | 6 | 0 | 10 | 0 | – | 0.000 | 0.015 | 0.015 | 0.044 | 0.045 | 0.045 | 0.083 | 0.045 | 0.045 |
| 12 *S. robustum* | 74 | 63 | 63 | 17 | 4 | 3 | 6 | 0 | 10 | 0 | 0 | – | 0.015 | 0.015 | 0.044 | 0.045 | 0.045 | 0.083 | 0.045 | 0.045 |
| 13 *S. spontaneum* C | 73 | 63 | 63 | 19 | 2 | 5 | 4 | 2 | 12 | 2 | 2 | 2 | – | 0.000 | 0.052 | 0.060 | 0.060 | 0.068 | 0.030 | 0.030 |
| 14 *S. spontaneum* S | 73 | 63 | 63 | 19 | 2 | 5 | 4 | 2 | 12 | 2 | 2 | 2 | 0 | – | 0.052 | 0.060 | 0.060 | 0.068 | 0.030 | 0.030 |
| 15 *Vetiveri* spp. | 71 | 66 | 66 | 22 | 9 | 6 | 11 | 5 | 15 | 6 | 6 | 6 | 7 | 7 | – | 0.082 | 0.082 | 0.120 | 0.082 | 0.082 |
| 16 *So.* B Cheribon | 76 | 66 | 66 | 15 | 10 | 9 | 12 | 6 | 14 | 6 | 6 | 6 | 8 | 8 | 11 | – | 0.000 | 0.083 | 0.060 | 0.060 |
| 17 *S. sinense* | 76 | 66 | 66 | 15 | 10 | 9 | 12 | 6 | 14 | 6 | 6 | 6 | 8 | 8 | 11 | 0 | – | 0.083 | 0.060 | 0.060 |
| 18 *Narengapor* | 79 | 69 | 69 | 24 | 11 | 14 | 13 | 11 | 18 | 11 | 11 | 11 | 9 | 9 | 16 | 11 | 11 | – | 0.068 | 0.068 |
| 19 *Miscanthus* sp. | 73 | 63 | 63 | 19 | 6 | 9 | 8 | 6 | 14 | 6 | 6 | 6 | 4 | 4 | 11 | 8 | 8 | 9 | – | 0.030 |
| 20 *M. sinensis* Z | 74 | 64 | 64 | 21 | 6 | 9 | 8 | 6 | 14 | 6 | 6 | 6 | 4 | 4 | 11 | 8 | 8 | 9 | 4 | – |

[a] Below diagonal: absolute distances; above diagonal: mean distances (adjusted for missing data). Distances were calculated as described in PAUP 3.1.1 (Swofford 1991)

equally parsimonious trees. A 50% majority rule consensus of these trees is shown in Fig. 4A. Topology of the consensus tree was unaffected by weighting transitions in relation to transversions. The phylogenetic hypothesis shown indicates that sorghum, maize, rice, wheat, and barley occupied outgroup positions, as expected, and that the ingroup, composed of *Saccharum* complex accessions, was monophyletic and displayed extremely low sequence variability. Furthermore, a strict consensus tree showed "raking" of the ingroup although a clear separation of *Erianthus* and *Vetiveria* was maintained (Fig. 4B; see also Table 4). If indels were ignored, then sequences from *E. kanashiori*, *E. trinii*, *S. barberi*, *S. edule*, *S. officinarum* NG 51–131, *S. robustum*, both *S. spontaneum* accessions, and *Vetiveria* sp. were the same. Phylogenetic analysis was also conducted using only maize as an outgroup. Under these conditions, six equally parsimonious trees were obtained (not shown). These trees preserved the general topology of the trees shown in Fig. 4.

Forty point mutations between maize and barley allowed us to calculate the nucleotide substitution rate in this region. Assuming a divergence time between maize and barley of 50 million years ago (mya) (Zurawski et al. 1984), we calculated a rate of $1.2\times10^{-9}$ substitutions/site per year. This is almost the same as that calculated by others (Zurawski et al. 1984; Doebley and Stec 1992). If we assume a constant rate of substitution among all the lineages in this study, then it would appear that the entire *Saccharum* complex is very young and that the common maternal ancestor of *Sorghum* and the *Saccharum* complex may have existed about 5 mya (Fig. 4C).

