

RESEARCH ARTICLE

Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae

Ya-Yi Huang, Shu-Ting Cho, Mindia Haryono, Chih-Horng Kuo*

Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

* chk@gate.sinica.edu.tw



OPEN ACCESS

Citation: Huang Y-Y, Cho S-T, Haryono M, Kuo C-H (2017) Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. PLoS ONE 12(6): e0179055. <https://doi.org/10.1371/journal.pone.0179055>

Editor: Hector Candela, Universidad Miguel Hernández de Elche, SPAIN

Received: November 3, 2016

Accepted: May 23, 2017

Published: June 15, 2017

Copyright: © 2017 Huang 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: Illumina reads are publicly available from the NCBI SRA under the accession number SRR5457035 (<https://www.ncbi.nlm.nih.gov/sra/SRR5457035>). The complete genome sequence is available from the NCBI GenBank under the accession number KY024482.1 (<https://www.ncbi.nlm.nih.gov/nuccore/KY024482.1>).

Funding: This study is supported by the Institute of Plant and Microbial Biology at Academia Sinica. The funder had no role in study design, data

Abstract

Common bermudagrass (*Cynodon dactylon* (L.) Pers.) belongs to the subfamily Chloridoideae of the Poaceae family, one of the most important plant families ecologically and economically. This grass has a long connection with human culture but its systematics is relatively understudied. In this study, we sequenced and investigated the chloroplast genome of common bermudagrass, which is 134,297 bp in length with two single copy regions (LSC: 79,732 bp; SSC: 12,521 bp) and a pair of inverted repeat (IR) regions (21,022 bp). The annotation contains a total of 128 predicted genes, including 82 protein-coding, 38 tRNA, and 8 rRNA genes. Additionally, our in silico analyses identified 10 sets of repeats longer than 20 bp and predicted the presence of 36 RNA editing sites. Overall, the chloroplast genome of common bermudagrass resembles those from other Poaceae lineages. Compared to most angiosperms, the *accD* gene and the introns of both *clpP* and *rpoC1* genes are missing. Additionally, the *ycf1*, *ycf2*, *ycf15*, and *ycf68* genes are pseudogenized and two genome rearrangements exist. Our phylogenetic analysis based on 47 chloroplast protein-coding genes supported the placement of common bermudagrass within Chloridoideae. Our phylogenetic character mapping based on the parsimony principle further indicated that the loss of the *accD* gene and *clpP* introns, the pseudogenization of four *ycf* genes, and the two rearrangements occurred only once after the most recent common ancestor of the Poaceae diverged from other monocots, which could explain the unusual long branch leading to the Poaceae when phylogeny is inferred based on chloroplast sequences.

Introduction

The Poaceae family, also known as the grass family, is one of the most important plant families, both economically and ecologically. The best-known examples are cereal grasses that provide staple food for humans around the world, e.g., rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), corn (*Zea mays* L.) and sorghum (*Sorghum bicolor* (L.) Moench). Systematically the

collection and analysis, decision to publish, or preparation of the manuscript.

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

Poaceae family is composed of 12 subfamilies. In addition to three basal lineages (Anomochlooideae, Pharioideae, and Puelioideae), the remaining subfamilies are divided into two major lineages: the BOP clade, which consists of the subfamilies Bambusoideae, Oryzoideae and Pooideae [1], and the PACMAD clade, which includes Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae [2]. Most of the subfamilies contain fewer than 400 species, except for the following four [3]: the Pooideae (4,234 spp.), the Panicoideae (3,560 spp.), the Bambusoideae (1,641 spp.), and the Chloridoideae (1,601 spp.).

The Poaceae family is also evolutionary significant in terms of the chloroplast genome. In 1982, a comparative study of corn, spinach, petunia, cucumber and mung bean using DNA hybridizations indicated two possible rearrangements specific to corn [4]. The rearrangements were further confirmed after the chloroplast genome of the rice was published [5]. To date there are more than 300 chloroplast genomes of the Poaceae deposited in GenBank. More than 88% of the sequenced taxa are from the four subfamilies closely related to human culture: Panicoideae (26.5%), Oryzoideae (23.34%), Bambusoideae (19.87%), and Pooideae (19.24%). Compared with the top three largest subfamilies, the subfamily Chloridoideae is relatively understudied. There are only 16 chloroplast genomes (5%) in GenBank. In order to expand the data pool, we sequenced the chloroplast genome of common bermudagrass (*Cynodon dactylon* (L.) Pers.), a species that also has a long term connection with human culture yet a complete chloroplast genome has not been determined and is often absent from family phylogeny.

Common bermudagrass is a perennial grass belonging to the largest tribe Cynodonteae (839 spp.) of the subfamily Chloridoideae [3]. Originating from south eastern Africa [6], common bermudagrass not only has a deep cultural history in south Asia but also has colonized six continents and sub-Antarctic islands [7]. Its ability of propagation in diverse environments through rhizomes and stolons makes it a cosmopolitan invasive weed. Nonetheless, the same characteristics also make it a valuable lawn/turf grass that is widely applied in golf courses, roadside slopes or sport grounds. We investigated the chloroplast genome of common bermudagrass and conducted a genome wide comparison to study the rearrangements, gene loss/pseudogenization, and IR expansions and contractions in Poaceae.

Materials and methods

Whole genome sequencing and *de novo* assembly

Fresh plant was collected from a parking lot at Guanyin Beach in western Taoyuan City in Taiwan (25°02'47.6", 121°04'28.7"), a public space where no specific permissions were required for collecting common bermudagrass, one of the common weeds in Taiwan. The procedures for sample preparation, sequencing, and assembly were based on those described in our previous studies [8–10]. Briefly, total genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, USA) following the manufacturer's protocol. High quality DNA (concentration >100 ng/μl; A260/230>1.7; A260/280 = 1.8~2.0) was prepared for Illumina MiSeq paired-end sequencing at the core facilities of our institution (see Acknowledgements). The read length is 301 bp and the insert size is ~550 bp. The raw reads were trimmed from the 5' end at the first bp that has a quality score lower than 20. The reads shorter than 200 bp or do not have a paired read after trimming were excluded from the initial *de novo* assembly. The Perl scripts that were used to trim and filter the Illumina reads are publicly available at GitHub (<https://github.com/chihhorngkuo/perl>). The *de novo* assembly was performed using Velvet version 1.2.10 [11] with the following settings: k = 151, scaffolding = no, exp_cov = auto, -cov_cutoff = 10, -max_coverage = 500, and -min_contig_lgth = 2000. All resulting contigs were used as the query to run TBLASTX [12, 13] searches against the complete chloroplast genome of *Spartina maritima* [14] with the following cutoff: e-value = 10⁻¹⁵ and sequence

identity = 0.8. A total of 15 putative chloroplast contigs were identified and used as the first version of the draft assembly for further improvement until the complete chloroplast genome sequence was obtained.

