Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species

Erxi Liu1,2, Chaozhu Yang2, Jiangdong Liu1, Surong Jin3, Nunung Harijati4, Zhongli Hu1, Ying Diao1 & Lingling Zhao1

*Amorphophallus* (Araceae) contains more than 170 species that are mainly distributed in Asia and Africa. Because the bulbs of *Amorphophallus* are rich in glucomannan, they have been widely used in food, medicine, the chemical industry and so on. To better understand the evolutionary relationships and mutation patterns in the chloroplast genome of *Amorphophallus*, the complete chloroplast genomes of four species were sequenced. The chloroplast genome sequences of *A. albus*, *A. bulbifer*, *A. konjac* and *A. muelleri* ranged from 162,853 bp to 167,424 bp. The *A. albus* chloroplast (cp) genome contains 113 genes, including 79 protein-coding genes, 30 tRNA genes and 4 rRNA genes. The *A. bulbifer* cp genome contains 111 genes, including 78 protein-coding genes, 29 tRNA genes and 4 rRNA genes. *A. muelleri* contains 111 and 113 genes, comprising 78 and 80 protein-coding genes, respectively, 29 tRNA genes and 4 rRNA genes. The IR (inverted repeat) region/LSC (long single copy) region and IR/SSC (short single copy) region borders of the four *Amorphophallus* cp genomes were compared. In addition to some genes being deleted, variations in the copy numbers and intron numbers existed in some genes in the four cp genomes. One hundred thirty-four to 164 SSRs (simple sequence repeats) were detected in the four cp genomes. In addition, the highest mononucleotide SSRs were composed of A and T repeat units, and the majority of dinucleotides were composed of AT and TA. SNPs (single nucleotide polymorphisms) and indels (insertion-deletions) were calculated from coding genes and noncoding genes, respectively. These divergences comprising SSRs, SNPs and indel markers will be useful in testing the maternal inheritance of the chloroplast genome, identifying species differentiation and even in breeding programs. Furthermore, the regression of *ndhK* was detected from four *Amorphophallus* cp genomes in our study. Complete cp genome sequences of four *Amorphophallus* species and other plants were used to perform phylogenetic analyses. The results showed that *Amorphophallus* was clustered in *Araceae*, and *Amorphophallus* was divided into two clades; *A. albus* and *A. konjac* were clustered in one clade, and *A. bulbifer* and *A. muelleri* were clustered in another clade. Phylogenetic analysis among the *Amorphophallus* genus was conducted based on *matK* and *rbcL*. The phylogenetic trees showed that the relationships among the *Amorphophallus* species were consistent with their geographical locations. The complete chloroplast genome sequence information for the four *Amorphophallus* species will be helpful for elucidating *Amorphophallus* phylogenetic relationships.

The *Amorphophallus* (Araceae) genus contains more than 170 species, mainly distributed throughout Asia and Africa. Twenty-six species were found in Sichuan, Chongqing, Yunan, Guizhou and Hubei Provinces in China. Because the bulbs of *Amorphophallus* are rich in glucomannan, they have been widely used in food, medicine, the chemical industry and so on. In general, the *Amorphophallus* genus produces starch and glucomannan, depending on the species. Much research has focused on in vitro propagation systems, due to the accumulation of pathogens from normal asexual reproduction, to increase the yield of *Amorphophallus*. The nucleotide sequences

1State Key Laboratory of Hybrid Rice, Lotus Engineering Research Center of Hubei Province, College of Life Science, Wuhan University, Wuhan, Hubei, 430072, P. R. China. 2Institute of Konjac, Enshi Academy of Agricultural Sciences, Enshi, P. R. China. 3School of Chemistry, Chemical Engineering and Life Science, Wuhan University of Technology, Wuhan, 430070, P. R. China. 4Department of Biology, Faculty of Mathematics and Natural Sciences, Brawijaya University, Jl. Veteran, Malang, 65145, Indonesia. Correspondence and requests for materials should be addressed to L.Z. (email: zhaolingling@whu.edu.cn)
(ITS1) and plastid sequences (rbcL and matK) revealed a new subgeneric delineation by large-scale phylogenetic analysis of *Amorphophallus*. The genome size of *Amorphophallus* is quite large, approximately 20 times larger than the rice genome. Furthermore, large variation exists in the genomic sequences of *Amorphophallus* species. Therefore, sequencing the whole genome of *Amorphophallus* species is very difficult. Complete sequencing of chloroplast (cp) genomes is much easier to achieve in *Amorphophallus* species. The plant chloroplast is a key plastid involved in photosynthesis and carbon fixation. Chloroplast genomes are more conserved than nuclear genomes and contain four important regions: a large single-copy (LSC) region, a small single-copy (SSC) region and a pair of inverted repeats (IRA, IRB). The cp genome contains important information and genetic markers for phylogenetic and taxonomic analyses between plant species and individuals because of the low rates of polymorphisms, indels and SNPs in cpss. More than 800 cp genomes have been sequenced and deposited in the NCBI. The first cp genome was discovered in *Zea mays*, and a complete sequence was determined in *Nicotiana tabacum* and *Marchantia polymorpha*. A circular cp genome of *Aquilaria sinensis* was found to be 159,565 bp long and contained 82 protein-coding genes. Zhang et al. reported sequences for five *Epimedium* species cp genomes, which provided valuable genetic information for accurately identifying species and assisted in the utilization of *Epimedium* plants. These complete cp genome sequences have been widely used in the development of molecular markers for phylogenetic research. Because of the ability for intracellular gene transfer and the conservation, diversity, and genetic basis of chloroplasts, transgene development has allowed for the engineering of high-value agricultural or biomedical products. With the advent of high-throughput sequencing technology, it has become both standard practice and inexpensive to obtain cp genome sequences.

