

RESEARCH ARTICLE

# Complete Chloroplast Genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: Comparative Genomics and Evolution of the *Saccharum* Complex

Shin-ichi Tsuruta<sup>1\*</sup>, Masumi Ebina<sup>2</sup>, Makoto Kobayashi<sup>2</sup>, Wataru Takahashi<sup>2</sup>

**1** Japan International Research Center of Agricultural Sciences, Tropical Agriculture Research Front (JIRCAS-TARF), Okinawa, Japan, **2** National Agriculture and Food Research Organization, Institute of Livestock and Grassland Science (NARO-ILGS), Tochigi, Japan

\* [stsuru@affrc.go.jp](mailto:stsuru@affrc.go.jp)



**OPEN ACCESS**

**Citation:** Tsuruta S-i, Ebina M, Kobayashi M, Takahashi W (2017) Complete Chloroplast Genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: Comparative Genomics and Evolution of the *Saccharum* Complex. PLoS ONE 12(1): e0169992. doi:10.1371/journal.pone.0169992

**Editor:** Berthold Heinze, Austrian Federal Research Centre for Forests BFW, AUSTRIA

**Received:** July 15, 2016

**Accepted:** December 27, 2016

**Published:** January 26, 2017

**Copyright:** © 2017 Tsuruta et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequence data are available from the GenBank/DDBJ/EMBL database (accession numbers: LC160130 and LC160131) or from the following URLs: *Erianthus*: <https://www.ncbi.nlm.nih.gov/nuccore/LC160130.1> *Miscanthus*: <https://www.ncbi.nlm.nih.gov/nuccore/LC160131.1>

**Funding:** This work was supported by research grants from the National Agriculture and Food Research Organization (#220a0) and the Japan International Research Center of Agricultural

## Abstract

The genera *Erianthus* and *Miscanthus*, both members of the *Saccharum* complex, are of interest as potential resources for sugarcane improvement and as bioenergy crops. Recent studies have mainly focused on the conservation and use of wild accessions of these genera as breeding materials. However, the sequence data are limited, which hampers the studies of phylogenetic relationships, population structure, and evolution of these grasses. Here, we determined the complete chloroplast genome sequences of *Erianthus arundinaceus* and *Miscanthus sinensis* by using 454 GS FLX pyrosequencing and Sanger sequencing. Alignment of the *E. arundinaceus* and *M. sinensis* chloroplast genome sequences with the known sequence of *Saccharum officinarum* demonstrated a high degree of conservation in gene content and order. Using the data sets of 76 chloroplast protein-coding genes, we performed phylogenetic analysis in 40 taxa including *E. arundinaceus* and *M. sinensis*. Our results show that *S. officinarum* is more closely related to *M. sinensis* than to *E. arundinaceus*. We estimated that *E. arundinaceus* diverged from the subtribe Sorghinae before the divergence of *Sorghum bicolor* and the common ancestor of *S. officinarum* and *M. sinensis*. This is the first report of the phylogenetic and evolutionary relationships inferred from maternally inherited variation in the *Saccharum* complex. Our study provides an important framework for understanding the phylogenetic relatedness of the economically important genera *Erianthus*, *Miscanthus*, and *Saccharum*.

## Introduction

The Poaceae is the grass family comprised of approximately 700 genera and more than 10,000 species and grouped into two major clades, BEP (the subfamilies Bambusoideae, Ehrhartoideae, and Pooideae) and PACMAD (the subfamilies Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae) [1–3]. The Andropogoneae is one of the tribes of the Panicoideae that includes many economically important C<sub>4</sub> grasses such as

Science (#001750). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

maize (*Zea mays* L.), sorghum (*Sorghum bicolor* L. Moench), and sugarcane (*Saccharum* spp.). The genera *Saccharum*, *Erianthus*, and *Miscanthus* are members of the subtribe Saccharinae within the Andropogoneae [4]. *Erianthus* and *Miscanthus* exhibit diverse important agricultural traits such as high productivity, high percentage of dry matter, good ratooning ability, vigor, and resistance to environmental stresses [5–7]. These genera are cross-compatible with *Saccharum* species [8], and sugarcane breeders have created intergeneric hybrids between commercial *Saccharum* spp. hybrids and these genera [9–12]. Thus, *Erianthus* and *Miscanthus* have attracted attention as potential genetic resources for sugarcane improvement [13–15]. In addition to their favorable agricultural traits, the low ash content and high heating value make *Erianthus* and *Miscanthus* promising cellulosic feedstocks at energy conversion plants; they can be used for methanol synthesis by gasification and for direct combustion [6, 7, 16]. Ongoing studies focus on the conservation and use of wild *Erianthus* and *Miscanthus* accessions as breeding materials [17, 18].

Despite current interest, the taxonomy and phylogenetic relatedness of *Saccharum* and these related genera have been controversial until recently, because the common criterion, variation of the awn on the lemma, used for differentiation within these genera does not clearly distinguish between the genera [12]. Therefore, *Erianthus* and *Miscanthus* have been regarded by some taxonomists as being synonymous with *Saccharum* and have been grouped into the so-called ‘*Saccharum* complex’ [19], which includes the members of *Saccharum* L., *Erianthus* Michx., *Miscanthus* Anderss., *Narenga* Bor, *Sclerostachya* A. Camus. This theory is widely accepted by sugarcane breeders [20].

Phylogenetic analyses based on molecular data have been employed to reconstruct the phylogeny of the *Saccharum* complex. In these studies, DNA variation detected by using DNA markers developed from nuclear genomes [8, 10, 14, 17, 21–24], was used to assess genetic diversity among wild accessions in these genera. Welker *et al.* [25] showed that a phylogenetic tree inferred from low-copy nuclear loci was useful for understanding the relationships between polyploid taxa and identifying allopolyploidization events in *Saccharum* and related genera. In addition, the data sets of partial sequences [21] and DNA markers [8, 14, 26–28] developed from organelle genomes were also used to estimate the phylogenetic relationships between the species and genera of the *Saccharum* complex. These studies have provided valuable insight into the phylogenetic relations within the *Saccharum* complex: (1) *Saccharum* is more closely related to *Miscanthus* than to *Erianthus*; (2) *Erianthus* is more closely related to *Sorghum* than to the other members of the *Saccharum* complex; (3) the evolutionary history of *Erianthus* may differ from that of other members of the *Saccharum* complex. These results have indicated the potential of this approach to elucidate the phylogenetic relationships within the *Saccharum* complex.

Because the chloroplast (cp) genome has conserved gene content and uniparental inheritance [29], polymorphism within the chloroplast genome is a valuable tool for phylogenetic and evolutionary studies [30]. To date, only 12 cpDNA markers [14] and 28 partial sequences are registered for *Erianthus arundinaceus* in GenBank; therefore, there is a clear need for additional sequence information on the *E. arundinaceus* cp genome. Comparison of the complete cp genome sequences could reveal novel genome features such as single-nucleotide polymorphisms (SNPs), insertions/deletions (indels), and microsatellites. This information would improve the analyses of the relationships in the *Saccharum* complex, especially for *Erianthus*. Multiple alignments of complete cp genomes reveal sequence variability, which is needed for the development of DNA markers for taxonomic and evolutionary studies. In the *Saccharum* complex, the complete cp genome sequences were first reported for *Saccharum officinarum* in 2004 [31, 32] and more recently for *Miscanthus sinensis* [33]. However, as the *E. arundinaceus* cp genome has not been fully sequenced, the whole-genome comparison between these major genera of the *Saccharum* complex has not yet been possible.

Recent advances in pyrosequencing, which allows high-throughput sequence analysis of a wide range of genomes, has simplified sequencing, considerably increased its speed, and reduced the cost. This approach enables faster and more efficient determination of whole cp genome sequences, and has been applied to many plant species, including those in the Poaceae [33, 34]. In this study, we present the complete cp genome sequence of *Erianthus arundinaceus* determined using pyrosequencing. On the basis of this sequence, we designed a primer set that is useful for validation of ambiguous sites such as homopolymeric and gap regions in Poaceae cp genomes, and also for sequencing of the entire cp genomes; we used these primers to sequence the whole cp genome of *Miscanthus sinensis*. Our analysis of these cp genomes provides detailed data on the distribution of SNPs, indels, and microsatellites in *Saccharum* and related genera. We also discuss the evolution of the *Saccharum* complex based on the sequence variations of these cp genomes.

## Results

### Assembly and annotation of the chloroplast genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*

The *E. arundinaceus* cp genome was sequenced using pyrosequencing on the 454 GS FLX system. A total of 481,406 sequence reads (average, 336 bp; range, 30–897 bp) were generated, representing a 162-Mbp sequence. After filtering the reads by local BLASTN analysis with the *S. officinarum* cp genome (GenBank accession No. NC006084) as a reference, 5,052 reads (average, 362 bp) were retained; a 12-fold coverage of the cp genome was reached. There were 30 homopolymeric stretches ( $\geq 10$  bp), which may lead to errors in the assembled sequences [35]. The accuracy of these regions and the inverted repeat (IR) junction regions in assembled sequences was confirmed by using PCR-based sequencing. Thus, the complete *E. arundinaceus* cp genome sequence was obtained. To determine the complete sequence of the *M. sinensis* cp genome, we used the Sanger sequencing with primers designed from the *E. arundinaceus* cp genome sequence. Sixteen overlapping regions were amplified with specific primers (Table 1) and a total of 320 sequence reads were obtained by using 258 primers, among which 253 primers (98.1%) were identical to both *M. sinensis* and *S. officinarum* cp genome sequences and 251 (97.3%) were also identical to that of *Sorghum bicolor* (S1 Table).