To improve the assembly, the complete chloroplast genome of *Spartina maritima* [14] was used as a guide for scaffolding. In each round of our iterative process, all Illumina raw reads were mapped to the draft assembly using BWA version 0.7.12 [15], programmatically checked using the MPileUP program in SAMTOOLS package version 1.2 [16], and visually inspected using IGV version 2.3.41 [17]. The polymorphic sites and gaps were manually corrected using the mapped reads when appropriate. For regions that could not be determined confidently using Illumina reads, such as regions with low coverage or the junctions between the single copy regions and the repeats, PCR and Sanger sequencing were used for validation. A complete list of the primer sequences used for this part is provided in Supporting Information S1 Table.

Genome annotation and visualization

Preliminary gene prediction was performed with the online program DOGMA [18], followed by manual inspection. Sequence alignment with homologous genes was implemented to identify the exact boundaries of genes and introns. All tRNA genes were predicted by tRNAscan-SE search server [19]. Annotated genome was submitted to online server GenomeVx [20] for visualization.

Repeat structure and RNA editing

Repeat sequences longer than 20 nucleotides were predicted by Tandem Repeats Finder [21] with the following parameters: (2, 7, 7) for alignment parameters (match, mismatch, indels), 80 for minimum alignment score to report repeat, and maximum period size of 500. Potential RNA editing sites in protein-coding genes were predicted by Predictive RNA Editor for Plants (PREP) suite with a cutoff value of 0.8 [22].

Phylogenetic analysis and character evolution

Our phylogenetic analysis contains 28 taxa, including 24 species of Poaceae that represent all 12 subfamilies. Two species of basal Poales (*Typha latifolia* L. and *Ananas comosus* (L.) Merr.) along with two non-Poales monocots (*Cocos nucifera* L. and *Kingia australis* R. Br.) were used as outgroups. GenBank accession numbers of the sampled taxa are listed in Table 1. Amino acid sequences of 47 conserved protein-coding genes were aligned by MUSCLE [23] and concatenated for phylogenetic analysis using PhyML with parameters estimated from the data [24]. Bootstrap re-sampling with 1,000 replicates was used to evaluate the branch supports. Events of genome rearrangement, gene loss, pseudogenization, and duplication were mapped onto the phylogenetic tree based on the parsimony principle.

Result and discussion

Characteristics and rearrangements of the genome

The complete chloroplast genome of common bermudagrass is a circular quadripartite molecule with a length of 134,297 bp (GenBank accession number KY024482.1). The Illumina reads used to generate this assembly were deposited at the NCBI Sequence Read Archive under the accession number SRR5457035. This genome comprises a large single copy (LSC) region (79,732 bp), a small single copy (SSC) region (12,521 bp), and a pair of inverted repeat (IR) regions (42,044 bp). There are 128 genes predicted, including 82 protein-coding, 38

Table 1. Sampled taxa along with their accession numbers in this study.

Taxon	GenBank accession number
Monocots: Poaceae	
Subfamily Anomochlooideae	
<i>Anomochloa marantoidea</i>	GQ329703
Subfamily Pharoideae	
<i>Pharus latifolius</i>	JN032131
Subfamily Puelioideae	
<i>Puelia olyriformis</i>	KC534841
Subfamily Oryzoideae	
<i>Leersia tisserantii</i>	JN415112
<i>Oryza rufipogon</i>	KF428978
Subfamily Bambusoideae	
<i>Bambusa bambos</i>	KJ870988
<i>Phyllostachys edulis</i>	HQ337796
Subfamily Pooideae	
<i>Agrostis stolonifera</i>	EF115543
<i>Brachypodium distachyon</i>	EU325680
<i>Poa palustris</i>	KM974749
<i>Triticum aestivum</i>	AB042240
Subfamily Aristidoideae	
<i>Aristida purpurea</i>	KJ920224
Subfamily Panicoideae	
<i>Centotheca lappacea</i>	KJ920225
<i>Coleataenia prionitis</i>	KJ920228
<i>Panicum virgatum</i>	HQ822121
<i>Thysanolaena latifolia</i>	KJ920236
Subfamily Arundinoideae	
<i>Elytrophorus spicatus</i>	KJ920230
<i>Monachather paradoxus</i>	KJ920235
Subfamily Micrairoideae	
<i>Isachne distichophylla</i>	KJ920233
<i>Micraira</i> sp.	KJ920234
Subfamily Danthonioideae	
<i>Chionochloa macra</i>	KJ920227
<i>Danthonia californica</i>	KJ920229
Subfamily Chloridoideae	
<i>Cynodon dactylon</i>	KY024482
<i>Neyraudia reynaudiana</i>	KF356392
Monocots: non-Poaceae	
<i>Typha latifolia</i>	GU195652
<i>Ananas comosus</i>	KR336549
<i>Cocos nucifera</i>	KF285453
<i>Kingia australis</i>	JX051651
Dicots	
<i>Amborella trichopoda</i>	AJ506156
<i>Catharanthus roseus</i>	KC561139

<https://doi.org/10.1371/journal.pone.0179055.t001>

tRNA, and eight rRNA genes. Of these genes, 18 genes along with the second and third exons of the *rps12* gene (3'*rps12*) have two copies, one in each of the IR regions. In addition, there were nine pseudogenes identified. Except the pseudo *rpl23* located in the LSC region, the remaining eight pseudogenes were all in the IR region, i.e. two copies each of the *ycf1*, *ycf2*, *ycf15*, and *ycf68* genes. Pseudo *rpl23* is one truncated fragment (240 bp) without a start or a stop codon; pseudo *ycf15* has an internal stop codon; pseudo *ycf1*, *ycf2* and *ycf68* genes have degraded to residual fragments with multiple internal stop codons. Particularly, the *ycf1* and *ycf2*, which normally have more than 5,000 bp in size, have reduced to 840 bp and 1,135 bp respectively in the chloroplast genome of common bermudagrass. The *accD* gene normally locating between *rbcL* and *psbI* genes in the LSC region in most angiosperms was lost, along with the introns of both *clpP* and *rpoC1* genes (Fig 1).

Structurally the chloroplast genome of common bermudagrass is similar to those of other Poaceae but distinct from most angiosperms. Sequence alignment with *T. latifolia*, one of the basal taxa of the Poales, indicated two significant rearrangements in the LSC region. The first rearrangement was an inversion of a fragment ca. 28 kb between the *rps14* and *trnS-GCU* genes, resulting in the relocation and reorientation of 23 genes (hypothetical Intermediate I in Fig 2). After the first inversion, a second inversion occurred between the *trnS-GCU* and *trnT-GGU* genes, affecting seven genes (hypothetical Intermediate II). Moreover, except for *Anomochloa*, the *rpoC1* intron was absent from all sampled Poaceae, represented by *Pharus latifolius* L. and *C. dactylon* in Fig 2.

Chloroplast *rpoC1* along with *rpoC2* genes of the angiosperms are homologous to the bacterial β' subunit [25, 26]. Together with the *rpoB* gene, they form an operon analogous to the *rpoBC* operon of *Escherichia coli* that encodes subunits of the RNA polymerase [26]. The presence of an intron in the *rpoC1* gene has been reported in most land plants, including the earliest bryophytes [27, 28]. Nonetheless, the absence of the *rpoC1* intron has also been observed sporadically in several angiosperm lineages, i.e. most Poaceae [5, 29, 30], the subfamily Cactoidae of the Cactaceae [31] and some species within the families of Passifloraceae, Aizoaceae, Goodeniaceae, and Fabaceae [30]. Our survey of the 12 subfamilies of the Poaceae further confirmed the intron loss of the *rpoC1* in all sampled Poaceae, except for the *Anomochloa marantoides* Brongn., within which the *rpoC1* intron has been retained (Fig 2).