In this study, for the first time, we sequenced the complete cp genomes of four major *Amorphophallus* species using high-throughput sequencing technology and the Illumina HiSeq2500 platform. This study had four aims: (1) determine the size range and structure of four *Amorphophallus* species cp genomes; (2) compare the variations of simple sequence repeats (SSRs) among four major *Amorphophallus* cp genomes; (3) examine the indels and SNPs among four major *Amorphophallus* cp genomes; (4) confirm the phylogenetic relationship among four *Amorphophallus* species, as well as other species, using the complete cp genomes. These results will provide valuable and basic sequence information for taxonomic study and the development of molecular markers for further species identification of *Amorphophallus*. After the completion of the whole cp genome sequence, it is possible to build a database of the species. Based on the differences in the gene sequences of the four cp genomes, a DNA barcode can easily be developed to allow for the building of an efficient platform for postgenomics species research, such as subsequent gene excavation and functional verification of DNA sequence information.

**Results and Discussions**

**Organization of four chloroplast genomes.** Approximately 2G of data for each cp genome was obtained with a 300 bp read length. Gap closing was based on the sequence of the complete cp genome from *Colocasia esculenta* (NC_016753). The chloroplast genome sequences of the four genomes ranged from 162,853 bp (*A. bulbifer*) to 167,424 bp (*A. konjac*) (Fig. 1, Table 1). The same typical quadripartite structure was displayed in the cp genomes. Two IR regions (25,379-26,120 bp) were separated by an LSC region (90,467-92,660 bp) and an SSC region (21,628-22,839 bp) (Table 1). The IRB region was 39 bp longer than the IRA region in the *A. konjac* cp genome. The IR/LSC and IR/SSC borders of the four Amorphophallus cp genomes were compared (Fig. S1). The variation of the IR/LSC and IR/SSC borders was considered to be the primary mechanism causing the length differences of angiosperm cp genomes. The GC content ranged from 35.39% to 35.90% for the four cp genomes (Table 1). These four *Amorphophallus* cp genome data were deposited in GenBank.

**Divergence hotspots in four chloroplast genomes.** The *A. albus* cp genome contains 113 genes, including 79 protein-coding genes, 30 tRNA genes and 4 rRNA genes. The *A. bulbifer* cp genome contains 111 genes, including 78 protein-coding genes, 29 tRNA genes and 4 rRNA genes. Both the *A. konjac* and *A. muelleri* cp genomes contain 112 genes, comprising 79 protein-coding genes, 29 tRNA genes and 4 rRNA genes. All of the features are shown in Table 1 and annotated in Fig. 1. All these genes play different roles in the chloroplast, and the classification is shown in Table 2.

The estimated deletion of some genes was detected in some *Amorphophallus* cp genomes (Table 3). The *ycf1* gene was present in three cp genomes but not in the *A. bulbifer* cp genome. Another gene named *trnL-C* appeared in the *A. albus* and *A. konjac* cp genomes. The *trnG-GCC* gene was lost in the *A. konjac* cp genome. The *accD* gene was found only in the *A. muelleri* cp genome, and the *psbE* was missing only in the *A. konjac* cp genome. The *rp12* and *rp23* genes were annotated in the IRA and IRB regions of the four cp genomes, but they were only found in the IRA region and were lost in the IRB region in the *A. albus* cp genome.

In addition to some genes being deleted, variations in the copy numbers and intron numbers of some genes were also found in the four cp genomes. Eight protein-coding genes, four rRNA genes, nine tRNA genes and two putative genes were present in two copies. Moreover, two introns were found to have two copies only in the *A. bulbifer* and *A. muelleri* cp genomes. In addition, three copies of the *rps12* gene were found. Fifteen genes contained introns, including four tRNA genes, ten protein-coding genes and one putative gene. The *psbF* and *ycf2* genes containing one intron were only found in the *A. muelleri* cp genome. There were no introns in the *clpP* gene in the *A. bulbifer* and *A. muelleri* cp genomes, but there were two introns in this gene in the *A. albus* and *A. konjac* cp genomes, while *ycf3* and *infA* had two introns in each of the four cp genomes. All of the above divergences are shown in Table 2. The development of molecular markers for the identification of *Amorphophallus* species was much easier based on the divergence hotspot regions of the four *Amorphophallus* cp genomes.

