

## Article

# Complete Chloroplast Genome Features of *Dendrocalamus farinosus* and Its Comparison and Evolutionary Analysis with Other Bambusoideae Species

Jialong Pei <sup>1</sup>, Yong Wang <sup>2</sup>, Juan Zhuo <sup>1</sup>, Huibin Gao <sup>2</sup>, Naresh Vasupalli <sup>1</sup>, Dan Hou <sup>1</sup> and Xinchun Lin <sup>1,\*</sup>

<sup>1</sup> State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, 666 Wusu Street, Hangzhou 311300, China

<sup>2</sup> Yibin Forestry and Bamboo Industry Research Institute, Yibin 644599, China

\* Correspondence: linxcx@163.com

**Abstract:** *Dendrocalamus farinosus* is one of the essential bamboo species mainly used for food and timber in the southwestern region of China. In this study, the complete chloroplast (cp) genome of *D. farinosus* is sequenced, assembled, and the phylogenetic relationship analyzed. The cp genome has a circular and quadripartite structure, has a total length of 139,499 bp and contains 132 genes: 89 protein-coding genes, eight rRNAs and 35 tRNAs. The repeat analyses showed that three types of repeats (palindromic, forward and reverse) are present in the genome. A total of 51 simple sequence repeats are identified in the cp genome. The comparative analysis between different species belonging to *Dendrocalamus* revealed that although the cp genomes are conserved, many differences exist between the genomes. The analysis shows that the non-coding regions were more divergent than the coding regions, and the inverted repeat regions are more conserved than the single-copy regions. Moreover, these results also indicate that *rpoC2* may be used to distinguish between different bamboo species. Phylogenetic analysis results supported that *D. farinosus* was closely related to *D. latiflorus*. Furthermore, these bamboo species' geographical distribution and rhizome types indicate two evolutionary pathways: one is from the tropics to the alpine zone, and the other is from the tropics to the warm temperate zone. Our study will be helpful in the determination of the cp genome sequences of *D. farinosus*, and provides new molecular data to understand the Bambusoideae evolution.

**Keywords:** *Dendrocalamus farinosus*; Bambusoideae; chloroplast genome; phylogenomics



**Citation:** Pei, J.; Wang, Y.; Zhuo, J.; Gao, H.; Vasupalli, N.; Hou, D.; Lin, X. Complete Chloroplast Genome Features of *Dendrocalamus farinosus* and Its Comparison and Evolutionary Analysis with Other Bambusoideae Species. *Genes* **2022**, *13*, 1519. <https://doi.org/10.3390/genes13091519>

Academic Editor: Jacqueline Batley

Received: 28 July 2022

Accepted: 22 August 2022

Published: 24 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Bambusoideae is one of the largest subfamilies in the grass family (Poaceae), including 75 genera and 1680 species, which are widely distributed in tropical, subtropical and warm temperate zones [1,2]. Bamboo has a high economic value and is widely used in building materials, in making paper, furniture and musical instruments, and as food, etc. [3]. The traditional bamboo species identification is mainly based on morphological characteristics, such as bamboo shoots and flowers [4]. However, the identification and classification of bamboo has always been a complex problem because many species contain similar morphological characteristics and the flowering interval of most Bambusoideae species is very long. For example, *Phyllostachys edulis* is reported to have a flowering interval of 67 years, *Ph. bambusoides*, 120 years, and *Melocanna baccifera*, 48 years [5]. Therefore, it is difficult to classify Bambusoideae species by morphological characteristics.

*Bambusa*, *Dendrocalamus* and *Gigantochloa* are morphologically similar genera in Bambusoideae. *Bambusa* has more than 100 species, and nearly 80 species are found in southern and southwestern China [1]. *Dendrocalamus* contains 66 species, mainly distributed in tropical and subtropical Asia. Among them, 30 species are found in southwestern China [1,6]. In addition, out of 30 species of *Gigantochloa*, seven are in China. The classification of

bamboo species of *Bambusa*, *Dendrocalamus*, and *Gigantochloa* has long been controversial. In particular, *Gigantochloa* and *Dendrocalamus* can only be distinguished by their filaments [1]. These characteristics make it challenging to differentiate these two genera. In recent years, researchers have continuously provided evidence to prove a difference between them, evidence which is used to classify bamboo species. The flora of China (<http://www.iplant.cn/foc/AboutFoc>, accessed on 11 January 2022) distinguished the three genera from the size of culm sheath blades and auricles, the length of culm internodes and the branchlets and the morphological characteristics of caryopsis [1]. However, the flowering interval of Bambusoideae is long, and caryopsis acquisition is difficult. Therefore, we need an effective method to assist in the classification of Bambusoideae species.

The chloroplast (cp) is regarded as the energy source for the evolution of life [7]. Photosynthesis is one of the most important chemical reactions and provides essential energy for most living things on earth [8]. The cp has an independent semi-autonomous extranuclear genome [9]. The cp genome is the crucial genetic material of plants, encoding 100–200 essential genes which encode proteins involved in photosynthesis and other chloroplast functions [10]. Moreover, the cp genomes are ideal tools for studying evolutionary relationships between species because cp genomes are very conserved, and their heredity is independent of the nucleus [11,12]. With the development of high-throughput sequencing technology, every year, numerous plant cp genomes have been sequenced [1]. Therefore, cp genomes have been widely used in plant classification and evolution [9]. Wu et al. (2009) sequenced the cp genomes of *D. latiflorus* and *Bambusa oldhamii*, and believed that cp genomes could be used to provide auxiliary information for the identification and evolution of bamboo species [13]. Many researchers have tried to use the cp genome to assist in bamboo species classification and believe that using the cp genome can provide valuable information in the classification of bamboo [14–16].

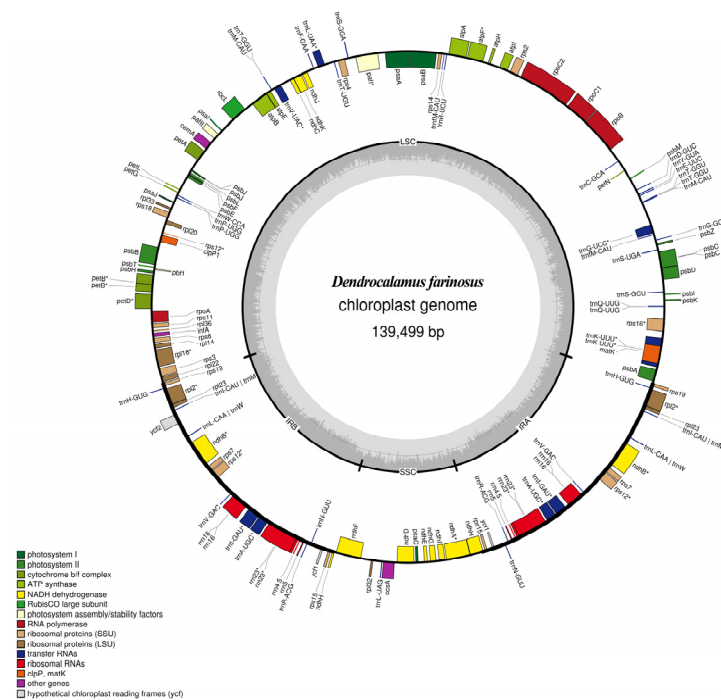
*D. farinosus* belongs to *Dendrocalamus*, which is mainly distributed in southwest China [17]. Compared with other species of *Dendrocalamus*, *D. farinosus* is resistant to cold and drought [18]. The bamboo shoots of *D. farinosus* have a high nutritive content, and *D. farinosus* is a high-quality material used for weaving and papermaking. With the continuous expansion of bamboo resources and the scale of development, further study of the cp genome will improve our understanding of this species, as well as assist in future breeding experiments. In this study, the complete cp genome of *D. farinosus* was assembled, and the features of this cp genome were fully elucidated. Moreover, 34 cp genomes of Bambusoideae were downloaded from GenBank and used to analyze the genealogical relationships and evolutionary history, which may allow us to better understand the identity of *D. farinosus* and the taxonomy of Bambusoideae. Moreover, this work gives us new information on the evolution and origin of Bambusoideae.