Similarly, the *clpP* gene, a proteolytic subunit of the ATP-dependent Clp protease, normally contains two introns in most land plants. However, the loss of intron 1 has been recorded in distantly related eudicots such as the IR lacking clade of the Papilionoids [32] and *Cuscuta* of the Convolvulaceae [33]. In others, i.e. *Pinus* of the Pinaceae [34, 35], all the Poaceae [36], including common bermudagrass reported in this study, some *Oenothera* of the Onagraceae [37], some *Lychnis* and *Silene* of the Caryophyllaceae [37, 38], and *Jasminum* and *Menodora* of the Oleaceae [39], both introns have been lost independently from the *clpP* gene.

In the chloroplast genomes of most land plants, the *accD* gene encodes a component of acetyl-CoA carboxylase (ACCase) equivalent to bacterial β subunit [40]. This gene, however, has either pseudogenized or completely lost from some species of Campanulaceae [41, 42], Geraniaceae [43], Oleaceae [39], and all Poaceae [40, 44], including common bermudagrass examined in this study. An experiment of streptavidin probe combined with Southern hybridization demonstrated that instead of the presence of prokaryotic ACCase in plastids and eukaryotic ACCase in cytosol as found in most vascular plants, members of the Poaceae possess only eukaryotic ACCase, both in plastids and in cytosol, suggesting the nucleus-encoded substitute for a plastid-encoded protein [44]. In contrast to the relocation of the chloroplast genes into the nucleus to reduce the size of the genome suffering selection pressure, e.g., the *tufA* gene of the angiosperms [45], the *accD* gene has been lost from chloroplast genomes and the ACCase in plastids is encoded by a nuclear gene in Poaceae. Plastid proteins encoded by

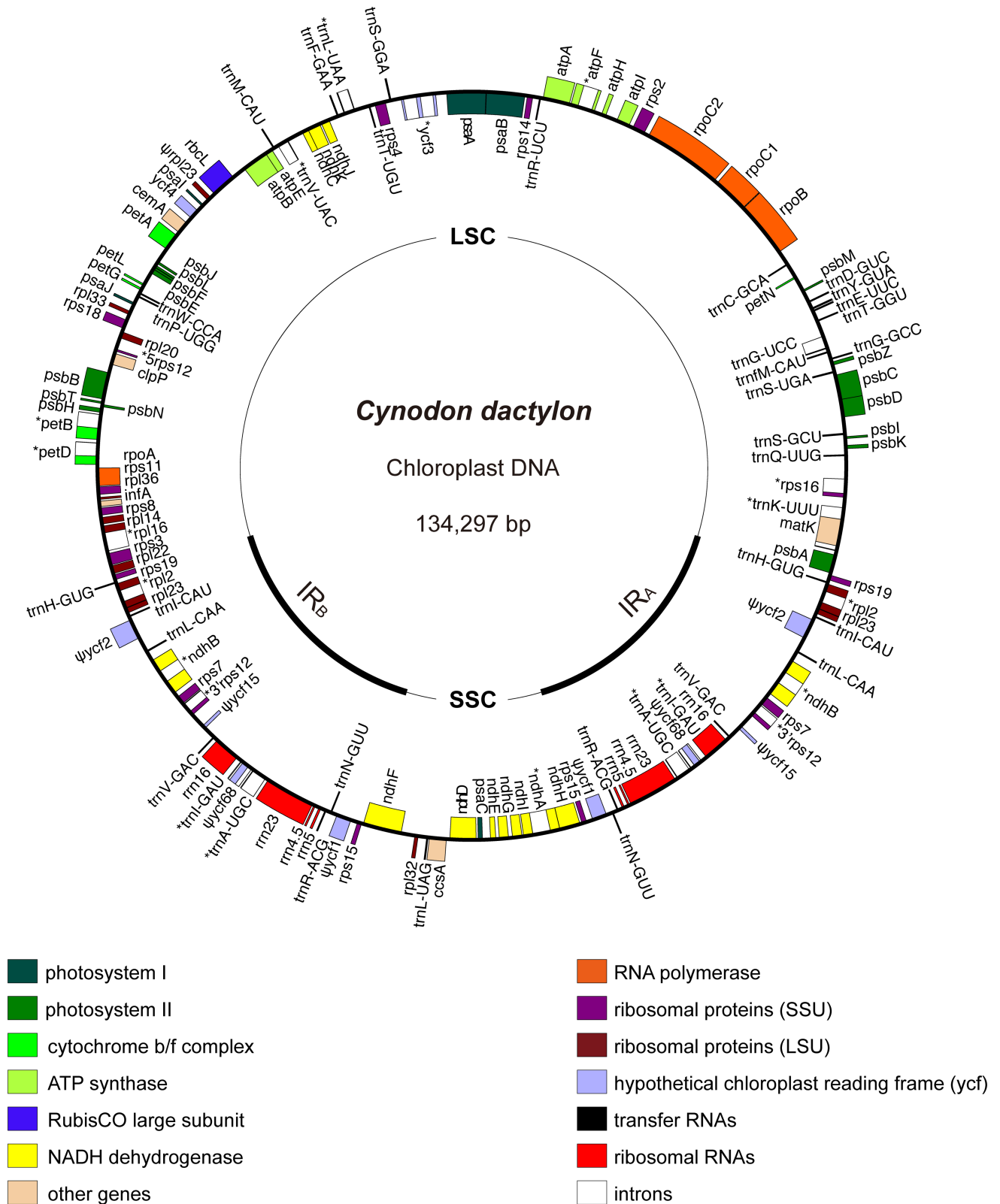


Fig 1. Chloroplast genome map of common bermudagrass. Genes shown on the outside of the large circle are transcribed clockwise, while genes shown on the inside are transcribed counterclockwise. The small circle indicates IRs. Genes with intron are marked with “*”. Pseudogenes are marked with “Ψ”.