The complete cp genomes of *E. arundinaceus* (141,210 bp) and *M. sinensis* (141,416 bp) had typical circular structures (Fig 1). The cp genome of *E. arundinaceus* included a large single-copy (LSC) region (83,170 bp) and a small single-copy (SSC) region (12,516 bp), which were separated by a pair of IRs (IRa and IRb; 22,762 bp each); that of *M. sinensis* consisted of an LSC (83,141 bp), an SSC (12,681 bp), and two IRs (22,797 bp each). The GC content was 38.5% in the *E. arundinaceus* genome and 38.4% in the *M. sinensis* genome; these values were similar to those of other Panicoideae including *S. officinarum* [31], *M. sinensis* [33], and *S. bicolor* [36]. The number of genes was 143 in *E. arundinaceus* and 141 in *M. sinensis*, including 86 and 84 protein-coding genes, respectively. Each genome contained 8 ribosomal RNA (rRNA) genes and 49 transfer RNA (tRNA) genes. Coding genes accounted for 58.9% (*E. arundinaceus*) and 58.4% (*M. sinensis*) of the genomes (Table 2). The difference in the gene number was due to a difference in *ycf68* in the IR regions, which appeared to be a pseudogene in *M. sinensis* because of a frame-shift mutation. *S. officinarum* and *E. arundinaceus* have the complete *ycf68* open reading frame, whereas *S. bicolor* has a frame-shift mutation at the same position as in *M. sinensis*. The members of the *Saccharum* complex also have lost *accD*, *ycf1*, and *ycf2*, which are absent in the cp genomes of other Panicoideae grasses [33, 36, 37]. We also found that the start codons of the *rpl2* and *rps19* genes are likely to convert to ACG and GTG via RNA editing during translation both in *E. arundinaceus* and *M. sinensis*, as reported in other species [37–39].

**Table 1. Primer pairs used for amplification of *Miscanthus sinensis* cp genome.**

Primer pair	Primer sequence (5' to 3')		T <sub>m</sub> (°C)	PCR product			Length (bp)
	Forward	Reverse		Location <sup>1</sup>	Start <sup>2</sup>	End <sup>2</sup>	
ES01	TTGTGAGCATTACGTTTCGTGC	GCTGAGTGGTTGATAGCTCCG	60	LSC	140	12113	11974
ES02	TGATCGTGATTTGGAACCTGTTTC	ATTGAAGCATCTCGCACCTT	58	LSC	11835	19267	7433
ES03	AATGAAAGGGTCTGGTTGGA	CCAATTGCATGCGTCTAATC	58	LSC	18929	26001	7073
ES04	AGAGTGCCTAATCACGAGGATCC	CCTCTTGATCATCAACCCATCG	60	LSC	25073	37410	12338
ES05	AACAAGGGCGATGAATCAG	AACCGTTCAGCTGTTTCCTG	56	LSC	36910	38656	1747
ES06	GTCGAATTTGCAGAAGGGACGAG	GAGTCTTGTGCGACTCCTTTGTG	60	LSC	37314	50226	12913
ES07	GTGGATTAATCGGACGAGGA	ACTGCAGCTCCTGCTTCTTC	58	LSC	49983	57728	7746
ES08	GCAGGCGCAGATCTATGAAT	CCTTGTCTCTGATGGTTGGAATC	58	LSC	56347	63939	7593
ES09	GGCTAGTTGAGTAGTTTTGATTAAGG	AGACCGTGGAGGATCCACAATAG	60	LSC	63690	76120	12431
ES10	CCATGAACAGGCTCCGTAAG	CGTTATGATACTGAATCTCATGCC	58	LSC	75648	82881	7234
ES11	TGGATTATGACGTGGATTGTATCG	GTAGGACTGGTGCCGACAGTTCATC	58	LSC-IRA	82318	94864	12547
ES12	CCAAACATATGCGGATCAAATCACG	GAATATTGGAGTTAACCATATTATC	56	IRA-SSC	94130	106419	12290
ES13	CCAAATTCAGATTCCAGCA	AAACCATTGCTTCGTCTGGT	54	SSC	105400	112986	7587
ES14	CCCATGTGAGATACGGAGGA	TGAAATTCGAGCCCAAAG	56	SSC	111573	119368	7796
ES15	TGTAATAACCCATAATATAGGTTTCG	CCAAACATATGCGGATCAAATCACG	56	SSC-IRB	118477	130428	11952
ES16	GTAGGACTGGTGCCGACAGTTCATC	TAGGTATTAGTACTATGGCATTTC	60	IRB-LSC	129694	290	12013

<sup>1</sup> LSC: Large single-copy, SSC: Small single-copy, IRA: Inverted repeat A, IRB: Inverted repeat B.

<sup>2</sup> Position (base pairs) in the *M. sinensis* chloroplast genome sequence.

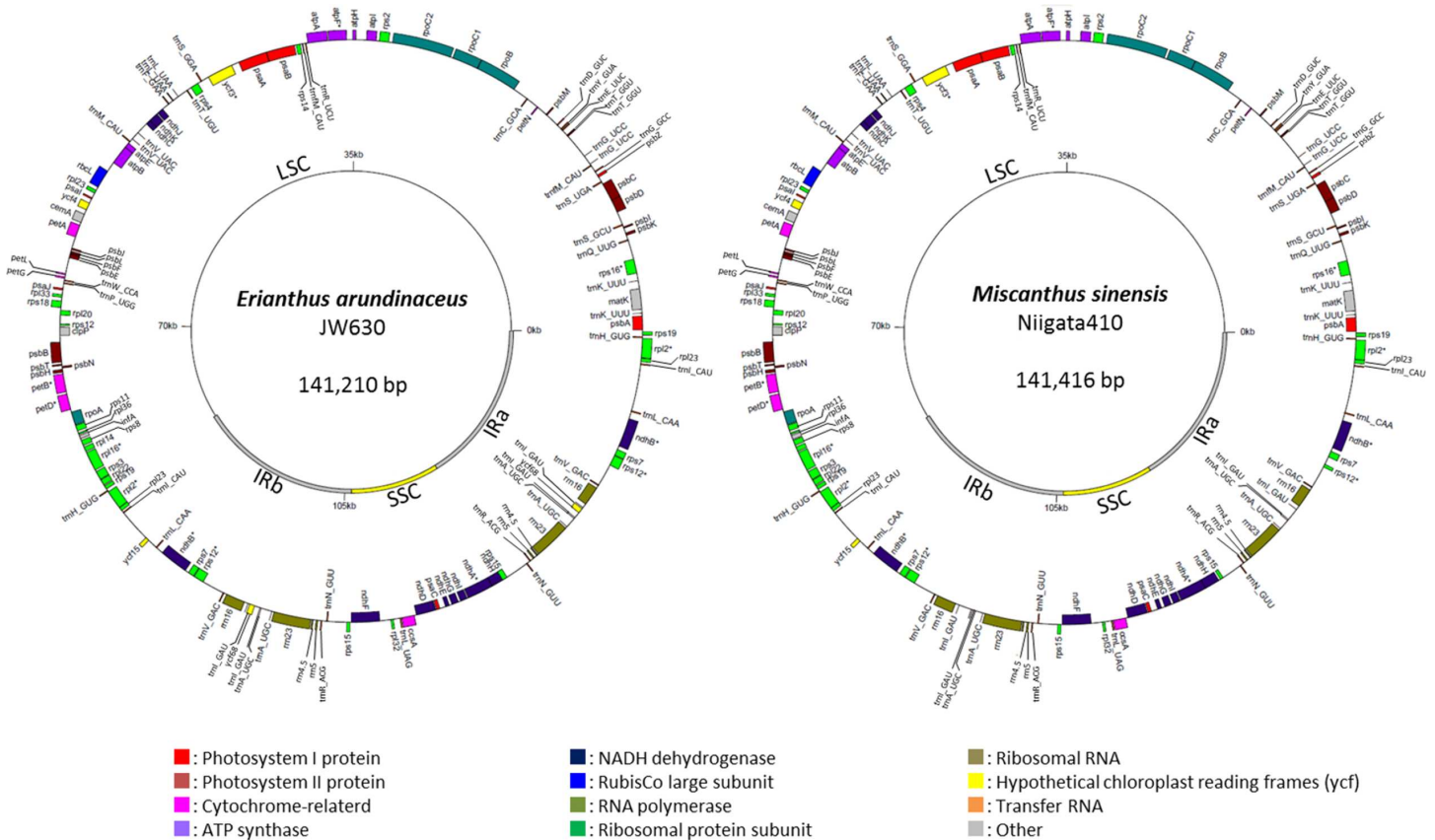
doi:10.1371/journal.pone.0169992.t001

## Sequence variations in cp genomes

We compared sequences determined in this study with those previously registered in GenBank (28 partial sequences including 10 regions for *E. arundinaceus* and the whole cp genome sequence for *M. sinensis*). For *E. arundinaceus*, sequence variations were identified at ten sites in seven regions, of which four sites (in *trnG-trnfM*, *atpB-rbcL*, *trnK* intron, and *rpl16* intron) were mutated in repeat regions (poly A or T). Base substitutions were detected at six sites (*atpA-rps14*, three sites in the *rpl16* intron, *rps16-trnQ*, and *rps3*). In the *atpA-rps14* intergenic spacer region we found an adenine-to-cytosine transition (A-to-C; A in Japanese accessions and C in Indonesian accessions), which could reflect geographical variation (S1 Fig). Detailed comparison between the *M. sinensis* sequence determined in this study and the previously reported one [33] (NC028721) detected three SNPs and nine indels. Of these, an indel in *rpoC2* and a SNP in *ycf3* resulted in amino acid sequence changes (S2 Table).