## 2. Results and Discussion

### 2.1. Characteristics of the *D. farinosus* cp Genome

In the current study, 3665 Mb of whole-genome raw data were developed from *D. farinosus*. After quality trimming, the data were reduced to 3128 Mb, with an average read length of 150 bp. Furthermore, we generated the complete cp genome by mapping the quality reads to the reference cp genome *B. emeiensis* (HQ337797), and 14.65% of the reads were mapped. The complete plastome sequence of *D. farinosus* has a circular and quadripartite structure, with a total length of 139,499 bp (GeneBankid: OM177223). The plastome has four distinct regions: a small single copy (SSC), a large single copy (LSC) and a pair of inverted repeats (IRa and IRb) (Figure 1). The gene coding region is 64,924 bp, constituting 46.54% of the genome. The length of the SSC, LSC, IRa and IRb regions were 12,879 bp (9.23%), 83,030 bp (59.52%), 21,795 bp (15.62%) and 21,795 bp (15.62%), respectively. Furthermore, the GC contents of the SSC, LSC, IRa and IRb regions were 33.2%, 37%, 44.3% and 44.3%, respectively (Table S1). The results are similar to those found in *D. sinicus*, *Ph. heteroclada*, *Ph. nidularia* and *Thyrsostachys siamensis*. The cp genomes of all five species

were composed of SSC, LSC and IR regions, and the proportions of those regions in the cp genomes for all five species, were similar [19–22].



**Figure 1.** The circle gene map of the *D. farinosus* cp genome. Genes on the outside and inside the map are transcribed clockwise and counterclockwise, respectively. Genes belonging to different functional groups are color-coded. The darker grey in the inner circle corresponds to GC content. The SSC region, LSC region, and inverted repeats (IRa and IRb) are indicated.

The complete cp genome of *D. farinosus* contains a total of 132 genes. Among the 132 genes are 89 protein-coding genes, eight rRNAs, and 35 tRNAs (Table 1). Furthermore, out of these 132 genes, 90 genes are present in the single-copy regions, 42 in the inverted repeat region, and the *ndhH* gene is present partly in both the SSC and IRa regions. In addition, the inverted repeat region contains 19 protein-coding genes, eight rRNAs and 20 tRNAs. In addition, the LSC includes 60 protein-coding genes and 23 tRNAs, while the SSC contains nine protein-coding genes and one tRNA. Hecht et al. (2017) identified that besides the canonical start codon (ATG), translation initiation with the non-canonical start codons might contribute to peptide synthesis [23]. The *D. farinosus* cp genome contains 14 genes with non-canonical start codons (Table S2). Among the 14 genes, eight genes are located in the LSC region, and six are located in the IR region. Furthermore, out of these 132 genes, 13 genes contain exons and introns. Among the 13 genes, 12 genes contain two exons and one intron. In addition, the *pafl* gene has three exons and two introns. Moreover, among these 13 genes, seven are distributed in the LSC, five genes occur in both the IRs and one in the SSC (Table 2). The longest intron of *trnK-UUU* is 2505 bp, while the shortest intron of *rps12* is 537 bp. Most of the genes in the cp genome are protein-coding genes. This result is present in Bambusoideae, such as *D. sinicus*, *Ph. heteroclada*, *Ph. nidularia*, *T. siamensis*, etc., and also in other species such as *Abutilon fruticosum* and *Catha edulis* [19–22,24,25]. The *trnK-UUU* gene also has the longest intron in the cp genome of other species (*A. fruticosum* and *C. edulis*) [24,25].

**Table 1.** Genes present in the cp genome of *D. farinosus*.

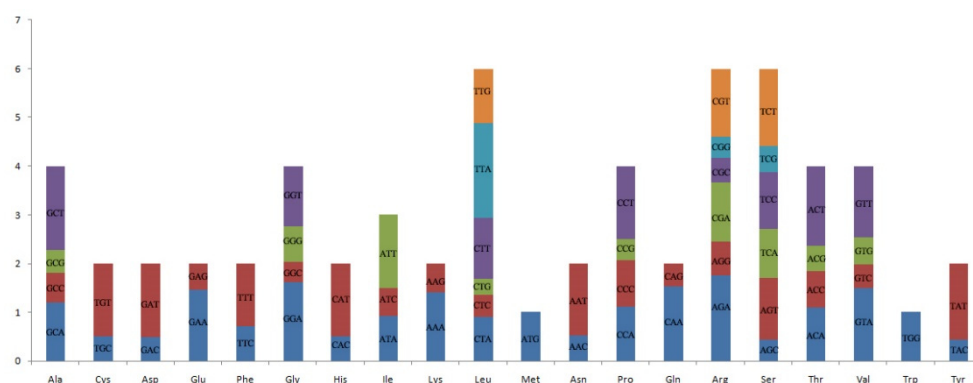
Category	Group of Genes	Name of Genes
Genes for photosynthesis	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF<sup>+</sup>, atpH, atpI</i>
	Subunits of NADH-dehydrogenase	<i>ndhA<sup>+</sup>, ndhB<sup>+,a</sup>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhJ, ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA, petB<sup>+</sup>, petD<sup>+</sup>, petG, petL, petN</i>
	Subunits of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Subunits of photosystem II	<i>psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ</i>
	Subunit of rubisco	<i>rbcL</i>
	photosystem assembly/stability factors	<i>pafI<sup>++</sup>, pafII, pbfI</i>
Self replication	Large subunit of ribosome	<i>rpl2<sup>+,a</sup>, rpl14, rpl16, rpl20, rpl22, rpl23<sup>a</sup>, rpl32, rpl33, rpl36</i>
	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7<sup>a</sup>, rps8, rps11, rps12<sup>++a</sup>, rps12-fragment<sup>+</sup>, rps14, rps15<sup>a</sup>, rps16<sup>+</sup>, rps18, rps19<sup>a</sup></i>
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1<sup>+</sup>, rpoC2</i>
	tRNA Genes	<i>trnA-UGC<sup>+,a</sup>, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC<sup>+</sup>, trnH-GUG<sup>a</sup>, trnI-CAU<sup>a</sup>, trnI-GAU<sup>+,a</sup>, trnK-UUU<sup>+</sup>, trnL-CAA<sup>a</sup>, trnL-UAA<sup>+</sup>, trnL-UAG, trnM-CAU, trnN-GUU<sup>a</sup>, trnP-UGG, trnQ-UUG, trnR-ACG<sup>a</sup>, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC<sup>a</sup>, trnV-UAC<sup>+</sup>, trnW-CCA, trnY-GUA</i>
	rRNA Genes	<i>rrn4.5<sup>a</sup>, rrn5<sup>a</sup>, rrn16<sup>a</sup>, rrn23<sup>+,a</sup></i>
Other genes	c-type cytochrom synthesis gene	<i>ccsA</i>
	Envelop membrane protein	<i>cemA</i>
	Protease	<i>clpP1</i>
	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
	Component of TIC complex	<i>ycf2, ycf15<sup>a</sup></i>
	Subunit acetyl-coA carboxylase	<i>accD</i>