<https://doi.org/10.1371/journal.pone.0179055.g001>

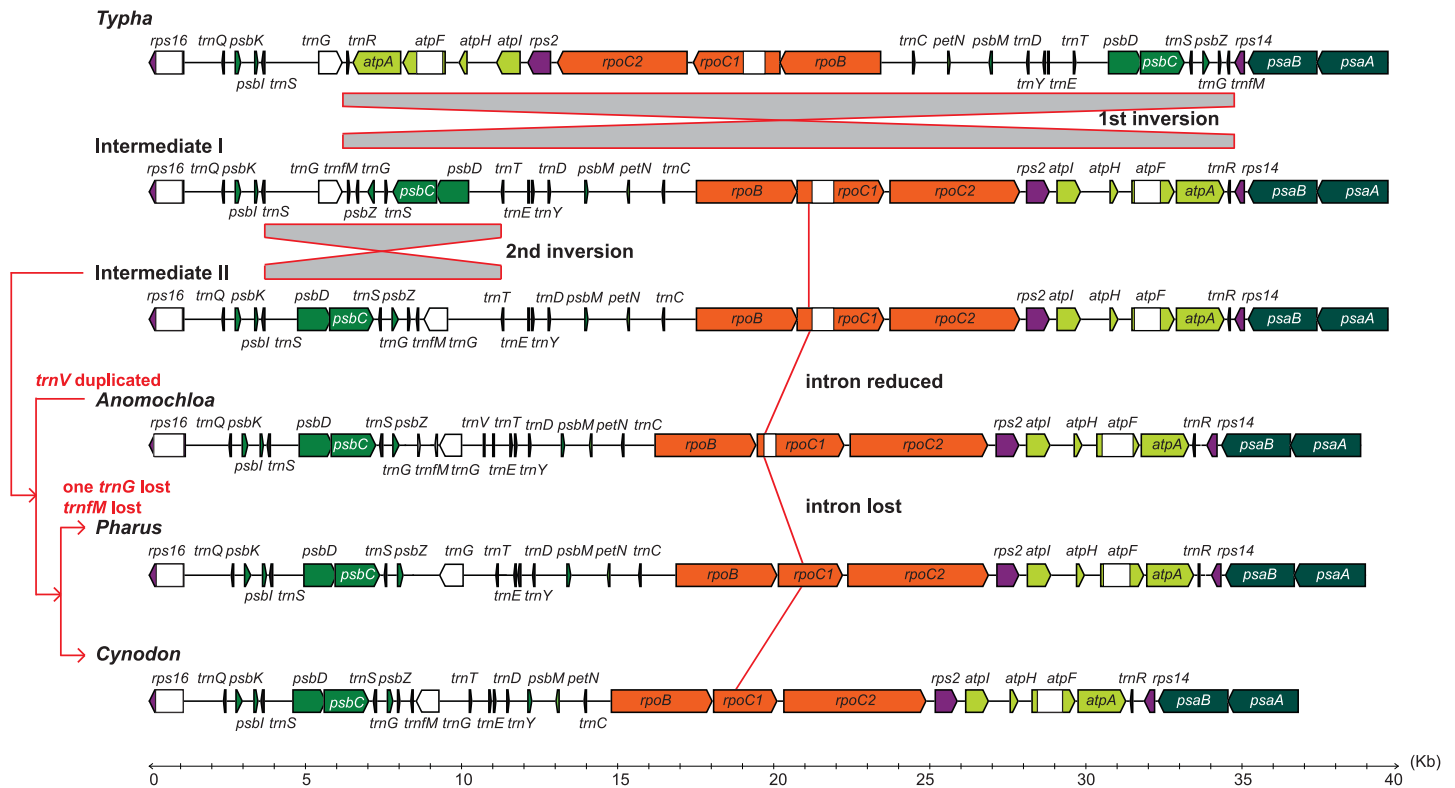


Fig 2. Simple evolutionary model for LSC rearrangements of the Poaceae. Ancestral chloroplast genome of the Poaceae (*Typha*-like) first underwent a major inversion (hypothetical Intermediate I) and then a small inversion (hypothetical Intermediate II).

<https://doi.org/10.1371/journal.pone.0179055.g002>

nuclear genes are not unprecedented. It is found in the case of a ribosomal protein of spinach [46]. However, the deletion of the *accD* gene from chloroplast genomes and its related protein encoded by a nuclear gene in Poaceae perhaps is the first example for a non ribosomal component [44]. In conclusion, to reduce the size of a chloroplast genome, in addition to gene transfer from chloroplast to nucleus, another possibility is to delete a chloroplast gene and let a nuclear gene encode the related protein.

IR fluctuation

In addition to major rearrangements that unify the entire grass family, minor variations among species were also detected at/near the junctions between the IRs and LSC/SSC regions. For example, comparative analysis showed that, in terms of size, the chloroplast genome of common bermudagrass has the shortest chloroplast genome and the second shortest IRs regions among sampled Poaceae while *P. latifolius* possesses the largest chloroplast genome with the longest IRs (Table 2). Graphical alignment showed that the *ndhH* gene was duplicated and invaded into the IRB region in the BOP clade, represented by *Oryza* and *Poa* in Fig 3. In summary, the IRs experienced constant expansion and contraction during the evolutionary process of the Poaceae chloroplast genomes (Fig 3).

The chloroplast genomes of the Poaceae in general are smaller than most angiosperms due to the pseudogenization of the *ycf1* and *ycf2* genes, two of the longest open reading frames in angiosperms with a size of ca. 5,000 bp and ca. 7,000 bp respectively. Compared to other angiosperms, pseudo *ycf1* and *ycf2* genes have reduced to ca. 840 bp and ca. 1,100 bp in Poaceae.

Table 2. Comparative genomics of Poaceae and other angiosperms.

Characteristics	Monocots						Dicots	
	Poaceae				Non-Poaceae		<i>Amborella trichopoda</i>	<i>Catharanthus roseus</i>
	<i>Cynodon dactylon</i>	<i>Oryza rufipogon</i>	<i>Pharus latifolius</i>	<i>Anomochloa marantoidea</i>	<i>Typha latifolia</i>	<i>Cocos nucifera</i>		
Size (bp)	134,297	134,557	142,077	138,412	161,572	154,731	162,686	154,950
LSC	79,732	80,604	83,340	82,274	89,140	84,230	90,970	85,765
SSC	12,521	12,347	12,527	12,162	19,652	17,391	18,414	17,997
IRs	42,044	41,606	46,210	43,976	52,780	53,110	53,302	51,188
Ratio of LSC (%)	59.4	59.9	58.7	59.4	55.2	54.4	55.9	55.4
Ratio of SSC (%)	9.3	9.2	8.8	8.8	12.2	11.2	11.3	11.6
Ratio of IRs (%)	31.3	30.9	32.5	31.8	32.7	34.3	32.8	33.0

<https://doi.org/10.1371/journal.pone.0179055.t002>

Although both genes are functional and essential for cell survival in chloroplast genomes of dicots [47, 48], in addition to the Poaceae, the pseudogenization of *ycf1* was also observed in chloroplast genome of *Pinus thunbergii* Parl. [34] and *Medicago truncatula* Gaertn. The latter also contains only a pseudogenized *ycf2*. It is possible that either the two genes are not essential as was assumed, or, it is possible that, similar to the case of the *tufA* gene in angiosperms [45], functional *ycf1* and *ycf2* genes were transferred to the nuclei of those taxa that have two pseudogenized *ycf* genes. It is also likely that proteins encoded by these two genes are now encoded by nuclear genes, similar to the case of the *accD* gene in Poaceae. Another interesting phenomenon that we observed in the chloroplast genomes of the Poaceae is the trend of LSC expansion accompanied by the contractions of both IRs and the SSC regions. Consequently the ratios of the LSC regions of the Poaceae are higher than those of other angiosperms whereas the ratios of both the IRs and SSC regions are lower (Table 2).