**COG analysis.** COG (clusters of orthologous groups of proteins) and KOG (eukaryotic ortholog groups) are based on the relationship between orthologous genes in the NCBI annotation system for prokaryotes and
Homologous genes from different species can be divided into different ortholog clusters combining evolutionary relationships. There are 4,873 categories in COG and 4,852 in KOG. Genes that are orthologs have the same function, and the functional annotation can be inferred to other members of the same COG/KOG clusters. All of the genes from the four cp genomes were classified into six categories: energy production and conversion; translation, ribosomal structure and biogenesis; posttranslational modification, protein turnover and chaperones; transcription; carbohydrate transport and metabolism; and lipid metabolism. The number of genes classified under each function in the four Amorphophallus genomes is shown in Fig. S2.

SSRs are important molecular markers for plant evolutionary and ecological studies, and they are widely present in the cp genome. With MISA analysis, 134–164 SSRs were detected in the four cp genomes (Table 4). Among these SSRs, mono-, di-, tri-, tetra-, and hexanucleotides were detected. The mononucleotide SSRs were most common, with 70.15% of the SSRs observed in A. bulbifer. In addition, most of the mononucleotide SSRs were composed of A and T repeat units, and the majority of the dinucleotides were composed of AT and TA. The cp SSRs are normally composed of short polyA or polyT repeats. Higher contents of A/T and AT/TA repeats in cp SSRs were also detected in the Metasequoia glyptostroboides.
Table 1. Summary of the sequencing data for the four Amorphophallus species.

| Species              | Raw reads no. | Clean reads no. | Gene no. | Protein coding genes no. | rRNA genes no. | Cp genome length (bp) | LSC length (bp) | IRa length (bp) | SSC length (bp) | IRbLength (bp) | GC content (%) |
|----------------------|---------------|-----------------|----------|--------------------------|---------------|-----------------------|----------------|----------------|----------------|----------------|----------------|
| Amorphophallus albus | 7383690       | 6572597         | 113      | 79                       | 30            | 166867               | 90789          | 21640          | 26120          | 35.69          |
| Amorphophallus konjac| 8590079       | 7645378         | 111      | 78                       | 29            | 162853               | 90467          | 25973          | 22839          | 35.39          |
| Amorphophallus albus | 7282525       | 6532968         | 111      | 78                       | 29            | 167424               | 92660          | 26012          | 22839          | 35.39          |
| Amorphophallus muelleri | 10581954     | 8610905         | 113      | 80                       | 29            | 164669               | 90789          | 21640          | 26120          | 35.69          |

Table 2. List of genes encoded by the four Amorphophallus chloroplast genomes. *Gene with two copies; †Gene with three copies; ‡Gene with one intron; §Gene with two introns. () Gene existed in which species cp genome as well as copy number and intron number in each cp genome. Aa, Amorphophallus albus cp genome; Ab, Amorphophallus bulbifer cp genome; Ak, Amorphophallus konjac cp genome; Am, Amorphophallus muelleri cp genome.

| Category for genes | Group of genes | Name of genes |
|--------------------|----------------|--------------|
| Self-replication   | rRNA genes     | rRNA genes   |
| tRNA genes         | trnA-UGC* , trnC-GCA , trnD-GUC , trnE-UCU , trnF-GAA , trnM-CAU , trnG-GCC (Aa, Ab, Am) , trnI-UCG , trnH-UGU , trnK-CAU* , trnG-GAU* , trnK-UUU , trnL-CAA* (Aa, Ak) , trnL-UAA , trnL-UAG , trnM-CAU* , trnN-GUU* , trnP-GGG , trnP-UGG , trnQ-UUG , trnR-AGC² , trnR-UCE , trnS-GCU , trnS-GGA , trnS-UGA , trnT-UGU , trnT-GGU (Aa, Ab* , Ak , Am* ) , trnV-GAC* , trnV-UAC* , trnW-CCA , trnY-GUA |
| Small subunit of ribosome | rps2 , rps3 , rps4 , rps7* , rps8 , rps11 , rps12* , rps14 , rps15 , rps16 , rps18 , rps19 |
| Large subunit of ribosome | rpl2 (Aa* , Ab* , Am* , Ak* ) , rpl20 , rpl23 (Aa* , Ab* , Am* , Ak* ) , rpl33 , rpl36 , rpl14 , rpl16 , rpl22 , rpl32 |
| RNA polymerase      | rpoA , rpoB , rpoC¹ , rpoC2 |
| Phytosynthesis      | NADH-dehydrogenase | ndhA* , ndhB* , ndhC , ndhD , ndhE , ndhF , ndhG , ndhH , ndhI , ndhJ , psbG |
| Photosystem I       | psaA , psaB , psaC , psaL , psaL , psaL , psaL |
| Photosystem II      | psbA , psbB , psbC , psbD , psbE (Aa , Ab , Am) , psbF (Aa , Ab , Am , Ak* ) , psbH , psbI , psbJ , psbK , psbL , psbM , psbN , psbP , psbZ |
| Cytochrome b/f complex | petA , petB , petD , petL , petG , petN |
| ATP synthase        | atpA , atpB , atpE , atpF |
| Large subunit to rubisco | rbcL |
| Other genes         | Maturase | matK |
| Protease            | clpP (Aa* , Ab , Am , Ak* ) |
| Envelope membrane protein | cemA |
| Subunit of Acetyl-CoA Carboxylase | accD (Am) |
| e-type cytochrome synthesis gene | ccsA |
| Translational initiation factor | infA |
| Genes of unknown function | Open Reading Frames | ycf1 (Aa , Ak , Am) , ycf2 (Aa* , Ab* , Am* , Ak* ) , ycf4 |
| Putative pseudogenes | ycf15 (Aa* , Ab* , Am* , Ak* ) , ycf8 |