## Whole-genome comparison in the *Saccharum* complex

A global alignment of the *Saccharum* complex cp genomes with the *Zea mays* cp genome (NC001666) as a reference is shown in Fig 2. High sequence similarities in the protein-coding regions were detected. The IR regions showed lower levels of sequence divergence than the single-copy regions, although there was some gene loss. The gene order was identical in *E. arundinaceus*, *M. sinensis*, and *S. officinarum*. However, detailed comparisons within the *Saccharum* complex revealed a number of SNPs and indels (Table 3). The rates of SNP substitutions (non-synonymous [dN] and synonymous [dS]) and their ratio (dN / dS) among the 76 protein-coding genes in comparison with those of *Z. mays* are shown in Table 4. The dN (0.0039) and dS (0.0170) values of *E. arundinaceus* were slightly higher than those of the other genera. The dN/dS values of the *Saccharum* complex were smaller than 1.0, similar to those of other Poaceae [40–42]; these values suggest purifying selection of the cp protein-coding genes in these genera.



**Fig 1. Chloroplast genome maps of *Erianthus arundinaceus* and *Miscanthus sinensis*.** The genes of different functional groups are indicated in different colors. Genes on the inside and outside of the maps are transcribed clockwise and counter-clockwise, respectively. The thick lines on the inner circles indicate inverted repeats (IRa and IRb), which separate the genomes into the small single-copy (SSC) and large single-copy (LSC) regions.

doi:10.1371/journal.pone.0169992.g001

**Table 2. Characteristics of the chloroplast genomes in three genera of the *Saccharum* complex and *Sorghum bicolor*.**

Species	Genome organization <sup>1</sup>					Number of genes <sup>2</sup>				GenBank accession No.
	Length of total genome	Length of LSC	Length of SSC	Length of IR	GC/AT contents (%)	Total	CDS <sup>3</sup>	rRNA	tRNA	
<i>S. officinarum</i>	141,182	83,048	12,544	22,795	38.4/61.6	136	87 (8) <sup>4</sup>	8 (4) <sup>4</sup>	41 (8) <sup>4</sup>	NC006084 [31]
<i>E. arundinaceus</i>	141,210	83,170	12,516	22,762	38.5/61.5	136	87 (8)	8 (4)	41 (8)	LC160130 [This study]
<i>M. sinensis</i>	141,416	83,141	12,681	22,797	38.4/61.6	134	85 (7)	8 (4)	41 (8)	LC160131 [This study]
<i>M. sinensis</i>	141,372	83,163	12,659	22,775	38.4/61.6	134	85 (7)	8 (4)	41 (8)	NC028721 [33]
<i>M. sacchariflorus</i>	141,332	83,207	12,575	22,775	38.4/61.6	134	85 (7)	8 (4)	41 (8)	NC028720 [33]
<i>S. bicolor</i>	140,754	82,688	12,502	22,782	38.5/61.5	135	86 (7)	8 (4)	41 (8)	NC008602 [36]

<sup>1</sup> Length is indicated in base pairs.

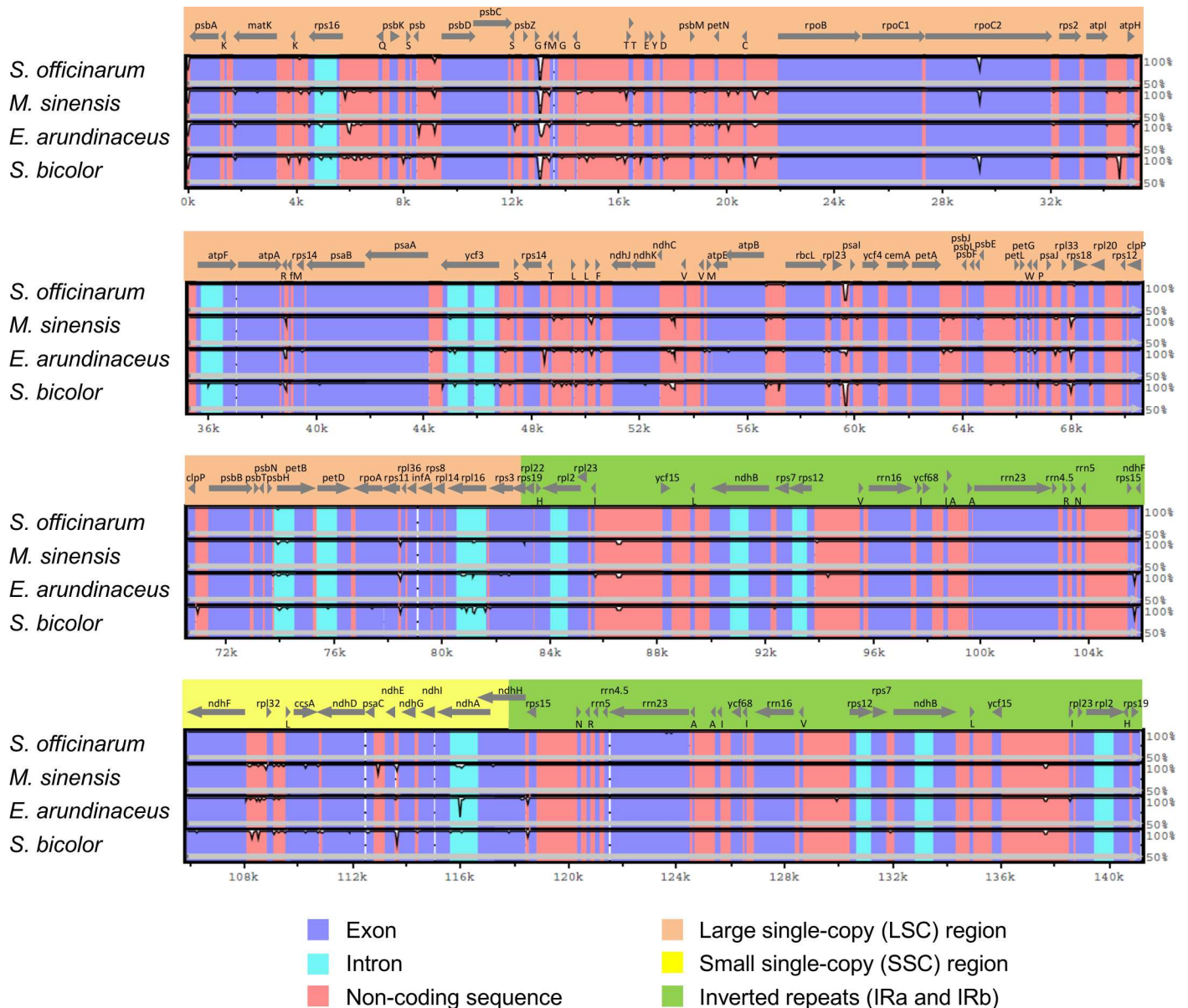
<sup>2</sup> Including genes detected in this study (not annotated in GenBank).

<sup>3</sup> Including *ycf15* and *ycf68*.

<sup>4</sup> The numbers of duplicated genes are shown in parentheses.

doi:10.1371/journal.pone.0169992.t002





**Fig 2. Alignment of whole chloroplast genome sequences from four Panicoidae species.** Chloroplast genomes were aligned by using the mVISTA program with the *Zea mays* sequence as a reference. The X- and Y-scales indicate the coordinates within cp genomes and the percentage of identity (50%–100%), respectively. Genome regions (exons, introns, and conserved non-coding sequences) are color-coded. Gray arrows indicate the direction of transcription of each gene. The genes encoding transfer RNAs (*trn*) are indicated under gray arrows using the single-letter amino acid code (e.g., K: *trnK*).

doi:10.1371/journal.pone.0169992.g002

The distribution of microsatellites (also called simple sequence repeats) in the cp genomes of *E. arundinaceus* and *M. sinensis* is shown in Table 5. A total of 40 microsatellite regions ( $\geq 8$  bp) were identified in *E. arundinaceus*, including 36 mono-, 3 tri-, and one tetranucleotide repeats. In *M. sinensis*, a total of 38 regions were identified, including 36 mono-, one tri-, and one tetranucleotide repeats. The majority of repeats were located in non-coding regions, whereas some were found in genes such as *psbC*, *rpoB*, *ndhK*, *infA*, and *rpl22*. Two microsatellites (in *rps16-trnQ/UUG* and *trnR/UCU-trnM/CAU*) were found in *E. arundinaceus* but not in *M. sinensis*.

**Table 3. SNPs and indels between *Erianthus arundinaceus* and *Miscanthus sinensis* chloroplast genomes.**

Category	SNP	Indel	Total
Photosystem I	<i>psaA</i> (2), <i>psaC</i> (1)		3
Photosystem II	<i>psbB</i> (4), <i>psbC</i> (4), <i>psbD</i> (1), <i>psbE</i> (4), <i>psbM</i> (1), <i>psbN</i> (1), <i>psbT</i> (1), <i>psbZ</i> (1)		17
ATP synthase	<i>atpA</i> (1), <i>atpB</i> (4), <i>atpI</i> (2)		7
Cytochrome	<i>petA</i> (2), <i>petB</i> (2), <i>petD</i> (3),	<i>petB</i> (1)	8
NADPH	<i>ndhA</i> (1), <i>ndhB</i> (1), <i>ndhC</i> (2), <i>ndhD</i> (6), <i>ndhF</i> (7), <i>ndhG</i> (3), <i>ndhH</i> (6), <i>ndhI</i> (1), <i>ndhJ</i> (3), <i>ndhK</i> (2)	<i>ndhA</i> (1)	33
Transcription	<i>rpoA</i> (4), <i>rpoB</i> (11), <i>rpoC1</i> (7), <i>rpoC2</i> (14)		36
Ribosomal proteins (Small subunit)	<i>rps2</i> (1), <i>rps3</i> (4), <i>rps8</i> (1), <i>rps11</i> (1), <i>rps14</i> (2), <i>rps15</i> (1), <i>rps18</i> (2)		12
Ribosomal proteins (Large subunit)	<i>rpl14</i> (1), <i>rpl16</i> (1), <i>rpl20</i> (2), <i>rpl22</i> (1), <i>rpl32</i> (1), <i>rpl33</i> (1)		7
Other	<i>infA</i> (1), <i>ycf3</i> (2), <i>ycf4</i> (1), <i>ycf68</i> (1), <i>matK</i> (12), <i>ccsA</i> (5)	<i>matK</i> (1), <i>ccsA</i> (1)	24
Rubisco	<i>rbcl</i> (5)		5
Non-coding	Intron (38), IGS <sup>1</sup> (360)	124	522
Total	546 (148) <sup>2</sup>	128	674

<sup>1</sup> Intergenic spacer region.