<sup>++</sup>—Gene with one intron, <sup>+++</sup>—gene with two introns and <sup>a</sup>—gene with copies.

**Table 2.** The introns in the genes of the *D. farinosus* cp genome.

S. No	Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
1	<i>rps16</i>	LSC	36	846	222		
2	<i>atpF</i>	LSC	158	837	409		
3	<i>pafI</i>	LSC	132	732	226	724	161
4	<i>rpl2</i>	IR	405	660	432		
5	<i>ndhB</i>	IR	777	712	756		
6	<i>rps12</i>	IR	232	537	36		
7	<i>ndhA</i>	SSC	548	1026	541		
8	<i>trnK-UUU</i>	LSC	38	2505	34		
9	<i>trnG-UCC</i>	LSC	23	673	48		
10	<i>trnL-UAA</i>	LSC	35	539	50		
11	<i>trnV-UAC</i>	LSC	39	596	37		
12	<i>trnI-GAU</i>	IR	42	946	35		
13	<i>trnA-UGC</i>	IR	38	811	35		

Codon usage is believed to play a crucial role in the cp genome [26]. Codon usage is a factor shaping the evolution of cp genomes, and it varies across different species [27]. In this study, we analyzed the codon usage bias and relative synonymous codon usage (RSCU) based on the nucleotide sequence of protein-coding genes 64,924 bp, which encoded 21,023 amino acids (Table S3). The most popular type was leucine (10.85%), followed by isoleucine (8.18%) and serine (7.47%), whereas the least popular was cysteine (1.14%). The RSCU values of 30 codons were >1, and most of these codons ended with A/T (U), while for 29 codons, the values were <1, and most of them ended with G/C (Figure 2). Only two amino acids, tryptophan and methionine, showed no biased usage (RSCU = 1), which agrees with the analysis of cp genomes from other species [26,28,29].



**Figure 2.** Codon content of 20 amino acids in all protein-coding genes of the *D. farinosus* cp genome.

## 2.2. Repeat Analyses

### 2.2.1. Long Repeats

The *D. farinosus* cp genome contains three kinds of long repeats (palindromic, forward and reverse) (Table S4). The analysis showed 14 palindromic repeats, 29 forward repeats and six reverse repeats. The longest repeat was 60 bp in length, and it was a forward-type repeat located in the LSC. While the two shortest repeats, a forward and a reverse repeat, both 19 bp in length, were located in the LSC. Most of the repeats were between 20–30 bp (85.71%) in length, with 10.2% of the repeats being longer than 30 bp, and 4.08% of the repeats being shorter than 20 bp. A total of thirty-six long repeats were present in the LSC, two repeats were present in the SSC, seven long repeats were detected in two different regions, and four repeats were present in the IR regions.

### 2.2.2. Simple Sequence Repeats (SSRs)

Simple sequence repeats (SSRs) are essential for genetic mapping, variety identification and molecular marker-assisted breeding. The *D. farinosus* cp genome contains forty-two SSRs, including seventeen mononucleotides (40.48%), four dinucleotides, three trinucleotides, nine tetranucleotides, and one pentanucleotide with a length of at least 10 bp (Table 3). In addition, the *D. farinosus* cp genome also contains eight compound SSRs (19.05%). All seventeen mononucleotides consisted of poly-A and poly-T. The AT/TA/TC types contributed to four dinucleotides, and the length of these SSRs was 10 bp. Three trinucleotide repeats were detected, namely AAT (12 bp), CAG (12 bp), and TTC (12 bp), which were located in the LSC. Moreover, the eight compound formations were also located in the LSC.

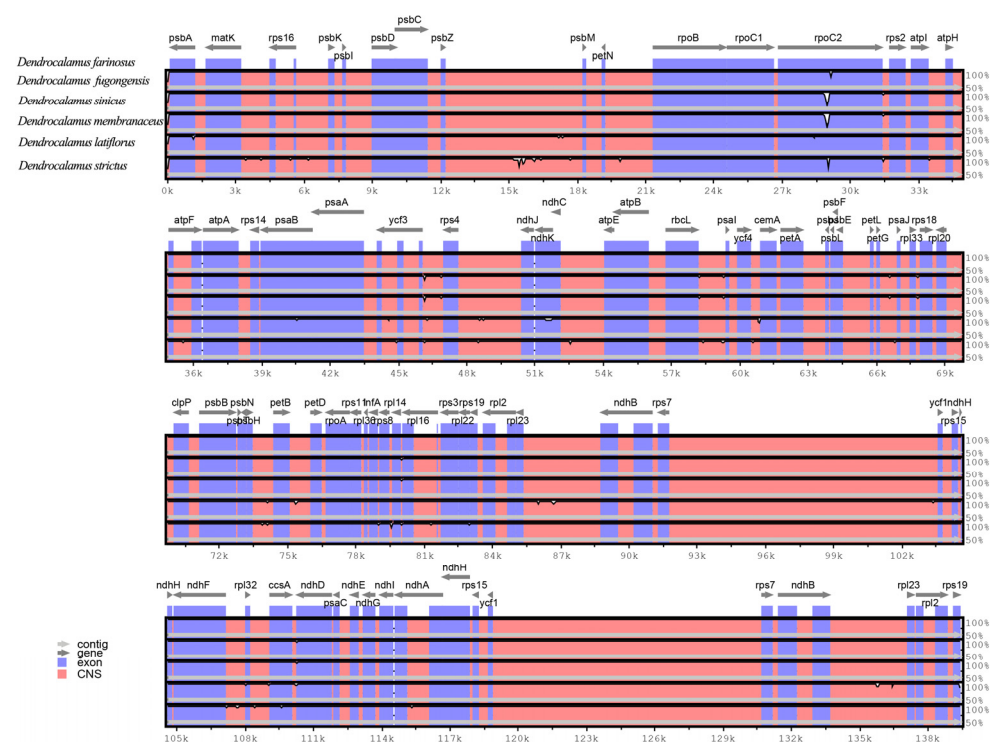
**Table 3.** Distribution of SSRs in the *D. farinosus* cp genome.