Using the A. albus cp genome as the reference sequence, we compared the SNP/indel loci of the four cp genomes. SNP markers were detected in 65 protein-coding genes in A. bulbifer, A. konjac and A. muelleri cp genomes. Eleven genes were in the SSC region, and 54 genes were in the LSC region, indicating that the protein-coding genes in the IR region were more conserved. These 65 genes were divided into four categories according to their different functions in plant chloroplasts, including photosynthetic apparatus, photosynthetic metabolism, gene expression, and other genes. Nine hundred sixty-nine and 943 SNP markers were detected between A. albus and A. bulbifer in protein-coding genes and noncoding regions, respectively. One hundred and four and 176 SNP markers were detected between A. albus and A. konjac in protein-coding genes and noncoding regions, respectively. Nine hundred and seventy-eight and 926 SNP markers were detected between A. albus and A. muelleri in protein-coding genes and noncoding regions, respectively. The SNPs in the A. konjac cp genome were significantly fewer than those in the A. bulbifer and A. muelleri cp genomes. One hundred and fifty-nine SNPs were found in Oryza sativa and Oryza nivara chloroplast genomes. 159, 919 SNP markers were detected between Solanum tuberosum and S. bulbocastanum plastomes, and 464 were detected between the plastomes of P. ginseng and P. notoginseng.
hotspots for SNPs and indels, and they were usually used for investigating sequence variation in seed plants28,29. and other plants (Table S1) were used to perform phylogenetic analyses (Fig. 4). The clade of the four species of Amorphophallus for phylogenetic analysis. In the present study, complete cp genomes sequences of four species that are the major commercial cultivation species used in our study. In addition, whole chloroplast sequences were much more accurate than individual gene sequences presented in the four Amorphophallus species cp genomes.

Table 3. Summary of genes estimate deletion in the four Amorphophallus cp genomes. + Gene existing. − Gene estimate deletion.

| Gene                  | A. albus | A. bulbifer | A. konjac | A. muelleri | Region |
|-----------------------|----------|-------------|-----------|-------------|--------|
| ycf1                  | +        | −           | +         | +           | SSC    |
| rpl23                 | IRA+; IRB− | +         | +         | +           | IR     |
| rpl2                  | IRA+; IRB− | +         | +         | +           | IR     |
| trnL-CAA              | +        | −           | +         | −           | IR     |
| trnG-GCC              | +        | +           | −         | +           | LSC    |
| accD                  | −        | −           | −         | +           | LSC    |
| psbE                  | +        | +           | −         | +           | LSC    |

Table 4. Simple sequence repeats (SSRs) in the four Amorphophallus species cp genomes.

| Species                  | SSR loci no. | P1 loci no. | P2 loci no. | P3 loci no. | P4 loci no. | P5 loci no. | P6 loci no. |
|--------------------------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Amorphophallus albus     | 163          | 86          | 49          | 27          | 1           | /           | /           |
| Amorphophallus bulbifer  | 134          | 94          | 25          | 14          | 1           | /           | /           |
| Amorphophallus konjac    | 164          | 82          | 51          | 30          | 1           | /           | /           |
| Amorphophallus muelleri  | 144          | 80          | 45          | 15          | 1           | /           | 3           |

All of the SNPs were classified into two types, including synonymous (S) and nonsynonymous (N) (Table 5, Fig. 2). For the 969 and 978 SNP markers in the gene coding regions of the A. bulbifer and A. muelleri cp genomes, respectively, 696 and 708 belonged to the nonsynonymous type, and 273 and 270 belonged to the synonymous type. Synonymous and nonsynonymous SNP makers from the gene coding genes shared very similar numbers in these two cp genomes. There were 32 synonymous SNPs and 72 nonsynonymous SNPs in the protein coding regions of the A. konjac cp genome. Forty-eight nonsynonymous and 47 synonymous SNP sites were detected in the Machilus cp genome; implying that a substitution constraint mechanism existed32. Genes ycf3, rpoC1 and clpP were detected with SNP markers in their introns. Six, 1 and 5 SNP markers were found in one intron from rpoC1; 6, 1 and 5 SNP markers were found in two introns from clpP in the A. bulbifer, A. konjac and A. muelleri cp genomes, respectively. ClpP and ycf1 were the variation hotspots for SNPs and indels, and they were usually used for investigating sequence variation in seed plants28,29.

cpSSR and SNP markers will be useful in testing maternal inheritance of the cp genome, identifying species differentiation and even in breeding programs60. cpSSRs have been demonstrated to be useful in gene flow studies to estimate seed and pollen contribution11 and in phylogeographic analyses32. Twenty-two protein-coding genes from three Amorphophallus cp genomes contained indels (Table 6). Only two coding genes were detected to contain indels in the A. konjac cp genome; one indel was in rps15, and two indels existed in ycf1. The indel numbers of each coding gene from the A. bulbifer and A. muelleri cp genomes are shown in Fig. 3A,B. The gene ycf1 was a hotspot for indel variation, and almost half of the number of indels existed in this gene (Fig. 3). Such mutational loci in cp genomes showed highly variable regions in the genomes.