<sup>2</sup> Parenthesis shows SNPs in protein-coding genes.

doi:10.1371/journal.pone.0169992.t003

### Phylogenetic analyses

Phylogenetic analyses were performed on an alignment of concatenated nucleotide sequences of 76 protein-coding genes from 40 angiosperm species (39 monocots and one dicot). After all positions containing gaps and missing data were excluded, the final dataset contained a total of 17,396 nucleotide sequences. Maximum likelihood (ML) analysis resulted in a single tree with the highest log-likelihood (lnL) of -89413.4029. Of the 37 nodes, 29 had bootstrap values of  $\geq 95\%$  and 24 of these had bootstrap values of 100% (Fig 3). Maximum parsimony (MP) analysis generated a single most parsimonious tree with a length of 11,454 (consistency index, 0.57; retention index, 0.86; data not shown). The ML and MP trees had similar topology, which was also similar to those of the published phylogenetic trees of grasses based on complete cp genomes [37, 43]. The 39 monocot taxa were divided into two major groups, one containing Poales, including the *Saccharum* complex, and the other one containing all other monocots. *E. arundinaceus*, *M. sinensis*, and *S. officinarum* were grouped into the PACMAD clade, which is one of the major Poaceae lineages. *S. officinarum* was more closely related to *M. sinensis* than to *E. arundinaceus*, in line with previous phylogenetic analyses [14, 44].

**Table 4. Substitution rates on 76 protein-coding chloroplast genes in three genera of the *Saccharum* complex and *Sorghum bicolor*.**

Species	Substitution rate <sup>1</sup>		dN/ dS
	dN	dS	
<i>S. officinarum</i>	0.0030±0.0007	0.0144±0.0030	0.2460
<i>E. arundinaceus</i>	0.0039±0.0010	0.0170±0.0030	0.2433
<i>M. sinensis</i>	0.0036±0.0010	0.0152±0.0030	0.2548
<i>S. bicolor</i>	0.0038±0.0009	0.0183±0.0028	0.2140

<sup>1</sup> dN: the rates of nonsynonymous, dS: the rates of synonymous substitutions, *Zea mays* was used as a reference.

doi:10.1371/journal.pone.0169992.t004

**Table 5. Microsatellites in *Erianthus arundinaceus* and *Miscanthus sinensis* chloroplast genomes.**

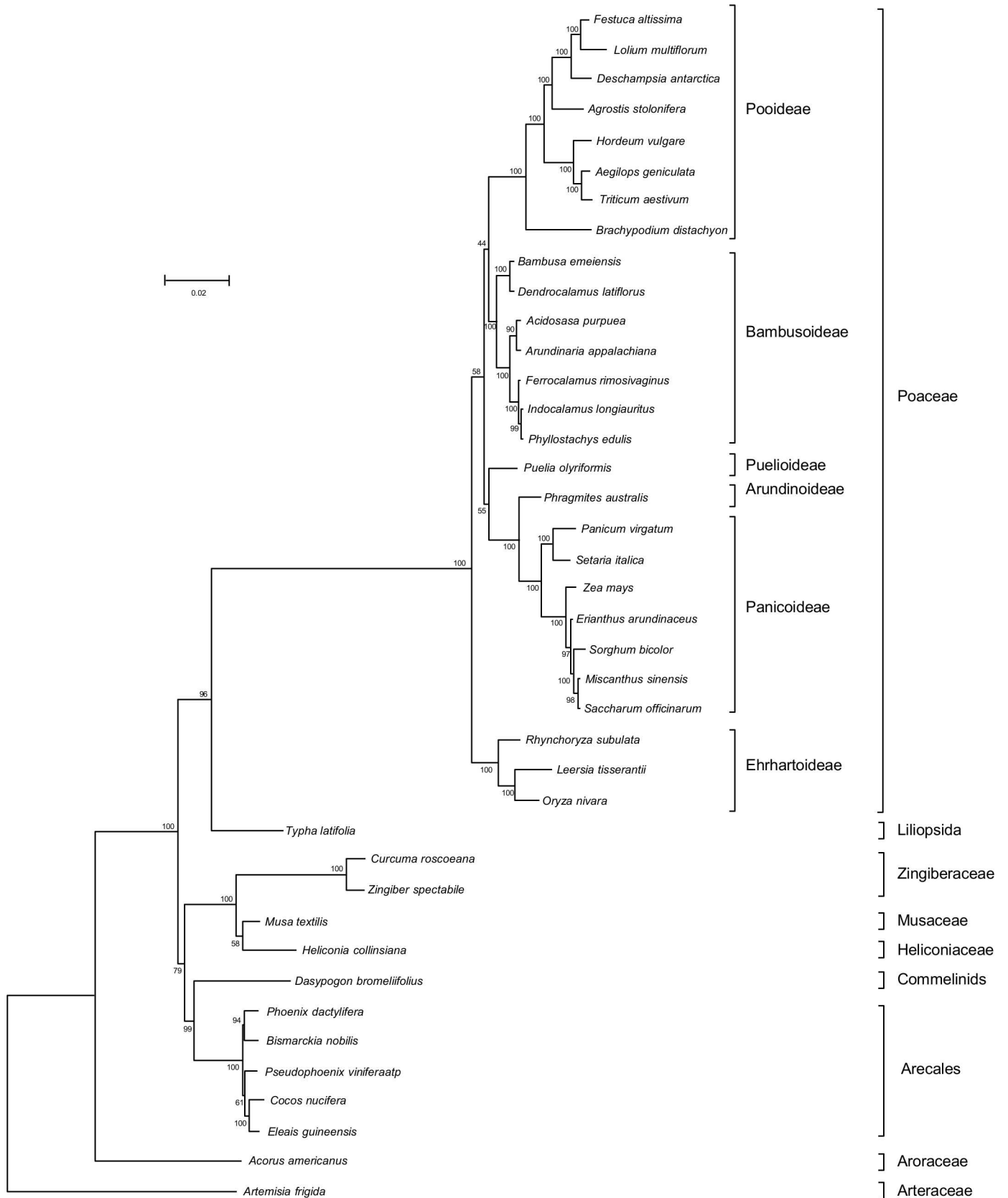
Location	Motif	<i>E. arundinaceus</i>			<i>M. sinensis</i>		
		Sequence	Start	End	Sequence	Start	End
<i>matK-trnK/UUU</i>	Mono	(A/T)8	3548	3555	(A/T)10	3550	3560
<i>matK-trnK/UUU</i>	Mono	(A/T)15	3756	3771	(A/T)13	3760	3773
<i>trnK/UUU-rps16</i>	Mono	(A/T)11	4118	4129	(A/T)10	4120	4130
<i>rps16-trnQ/UUG</i>	Tri	(ATT)4	5844	5856	-	-	-
<i>rps16-trnQ/UUG</i>	Mono	(A/T)11	6417	6428	(A/T)13	6214	6227
<i>psbK-psbI</i>	Mono	(A/T)10	7757	7767	(A/T)14	6471	6485
<i>trnS/GCU-psbD</i>	Mono	(A/T)12	9067	9079	(A/T)10	8746	8756
<i>psbC</i>	Mono	(G/C)10	11033	11043	(G/C)10	11032	11042
<i>trnG/GCC-trnM/CAU</i>	Mono	(A/T)10	13447	13457	(A/T)12	13440	13452
<i>trnT/GGU-trnE/UUC</i>	Mono	(A/T)15	16614	16629	(A/T)11	16638	16649
<i>trnT/GGU-trnE/UUC</i>	Mono	(A/T)11	16708	16719	(A/T)9	16728	16736
<i>trnD/GUC-psbM</i>	Mono	(A/T)14	18717	18731	(A/T)11	18736	18747
<i>psbM-petN</i>	Mono	(A/T)14	19267	19281	(A/T)11	19279	19290
<i>trnC/GCA-rpoB</i>	Mono	(A/T)11	21124	21135	(A/T)13	21111	21124
<i>rpoB</i>	Mono	(A/T)10	31970	31980	(A/T)10	31961	31971
<i>atpI-atpH</i>	Mono	(A/T)10	34148	34158	(A/T)14	34140	34154
<i>atpI-atpH</i>	Mono	(A/T)9	34684	34692	(A/T)12	34696	34708
<i>atpF intron</i>	Mono	(A/T)9	35871	35879	(A/T)10	35892	35902
<i>atpA-trnR/UCU</i>	Mono	(A/T)9	38743	38751	(A/T)10	38764	38774
<i>trnR/UCU-trnM/CAU</i>	Tri	(ATT)7	38901	38922	-	-	-
<i>psaA-ycfIII</i>	Mono	(A/T)13	44325	44338	(A/T)13	45675	45688
<i>trnT/UGU-trnL/UAA</i>	Mono	(A/T)11	48910	48921	(A/T)8	48896	48903
<i>trnL/UAA-trnF/GAA</i>	Mono	(A/T)10	50274	50284	(A/T)8	50282	50289
<i>ndhK</i>	Mono	(A/T)14	52432	52446	(A/T)14	52438	52452
<i>trnM/CAU-atpE</i>	Tetra	(AGGT)4	54731	54747	(AGGT)3	54736	54747
<i>atpE-rbcL</i>	Mono	(A/T)12	56799	56811	(A/T)11	56982	56993
<i>atpE-rbcL</i>	Mono	(A/T)12	57457	57469	(A/T)11	57454	57465
<i>rpl23-psaI</i>	Mono	(A/T)10	59774	59784	(A/T)8	59777	59784
<i>psaI-ycf4</i>	Mono	(A/T)10	60225	60235	(A/T)9	60226	60234
<i>petA-psbJ</i>	Mono	(A/T)11	63645	63656	(A/T)9	63644	63652
<i>psbE-petL</i>	Mono	(A/T)14	65805	65819	(A/T)10	65604	65614
<i>rpl33-rps18</i>	Mono	(A/T)12	68260	68272	(A/T)12	68219	68231
<i>petB intron</i>	Mono	(A/T)12	74056	74068	(A/T)11	78555	78566
<i>infA</i>	Mono	(A/T)10	79145	79155	(A/T)10	79104	79114
<i>infA</i>	Mono	(A/T)10	79163	79173	(A/T)10	79122	79132
<i>rpl16 intron</i>	Mono	(A/T)13	81363	81376	(A/T)15	81328	81343
<i>rps3-rpl22</i>	Mono	(A/T)10	82624	82634	(A/T)10	81698	81708
<i>rpl22</i>	Tri	(CTT)4	83063	83075	(CTT)4	83028	83040
<i>rpl32-trnL/UAG</i>	Mono	(A/T)9	109365	109373	(A/T)11	109392	109403
<i>ndhA intron</i>	Mono	(A/T)7	116146	116152	(A/T)12	116318	116330