RNA editing and repeats

Overall 36 RNA editing sites were predicted in 15 genes of common bermudagrass, five of which were species specific (Table 3). All the editing events were non-silent C-to-U, of which seven (19.4%) were at the first position of the codon, including one that altered the initiator codon ACG to AUG in the *rpl2* gene. The remaining 29 (80.6%) were at the second and none was at the third position of the codon. The conversions of amino acids include 25 hydrophilic to hydrophobic (H to Y, S to L, S to F, T to M, and T to I) and 11 hydrophobic to hydrophobic (L to F and A to V, and P to L). The majority of editing sites were predicted in the *ndhB* gene (7 editing sites), followed by the *ndhA* gene (5 editing sites). Comparison of predicted RNA editing among 12 species of the Poaceae, representing 12 subfamilies, showed that all the editing in those sampled taxa were non-silent C-to-U and at either the first or the second positions of the codons. There is a trend of decline in the number of the total editing sites and the first-codon position editing through the evolution of the Poaceae (Table 4), which concurs with the observations of the editing events across land plant lineages [49].

Our repeat search identified 10 sets of repeats longer than 20 bp from the chloroplast genome of common bermudagrass, including three direct repeats and seven tandem repeats (Table 5). All of the repeats were in the LSC region. Seven were in intergenic spacers, two in *rpoC2* and one in *rps18* genes. The length of the repeats ranges between 20 and 67 bp. Repeated sequences are known to correlate with genome rearrangements, which can be demonstrated through chloroplast genome comparison between Poaceae and the palm family, both are

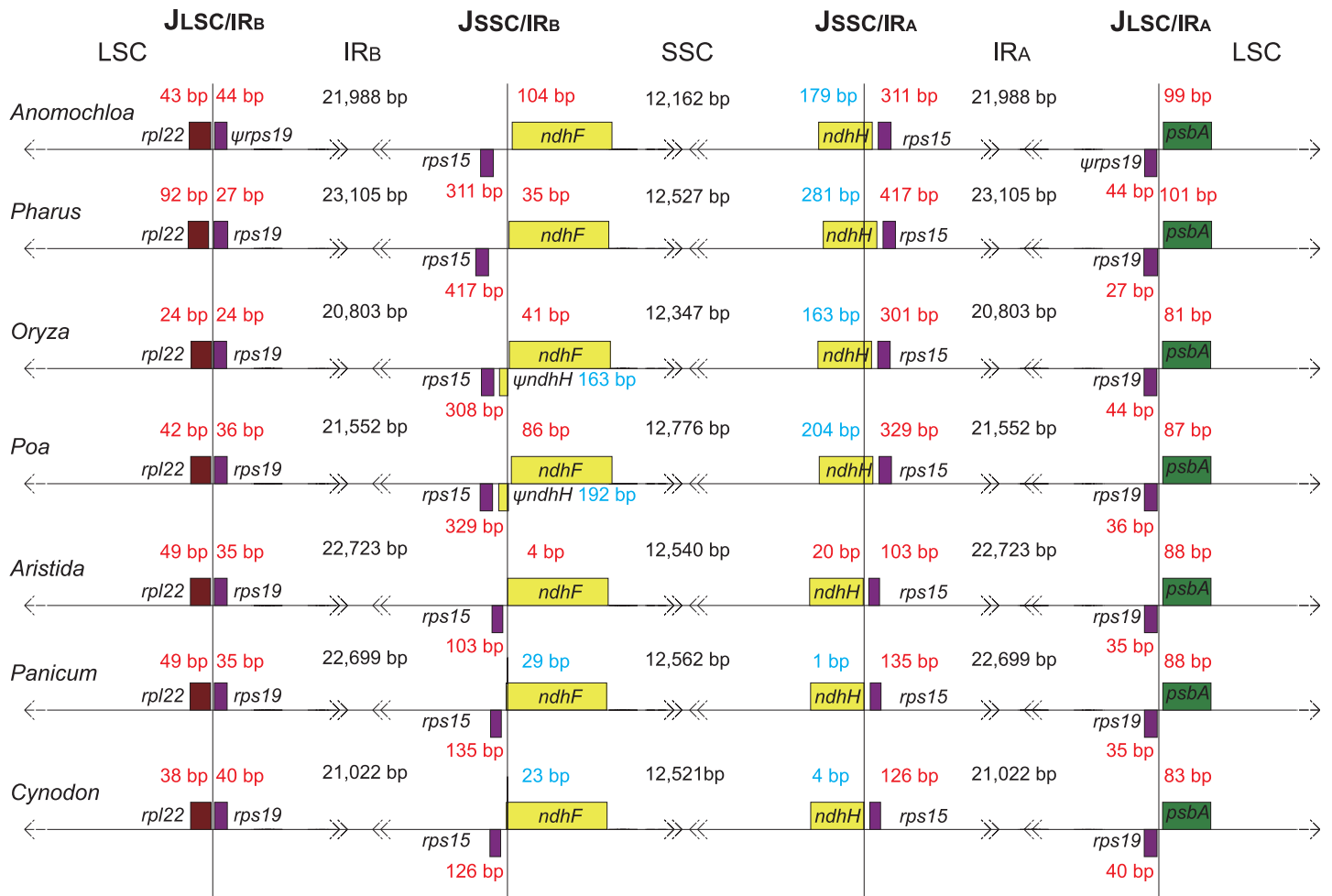


Fig 3. Comparison of IR boundaries among seven grass species. Numbers in red denote distance between border genes (*rpl22*, *rps19*, *rps15*, *ndhF*, and *ndhH*) and junctions of LSC/SSC and IRs. Numbers in blue denote invasion of SSC border genes (*ndhF*, *ndhH* and Ψ *ndhH*) to the IRs.

<https://doi.org/10.1371/journal.pone.0179055.g003>

monocots. The former is one of the best-known plant families with significant rearrangements while the latter is relatively conserved without known dramatic variations. The repeats found in the palm family, representing by coconut, oil palm and date palm range between seven to 13 and the longest one is 39 bp in length [50, 51, 52]. In contrast, repeats in Poaceae are more abundant (>30 in some taxa) and longer in size (>100 bp) [53, 54].

Phylogeny and character mapping

Our phylogenetic analysis based on 47 protein-coding genes of the chloroplast genomes showed that common bermudagrass is sister to *Neyraudia reynaudiana* (Kunth) Keng ex Hitchc., a member of the Chloridoideae. The overall topology is also congruent with current classification of Poaceae [3]: consecutive divergence of three basal lineages (Anomochloideae, Pharioideae, and Puelioideae) followed by the split of the BOP (Bambusoideae, Oryzoideae and Pooideae) and the PACMA (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae) clades. Within the BOP clade, the Oryzoideae is sister to Pooideae and Bambusoideae while in the PACMA clade, the Aristidoideae diverged first, followed by

Table 3. Predicted RNA editing sites and amino acid change.