The re-coding of mitochondrial data in D'Hont et al. (1993) as 16 binary characters (14 polymorphic) (FDA, Bremer 1991) allowed collapsing of the 53 accessions studied into 10 taxa, defined by 9 informative characters and 5 uninformative ones. Representative accessions, for which there were no missing data, were used as input for phylogenetic analyses in which characters were unordered (Wagner tree, not shown). Phylogenetic analysis yielded a single most parsimonious unrooted tree found by an exhaustive PAUP search. There were 20 trees that were one step longer and 178 trees that were two steps longer (not shown). Five thousand bootstrap resamplings of these data yielded very low confidence limits (not shown) because very few characters were informative within *Saccharum*. Unfortunately, accessions studied by D'Hont et al. (1994) and Sobral et al. (1993) were different, so that a joint analysis of all the data could not be conducted.

## Discussion

Sobral et al. (1994) showed that there is little variation within the chloroplast genomes of the *Saccharum* complex by scanning the genome with 15 restriction enzymes and 12 probes that covered the entire chloroplast genome. The present study, although limited to a few loci, confirms those results and is also in agreement with the results of D'Hont

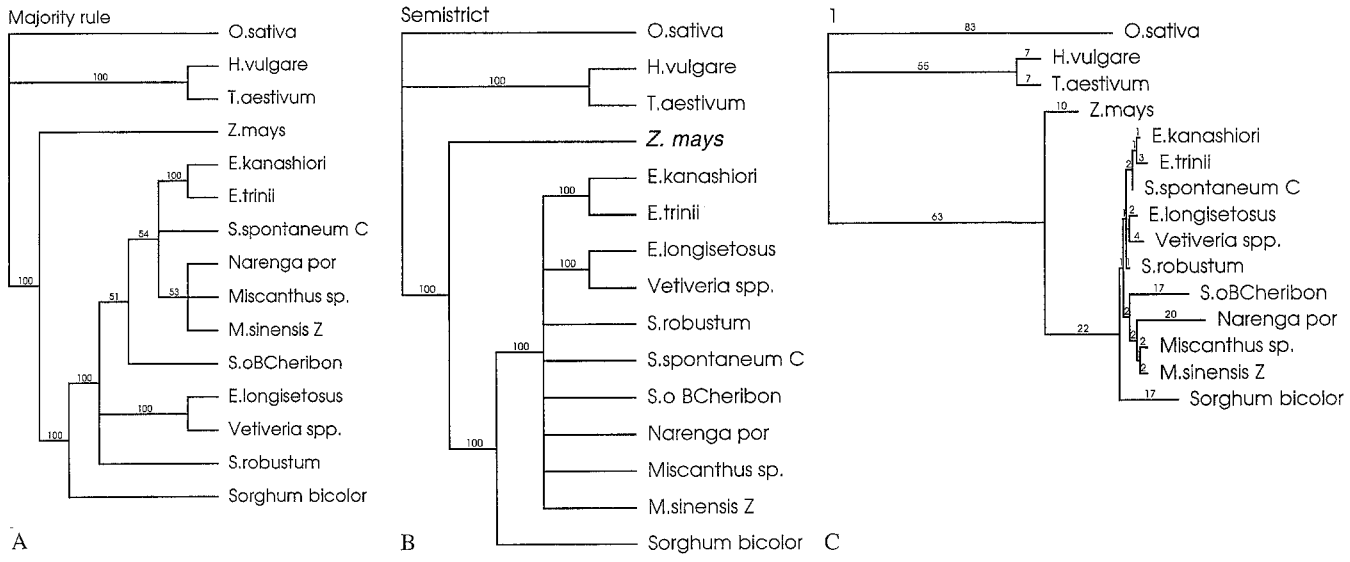


**Fig. 4A–C** Phylogenetic hypothesis generated by analysis of *rbcL-atpB* sequences. **A** Fifty percent majority rule of 79 equally parsimonious trees generated from analysis of 664 nucleotides and 55 insertion/deletion events scored as unordered binary characters (1,0); *numbers* on branches refer to number of times (in percentage) in the 79 trees in which the bifurcation was supported. **B** Semistrict consensus of 79 trees, as in **A**. **C** Example of 1 of the 79 equally parsimonious trees, represented as a phylogram in which branch lengths (shown above lines) are proportional to genetic distances calculated in PAUP. In these trees, the following terminal taxa represent more than one accession: *Z. mays* (2 genotypes sequenced); *S. robustum* = *S. barberi* = *S. edule* = *S. officinarum* NG 51-131; *S. officinarum* Black Cheribon = *S. sinense*