**ndhK regression in four Amorphophallus species cp genomes.**  The ndhK gene was a new gene represented in the four Amorphophallus species cp genomes. It was 744 bp in length in the A. albus and A. konjac cp genomes, and 741 bp in length in the A. bulbifer and A. muelleri cp genomes. The gene ndhK is present in a novel protein complex of the thylakoid membrane and shows homology to a mitochondrial gene that encodes a subunit of the NADH-ubiquinone oxidoreductase of the mitochondria30. ndhK was reported as a gene encoding a subunit of PSIi, but later, this protein was classified as a subunit of NADH dehydrogenase, and the gene has been renamed ndhK34,35. In many plants, such as Glycine max, Epipedium acuminatum, Psilotum nudum, Machilus yunnanensis, Actinidia chinensis, Veronica persica, and Aquilaria sinensis (Lour.) Gilg., ndhK was lost from their cp genome mes15,16,27,36–39. ndhK has been found in the Paramecium aurelia mitochondrial (mt) genome80. The presence of this gene in the mt genome raises interesting questions concerning its evolutionary origin. The gene ndhK may play a crucial role in photosynthesis in four Amorphophallus species, and its presence in the cp genomes can be used as a marker for distinguishing them from other family species.

**Phylogenetic analysis.**  Phylogenetic analysis of Amorphophallus species has been reported using different aspects, such as several chloroplast genes41, two chloroplast genes, leafy intron sequences42, plastid DNA markers and fingerprinting43. These studies simply demonstrated Amorphophallus sample species relationships and did not include the four Amorphophallus species that are the major commercial cultivation species used in our study. In addition, whole chloroplast sequences were much more accurate than individual gene sequences for phylogenetic analysis. In the present study, complete cp genomes sequences of four Amorphophallus species and other plants (Table S1) were used to perform phylogenetic analyses (Fig. 4). The clade of the four species of Amorphophallus was grouped with other Araceae species as expected. A. albus and A. konjac were clustered into one clade, and A. bulbifer and A. muelleri were clustered into another clade. These results showed that A. albus and
A. konjac had a close relationship, and A. bulbifer and A. muelleri were closely related. The matK and rbcL genes were also used for phylogenetic analysis among the Amorphophallus genus (Figs S3 and S4). Both of the phylogenetic trees indicated that the Amorphophallus species were grouped into three major clades named Africa, southeast Asia, and Continental Asia. The Continental Asia clade covered the taxa distributed from India to China and Thailand, which were subdivided into two subclades, Continental Asia I and II. The four Amorphophallus species in our study were all derived from the Chinese mainland; A. albus and A. konjac were grouped as Continental Asia I, and A. bulbifer and A. muelleri were grouped as Continental Asia II. The first two species came from the central region of China, and the other two species were collected from the southern region of China near Burma. The matK and rbcL genes well supported clades in consensus trees and the resolution of ingroup relationships within Amorphophallus⁴⁴. All the results suggested that the relationship in Amorphophallus was consistent with the biogeographical distribution. A. konjac and A. bulbifer were also classified in two different clades by Sedayu⁴². A. albus and A. konjac have the same chromosome number (2N = 2X = 26), while A. bulbifer and A. muelleri are triploid (3N = 3X = 39). The propagation coefficient of A. albus and A. konjac did not exceed single digits, while

Figure 2. SNPs statistics of A. bulbifer, A. konjac and A. muelleri cp genomes. The Amorphophallus albus cp genome was used as the reference sequence for SNPs analyses for the other three cp genomes. SNPs belonging to different type groups are color coded. (A) Number of SNPs in the A. bulbifer cp genome sequence. (B) The number of SNPs in the A. konjac cp genome sequence. (C) The number of SNPs in the A. muelleri cp genome sequence.
the propagation coefficient in *A. bulbifer* and *A. muelleri* increased significantly because of aerial bulbs growing in the stems. The aerial bulbs diminish the need for sexual reproduction and lead to a significantly increased reproductive capacity. In many cases, the evolutionary process is closely linked with the reproduction system of the species. *A. muelleri* and *A. bulbifer* reproduce, thus far, through apomictic processes. The corm of *A. bulbifer* is light red, and that of *A. muelleri* is light yellow. These phenotypes also demonstrated the relationship among the four *Amorphophallus* species. The sequenced cp genomes of the four *Amorphophallus* species provide a large amount of genetic information for phylogenetic analysis and taxonomic study.

**Conclusion**

We sequenced the chloroplast genomes of four *Amorphophallus* plants: *A. albus*, *A. bulbifer*, *A. konjac*, and *A. muelleri*. We annotated the four cp genomes and analyzed the structural divergence among the four cp genomes; moreover, we identified the SSR loci and SNPs in protein-coding genes. These SSRs and SNPs could be selected for use in developing markers and in phylogenetic analysis. Comparing the cp genomes among some plants suggested that *psbG* regressed in the *A. albus*, *A. konjac*, *A. bulbifer* and *A. muelleri* cp genomes. We also detected that some genes and introns were lost, in addition to copy differences of some genes among the four cp genomes.