Gene	Nucleotide position	Codon change	Editing position within codon	Amino acid change
<i>matK</i>	635*	TCT-TTT	2	S-F
	1258	CAT-TAT	1	H-Y
<i>atpA</i>	1148	TCA-TTA	2	S-L
<i>ccsA</i>	647	ACT-ATT	2	T-I
<i>ndhA</i>	50	TCG-TTG	2	S-L
	473	TCA-TTA	2	S-L
	563	TCA-TTA	2	S-L
	919	CTT-TTT	1	L-F
	1070	TCC-TTC	2	S-F
<i>ndhD</i>	637	CTT-TTT	1	L-F
	878	TCA-TTA	2	S-L
<i>ndhF</i>	62	TCA-TTA	2	S-L
	1420*	CAT-TAT	1	H-Y
	1915	CTT-TTT	1	L-F
<i>ndhB</i>	467	CCA-CTA	2	P-L
	586	CAT-TAT	1	H-Y
	611	TCA-TTA	2	S-L
	737	CCA-CTA	2	P-L
	830	TCA-TTA	2	S-L
	836	TCA-TTA	2	S-L
	1481	CCA-CTA	2	P-L
<i>ycf3</i>	44	TCC-TTC	2	S-F
	185	ACG-ATG	2	T-M
<i>petB</i>	611	CCA-CTA	2	P-L
<i>psbE</i>	235*	CTT-TTT	1	L-F
<i>rpoB</i>	467	TCG-TTG	2	S-L
	545	TCA-TTA	2	S-L
	560	TCG-TTG	2	S-L
	617	CCG-CTG	2	P-L
<i>rpoC1</i>	1940	ACC-ATC	2	T-I
<i>rpoC2</i>	2741	TCA-TTA	2	S-L
	3080	CCT-CTT	2	P-L
	3347*	ACA-ATA	2	T-I
<i>rpl2</i>	2	ACG-ATG	2	T-M
<i>rps14</i>	80	TCA-TTA	2	S-L
	104*	GCT-GTT	2	A-V

*' denotes editing sites specific to common bermudagrass.

<https://doi.org/10.1371/journal.pone.0179055.t003>

Panicoideae that is sister to a clade consisting of two subclades: one with Arundinoideae sister to Micrairoideae and the other with Chloridoideae sister to Danthonioideae (Fig 4).

Although similar events of rearrangements, gene pseudonization/loss, and intron loss occurred independently in several angiosperm lineages, our character mapping indicated that the two inversions in the LSC regions (character 1 and 2 in Fig 4), the loss of the *accD* gene (character 3), the pseudogenization of the four *ycf* genes (characters 4–7), and the intron loss of the *clpP* (character 8) gene occurred only once before the Poaceae evolved and diverged from other monocots. Those events may offer an explanation for the unusual long branch

Table 4. Comparison of RNA editing in 12 species of the Poaceae.

Taxa	Total editing sites	1 st codon editing(%)	2 nd codon editing (%)	3 rd codon editing (%)
<i>Anomochloa marantoidea</i>	51	29	71	0
<i>Pharus latifolius</i>	42	12	88	0
<i>Puelia olyrififormis</i>	39	15	85	0
<i>Poa palustris</i>	35	17	83	0
<i>Bambusa bambos</i>	37	19	81	0
<i>Oryza rufipogon</i>	27	7	93	0
<i>Aristida purpurea</i>	37	27	73	0
<i>Panicum virgatum</i>	30	20	80	0
<i>Monachather paradoxus</i>	34	21	79	0
<i>Isachne distichophylla</i>	38	21	79	0
<i>Danthonia californica</i>	31	19	81	0
<i>Cynodon dactylon</i>	36	19	81	0

<https://doi.org/10.1371/journal.pone.0179055.t004>

leading to the Poaceae when phylogenetic reconstruction was built upon chloroplast genes. The loss of the *rpoC1* intron occurred after the divergence of Anomochlooideae (character 11). After the divergence of the Poaceae, other events also occurred independently in several lineages, e.g., the duplication of a *trnV-GAC* gene between *trnG-UCC* and *trnT-GGU* in the LSC region (character 9) and the pseudogenization of *rps19* gene at/near the IR and LSC/SSC junctions in Anomochlooideae (character 10), the loss of both *trnfM* and *trnG-GCC* genes from the LSC region (characters 12–13) in Pharoideae, and the duplication of *ndhH* gene at the IRB and SSC junction in the Oryzoideae, Bambusoideae and Pooideae (character 14).

Table 5. Distribution of repeat sequences in chloroplast genome of common bermudagrass.

No	Size (bp)	Copy number	Type	Repeat sequence	Region
1	67	1.9	D	AGTGGTAGAGTAATGCCATGGTAAGGCATAAGTCATCGGTTCAAATCCGATAAAGGGCTTTTCCTT	LSC; partial <i>trnT</i> , spacer between <i>trnT</i> and <i>trnE</i>
2	42	2.9	D	ACTCTAGAAGACGAATATGAAACCCTAGAAGACGAATATAGG	LSC; <i>rpoC2</i> gene
3	35	1.9	D	TTTTCTTTCTATTCTATTGAAATGGCAAAGGA	LSC; spacer between <i>trnC</i> and <i>rpoB</i>
4	22	2	T	AATCAGATATATTTTCTCTTCA	LSC; spacer between <i>trnS</i> and <i>psbD</i>
5	22	4	T	TCTAAAAGAGGATTAATCCT	LSC; spacer between <i>psbE</i> and <i>petL</i>
6	22	2	T	TTTACTCTGTTTCGACATAAG	LSC; spacer between <i>petG</i> and <i>trnW</i>
7	21	1.9	T	ATTGTCGAATCCTACTCAGCA	LSC; spacer between <i>rpoC1</i> and <i>rpoC2</i>
8	21	3.9	T	ATATAGGGCCCTAGAGGAAGA	LSC; <i>rpoC2</i> gene
9	21	5.8	T	AAACAACCTTTTCGTAAATCC	LSC; <i>rps18</i> gene
10	20	2	T	TATTTTATATTTTATAT	LSC; spacer between <i>rpl33</i> and <i>rps18</i>