doi:10.1371/journal.pone.0169992.t005

### Divergence time estimates

Using 76 concatenated protein-coding genes from the PACMAD clade, including the *Saccharum* complex, we estimated the divergence time with the Bayesian approach assuming a relaxed lognormal clock with the constrained calibration point of the oldest *C<sub>4</sub>* lineage in Chloridoideae. As shown in Fig 4, the BEP and the PACMAD clades diverged 81.97 million





**Fig 3. Phylogenetic analysis of 40 species including three genera of the *Saccharum* complex.** A phylogenetic tree was generated using the maximum-likelihood method based on the concatenated nucleotide sequences of 76 protein-coding chloroplast genes. Numbers beside the nodes indicate the bootstrap values (%) from 1,000 replicates.

doi:10.1371/journal.pone.0169992.g003

years ago (mya). Within the PACMAD clade, *Panicum virgatum* (Paniceae) diverged from the other species 24.50 mya (range, 20.04–44.20 mya). *E. arundinaceus* was estimated to have diverged from the other genera of the *Saccharum* complex 9.14 mya (range, 0.91–17.99 mya), whereas *M. sinensis* and *S. officinarum* diverged approximately 3.64 mya (range, 0.01–9.01 mya).

## Discussion

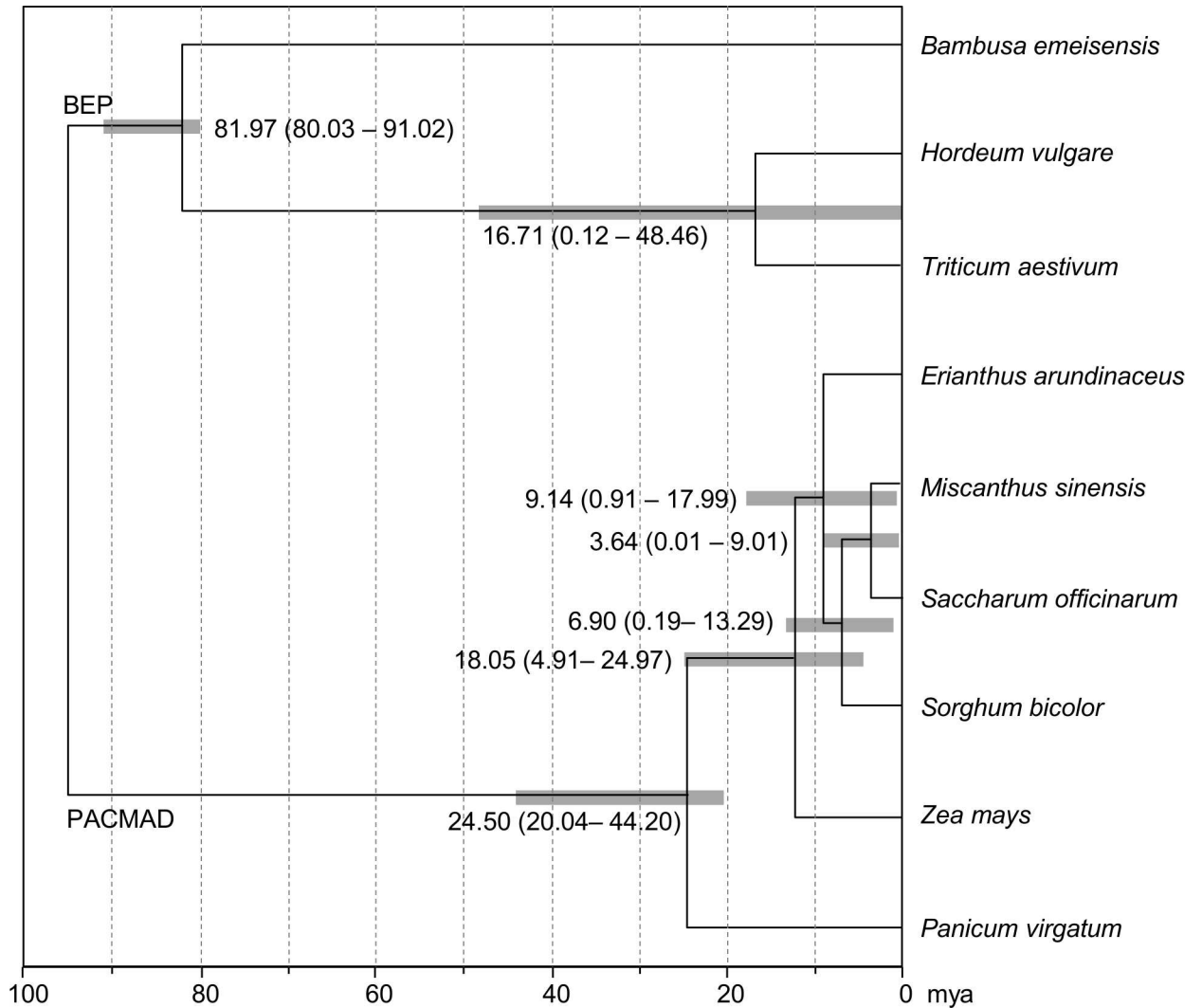
### Features of the chloroplast genomes of *E. arundinaceus* and *M. sinensis*

In this study, we determined the complete cp genome sequences of the members of the *Saccharum* complex, *E. arundinaceus* and *M. sinensis*, using 454 GS FLX pyrosequencing and Sanger sequencing. Pyrosequencing has been increasingly used for the sequencing of entire cp genomes, including those of species from several genera of the Poaceae family [33, 36, 37], because of its high throughput and low cost. However, homopolymer stretches (mononucleotide repeats) cause errors in pyrosequencing data; these errors are generally difficult to correct by increasing sequence read depth [45, 46]. In addition, alignment gaps are often allowed in the assembled sequences [45]. In this study, we designed 258 primers, which made it possible to complete sequencing of the entire *E. arundinaceus* cp genome, and applied these primers to *M. sinensis*. These primers have high identity with other plant cp genome sequences such as those of *S. officinarum* and *S. bicolor* (S1 Table), and could be used, together with pyrosequencing, for resequencing of ambiguous sites such as homopolymeric and gap regions in Poaceae cp genomes, but also for sequencing of entire cp genomes.

Homopolymers are often present in cp genomes and may be used as microsatellite markers. Because the cp genome sequences are highly conserved among grasses, microsatellite primers for cp genomes are transferable across species and genera. In addition, homopolymers are highly polymorphic, and are valuable markers for the analysis of differentiation and population structure, although overall the cp genome sequences are highly conserved. Inter- and intraspecific variations of cp microsatellites have been used to estimate the genetic diversity and phylogenetic relationships among species and genera [47]. With a threshold of  $\geq 8$  bp, we found 40 microsatellite loci for *E. arundinaceus* and 38 for *M. sinensis*, including 3 tri- and one tetranucleotide repeats, which were located mostly in non-coding regions. This information could be useful for the development of microsatellite markers for the analysis of genetic diversity in *Erianthus*, *Miscanthus*, and related genera.

### Comparison of the sequences within and among *Saccharum* complex species

Comparison of the sequences determined in this study and the sequences previously registered in GenBank identified some polymorphisms. Most of them were found in homopolymeric regions in *E. arundinaceus*. A base substitution identified in the *atpA-rps14* intergenic spacer region reflects geographic heterogeneity. Comparison of the whole cp genome sequences of two *M. sinensis* accessions detected SNPs and indels at 12 sites. These results indicated the presence of intraspecific mutations in the highly conservative cp genome and could be useful for the analysis of genetic diversity and evolution of *Erianthus*, *Miscanthus*, and related genera. However, Yook *et al.* [48] have reported (on the basis of phenotypic and nuclear SSR genotypic



**Fig 4. Divergence times of the PACMAD clade.** A Bayesian relaxed-clock approach based on 76 concatenated protein-coding chloroplast genes was used to estimate divergence times.

doi:10.1371/journal.pone.0169992.g004

analyses) that some *M. sinensis* accessions, including those used for cp genome sequencing, might be hybrids with *M. sacchariflorus*. Further studies are required to validate intraspecific mutations in *M. sinensis*.