SSR Type	Unit	Length	Number	Position on Genome (bp)
P1	A	10	4	7647–7656, 12,227–12,236, 31,304–31,313, 39,008–39,017
	A	11	3	49,902–49,912, 107,247–107,257, 139,481–139,491
	A	12	2	45,950–45,961, 46,921–46,932
	A	13	1	110,286–110,298
	T	10	4	38,212–38,221, 43,628–43,637, 67,161–67,170, 82,514–82,523
	T	11	2	34,229–34,239, 79,557–79,567
	T	12	1	51,765–51,776
P2	AT	10	1	27,587–27,596
	TA	10	2	87,344–87,353, 135,177–135,186
	TC	10	1	117,486–117,495
p3	AAT	12	1	26,292–26,303
	CAG	12	1	684–695
	TTC	12	1	67,459–67,470
P4	AAAG	12	1	18,988–18,999
	AACG	12	1	101,658–101,669
	AATA	12	1	110,503–110,514
	ATAC	12	1	17,899–17,910
	GAAA	12	1	65,129–65,140
	GTAG	16	1	53,971–53,986
	TATT	12	1	50,161–50,172
	TCCT	12	1	44,873–44,884
	TCGT	12	1	120,860–120,871
p5	TTTTA	15	1	31,515–31,529
C	(A)10/(A)11	50	1	35,536–35,585
	(T)10/(T)10	28	1	78,948–78,975
	(T)10/(T)10	32	1	8601–8632
	(T)10/(A)13	106	1	48,475–48,580
	(AGAA)3/(T)11	39	1	70,751–70,789
	(TTTA)3/(A)13	61	1	74,099–74,159
	(TCT)4/(T)11	95	1	82,955–83,049
	(T)10/(AATA)3/(TATT)3	143	1	58,342–58,484

### 2.3. Comparative Analysis of Plastomes of Bambusoideae Species

To detect divergence in the cp genome of the six species of *Dendrocalamus* (*D. farinosus*, *D. fugongensis*, *D. sinicus*, *D. membranaceus*, *D. latiflorus* and *D. strictus*), a comparative analysis was conducted by aligning the cp genome sequences using *D. farinosus* as a reference genome (Figure 3). The results showed that the cp genomes of *Dendrocalamus* were conserved primarily. The most divergent gene among the six cp genomes is *rpoC2*. Furthermore, a lower level of variability was also observed in non-coding regions. Moreover, we also compared the divergence of six cp genomes belonging to five genera in *Bambusoideae* (*D. farinosus*, *D. bambusoides*, *Yushania brevipaniculata*, *Ph. edulis*, *Fargesia edulis* and *B. flexuosa*) (Figure 4). The analysis showed that the non-coding regions were more divergent than the coding regions, and the inverted repeat regions were more conserved than the single-copy regions. Again, the most divergent gene was *rpoC2*. In addition, a lower level of variability was also observed in the following genes: *psbA*, *matK*, *psbD*, *rpoB*, *rpoC1*, *rps2*, *atpA*, *psaB*, *rps4*, *ndhK*, *atpB*, *rbcL*, *cemA*, *rpl22*, *ndhF*, *ccsA*, *ndhD* and *ndhH*. The results suggest that these genes can be used as identification loci for different genera in *Bambusoideae*, and *rpoC2* may be used to distinguish different bamboo species. Zhang et al. (2011) found that the genetic divergence is very low among the cp genomes of *Bambusoideae*, but the *rpoC2* gene is an exception [30]. This result is consistent with our study.

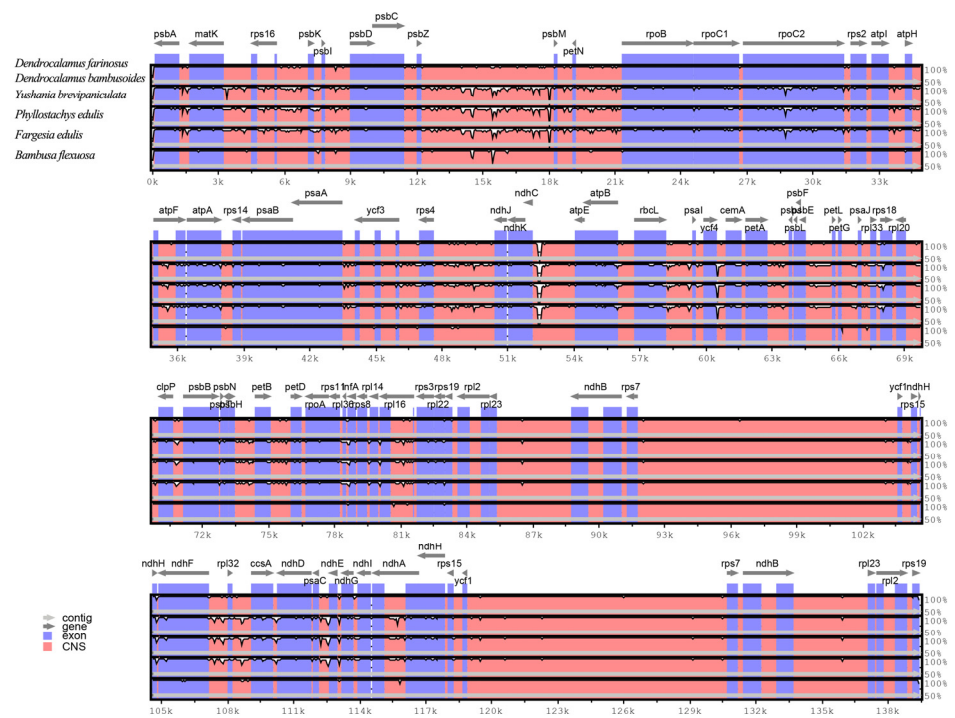




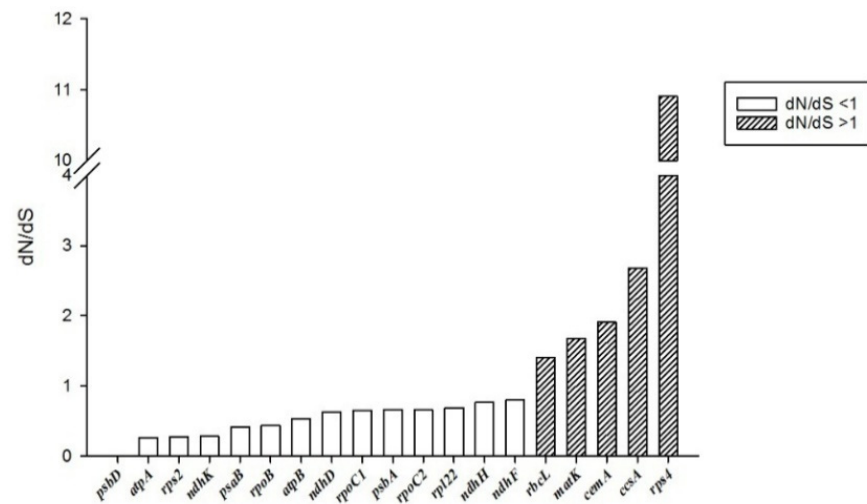
**Figure 3.** Sequence alignment of the six cp genomes of *Dendrocalamus* by mVISTA, using *D. farinosus* as the reference. The six genomes of *Dendrocalamus* are *D. farinosus*, *D. fugongensis*, *D. sinicus*, *D. membranaceus*, *D. latiflorus* and *D. strictus*. The horizontal axis indicates the coordinates within the cp genome. Genome regions are color-coded as follows; genes are represented in blue; conserved non-coding sequences are represented in dark pink; forward and backward arrows represent the direction of the gene.