'D': direct repeat

'T': tandem repeat

<https://doi.org/10.1371/journal.pone.0179055.t005>

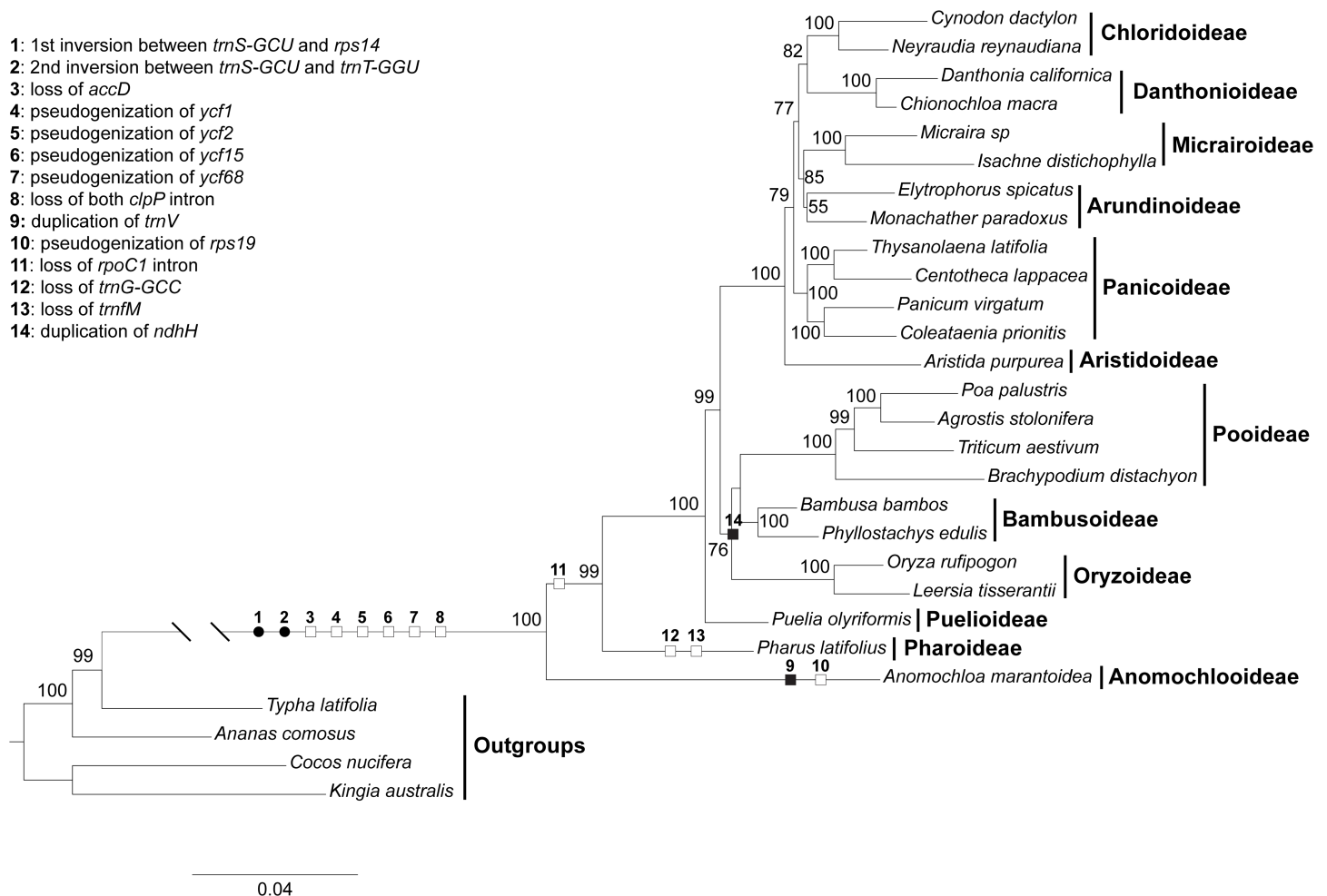


Fig 4. Phylogenetic tree of the Poaceae. Numbers above/below the branches are bootstrap value (only values higher than 50% are shown). Black circle denotes rearrangement events, white square denotes gene loss/pseudogenization and intron loss, and black square denotes gene duplication.

<https://doi.org/10.1371/journal.pone.0179055.g004>

Conclusions

Common bermudagrass has a long term connection with human culture but is relatively understudied in terms of systematics, which may explain its frequent absence from family-level phylogeny of Poaceae. We sequenced and investigated the chloroplast genome of common bermudagrass to fill this gap. Our results showed that the chloroplast genome of common bermudagrass resembles those of other Poaceae in their overall organization and gene content, while distinct from most of the other angiosperms. Our phylogenetic analysis confirmed the position of common bermudagrass within the subfamily Chloridoideae and showed congruent relationships among 12 subfamilies of the Poaceae. This study enriches the genomic resources available for the study of the Poaceae.

Supporting information

S1 Table. Primers for PCR validation.
(DOCX)

Acknowledgments

We thank the DNA Analysis Core Laboratory (Institute of Plant and Microbial Biology, Academia Sinica) for Sanger sequencing service, the DNA Microarray Core Laboratory (Institute of Plant and Microbial Biology, Academia Sinica) for Illumina sequencing library preparation, and the DNA Sequencing Core Facility (Institute of Molecular Biology, Academia Sinica) for Illumina MiSeq sequencing service.

Author Contributions

Conceptualization: YYH CHK.

Data curation: YYH CHK.

Formal analysis: YYH STC.

Funding acquisition: CHK.

Investigation: YYH STC MH.

Methodology: YYH CHK.

Project administration: YYH.

Resources: CHK.

Software: CHK.

Supervision: CHK.

Validation: STC MH.

Visualization: YYH.

Writing – original draft: YYH.

Writing – review & editing: YYH CHK.

References

1. Clark LG, Zhang W, Wendel JF. A phylogeny of the grass Family (Poaceae) based on *ndhF* sequence data. *Syst Bot.* 1995; 20: 436–460. <https://doi.org/10.2307/2419803>
2. Sánchez-Ken JG, Clark LG. Phylogeny and a new tribal classification of the Panicoideae s.l. (Poaceae) based on plastid and nuclear sequence data and structural data. *Am J Bot.* 2010; 97: 1732–1748. <https://doi.org/10.3732/ajb.1000024> PMID: 21616806
3. Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, et al. A worldwide phylogenetic classification of the Poaceae (Gramineae). *J Syst Evol.* 2015; 53: 117–137. <https://doi.org/10.1111/jse.12150>
4. Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 1982; 29: 537–550. [https://doi.org/10.1016/0092-8674\(82\)90170-2](https://doi.org/10.1016/0092-8674(82)90170-2) PMID: 6288261
5. Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, et al. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol Gen Genet.* 1989; 217: 185–194. <https://doi.org/10.1007/BF02464880> PMID: 2770692
6. Harlan JR, de Wet JMJ. Sources of variation in *Cynodon dactylon* (L). *Pers. Crop Sci.* 1969; 9: 774–778. <https://doi.org/10.2135/cropsci1969.0011183X000900060031x>
7. Way AG “A cosmopolitan weed of the world”: following Bermudagrass. *Agric Hist.* 2014; 88: 354–367. <https://doi.org/10.3098/ah.2014.88.3.354>

8. Ku C, Hu JM, Kuo CH. Complete plastid genome sequence of the basal asterid *Ardisia polysticta* Miq. and comparative analyses of asterid plastid genomes. PLOS ONE 2013; 8 (4): e62548. <https://doi.org/10.1371/journal.pone.0062548> PMID: 23638113
9. Ku C, Chung WC, Chen LL, Kuo CH. The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: plastid genome evolution, molecular marker identification, and phylogenetic implications in Asterids. PLOS ONE 2013; 8(6): e68518. <https://doi.org/10.1371/journal.pone.0068518> PMID: 23825699
10. Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH. Complete chloroplast genome sequence of Omani lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. PLOS ONE 2014; 9 (11): e113049. <https://doi.org/10.1371/journal.pone.0113049> PMID: 25398081
11. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18: 821–829. <https://doi.org/10.1101/gr.074492.107> PMID: 18349386
12. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402. <https://doi.org/10.1093/nar/25.17.3389> PMID: 9254694
13. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009; 10: 421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
14. Rousseau-Gueutin M, Bellot S, Martin GE, Boutte J, Chelaifa H, Lima O, et al. The chloroplast genome of the hexaploid *Spartina maritima* (Poaceae, Chloridoideae): Comparative analyses and molecular dating. Mol Phylogenet Evol. 2015; 93: 5–16. <https://doi.org/10.1016/j.ympev.2015.06.013> PMID: 26182838
15. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
17. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotech. 2011; 29: 24–26. <https://doi.org/10.1038/nbt.1754> PMID: 21221095
18. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 2004; 20: 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352> PMID: 15180927
19. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucl Acids Res. 2005; 33: W686–W689. <https://doi.org/10.1093/nar/gki366> PMID: 15980563
20. Conant G, Wolfe K. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 2008; 24: 861–862. <https://doi.org/10.1093/bioinformatics/btm598> PMID: 18227121
21. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucl Acids Res. 1999; 27: 573–580. <https://doi.org/10.1093/nar/27.2.573> PMID: 9862982
22. Mower JP. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucl Acids Res. 2009; 37: W253–W259. <https://doi.org/10.1093/nar/gkp337> PMID: 19433507
23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res. 2004; 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
24. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52: 696–704. <https://doi.org/10.1080/10635150390235520> PMID: 14530136
25. Cozens AL, Walker JE. Pea chloroplast DNA encodes homologues of *Escherichia coli* ribosomal subunit S2 and the β -subunit of RNA polymerase. Biochem J. 1986; 236: 453–460. <https://doi.org/10.1042/bj2360453> PMID: 3530249
26. Hudson GS, Holton TA, Whitfield PR, Bottomley W. Spinach chloroplast *rpoBC* genes encode three subunits of the chloroplast RNA polymerase. J Mol Biol. 1988; 200: 639–654. [https://doi.org/10.1016/0022-2836\(88\)90477-9](https://doi.org/10.1016/0022-2836(88)90477-9) PMID: 3045324
27. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 1986; 322: 572–574. <https://doi.org/10.1038/322572a0>
28. Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucl Acids Res. 2003; 31: 716–721. <https://doi.org/10.1093/nar/gkg155> PMID: 12527781