Figure 3. InDels statistics of *A. bulbifer* and *A. muelleri* cp genomes. The *Amorphophallus albus* cp genome was used as the reference sequence for InDels analyses for the other three cp genomes. InDels belonging to different coding genes are color coded. Only three InDels were detected in the *A. konjac* cp genome, so the statistics results are shown in the main text. (A) The number of InDels of each coding gene in the *A. bulbifer* cp genome sequence, (B) The number of InDels of each coding gene in the *A. muelleri* cp genome sequence.
**Figure 4.** Phylogenetic tree based on 30 complete cp genome sequences.

| Gene   | A. bulbifer | A. konjac | A. muelleri | Region |
|--------|-------------|-----------|-------------|--------|
| psbA   | 9           | 0         | 0           | LSC    |
| psbB   | 7           | 2         | 0           | LSC    |
| psbD   | 2           | 1         | 0           | LSC    |
| psbC   | 3           | 1         | 0           | LSC    |
| psbF   | 2           | 0         | 1           | LSC    |
| psbG   | 2           | 5         | 1           | LSC    |
| psbH   | 2           | 0         | 0           | LSC    |
| psbI   | 3           | 1         | 0           | LSC    |
| psbL   | 1           | 0         | 1           | LSC    |
| psbN   | 1           | 0         | 0           | LSC    |
| psaA   | 10          | 4         | 0           | LSC    |
| psaB   | 10          | 2         | 2           | SSC    |
| psaC   | 1           | 0         | 0           | SSC    |
| psaI   | 1           | 0         | 0           | SSC    |
| psaJ   | 0           | 1         | 0           | SSC    |
| petA   | 3           | 3         | 1           | LSC    |
| petB   | 3           | 0         | 0           | LSC    |
| petD   | 2           | 1         | 0           | LSC    |
| petG   | 1           | 0         | 1           | LSC    |
| petL   | 0           | 0         | 0           | LSC    |
| ycf3*  | 3           | 1         | 0           | LSC    |
| Total  | 66          | 22        | 7           |        |

Continued
The results of SNP detection demonstrated that very few of the SNPs were identified between the *A. albus* and *A. konjac* cp genomes; on the contrary, a large number of SNPs between *A. bulbifer* and *A. muelleri* were identified when the *A. albus* cp genome was used as the reference sequence. Interestingly, the SNPs were almost the same in the *A. bulbifer* and *A. muelleri* cp genomes. The indel results were very similar between *A. albus* and *A. konjac* because only three indels were detected in the *A. konjac* cp genome. In addition, phylogenetic analysis using complete cp genome sequences showed that *A. albus* and *A. konjac* were in a clade and *A. bulbifer* and *A. muelleri* were

| Gene | *A. bulbifer* | *A. konjac* | *A. muelleri* | Region |
|------|--------------|-------------|---------------|--------|
| atpA | 9            | 5           | 1             | LSC    |
| atpB | 7            | 5           | 0             | LSC    |
| atpE | 2            | 1           | 0             | LSC    |
| atpF | 0            | 4           | 0             | LSC    |
| atpH | 1            | 1           | 0             | LSC    |
| atpI | 3            | 3           | 1             | LSC    |
| ndhA | 10           | 12          | 2             | SSC    |
| ndhB | 1            | 2           | 1             | LSC    |
| ndhC | 17           | 6           | 2             | SSC    |
| ndhD | 4            | 0           | 0             | SSC    |
| ndhE | 13           | 14          | 2             | SSC    |
| ndhF | 2            | 4           | 1             | SSC    |
| ndhG | 11           | 7           | 3             | SSC    |
| ndhH | 1            | 0           | 0             | LSC    |
| rbcL | 12           | 2           | 0             | LSC    |
| Total | 96          | 67          | 15            | 93     | 69     |
| Gene expression | | | | |
| rpoA | 3            | 5           | 0             | LSC    |
| rpoB | 10           | 23          | 1             | LSC    |
| rpoC2 | 18          | 41          | 2             | LSC    |
| rpoC1* | 12         | 13          | 2             | LSC    |
| rpo2 | 4            | 30          | 1             | LSC    |
| rps3 | 1            | 40          | 0             | LSC    |
| rps4 | 3            | 22          | 0             | LSC    |
| rps8 | 1            | 5           | 0             | LSC    |
| rps11 | 7           | 30          | 0             | LSC    |
| rps12 | 2           | 5           | 0             | LSC    |
| rps14 | 2           | 11          | 0             | LSC    |
| rps15 | 0           | 13          | 0             | LSC    |
| rps16 | 0           | 6           | 0             | LSC    |
| rps18 | 1           | 14          | 0             | LSC    |
| rps19 | 2           | 27          | 1             | LSC    |
| rps14 | 1           | 19          | 0             | LSC    |
| rpl16 | 1           | 10          | 0             | LSC    |
| rps20 | 2           | 11          | 0             | LSC    |
| rps22 | 4           | 8           | 0             | LSC    |
| rpl32 | 1           | 2           | 0             | LSC    |
| rpl33 | 1           | 2           | 0             | LSC    |
| rpl36 | 1           | 9           | 0             | LSC    |
| Total | 77          | 346         | 7             | 77     | 351    |
| Other genes | | | | |
| ycf1 | 12          | 170         | 0             | SSC    |
| ycf4 | 2            | 3           | 0             | LSC    |
| clpP* | 15          | 73          | 2             | LSC    |
| infA | 1            | 2           | 0             | LSC    |
| rps18 | 1           | 3           | 0             | LSC    |
| Total | 34          | 261         | 3             | 34     | 266    |