The gene contents differ slightly among the three genera of the *Saccharum* complex because of a frame-shift mutation that resulted in a premature stop codon and loss of the hypothetical gene *ycf68* in these genera. Similar mutations have been reported in some other plant species [49]. Intact copies of another hypothetical gene, *ycf15*, were detected in both *E. arundinaceus* and *M. sinensis* cp genomes, although in some other species this gene contains several internal stop codons and is thus nonfunctional [49]. The validity of *ycf15* and *ycf68* as protein-coding genes is questionable: according to Raubeson *et al.* [50], their pattern of evolution is not consistent with them encoding proteins. Therefore, these genes were excluded from subsequent analysis in this study and further investigation is required to understand their functions.

## Phylogenetic relationships and evolution

Our phylogenetic analysis based on the variation of the nucleotide sequences of 76 protein-coding genes in cp genomes separated Poales from other monocot groups with a bootstrap value of 100%, which is largely consistent with a recent analysis of other cp genome sequences [40]. Our data suggest that *S. officinarum* is more closely related to *M. sinensis* than to *E. arundinaceus*. We estimated that *S. officinarum* and *M. sinensis* diverged 3.6 mya, which is in good agreement with divergence times previously estimated on the basis of nuclear (3.1–3.8 mya) genome diversity [51, 52]. A study based on restriction fragment length polymorphism analysis, which used 12 cp-specific probes and examined 32 *Saccharum* complex genotypes, showed that *Erianthus* diverged from other lineages early in the evolution of the subtribe Saccharinae [14]. Our analysis estimated the divergence time as 9.1 mya. In addition, *E. arundinaceus* diverged from the subtribe Sorghinae before the divergence of *S. bicolor* and the common ancestor of *S. officinarum* and *M. sinensis*. The present study showed that the cp genome of *E. arundinaceus* is more closely related to that of *S. bicolor* than to those of other members of the *Saccharum* complex. These data support the suggestion of Sobral *et al.* [14] that the evolutionary history of *Erianthus* may differ from that of other members of the *Saccharum* complex.

In the Old World, *Erianthus* species comprise four cytotypes: diploid ( $2n = 2x = 20$ ), triploid ( $2n = 3x = 30$ ), tetraploid ( $2n = 4x = 40$ ), and hexaploid ( $2n = 6x = 60$ ), with a basic number of  $x = 10$  [4]. The present study does not clarify how *Erianthus* was established, and additional investigations are required. Inclusion of different cytotypes in phylogenetic analysis based on cp genome sequences may provide useful information on the origin and establishment of this genus. Maternal origin of hybrids and polyploids of several species has been investigated using cpDNA variations [53–55]. The use of combined data on nuclear and cpDNA variations may help determine the origin and evolutionary history of polyploids [56]. In the subtribe Saccharinae, comparative analysis of nuclear genome variations in *Saccharum* and *Miscanthus* suggested that a whole-genome duplication occurred in their common ancestor [51]. This molecular phylogenetic approach, which is used to elucidate the origin and history of polyploidization, could also contribute to characterization of the phylogenetic relationships of *Erianthus*. Therefore, understanding nuclear genome variations, especially in low-copy nuclear loci [52, 57], together with cp genome variations would also be useful for clarifying the evolution of the *Erianthus* polyploid complex. Understanding its evolution could help us to gain more insight into the phylogenetic relationships of the *Saccharum* complex genera and provide useful information on their ancestor and polyploidization, which is critical for genetic studies and breeding in these genera.

## Conclusion

Comparison of the complete cp genomes provided detailed information on genetic variations among three economically important genera, *Saccharum*, *Erianthus*, and *Miscanthus*. Comparison of the sequences indicated that *S. officinarum* and *M. sinensis* are more closely related to each other than to *E. arundinaceus*. We suggest that *E. arundinaceus* diverged from the subtribe Sorghinae before the divergence of *S. bicolor* and the common ancestor of *S. officinarum* and *M. sinensis*. This is the first report of phylogenetic and evolutionary relationships among the three genera of the *Saccharum* complex inferred from maternally inherited variations in whole cp genomes and gene data sets. Our results provide an important framework for understanding the phylogeny and evolutionary history of the *Saccharum* complex. Molecular data for the other genera of the complex, *Narenga* and *Sclerostachya*, are limited and further studies on these genera are needed to improve our understanding of the phylogeny and evolution of the *Saccharum* complex.

## Materials and Methods

### Plant materials and DNA extraction

The *E. arundinaceus* accession JW630 (Genebank accession number JP173957 at the Genetic Resources Center of the National Institute of Agrobiological Sciences, Japan; [https://www.gene.affrc.go.jp/index\\_en.php](https://www.gene.affrc.go.jp/index_en.php)) is a wild hexaploid collected in Shizuoka prefecture, Japan (the northernmost area of the wild *E. arundinaceus* range in Japan). The *M. sinensis* accession Niigata 410 (JP177091) is a wild diploid collected in Niigata prefecture, Japan. Plants were cultivated in a greenhouse at the National Agriculture and Food Research Organization, Institute of Livestock and Grassland Science (NARO-ILGS), and genomic DNA was isolated from fresh green leaves using the CTAB method [58].

### *E. arundinaceus* chloroplast genome sequencing and assembly

The *E. arundinaceus* cp genome was sequenced by using pyrosequencing. Total *E. arundinaceus* genomic DNA was sheared by nebulization and then amplified by emulsion PCR. Amplification products were sequenced on a 454 GS FLX Titanium platform (Roche, Basel, Switzerland) [59]. Chloroplast sequence reads were extracted by local BLASTN searches using the cp genome of *S. officinarum* [31] as a reference and assembled with Newbler software (v 2.5; Roche). Homopolymer regions (poly A/T and poly G/C) and the junctions between single-copy regions (LSC and SSC) and IRs were amplified and confirmed using primers designed from the *E. arundinaceus* cp sequence (S1 Table) and PrimeSTAR HS DNA polymerase (TaKaRa, Shiga, Japan). PCR products were purified in a QuickStep2 PCR Purification system (Edge Biosystems, Gaithersburg, MD, USA). They were cycle-sequenced with a BigDye Terminator Cycle Sequence Kit v3.1 (Life Technologies, Foster city, CA, USA) and sequenced using an ABI3130xl genetic analyzer (Life Technologies) using primers described below (S1 Table).

### *M. sinensis* chloroplast genome sequencing

The *M. sinensis* cp genome was sequenced by using Sanger sequencing of PCR products. Sixteen primers to amplify overlapping products (1,747–12,913 bp) were designed from the *E. arundinaceus* cp genome sequence for initial amplification of the *M. sinensis* cp genome (Table 1). Amplification reactions and cycle-sequencing were performed as described above for *E. arundinaceus*. A total of 258 primers (S1 Table) were used to sequence the entire *M. sinensis* cp genome.

### Annotation, microsatellite analysis, and comparison of the chloroplast genomes

The entire sequences of the *E. arundinaceus* and *M. sinensis* cp genomes were annotated using Dual Organellar GenoMe Annotator (DOGMA) software [60]. The predicted annotations were manually checked and verified by comparison with sequences from other PAC-MAD clade species. The circular chloroplast genome maps were drawn by GenomeVx software [61].

Microsatellites were predicted using MSATCOMMANDER 1.03 software [62]. We defined microsatellites as  $\geq 10$  repeats (10 bases) for mononucleotides,  $\geq 8$  repeats (16 bases) for dinucleotides,  $\geq 5$  repeats (15 bases) for trinucleotides,  $\geq 4$  repeats (16 bases) for tetranucleotides,  $\geq 4$  repeats (20 bases) for pentanucleotides, and  $\geq 4$  repeats (24 bases) for hexanucleotides.

Genome structures among the genera of the *Saccharum* complex were compared using mVISTA software in Shuffle-LAGAN mode [63]; sequence annotation of *Z. mays* was used.



## Substitution rates

Substitution rates were calculated using the PAMLX package [64]. The program CODEML in PAMLX was employed to estimate the rates of nonsynonymous (dN) and synonymous (dS) substitutions and their ratio (dN / dS) in 76 cp protein-coding genes aligned by using PAL2-NAL [65]. The maximum likelihood (ML) tree (see below) was used as a topologically constrained tree. The F3 × 4 model was adopted for codon frequencies under the branch-site model (model = 2, NSsites = 2, and cleandata = 1).

## Phylogenetic analysis

Nucleotide sequences of 76 cp protein-coding genes of 37 monocot angiosperms and one dicot angiosperm (*Artemisia frigida*) available in the GenBank database, and those of *E. arundinaceus* and *M. sinensis* were concatenated and aligned using Clustal W [66]. After manual editing, phylogenetic analyses using ML and maximum parsimony (MP) were performed with MEGA6 [67] using subtree-pruning-regrafting and nearest-neighbor-interchange algorithms, respectively. The gaps in the alignment were treated as missing data and statistical support at each node was assessed by bootstrapping [68] with 1,000 replicates. Bootstrap values are indicated on the tree.