The nonsynonymous (dN)/synonymous (dS) ratio was calculated for 19 genes that were observed to be divergent in the cp genomes of the Bambusoideae genera (Figure 5). The dN/dS of 13 genes (68.42%) were <1, suggesting that most of them were in negative selection. The dN/dS value of five genes, *rbcL*, *matK*, *cemA*, *ccsA* and *rps4*, were >1, indicating that they were in positive selection, and may provide an evolutionary advantage. Furthermore, the *psbD* gene does not contain any nonsynonymous changes in the plastome of the species used in this study.

The cp genome of plants is conserved in structure and size. In contrast, the expansion and contraction in the cp genomes also occur, and the size and location of the boundaries of single-copy regions and inverted repeat regions changed slightly. This is generally considered the driving force in plant cp genome variation [26]. The boundaries of the IR-LSC and the IR-SSC of six bamboo species (five genera) were compared and analyzed (Figure 6). The size of the six cp genomes ranged from 139,355 bp (*B. flexuosa*) to 139,678 bp (*Ph. edulis*). The gene, *ndhH*, existed simultaneously at the junction of the SSC-IRa (JSA) and the junction of the SSC-IRb (JSB) in the cp genome of *D. farinosus*, while it only appeared at the JSA in the other five cp genomes, and expanded to 187–197 bp in the IRa region. The gene, *rps19*, existed at both the junction of the LSC-IRa (JLA) and the junction of the LSC-IRb (JLB), and varying degrees of contraction and expansion occurs in the cp genomes of the six species. The *rpl22* gene of *D. farinosus*, *Ph. edulis* and *F. edulis* contracted inward by 25 bp, 24 bp and 36 bp from the border, respectively. Kim et al. (2004) thought that the expansion and contraction in the cp genomes was probably mediated by the intramolecular recombination of two short direct repeat sequences which frequently occur within the genes located at the borders [31]. Therefore, the driving force for the expansion and contraction of the cp genome of *D. farinosus* may be greater than that of the other five cp genomes.

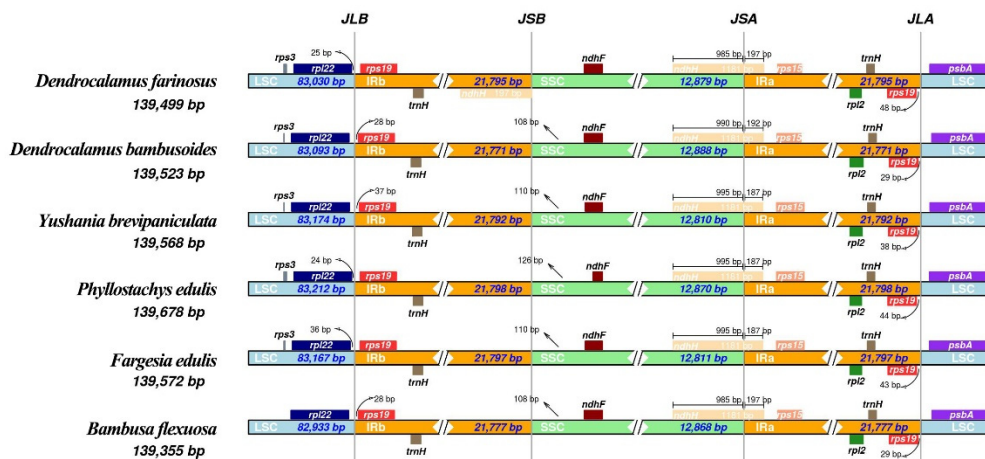


**Figure 4.** Sequence alignment of the six cp genomes of *Bambusoideae* with mVISTA using *D. farinosus* as a reference. The six genomes of *Bambusoideae* are *D. farinosus*, *D. bambusoides*, *Y. brevipaniculata*, *Ph. edulis*, *F. edulis* and *B. flexuosa*. The horizontal axis indicates the coordinates within the cp genome. Genome regions are color-coded as follows; genes are represented in blue; conserved non-coding sequences are represented in dark pink; forward and backward arrows represent the direction of the gene.



**Figure 5.** The nonsynonymous (dN)/synonymous (dS) ratio values of 19 genes from six *Bambusoideae* cp genomes (*D. farinosus*, *Y. brevipaniculata*, *Ph. edulis*, *F. edulis*, *D. bambusoides* and *B. flexuosa*).