et al. (1993). Therefore, it is clear that the limited variability of maternal genomes in the *Saccharum* complex is not due to the study of a limited number of accessions, limited number of loci, or an artifact of the studies. Furthermore, cultivated sugarcane displayed no detectable chloroplast diversity, also in agreement with Sobral et al. (1994), suggesting a world-wide cytoplasmic monoculture for sugarcane. This is in contrast to the situation in the closely related annual *Sorghum bicolor* (Duvall and Doebley 1990), for which cytoplasmic diversity has been demonstrated and shown to be at least as great as the diversity we have revealed among *Saccharum* complex members. The amount of chloroplast variation that the current and previous studies (Sobral et al. 1994; D'Hont et al. 1993) have revealed within the *Saccharum* complex is similar to the amount of chloroplast variation found within *Zea* or *Sorghum* (summarized in Soltis et al. 1992), two other Andropogoneae.

We observed frequent insertion/deletion events (indels) in the *rbcL-atpB* spacer region (Fig. 3), which is in agreement with Zurawski and Clegg (1987), who showed that this type of mutation is a common feature of chloroplast noncoding regions. Indels were more useful than point mutations in separating members within the *Saccharum* complex from each other but were insufficient to provide clear phylogenetic resolution, as can be seen from the trees in Fig. 4. Rapid rates of indels (with respect to site mutation) in noncoding regions of the chloroplast have been used to help obtain phylogenetic resolution among closely related genotypes of *Pennisetum* and *Cenchrus* (Zurawski and Clegg 1987). The *rbcL-atpB* spacer region should be particularly useful in providing phylogenetic resolution at the genus level in most cases, although within a cytoplasmically invariant group as the *Saccharum* complex, they did not provide sufficient resolution.

Sequenced mitochondrial loci were nearly invariant within *Saccharum* complex members, supporting conclusions drawn from the analysis of the chloroplast regarding the maternal lineages studied herein. In addition, *Miscanthus* sp. NG 77-193, which is presumed to be of hybrid origin because of its high chromosome number, displayed the same mitochondrial type as *Saccharum* species, whereas *M. sinensis* showed a different type (Tables 1, 2). This further substantiates the possibility of intergeneric hybridization in the wild, as now chloroplast and mitochondrial sequences of this accession have been shown to be the same as those of *Saccharum*, although we cannot exclude polymorphism within *Miscanthus* as an alternative explanation because of the limited number of accessions studied.

Mitochondrial sequences provided little resolution between the *Saccharum* complex and sorghum. Sorghum is one of the closest non-complex relatives of these plants. Data from D'Hont et al. (1993) was used for phylogenetic reconstruction following a scoring scheme proposed by Bremer (1991), called FDA, although the input data in this case violate the assumption of independence of characters. No outgroup species were included in the analysis of D'Hont et al. (1993). However, on the basis of the conclusions of Sobral et al. (1994), *Erianthus* may be an appropriate outgroup species and was therefore used in our analysis of their data. Mitochondrial probes revealed a larger amount of variability present in *S. spontaneum*, than in other *Saccharum* complex accessions, although many of the other accessions were probably hybrid in origin; either artificial or, presumably, naturally occurring. Large variability in *S. spontaneum* is in agreement with morphological, geographical, and chemotaxonomic data (Daniels and Roach 1987). One grouping within *S. robustum* accessions

from Papua New Guinea, representing different cytological, morphological, and chemotaxonomical types, was also identified by these probes and may support exclusion of 2n=60 forms of *S. robustum* in the evolution of *S. officinarum* (2n=80). More intra-complex variability needs to be uncovered before within-complex relationships can be understood. The mitochondrial genome may be useful in providing clues.