Table 5. Comparisons of mutation changes, number of synonymous (S) and nonsynonymous (N) substitutions per gene of protein coding cp genes among *A. bulbifer*, *A. konjac* and *A. muelleri*. *SNP markers were detected in their introns.*

The results of SNP detection demonstrated that very few of the SNPs were identified between the *A. albus* and *A. konjac* cp genomes; on the contrary, a large number of SNPs between *A. bulbifer* and *A. muelleri* were identified when the *A. albus* cp genome was used as the reference sequence. Interestingly, the SNPs were almost the same in the *A. bulbifer* and *A. muelleri* cp genomes. The indel results were very similar between *A. albus* and *A. konjac* because only three indels were detected in the *A. konjac* cp genome. In addition, phylogenetic analysis using complete cp genome sequences showed that *A. albus* and *A. konjac* were in a clade and *A. bulbifer* and *A. muelleri* were...
| Gene   | motif                     | size | Direction |
|--------|---------------------------|------|-----------|
| matK   | Ab: accaaatccaa           | 12   | Deletion  |
|        | Am: accaaatccaa           | 12   | Deletion  |
| psbB   | Ab: aaaa                  | 4    | Insertion |
|        | Am: aataa                 | 5    | Insertion |
| rps2   | Ab: cttttt                | 6    | Insertion |
|        | Am: ctcttt                | 6    | Insertion |
| rpoC2  | Ab: aat                  | 3    | Deletion  |
| rpoC2  | Ab: act                  | 3    | Insertion |
|        | Am: cat                  | 3    | Insertion |
| rpoC1  | Ab: a                    | 1    | Insertion |
|        | Am: tta                  | 3    | Insertion |
| rps14  | Ab: tttttt                | 6    | Insertion |
|        | Am: tttttt                | 6    | Insertion |
| psbG   | Ab: tt                   | 3    | Deletion  |
|        | Am: ttt                  | 3    | Deletion  |
| ndhK   | Ab: aaaaaaaaaa            | 8    | Insertion |
|        | Am: tggaattggaattacccca  | 22   | Insertion |
| atpB   | Ab: a                    | 1    | Insertion |
|        | Am: a                    | 1    | Insertion |
| comA   | Ab: tgat                 | 4    | Deletion  |
|        | Am: tgat                 | 4    | Deletion  |
| rpl33  | Ab: a                    | 1    | Insertion |
|        | Am: t                    | 1    | Deletion  |
| rps18  | Ab: aaaaaaaaaa            | 10   | Insertion |
|        | Ab: t                    | 1    | Insertion |
|        | Am: a                    | 1    | Deletion  |
| rps2   | Ab: c                    | 1    | Deletion  |
|        | Am: c                    | 1    | Deletion  |
| rps2   | Ab: g                    | 1    | Insertion |
|        | Am: g                    | 1    | Insertion |
| rps11  | Ab: a                    | 1    | Insertion |
|        | Am: a                    | 1    | Insertion |
| rps11  | Ab: t                    | 1    | Deletion  |
|        | Am: t                    | 1    | Deletion  |
| rps11  | Ab: c                    | 1    | Insertion |
|        | Am: c                    | 1    | Insertion |
| rpo8   | Ab: ttctcttag             | 9    | Deletion  |
|        | Am: ttctcttag             | 9    | Deletion  |
| rpo8   | Ab: cgggg                | 4    | Insertion |
|        | Am: ctggttg              | 7    | Insertion |
| rpl14  | Ab: aaaaaacg             | 6    | Insertion |
|        | Am: aatatacg             | 8    | Insertion |
| rpl14  | Ab: aatattt              | 6    | Insertion |
| rps3   | Am: ta                   | 2    | Insertion |
| rpl22  | Ab: c                    | 1    | Insertion |
| rpl22  | Am: c                    | 1    | Insertion |
| rps19  | Ab: aaa                  | 3    | Deletion  |
|        | Am: aaa                  | 3    | Deletion  |
| ndhF   | Ab: ttgaca               | 7    | Deletion  |
| ndhF   | Ab: aaa                  | 3    | Insertion |
|        | Am: aatccca              | 6    | Insertion |