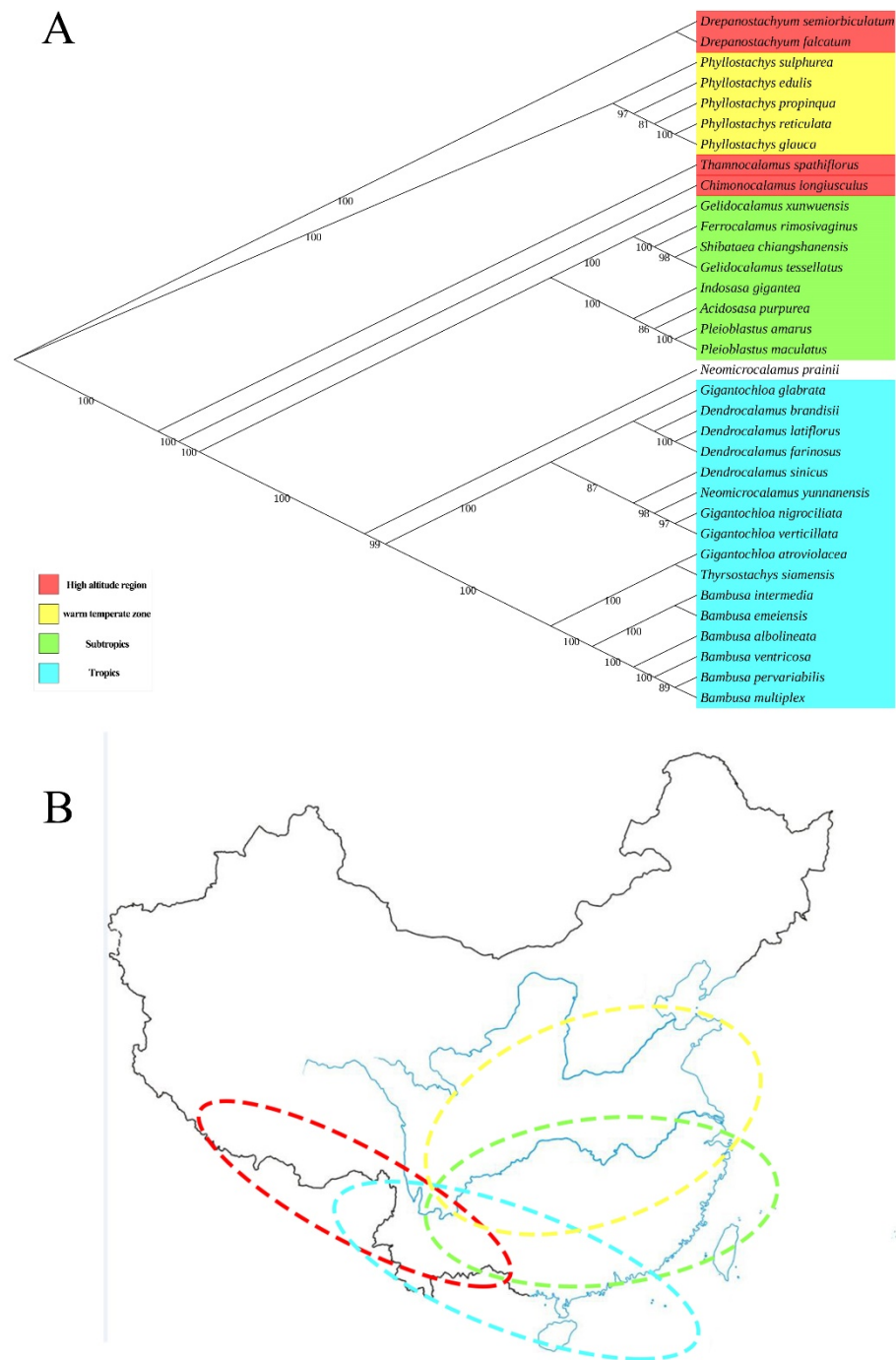


**Figure 6.** Comparison of the borders of the SSC, LSC, IRa and IRb regions among six cp genomes of Bambusoideae. JLB: junction of the LSC and the IRb. JSB: junction of the SSC and the IRb. JSA: junction of the SSC and the IRa. JLA: junction of the LSC and the IRa.

#### 2.4. Phylogenetic Analysis

The cp genomes containing a large amount of genetic information are a good resource for inferring evolutionary and phylogenetic relationships [26,32–34]. Many bamboo species have a similar morphology, which makes them difficult to classify. In this study, the cp genomes of 34 Bambusoideae species (15 genera) were analyzed to understand the evolutionary relationship (Figure 7A). The results showed that these 34 species were classified into four clades, and *Neomicrocalamus prainii* was an outgroup. *D. farinosus* was divided into the same clade as other members of *Dendrocalamus*. Members of *Dendrocalamus*, *Bambusa*, and *Gigantochloa* are in the same clade along with *Neomicrocalamus yunnanensis*. These results differed from the traditional classification method, while remaining consistent with recent studies [1,35]. Members of *Phyllostachys* are in a clade and members of *Drepanostachyum* are in a clade. The result is consistent with the classification according to morphology, indicating that the cp genome-assisted bamboo species classification may be able to solve the complex problem of bamboo classification. As increasing numbers of cp genomes of bamboo species are sequenced, it will be more advantageous to demonstrate that this method is feasible.

The evolutionary analysis of the cp genome may provide new information about the evolution and origin of bamboos. The evolution and origin of bamboo have long been controversial. The origin of bamboo is believed to be polycentric, with two major points of origin in Asia and South America. At the same time, Wen Taihui (1983) proposed that there is only one origin center for bamboo globally, and it is located in the Yunnan province of China [36]. In this study, out of 34 bamboo species, 33 were divided into four groups: high altitude region, warm temperate zone, Subtropics and tropics. These four groups overlap at the Sichuan and Yunnan provinces. At the same time, the 34th bamboo species used in this study, *Neomicrocalamus prainii*, is mainly distributed at the junction of four regions (Sichuan province, China, located in the subtropical region, altitude of 1000–3000 m) (Figure 7B). Furthermore, all 34 bamboo species used in this study were available in the Yunnan province, which we believe may be the origin of bamboo. The results are consistent with Wen [36]. The members of *Drepanostachyum* and *Thamnocalamus* are found mainly in the vicinity of the Himalayas, where the average height above sea level is 6000 m [37]. The *Chimonocalamus longiusculus* is primarily found in southeastern Tibet (average altitude: 4000 m) and southwestern Yunnan (average altitude: 2000 m) [37]. The members of *Phyllostachys* are found mainly in warm temperate zones and subtropical regions [37]. Nine species are distributed in the subtropical regions, and 16 are mainly distributed in the tropics [37].



**Figure 7.** The maximum-likelihood tree based on the 19 cp genomes of Bambuseae and the distribution of the main geographical provenances in China. **(A)** The bootstrap value based on 1000 replicates is shown on each node. **(B)** Distribution of the main geographical provenances of 34 bamboo species in China. Red circle: the distribution of bamboos in high altitude regions; Yellow circle: the distribution of bamboos in the warm temperate zone; Green circle: the distribution of bamboos in the subtropics; Blue circle: the distribution of bamboos in the tropics.

The comparison of the rhizomes of the 34 species revealed that all of the species of high-altitude regions and the tropics have sympodial rhizomes, and all of the species of the Subtropics involved in the comparison have amphipodial rhizomes and monopodial rhizomes. Based on the geographical distribution and rhizome types of these bamboo species, we speculate that the evolution of all the species involved in the comparison

in China may be divided into tropics to alpine and tropics to warm temperate zones. The rhizomes of bamboo species that evolve to a warm temperate zone transform from sympodial to amphipodial, and then from amphipodial to monopodial. This result is consistent with the prediction of bamboo phylogeny from morphology [38]. With the completion of sequencing and further analyses of the cp genomes of more bamboo species, new evidence for these hypotheses may be provided.

### 3. Materials and Methods

#### 3.1. Plant Material and DNA Extraction

Young and healthy leaves of *D. farinosus*, was collected from Changning, Sichuan province (28°29' N, 104°58' E). The total DNA was extracted using a modified CTAB protocol [39]. The voucher specimens are available at the College of Forestry and Biotechnology, Zhejiang A&F University (Accession No. BR01).

#### 3.2. Chloroplast Genome Assembly and Gene Annotation

The cp genome was sequenced using the Illumina NovaSeq PE150 platform (Novogene Bioinformatics Technology Co, Ltd. in Beijing, China). The complete cp genome was constructed using MITObim v1.3, and annotated using the online program, GeSeq [40]. Genes were annotated using CPGAVAS2 (Chloroplast Genome Annotation, Visualization, Analysis and GenBank Submission Tool) [41]. The circular cp genome maps were drawn using OGDRAW [42].