It may be surprising that taxa within the *Saccharum* complex, which contain plants with diverse chromosome number and proposed geographic origin actually display such little variation within their chloroplast genome. This may be caused by the relatively recent evolution of the complex, although other explanations are possible. It has been shown that morphological variation within a species and chloroplast DNA sequence variation do not necessarily correlate (Soltis et al. 1992), although some authors have considered many of these genera to be synonomous (Clayton and Renvoize 1986). On purely speculative grounds, it is tempting to imagine that some differences among these plants, despite large variations in ploidy, may be due to a few genes with large effects, such as has been observed for some morphological variation between cultivated maize and teosinte (Doebley and Stec 1991). If such were the case, it could be that small numbers of genes, possibly with large effects, account for the variation in sucrose levels, even though this is continuous in most commercial crosses (unpublished data).

Previous results revealed a distinct chloroplast genome in the *Erianthus/Eccoilopus* species (Sobral et al. 1994). The present results clearly separate *Erianthus/Eccoilopus* accessions from *Saccharum* for chloroplast and mitochondrial loci. Sobral et al. (1993) suggested that *Erianthus/Eccoilopus* should be removed from the informal taxonomic grouping known as the *Saccharum* complex based on a variety of nuclear and cytoplasmic evidence. If the entirety of the molecular data is examined, including other loci and the mitochondrial data of D'Hont et al. (1993), it is highly suggestive of evolutionary differentiation between lineages of *Erianthus/Eccoilopus* and those of the remainder of the complex. Separation may have occurred almost as long ago as the *Saccharum*-sorghum separation. This is interesting because current taxonomy suggests that *Erianthus* is synonymous with *Saccharum* (Clayton and Renvoize 1986). It is also notable that *Erianthus* is the only example within the *Saccharum* complex that has New World species.

Nucleotide sequence divergence rates calculated for the separation of the maize-barley clades were similar to those obtained by a more extensive analysis of this region (Zurawski and Clegg 1984). If a molecular clock is assumed, it suggests that the *Saccharum* complex is young indeed. Branch lengths in any of the 79 equally parsimonious solutions of the data, calibrated using the maize-barley divergence time of 50 mya, suggest that the common ancestor of *Sorghum* and *Saccharum* complex lineages existed less than 5 mya. However, rate differences have been shown between palms and annual species and vegetative growth was invoked as a possible explanation for the different evolutionary rates (Wilson et al. 1990). We cannot differentiate between different rates of evolution and the recent evolution of the complex using our data alone.

*S. spontaneum* has the widest geographic range of the *Saccharum* complex members, excluding *Erianthus*, and some of its populations may provide additional clues to the origin of the polyploid complex. With respect to *S. spontaneum*, it is perhaps interesting to note that at least 1 *S. spontaneum* genotype, and presumably other members of the species, displays random chromosome pairing and assortment, typical of autopolyploid species, and may be an auto-octaploid species (Al-Janabi et al. 1993; Da Silva et al. 1993). This suggests that repeated autopolyploidization, perhaps superimposed on early allopolyploidization events, may be an important evolutionary force in the speciation of plants within the *Saccharum* complex. Studies are underway to substantiate the nature of chromosome pairing and assortment in other species within *Saccharum*. Further collecting in strategic regions is warranted, especially because of eroding habitats.

**Acknowledgements** We thank James Irvine (Texas Agricultural Experiment Station, Weslaco) and David Burner (ARS-USDA Houma LA) for numerous conversations and suggestions and for harvesting and identifying most of the plant materials used herein, Elizabeth Kellogg (Harvard Arboretum) and Elizabeth Zimmer (Smithsonian Institution National Museum of Natural History, Washington DC) for encouragement, Paul Keim (Northern Arizona University, Flagstaff) for friendship and pinyon pine DNAs, Miguel Peinado (California Institute of Biological Research, La Jolla, Calif.) for magnificent technical assistance with cycle-sequencing, Rhonda Honeycutt (California Institute of Biological Research, La Jolla) for critical reviews, suggestions, and extraction of numerous DNA samples, and Craig Hobart (California Institute of Biological Research) for technical assistance. This work was supported by a grant from the International Consortium for Sugarcane Biotechnology to BWSS. MM and CP were partially supported by NIH grants AI 34829 and HG 00456.