29. Katayama H, Ogihara Y. Structural alterations of the chloroplast genome found in grasses are not common in monocots. *Curr Genet.* 1993; 23: 160–165. <https://doi.org/10.1007/BF00352016> PMID: 8431958
30. Downie SR, Llanas E, Katz-Downie DS. Multiple independent losses of the *rpoC1* intron in angiosperm chloroplast DNA's. *Syst Bot.* 1996; 21: 135–151. <https://doi.org/10.2307/2419744>
31. Wallace RS, Cota JH. An intron loss in the chloroplast gene *rpoC1* supports a monophyletic origin for the subfamily Cactoideae of the Cactaceae. *Curr Genet.* 1996; 29: 275–281. <https://doi.org/10.1007/BF02221558> PMID: 8595674
32. Jansen RK, Wojciechowski MF, Sanniyasi E, Lee S-B, Daniell H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol Phylogenet Evol.* 2008; 48: 1204–1217. <https://doi.org/10.1016/j.ympev.2008.06.013> PMID: 18638561
33. McNeal JR, Kuehl JV, Boore JL, Leebens-Mack J, dePamphilis CW. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS ONE* 2009; 4(6): e5982. <https://doi.org/10.1371/journal.pone.0005982> PMID: 19543388
34. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci USA.* 1994; 91: 9794–9798. PMID: 7937893
35. Fang MF, Wang YJ, Zu YM, Dong WL, Wang RN, Deng TT, et al. The complete chloroplast genome of the Taiwan red pine *Pinus taiwanensis* (Pinaceae). *Mitochondrial DNA* 2016; 27: 2732–2733. <https://doi.org/10.3109/19401736.2015.1046169> PMID: 26057016
36. 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. <https://doi.org/10.1007/s00239-009-9317-3> PMID: 20091301
37. Erixon P, Oxelman B. Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* Gene. *PLoS ONE* 2008; 3(1): e1386. <https://doi.org/10.1371/journal.pone.0001386> PMID: 18167545
38. Raman G, Park S. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus*, from a comparative genomics perspective. *PLoS ONE* 2015; 10(10): e0141329. <https://doi.org/10.1371/journal.pone.0141329> PMID: 26513163
39. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol Biol Evol.* 2007; 24: 1161–1180. <https://doi.org/10.1093/molbev/msm036> PMID: 17329229
40. Sasaki Y, Hakamada K, Suama Y, Nagano Y, Furusawa I, Matsuno R. Chloroplast-encoded protein as a subunit of acetyl-CoA carboxylase in pea plant. *J Biol Chem.* 1993; 268: 25118–25123. PMID: 7901221
41. Haberle R, Fourcade H, Boore J, Jansen R. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol.* 2008; 66: 350–361. <https://doi.org/10.1007/s00239-008-9086-4> PMID: 18330485
42. Cheon KS, Kim KA, Jang SK, Yoo KO. Complete chloroplast genome sequence of *Campanula takesimana* (Campanulaceae), an endemic to Korea. *Mitochondrial DNA* 2016; 27: 2169–2171. <https://doi.org/10.3109/19401736.2014.982610> PMID: 25423504
43. Blazier C, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol.* 2011; 76: 263–72. <https://doi.org/10.1007/s11103-011-9753-5> PMID: 21327834
44. Konishi T, Shinohara K, Yamada K, Sasaki Y. Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. *Plant Cell Physiol.* 1996; 37: 117–122. <https://doi.org/10.1093/oxfordjournals.pcp.a028920> PMID: 8665091
45. Baldauf SL, Manhart JR, Palmer JD. Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. *Proc Natl Acad Sci USA.* 1990; 87: 5317–5321. <https://doi.org/10.1073/pnas.87.14.5317> PMID: 2371274
46. Bubunenko MG, Schmidt J, Subramanian AR. Protein substitution in chloroplast ribosome evolution: a eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. *J Mol Biol.* 1994; 240: 28–41. <https://doi.org/10.1006/jmbi.1994.1415> PMID: 8021938
47. Drescher A, Ruf S, Calsa T, Carrer H, Bock R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 2000; 22: 97–104. <https://doi.org/10.1046/j.1365-3113x.2000.00722.x> PMID: 10792825
48. Steane DA. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum. *Eucalyptus globulus* (Myrtaceae). *DNA Res* 2005; 12: 215–220. <https://doi.org/10.1093/dnares/dsi006> PMID: 16303753

49. Chen H, Deng L, Jiang Y, Lu P, Yu J. RNA editing sites exist in protein-coding genes in the chloroplast genome of *Cycas taitungensis*. *J Integr Plant Biol*. 2011; 53: 961–970. <https://doi.org/10.1111/j.1744-7909.2011.01082.x> PMID: 22044752
50. Yang M, Zhang X, Liu G, Yin Y, Chen K, et al. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLOS ONE* 2010; 5(9): e12762. <https://doi.org/10.1371/journal.pone.0012762> PMID: 20856810
51. Uthaipaisanwong P, Chanprasert J, Shearman JR, Sangsrakru D, Yoocha T, Jomchai N, et al. Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.). *Gene* 2012; 500: 172–180. <https://doi.org/10.1016/j.gene.2012.03.061> PMID: 22487870
52. Huang Y-Y, Matzke AJM, Matzke M. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLOS ONE* 2013; 8(8): e74736. <https://doi.org/10.1371/journal.pone.0074736> PMID: 24023703
53. Saski C, Lee SB, 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. <https://doi.org/10.1007/s00122-007-0567-4> PMID: 17534593
54. Zhang Y-J, Ma P-F, Li D-Z. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLOS ONE* 2011; 6(5): e20596. <https://doi.org/10.1371/journal.pone.0020596> PMID: 21655229