#### 3.3. Repeat Analysis

The long repeat sequences of the *D. farinosus* cp genome were detected using REPuter [43] with default parameters. Simple sequence repeats (SSR) were detected using MISA [44] with the following settings: ten repeats for mono-types, five repeats for di-types, four repeats for tri-types, three repeats for tetra-types, three repeats for penta-types and three repeats for hexa-types.

#### 3.4. Codon Bias Usage Analysis

CPGAVAS2 with default settings was used to analyze the codon bias usage of the protein-coding genes [24]. The relative synonymous codon usage (RSCU) was analyzed.

#### 3.5. Comparison of Related cp Genomes

The mVISTA program was used to analyze sequence divergence between the *D. farinosus* cp genome and the cp genomes of nine related species [45]. IRscope (<https://irscope.shinyapps.io/irapp/>, accessed on 4 January 2022) was used to visualize the genes on the boundaries of the junction sites of the six cp genomes. MEGA 7.0 was used to estimate the dN/dS ratio to detect the genes under selection pressure [46].

#### 3.6. Phylogenetic Analysis

The complete cp genomes of 34 Bambusoideae species were downloaded from GenBank (Table S5) and aligned using MAFFT v7 [47]. The evolutionary history was inferred using RAxML-NG with maximum likelihood, and the bootstrap consensus tree was inferred from 1000 replicates [48].

### 4. Conclusions

Our results identified the complete cp genome of *D. farinosus* and the evolutionary studies with other Bambusoideae species. Our study identified the basic structure and gene content of the cp genome of *D. farinosus*. Further, our analysis also determined that the *rpoC2* gene might be used to distinguish between different bamboo species. Moreover, these bamboo species' geographical distribution and rhizome types indicate two evolutionary pathways: the tropics to the alpine zone and the tropics to the warm temperate zone. This

study provides the cp genome data for species identification and phylogenetic analysis of the Bambusoideae family.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13091519/s1>, Table S1: Base composition in the *D. farinosus* cp genome; Table S2: Genes with non-canonical start codons in the *D. farinosus* cp genome; Table S3: Summary of codon usage and amino acid patterns of the *D. farinosus* cp genome; Table S4: The long repeat sequences detected in the *D. farinosus* cp genome; Table S5: List of the cp genome of 34 Bambusoideae species used for phylogenetic analysis.

**Author Contributions:** Conceptualization, X.L. and Y.W.; methodology, J.P. and Y.W.; formal analysis, J.P. and J.Z.; investigation and resources, Y.W. and H.G.; writing—original draft preparation, J.P.; writing—review and editing, J.P., N.V., D.H. and X.L.; project administration, X.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Tissue culture system construction and molecular studies of excellent fiber production bamboo, grant number 2021YLZ005.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data which supports the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov>, accessed on 20 January 2022, reference number (*D. farinosus*, OM177223).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Liu, J.X.; Zhou, M.Y.; Yang, G.Q.; Zhang, Y.X.; Ma, P.F.; Guo, C.; Li, D.Z. ddRAD analyses reveal a credible phylogenetic relationship of the four main genera of *Bambusa-Dendrocalamus-Gigantochloa* complex (Poaceae: Bambusoideae). *Mol. Phylogenet. Evol.* **2020**, *146*, 106758. [[CrossRef](#)] [[PubMed](#)]
- Santosh, S.; Lalit, M.B.; Poonam, S.; Naik, S.N. Bamboo shoot processing: Food quality and safety aspect (a review). *Trends Food Sci. Technol.* **2010**, *21*, 181–189. [[CrossRef](#)]
- Li, Y.; Xu, B.; Zhang, Q.; Jiang, S. Present situation and the countermeasure analysis of bamboo timber processing industry in China. *J. For. Eng.* **2016**, *1*, 2–7. [[CrossRef](#)]
- Lichtenthaler, H.K. Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Meth. Enzymol.* **1987**, *148*, 350–382. [[CrossRef](#)]
- Janzen, D.H. Why Bamboos Wait So Long to Flower. *Annu. Rev. Ecol. Syst.* **1976**, *7*, 347–391. [[CrossRef](#)]
- Yang, H.Q.; Xie, N.; Sun, M.S.; Xu, T.; Li, D.Z. *Dendrocalamus atroviridis* (Poaceae: Bambusoideae, Bambuseae), a new species from Southwest China. *Phytotaxa* **2016**, *243*, 170. [[CrossRef](#)]
- Neuhaus, H.E.; Emes, M.J. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **2000**, *51*, 111–140. [[CrossRef](#)]
- Fan, S.J.; Guo, X.X. Advances in research and application of plant chloroplast genome. *J. Shandong Normal Univ. (Nat. Sci.)* **2022**, *37*, 22–31. [[CrossRef](#)]
- Li, Y.H.; Ren, Y.K.; Zhao, X.H.; Liu, J.; Han, B.; Wang, C.B.; Tang, Z.H. Research Progress on Chloroplast Genome of Major Gramineous Crops. *Biotechnol. Bull.* **2020**, *36*, 112–121. [[CrossRef](#)]
- Howe, C.J. Chloroplast Genome. In *eLS*; John Wiley & Sons, Ltd: Chichester, UK, 2012. [[CrossRef](#)]
- Wolfe, K.H.; Li, W.H.; Sharp, P.M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 9054–9058. [[CrossRef](#)]
- Jansen, R.K.; Cai, Z.Q.; Raubeson, L.A.; Daniell, H.; Depamphilis, C.W.; Mack, J.L.; Müller, K.F.; Bellian, M.G.; Haberle, R.C.; Hansen, A.K.; et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19369–19374. [[CrossRef](#)] [[PubMed](#)]
- Wu, F.H.; Kan, D.P.; Lee, S.B.; Daniell, H.; Lee, Y.W.; Lin, C.C.; Lin, C.S. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. *Tree Physiol.* **2009**, *6*, 847. [[CrossRef](#)] [[PubMed](#)]
- Burke, S.V.; Grennan, C.P.; Duvall, M.R. Plastome sequences of two New World bamboos—*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)—extend phylogenomic understanding of Bambusoideae. *Am. J. Bot.* **2012**, *99*, 1951–1961. [[CrossRef](#)] [[PubMed](#)]
- Burke, S.V.; Clark, L.G.; Triplett, K.J.; Grennan, C.P.; Duvall, M.R. Biogeography and phylogenomics of New World Bambusoideae (Poaceae), revisited. *Am. J. Bot.* **2014**, *101*, 886–891. [[CrossRef](#)] [[PubMed](#)]