## References

Albert VA, Mischler BD, Chase MW (1992) Character-state weighting for restriction site data in phylogenetic reconstruction, with an example from chloroplast DNA. In: Soltis PS, Soltis DE, and Doyle JJ (eds) Molecular systematics of plants, Chapman and Hall, New York, pp 349–403

Al-Janabi, SM, Honeycutt RJ, McClelland M, Sobral BWS. (1993) A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. Genetics 134:1249–1260

Artschwager E (1954) A taxonomic study of *Saccharum sinense* Roxb. and *S. barberi* Jeswiet. USDA Technical Bulletin No. 1089, Washington D.C.

Artschwager E, Brandes EW (1958) Sugarcane (*Saccharum officinarum* L.): origin, classification, characteristics, and descriptions of representative clones. USDA Handbook No. 122, Washington, D. C.

Bellwood P (1985) Prehistory of the Indo-Malaysian archipelago. Academic Press, Orlando, Fa.

Birky CW (1988) Evolution and variation in plant chloroplast and mitochondrial genomes. In Gottlieb LD, IAIN SK (eds): "Plant evolutionary biology". Chapman and Hall, London, pp. 23–53

Bowman CM, Bonnard G, Dyer TA (1983) Chloroplast DNA variation between species of *Triticum* and *Aegilops*. Location of the variation on the chloroplast genome and its relevance to the inheritance and classification of the cytoplasm. Theor Appl Genet 65:247–262