## Estimation of divergence time of the *Saccharum* complex

A set of 76 protein-coding genes was aligned and used for the estimation of divergence time. The analysis was performed with nine species including three species of the *Saccharum* complex with a focus on the PACMAD clade (Fig 4) using the BEAST2 program, which infers tree topology, branch lengths, and node ages by using Bayesian inference and Markov Chain Monte Carlo (MCMC) analysis [69]. The AIC (Akaike Information Criterion) analysis was performed by using jModelTest 2.1.6 [70] to identify the best fit of the substitution model for mutation rates. BEAUti in the BEAST2 program was used to set the criteria for the analysis. We used the GTR (general-time reversible) model of nucleotide substitution with five categories of gamma-distributed rate. An uncorrelated lognormal model of rate variation among branches was assumed and a Yule prior on the birth rate of new lineages was employed [71]. A single divergence time was previously estimated, assuming that the major diversification of the grass groups occurred 80 mya and the Andropogoneae crown diverged 20 mya [72, 73]; these two time points were used to calibrate the age of the stem nodes. Two independent MCMC runs were performed for 10 million generations with tree sampling every 1,000 generations. The results were checked with Tracer 1.6 [74], and the sampled trees were summarized by using TreeAnnotator v.2.1.2 available in the BEAST2 package, and edited by using FigTree v.1.4.2 [75]. The mean and the estimated 95% highest posterior density interval for the divergence time are given for the major PACMAD lineages.

## Supporting Information

**S1 Fig. Sequence variations detected among the whole cp genome (sequenced in this study) and partial sequences (registered in Genbank) from *Erianthus arundinaceus*.** Origin is indicated as follows: JPN, Japan; IDN, Indonesia; IND, India; THA, Thailand. (PPTX)

**S1 Table. Sequencing primers designed from the *Erianthus arundinaceus* cp genome.** (XLSX)

**S2 Table. Summary of sequence variations detected between chloroplast genome sequences of two *Miscanthus sinensis* accessions.**

(DOCX)

**Acknowledgments**

We thank all the staff of the Forage Plant Breeding and Genetics group at NARO-ILGS who provided technical assistance with the experiments. We also thank all the staff of the sugarcane improvement project team at JIRCAS-TARF for great help and contribution to the data analysis.

**Author Contributions****Conceptualization:** ST ME.**Data curation:** ST ME.**Formal analysis:** ST ME.**Funding acquisition:** ST MK WT.**Investigation:** ST ME.**Methodology:** ST ME.**Project administration:** ST MK.**Resources:** MK ME MK WT.**Supervision:** ST MK.**Validation:** ST ME.**Visualization:** ST.**Writing – original draft:** ST.**Writing – review & editing:** ST ME MK WT.**References**

1. GPWG. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann Mol Bot Gard.* 2001; 88: 373–457.
2. Duvall MR, Davis JI, Clark LG, Noll JD, Goldman DH, Sánchez-Ken JG. Phylogeny of the grasses (Poaceae) revisited. *Aliso: A Journal of Systematic and Evolutionary Botany.* 2007; 23: 237–247.
3. Sánchez-Ken JG, Clark LG, Kellogg EA, Kay EE. Reinstatement and emendation of subfamily Micrairoideae (Poaceae). *Syst Bot.* 2007; 32: 71–80.
4. Amalraj VA, Balasundaram N. On the taxonomy of the members of ‘*Saccharum* complex’. *Genet Resour Crop Evol.* 2005; 53: 35–41.
5. Berding N, Roach BT. Germplasm collection, maintenance, and use. In: Heinz DJ, editor. *Sugarcane improvement through breeding.* Amsterdam: Elsevier; 1987. pp. 143–210.
6. Clifton-Brown J, Stampfl PF, Jones MB. *Miscanthus* biomass production for energy in Europe and its potential contribution to decreasing fossil fuel carbon emissions. *Global change Biology.* 2004; 10: 509–518.
7. Clifton-Brown J, Chiang YC, Hodkinse T. *Miscanthus*: Genetic resources and breeding potential to enhance bioenergy production. In: Vermerris W editor. *Genetic Improvement of Bioenergy Crops.* New York: Springer; 2008. pp. 273–294.
8. Hodkinson TR, Chase MW, Lledó MD, Salamin N, Renvoize SA. Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from

- ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL*-F intergenic spaces. *J Plant Res.* 2002; 115: 381–392. doi: [10.1007/s10265-002-0049-3](https://doi.org/10.1007/s10265-002-0049-3) PMID: [12579363](https://pubmed.ncbi.nlm.nih.gov/12579363/)
9. Jackson P, Henry RJ. Plant breeding. In Kole C, editor. *Wild crop relatives: genomic and breeding resources: industrial crops.* Berlin: Springer; 2011. pp. 97–109.
  10. Cai Q, Aitken KS, Fan YH, Piperidis G, Jackson P, McIntyre CL. A preliminary assessment of the genetic relationship between *Erianthus rockii* and the “*Saccharum* complex” using microsatellite (SSR) and AFLP markers. *Plant Sci.* 2005; 169: 976–984.
  11. Piperidis N, Chen JW, Deng HH, Wang LP, Jackson P, Piperidis G. GISH characterization of *Erianthus arundinaceus* chromosomes in three generations of sugarcane intergeneric hybrids. *Genome.* 2010; 53: 331–336. doi: [10.1139/g10-010](https://doi.org/10.1139/g10-010) PMID: [20616864](https://pubmed.ncbi.nlm.nih.gov/20616864/)
  12. Clayton WD, Renvoize SA. *Genera graminum: grasses of the world.* Kew Bull Addit Ser. 1986; 13: 320–375.
  13. Selvi A, Nair NV, Noyer JL, Singh NK, Balasundaram N, Bansal KC, et al. AFLP analysis of the phonetic organization and genetic diversity in the sugarcane complex, *Saccharum* and *Erianthus*. *Genet Resour Crop Evol.* 2006; 53: 831–842.
  14. Sobral BWS, Braga DPV, LaHood ES, Keim P. Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb. subtribe of the Andropogoneae Dumort. tribe. *Theor Appl Genet.* 1994; 87: 843–853. doi: [10.1007/BF00221137](https://doi.org/10.1007/BF00221137) PMID: [24190471](https://pubmed.ncbi.nlm.nih.gov/24190471/)
  15. Burner DM, Tew TL, Harvey JJ, Belesky DP. Dry matter partitioning and quality of *Miscanthus*, *Panicum*, and *Saccharum* genotypes in Arkansas, USA. *Biomass Bioenerg.* 2009; 33: 610–619.
  16. Nakagawa H, Sakai M, Harada T, Ichinose T, Takeno K, Matsumoto S, et al. Biomethanol production from forage grasses, tree, and crop residues. In: dos Santos Bernardes MA, editor. *Biofuel's Engineering Process Technology.* Croatia: Rijeka; 2011. pp. 715–732.
  17. Zhang J, Yan J, Zhang Y, Ma Z, Bai S, Wu Y, et al. Molecular insights of genetic variation in *Erianthus arundinaceus* populations native to China. *PLoS ONE.* 2013; 8: e80388. doi: [10.1371/journal.pone.0080388](https://doi.org/10.1371/journal.pone.0080388) PMID: [24282538](https://pubmed.ncbi.nlm.nih.gov/24282538/)
  18. Clark LV, Stewart JR, Nishiwaki A, Toma Y, Kjeldsen JB, Jorgensen U, et al. Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. *J Exp Bot.* 2015; 66: 4213–4225. doi: [10.1093/jxb/eru511](https://doi.org/10.1093/jxb/eru511) PMID: [25618143](https://pubmed.ncbi.nlm.nih.gov/25618143/)
  19. Mukherjee SK. Origin and distribution of *Saccharum*. *Bot Gaz.* 1957; 119: 55–61.
  20. Daniels J, Roach BT. Taxonomy and evolution. In: Heinz DJ, editor. *Sugarcane improvement through breeding.* Amsterdam: Elsevier; 1987. pp. 7–84.
  21. Al Janabi SM, McClelland M, Petersen C, Sobral BWS. Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. *Theor Appl Genet.* 1994; 88: 933–944. doi: [10.1007/BF00220799](https://doi.org/10.1007/BF00220799) PMID: [24186245](https://pubmed.ncbi.nlm.nih.gov/24186245/)
  22. Besse P, McIntyre CL, Berding N. Ribosomal DNA variations in *Erianthus*, a wild sugarcane relative (Andropogoneae-Saccharinae). *Theor Appl Genet.* 1996; 92: 733–743. doi: [10.1007/BF00226096](https://doi.org/10.1007/BF00226096) PMID: [24166398](https://pubmed.ncbi.nlm.nih.gov/24166398/)
  23. Nair NV, Selvi A, Sreenivasan TV, Pushpalatha KN, Mary S. Molecular diversity among *Saccharum*, *Erianthus*, *Sorghum*, *Zea* and their hybrids. *Sugar Tech.* 2005; 7: 55–59.
  24. Suman A, Ali K, Arro J, Parco AS, Kimbeng CA, Baisakh N. Molecular diversity among members of the *Saccharum* complex assessed using TRAP markers based on lignin-related genes. *Bioenerg Res.* 2012; 5: 197–205.
  25. Welker CA, Souza-Chies T, Longhi-Wagner HM, Peichoto MC, McKain MR, Kellogg E. Phylogenetic analysis of *Saccharum* S. L. (Poaceae; Andropogoneae), with emphasis on the circumscription of the south American species. *Am J Bot.* 2015; 102: 248–263. doi: [10.3732/ajb.1400397](https://doi.org/10.3732/ajb.1400397) PMID: [25667078](https://pubmed.ncbi.nlm.nih.gov/25667078/)
  26. Besse P, McIntyre CL, Berding N. Characterisation of *Erianthus* sect. *Ripidium* and *Saccharum* germplasm (Andropogoneae—Saccharinae) using RFLP markers. *Euphytica.* 1997 93: 283–292.
  27. Besse P, Taylor G, Carroll B, Berding N, Burner D, McIntyre CL. Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. *Genetica.* 1998; 104: 143–153. doi: [10.1023/A:1003436403678](https://doi.org/10.1023/A:1003436403678) PMID: [16220373](https://pubmed.ncbi.nlm.nih.gov/16220373/)
  28. de Cesare M, Hodkinson TR, Barth S. Chloroplast DNA markers (cpSSRs, SNPs) for *Miscanthus*, *Saccharum* and related grasses (Panicoideae, Poaceae). *Mol Breed.* 2010; 26: 539–544.
  29. Birky CW Jr. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci USA.* 1995; 92: 11331–11338. PMID: [8524780](https://pubmed.ncbi.nlm.nih.gov/8524780/)
  30. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellite: new tools for studies in plant ecology and evolution. *Trend Ecol Evol.* 2001; 16: 142–147.

31. Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res.* 2004; 11: 93–99. PMID: [15449542](#)
32. Calsa JT, Carraro DM, Benatti MR, Barbosa AC, Kitajima JP, Carrer H. Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. *Curr Genet.* 2004; 46: 366–373. doi: [10.1007/s00294-004-0542-4](#) PMID: [15526204](#)
33. Nah G, Im JH, Lim SH, Kim K, Choi Y, Yook MJ, et al. Complete chloroplast genome of two *Miscanthus* species. *Mitochondrial DNA.* 2015; 27: 4359–4360. doi: [10.3109/19401736.2015.1089500](#) PMID: [26465710](#)
34. Hand ML, Spangenberg GC, Foster JW, Cogan NO. Plastome sequence determination and comparative analysis for members of the *Lolium-Festuca* grass species complex. *G3.* 2013; 3: 607–616. doi: [10.1534/g3.112.005264](#) PMID: [23550121](#)
35. Loman NJ, Constantinoiu C, Chan JZM, Halachev M, Sergeant M, Penn CW, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Rev Microbiol.* 2012; 10: 599–606.
36. Saski C, Lee S-B, Fjellheim S, Guda C, Jansen RK, Luo H, et al. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet.* 2007; 115: 571–590. doi: [10.1007/s00122-007-0567-4](#) PMID: [17534593](#)
37. Young HA, Lanzatella CL, Sarath G, Tobias CM. Chloroplast genome variation in upland and lowland switchgrass. *PLoS ONE.* 2011; 6: e23980. doi: [10.1371/journal.pone.0023980](#) PMID: [21887356](#)
38. Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA, et al. Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS ONE.* 2013; 8: e57533. doi: [10.1371/journal.pone.0057533](#) PMID: [23460871](#)
39. Neckermann K, Zeltz P, Igloi GL, Kossel H, Maier RM. The role of RNA editing in conservation of start codons in chloroplast genomes. *Gene.* 1994; 146: 177–182. PMID: [8076816](#)
40. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *J Mol Evol.* 2010; 70: 149–166. doi: [10.1007/s00239-009-9317-3](#) PMID: [20091301](#)
41. Wang XL, Fan X, Zeng J, Sha LN, Zhang HQ, Kang HY, et al. Phylogeny and molecular evolution of the *DMC1* gene within the StH genome species in Triticeae (Poaceae). *Gen Genomics.* 2012; 34: 237–244.
42. Schwerdt JG, MacKenzie K, Wright F, Oehme D, Wagner JM, Harvey AJ, et al. Evolutionary dynamics of the cellulose synthase gene superfamily in grasses. *Plant Physiol.* 2015; 168: 968–983. doi: [10.1104/pp.15.00140](#) PMID: [25999407](#)
43. Wu Z-Q, Ge S. The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Mol Phylogenet Evol.* 2012; 62: 573–578. doi: [10.1016/j.ympev.2011.10.019](#) PMID: [22093967](#)
44. Alix K, Paulet F, Glaszmann JC, D'Hont A. Inter-Alu-like species-specific sequences in the *Saccharum* complex. *Theor Appl Genet.* 1999; 99: 962–968.
45. Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, et al. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol.* 2006; 6: 17. doi: [10.1186/1471-2229-6-17](#) PMID: [16934154](#)
46. Huse S, Huber J, Morrison H, Sogin M, Weich D. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 2007; 8: R143. doi: [10.1186/gb-2007-8-7-r143](#) PMID: [17659080](#)
47. Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE. A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl plant Sci.* 2014; 12: 1400059.
48. Yook MJ, Lim S-H, Song J-S, Kim J-W, Zhang C-J, Lee EJ, et al. Assessment of genetic diversity of Korean *Miscanthus* using morphological traits and SSR markers. *Biomass Bioenergy.* 2014; 66: 81–92.
49. Chaw SM, Chang CC, Chen HL, Li W-H. Dating the monocot–dicot divergence and the origin of core Eudicots using whole chloroplast genomes. *J Mol Evol.* 2004; 58: 424–441. doi: [10.1007/s00239-003-2564-9](#) PMID: [15114421](#)
50. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics.* 2007; 8: 174. doi: [10.1186/1471-2164-8-174](#) PMID: [17573971](#)
51. Kim C, Wang X, Lee T-H, Jakob K, Lee G-J, Paterson AH. Comparative analysis of *Miscanthus* and *Saccharum* reveals a shared whole-genome duplication but different evolutionary fates. *The Plant J.* 2014; 26: 2420–2429.

52. Estep MC, McKain MR, Diaz DV, Zhong J, Hodge JG, Hodkinson TR, et al. Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc Natl Acad Sci USA*. 2014; 111: 15149–15154. doi: [10.1073/pnas.1404177111](https://doi.org/10.1073/pnas.1404177111) PMID: [25288748](https://pubmed.ncbi.nlm.nih.gov/25288748/)
53. Ni Y, Asamoah-Odei N, Sun G. Maternal origin, genome constitution and evolutionary relationships of polyploid *Elymus* species and *Hordelymus europaeus*. *Biologia Plantarum*. 2011; 55: 68–74.
54. Gao G, Tang Z, Wang Q, Gou X, Ding C, Zhang L. Phylogeny and maternal donor of *Kengyilia* (Triticeae: Poaceae) based on chloroplast *trnT-trnL* sequences. *Biochem Sys Ecol*. 2014; 57: 102–107.
55. Lózsa R, Xia N, Deák T, Bisztray GD. Chloroplast diversity indicates two independent maternal lineages in cultivated grapevine (*Vitis vinifera* L. subsp. *vinifera*). *Genet Resor Crop Evol*. 2015; 62: 419–429.
56. Li L, Wang HY, Zhang C, Wang X-F, Shi F-X, Chen W-N, et al. Origins and Domestication of Cultivated Banana Inferred from Chloroplast and Nuclear Genes. *PLoS ONE*. 2013; 8: e80502. doi: [10.1371/journal.pone.0080502](https://doi.org/10.1371/journal.pone.0080502) PMID: [24260405](https://pubmed.ncbi.nlm.nih.gov/24260405/)
57. Estep MC, Vela Diaz DM, Zhong J, Kellogg EA. Eleven diverse nuclear-encoded phylogenetic markers for the subfamily Panicoideae (Poaceae). *Am J Bot*. 2012; 99: e443–e446. doi: [10.3732/ajb.1200186](https://doi.org/10.3732/ajb.1200186) PMID: [23108465](https://pubmed.ncbi.nlm.nih.gov/23108465/)
58. Murray MG, Thompson WF. Rapid isolation of high-molecular-weight plant DNA. *Nucleic Acids Res*. 1980; 8: 4321–4325. PMID: [7433111](https://pubmed.ncbi.nlm.nih.gov/7433111/)
59. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437: 376–380. doi: [10.1038/nature03959](https://doi.org/10.1038/nature03959) PMID: [16056220](https://pubmed.ncbi.nlm.nih.gov/16056220/)
60. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20: 3252–3255. doi: [10.1093/bioinformatics/bth352](https://doi.org/10.1093/bioinformatics/bth352) PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
61. Conant GC, Wolfe KH. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*. 2008; 24: 861–862. doi: [10.1093/bioinformatics/btm598](https://doi.org/10.1093/bioinformatics/btm598) PMID: [18227121](https://pubmed.ncbi.nlm.nih.gov/18227121/)
62. Faircloth BC. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol Ecol Resor*. 2008; 8: 92–94.
63. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004; 32: W273–279. doi: [10.1093/nar/gkh458](https://doi.org/10.1093/nar/gkh458) PMID: [15215394](https://pubmed.ncbi.nlm.nih.gov/15215394/)
64. Xu B, Yang Z. PAMLX: a graphical user interface for PAML. *Mol Biol Evol*. 2013; 30: 2723–2724. doi: [10.1093/molbev/mst179](https://doi.org/10.1093/molbev/mst179) PMID: [24105918](https://pubmed.ncbi.nlm.nih.gov/24105918/)
65. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006; 34: 609–612.
66. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res*. 1997; 25: 4876–4882. PMID: [9396791](https://pubmed.ncbi.nlm.nih.gov/9396791/)
67. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*. 2013; 30: 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
68. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evol*. 1985; 39: 783–791.
69. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2014; 10: e1003537. doi: [10.1371/journal.pcbi.1003537](https://doi.org/10.1371/journal.pcbi.1003537) PMID: [24722319](https://pubmed.ncbi.nlm.nih.gov/24722319/)
70. Durraba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*. 2012; 9: 772.
71. Drummond A, Ho S, Phillips M, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006; 4: e88. doi: [10.1371/journal.pbio.0040088](https://doi.org/10.1371/journal.pbio.0040088) PMID: [16683862](https://pubmed.ncbi.nlm.nih.gov/16683862/)
72. Prasad V, Strömberg CAE, Leaché AD, Samant B, Patnaik R, Tang L, et al. Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature Communications*. 2011; 2: 480. doi: [10.1038/ncomms1482](https://doi.org/10.1038/ncomms1482) PMID: [21934664](https://pubmed.ncbi.nlm.nih.gov/21934664/)
73. Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA. The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biology*. 2008; 14: 2963–297.
74. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v.1.6. 2014. Available from: <http://beast.bio.ed.ac.uk/Tracer>.
75. Rambaut A. Figtree v.1.4.2. 2014. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>.