16. Wysocki, W.P.; Clark, L.G.; Attigala, L.; Sanchez, E.R.; Duvall, M.R. Evolution of the bamboos (Bambusoideae; Poaceae): A full plastome phylogenomic analysis. *BMC Evol. Biol.* **2015**, *15*, 50. [[CrossRef](#)]
17. Chen, P.; Zheng, R.H.; Wang, Y.; Hu, Y.B. Study on clonal selection of *Dendrocalamus farinosus* in south Sichuan. *J. Sichuan For. Sci. Technol.* **2021**, *42*, 109–113. [[CrossRef](#)]
18. Fu, M.D.W. Research Progress in Overview of Bio-ecological Characteristics and Cultivation Technology for *Dendrocalamus farinosus*. *World Bamboo Ratt.* **2014**, *12*, 25–28. [[CrossRef](#)]
19. Wang, Y.; Yuan, X.; Li, Y.; Zhang, J. The complete chloroplast genome sequence of *Dendrocalamus sinicus*. *Mitochondrial DNA B Resour.* **2019**, *4*, 2988–2989. [[CrossRef](#)]
20. Tang, S.L.; Xie, J.H.; Cai, J.Z. The complete plastid genome of *Thyrsostachys siamensis* (Poaceae, Bambusoideae). *Mitochondrial DNA B Resour.* **2021**, *6*, 1781–1783. [[CrossRef](#)]
21. Ya-Ping, H.; Jie, Z.; Zhao-Yan, Y.; Jia-Jia, L.; Ming-Ye, X.; Qi-Rong, G. The complete chloroplast genome of *Phyllostachys heteroclada* f. *solida* (Poaceae). *Mitochondrial DNA B Resour.* **2021**, *6*, 566–567. [[CrossRef](#)]
22. Jie, Z.; Yaping, H.; Zhaoyan, Y.; Jiajia, L.; Mingye, X.; Qirong, G. The complete chloroplast genome of a solid type of *Phyllostachys nidularia* (Bambusoideae: Poaceae), a species endemic to China. *Mitochondrial DNA B Resour.* **2021**, *6*, 978–979. [[CrossRef](#)]
23. Hecht, A.; Glasgow, J.; Jaschke, P.R.; Bawazer, L.A.; Munson, M.S.; Cochran, J.R.; Salit, M. Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Res.* **2017**, *45*, 3615–3626. [[CrossRef](#)] [[PubMed](#)]
24. Alzahrani, D.A. Complete Chloroplast Genome of *Abutilon fruticosum*: Genome Structure, Comparative and Phylogenetic Analysis. *Plants* **2021**, *10*, 270. [[CrossRef](#)]
25. Gu, C.; Tembrock, L.R.; Zheng, S.; Wu, Z. The Complete Chloroplast Genome of *Catha edulis*: A Comparative Analysis of Genome Features with Related Species. *Int. J. Mol. Sci.* **2018**, *19*, 525. [[CrossRef](#)] [[PubMed](#)]
26. Zhu, B.; Qian, F.; Hou, Y.; Yang, W.; Cai, M.; Wu, X. Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (Brassicaceae). *PLoS ONE* **2021**, *16*, e0248556. [[CrossRef](#)] [[PubMed](#)]
27. Srivastava, D.; Shanker, A. Identification of Simple Sequence Repeats in Chloroplast Genomes of Magnoliids Through Bioinformatics Approach. *Interdiscip Sci.* **2016**, *8*, 327–336. [[CrossRef](#)]
28. Du, X.; Zeng, T.; Feng, Q.; Hu, L.; Luo, X.; Weng, Q.; Zhu, B. The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species. *Gene* **2020**, *731*, 144340. [[CrossRef](#)]
29. Yan, C.; Du, J.; Gao, L.; Li, Y.; Hou, X. The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): Genome organization, adaptive evolution and phylogenetic relationships in Cardamineae. *Gene* **2019**, *699*, 24–36. [[CrossRef](#)]
30. 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*, e20596. [[CrossRef](#)]
31. Kim, K.J.; Lee, H.L. Complete Chloroplast Genome Sequences from Korean Ginseng (*panax schinseng* nees) and Comparative Analysis of Sequence Evolution among 17 Vascular Plants. *DNA Res.* **2004**, *4*, 247–261. [[CrossRef](#)]
32. Tong, W.; Kim, T.S.; Park, Y.J. Rice Chloroplast Genome Variation Architecture and Phylogenetic Dissection in Diverse *Oryza* Species Assessed by Whole-Genome Resequencing. *Rice* **2016**, *9*, 57. [[CrossRef](#)] [[PubMed](#)]
33. Du, Y.P.; Bi, Y.; Yang, F.P.; Zhang, M.F.; Chen, X.Q.; Xue, J.; Zhang, X.H. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **2017**, *7*, 5751. [[CrossRef](#)] [[PubMed](#)]
34. Zheng, Y.; Hou, D.; Zhuo, J.; Zheng, R.; Wang, Y.; Li, B.; Lin, X. Complete chloroplast genome sequence of *Bambusa rigida* (Bambuseae). *Mitochondrial DNA B Resour.* **2020**, *5*, 2972–2973. [[CrossRef](#)] [[PubMed](#)]
35. Goh, W.L.; Chandran, S.; Franklin, D.C.; Isagi, Y.; Koshy, K.C.; Sungkaew, S.; Wong, K.M. Multi-gene Region Phylogenetic Analyses Suggest Reticulate Evolution and a Clade of Australian Origin Among Palearctic Woody Bamboos (poaceae: Bambusoideae: Bambuseae). *Plant Syst. Evol.* **2013**, *299*, 239–257. [[CrossRef](#)]
36. Wen, T.H. Some ideas about the origin of bamboos. *J. Bamboo Res.* **1983**, *2*, 1–10.
37. Wu, Z.Y. *Flora of China*; Science Press: Beijing, China, 2013.
38. Hanjiao, G.; Cancan, Z.; Jinsong, W.; Xuewen, S.; Ruixue, X.; Bin, L.; Fusheng, C.; Wensheng, B. Variation in basic morphological and functional traits of Chinese bamboo. *Biodivers. Sci.* **2019**, *27*, 10. [[CrossRef](#)]
39. Doyle, J. Dna Protocols for Plants. In *Molecular Techniques in Taxonomy*; Hewitt, G.M., Johnston, A.W.B., Young, J.P.W., Eds.; Springer: Berlin, Germany, 1991; Volume 57, pp. 283–293. [[CrossRef](#)]
40. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [[CrossRef](#)]
41. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [[CrossRef](#)]
42. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [[CrossRef](#)]
43. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)]
44. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [[CrossRef](#)] [[PubMed](#)]
45. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [[CrossRef](#)] [[PubMed](#)]

46. Li, H.; Hughes, A.L.; Bano, N.; McArdle, S.; Livingston, S.; Deubner, H.; Gretch, D.R. Genetic diversity of near genome-wide hepatitis C virus sequences during chronic infection: Evidence for protein structural conservation over time. *PLoS ONE* **2011**, *6*, e19562. [[CrossRef](#)] [[PubMed](#)]
47. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* **2019**, *20*, 1160–1166. [[CrossRef](#)] [[PubMed](#)]
48. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453–4455. [[CrossRef](#)]