Brandes EW (1928) Into primeval Papua by seaplane. Natl Geographic 56:253–332
Brandes EW, Sartoris GB, Grassl CO (1939) Assembling and evaluating wild forms of sugarcane and closely related plants. In: Proc. 6th Int Soc Sugarcane Technol Congr., pp. 128–154
Bremer B (1991) Restriction site data from chloroplast DNA for phylogenetic reconstruction: is there only one accurate way of scoring? Plant Syst Evol 175:39–54
Burner DM (1991) Cytogenetic analyses of sugarcane relatives in the subtribe *Saccharinae*. Euphytica 54:125–133
Celarier RP (1956) Cytotaxonomy of the *Andropogoneae*. 1. Subtribes *Dimeriinae* and *Saccharinae*. Cytologia 21:272–291
Clayton WD, Renvoize SA (1986) Genera Graminum. Grasses of the world. Kew Bulletin Additional Series XIII, Royal Botanical Gardens, Kew, Her Majesty's Stationery Office, London
Da Silva J, Sorrells ME, Burnquist WL, Tanksley SD (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome, *in press*
Daniels J, Roach BT (1987) Taxonomy and evolution in breeding. In: Heinz DJ (ed). Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 7–84
Daniels J, Williams CA (1975) The origin of the genus *Saccharum*. ISSCT Sugarcane Breed Newsl 36:24–39
Daniels J, Paton H, Smith P, Williams C (1980) Further studies on leaf flavonoids as evolutionary indicators in *Saccharum officinarum* L. In: Proc 17th XVII ISSCT Congr., pp. 1317–1335
D'Hont A, Lu YH, Feldmann P, Glaszmann JC (1993) Cytoplasmic diversity in sugarcane revealed by heterologous probes. Sugar Cane 1993:12–15
Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. Genetics 129:285–295
Duvall MR, Doebley J (1990) Restriction site variation in the chloroplast genome of *Sorghum* (Poaceae). Syst Bot 15:472–480
Felsenstein J (1989) PHYLIP- phylogeny inference package. Cladistics 5:164–166
Glaszmann JC, Lu YH, Lanaud C (1990). Variation of nuclear ribosomal DNA in sugarcane. J Genet Breed 44:191–198
Hamby RK, Zimmer EA (1988) Ribosomal RNA sequences for inferring phylogeny within the grass family (*Poaceae*). Plant Syst Evol 160:29–37
Hartley W (1958) Studies on the origin, evolution, and distribution of the *Graminiae*. I. The tribe *Andropogoneae*. Aust J Bot 6: 116–128
Honeycutt RJ, Sobral BWS, Keim P, Irvine JE (1992) A rapid DNA extraction method for sugarcane and its relatives. Plant Mol Bio Rep 10:66–72
Liston A (1992) Variation in the chloroplast genes *rpoc1* and *rpoc2* of the genus *Astragalus* (Fabaceae): evidence from restriction site mapping of a PCR-amplified fragment. Am J Bot 79:953–961
Maniatis T , Fritch EF, Sambrook J (1982) Molecular cloning labrotary manual. Cold Spring Harbor Labrotary Press, Cold Spring Harbor, N.Y.
McIntosh L, Poulsen C, Bogorad L (1980) Chloroplast gene sequence for the large subunit of ribulose bisphosphatecarboxylase of maize. Nature 288:556–560
Mohan N, Sreenivasan TV (1983) Chromosome number in the genus *Erianthus*, Michx. (Poaceae) of Indonesian Arquipelago. Cell Chromosome Res 6:14–16
Moriya A (1940) List of chromosome numbers in the genus *Saccharum* and related genera. Jpn J Genet 16:126–136
Mukherjee SK (1957) Origin and distribution of *Saccharum*. Bot Gaz 119:55–61
Panje RR, Babu CN (1960) Studies in *Saccharum spontaneum* distribution and geographical association of chromosome numbers. Cytologia 25:152–172
Price S (1957) Cytological studies in *Saccharum* and allied genera. III. Chromosome numbers in interspecific hybrids. Bot Gaz (March):146–159
Price S (1968) Chromosome transmission by *Saccharum robustum* in interspecific crosses. J Hered 59:245–247
Ralph D McClleland, M Welsh J, Branton G, Perolat P (1993) Leptospira species categorized by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified rRNA genes. J Bacteriol 175:973–981
Ruanto G, Kidd KK (1991) Coupled amplification and sequencing of genomic DNA. Proc Natl Acad Sci USA 88:2815–2819
Salts Y, Herrmann RG, Peleg N, Lavi U, Izhar S, Frankel R, Beckmann JS (1984) Physical mapping of plastid DNA variation among eleven *Nicotiana* species. Theor Appl Genet 69:1–14
Sanger F, Coulsen AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441–448
Smith RM, Martin-Smith M (1978) Triterpene methyl ethers in leaf waxes of *Saccharum* and related genera. Phytochemistry 17:1307–1312
Sobral BWS, Braga DPV, LaHood E, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae:Andropogoneae. Theor Appl Genet (in press)
Soltis PS, Doyle JJ, Soltis DE (1992) Molecular data and polyploid evolution in plants. In: Soltis PS, Soltis DE, Doyle JJ (eds). Molecular systematics of plants. Chapman and Hall, New York, pp 177–201
Springer PS, Zimmer EA, Bennetzen JL (1989) Genomic organization of the ribosomal DNA of sorghum and its close relatives. Theor Appl Genet 77:844–850
Stebbins GL (1956) Cytogenetics and evolution of the grass family. Am J Bot 43:890–905
Swofford DL (1991) PAUP: Phylogenetic Analysis Using Parsimony, version 3.1 s. Computer program distributed by the Illinois Natural History Survey, Champaign, Ill.
Swofford DL, Olsen GJ (1990) Phylogeny reconstruction. In: Hillis DM, Moritz C (eds) Molecular Systematics. Sinauer Assoc, Sunderland, Mass., pp. 411–501
Waldron JC, Glasziou KT, Daniels J (1974) B-amylase isoenzymes as genetic markers in *Saccharum* and related genera. In: Proc. 15th ISSCT Congr.. pp. 145–152
Watson L, Clifford HT, Dallwitz MJ (1985) The classification of the *Poaceae*: subfamilies and supertribes. Aust J Bot 33:433–484
Williams CA, Harborne JB (1974) The taxonomic significance of leaf flavonoids in *Saccharum* and related genera. Phytochemistry 13:1141–1149
Wilson MA, Gaut B, Clegg MT (1990) Chloroplast DNA evolves slowly in the palm family (Arecaceae). Mol Biol Evol 7:303–314
Zurawski G, Clegg MT (1987) Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Annu Rev Plant
Physiol 38:391–418
Zurawski G, Clegg MT, Brown AHD (1984) The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735–749