Continued
| Gene | motif | size | Direction |
|------|-------|------|-----------|
| ndhF | Am: tttttt | 7 | Insertion |
| rps15 | Ab: g | 1 | Deletion |
| rps15 | Ab: gttgaatattacg | 12 | Deletion |
| rps15 | Am: egaga | 5 | Deletion |
| rps15 | Ab: gttttgtg | 7 | Deletion |
| rps15 | Am: aaaa | 3 | Deletion |
| rps15 | Am: g | 1 | Deletion |
| rps15 | Ab: aaaa | 4 | Insertion |
| ycf1 | Ab: agg | 3 | Insertion |
| ycf1 | Am: tagg | 4 | Insertion |
| ycf1 | Ab: ctttt | 5 | Insertion |
| ycf1 | Am: tcttt | 5 | Insertion |
| ycf1 | Am: gtttttaaa | 9 | Deletion |
| ycf1 | Am: tttta | 6 | Insertion |
| ycf1 | Ab: aaaaaaagggggggtttttt | 30 | Insertion |
| ycf1 | Am: tcgctacatagctatattttt | 30 | Insertion |
| ycf1 | Ab: aaccttt | 6 | Insertion |
| ycf1 | Am: ctcacct | 6 | Insertion |
| ycf1 | Ab: aagtt | 3 | Insertion |
| ycf1 | Am: gataat | 6 | Insertion |
| ycf1 | Ab: cat | 3 | Deletion |
| ycf1 | Am: cat | 3 | Deletion |
| ycf1 | Ab: cgttaataa | 9 | Deletion |
| ycf1 | Am: cgttaataa | 9 | Deletion |
| ycf1 | Ab: acc | 3 | Insertion |
| ycf1 | Am: tctatc | 6 | Insertion |
| ycf1 | Ab: aaaaaaagttttt | 15 | Insertion |
| ycf1 | Am: tga | 3 | Insertion |
| ycf1 | Am: ttc | 3 | Insertion |
| ycf1 | Am: tt | 2 | Insertion |
| ycf1 | Am: t | 1 | Insertion |
| ycf1 | Ab: act | 3 | Insertion |
| ycf1 | Am: cat | 3 | Insertion |
| ycf1 | Ab: c | 1 | Insertion |
| ycf1 | Am: ttc | 3 | Insertion |
| ycf1 | Ab: a | 1 | Insertion |
| ycf1 | Am: tta | 3 | Insertion |
| ycf1 | Am: g | 1 | Deletion |
| ycf1 | Ab: cctttt | 6 | Insertion |
| ycf1 | Am: tctttt | 6 | Insertion |
| ycf1 | Ab: gattttgccatgactcctt | 20 | Deletion |
| ycf1 | Am: gattttgccatgactcctt | 20 | Deletion |
| ycf1 | Ab: actgacttttagattttctgcc | 21 | Deletion |
| ycf1 | Ab: g | 1 | Deletion |
| ycf1 | Am: g | 1 | Deletion |
| ycf1 | Am: ttttttttcctttttactt | 23 | Insertion |
| ycf1 | Am: tt | 2 | Insertion |
| ycf1 | Am: tc | 2 | Insertion |
| ycf1 | Ab: ccecgatttattttt | 24 | Insertion |

Continued
in a different clade. The clustering analysis results verified the results of the SNP data. All the data will be very helpful in further research on Amorphophallus plants and chloroplasts and in expanding our understanding of the evolutionary history of the Amorphophallus cp genomes. All of these divergences in the four cp genomes were significant for taxonomic and evolutionary studies, as well as for genetic engineering developments in the future.

**Methods**

**Plant material preparation and sequencing.** Fresh young leaves of *A. albus*, *A. bulbifer*, *A. konjac* and *A. muelleri* were collected from live individuals at the greenhouse of Wuhan University in China. Five micrograms of cp DNA was isolated from leaves and sheared into 300 bp DNA fragments using a Covaris M220 (Covaris, United States). NEB Next UltraTM DNA Library Prep Kit for Illumina (NEB, United States) was used to build the library after DNA fragmentation. The genomic DNA of four species was sequenced on a single HiSeq2500 flow cell lane (Illumina Inc.) by the Chinese National Human Genome Center (http://www.chgc.sh.cn/), Shanghai, China.

**Plant cp genome assembly and annotation.** Trimmomatic v 0.32 was used for raw data processing, and the resulting clean data were used for assembly and post analysis. Fastqc v0.10.0 was used to evaluate the quality of the data visually. Velvet v1.2.07 was used to assemble the clean data, and the complete chloroplast genome sequence was obtained after gap closing. DOGMA was used to annotate the cp genomes and predict the rRNA/tRNA of *A. albus*, *A. bulbifer*, *A. konjac*, and *A. muelleri*. COGs (clusters of orthologous groups of proteins) were analyzed through rpsblast v2.2.30. The circular cp genome maps were drawn using the OrganellarGenomeDRAW program.

**Mutation events analysis.** To compare the mutations among the four complete cp genomes, MISA and MUMMER 3.23 software was used for SSR and SNP/indel analyses, respectively. The *A. albus* cp genome was used as a reference sequence for SNP/indel analyses. Definition of microsatellites (unit size/minimum number of repeats): (1/10) (2/5) (3/4) (4/4) (5/4) (6/4).

**Phylogenetic analysis.** We selected twenty-six cp genomes (Table S1), representing the nine families, for phylogenetic analysis. The *matK* and *rbcL* genes were used for phylogenetic analysis among the *Amorphophallus* genus, and the selected species are shown in Tables S2 and S3. MEGA 6.06 software was used for building the evolutionary tree. The analysis was carried out based on the complete cp DNA sequences.

| Gene | motif | size | Direction |
|------|-------|------|-----------|
| ycf1 | Ac: tcagaa | 6 | Insertion |
| ycf1 | Ab: ctt | 3 | Insertion |
| ycf1 | Am: gtcattattcattattcattatt | 24 | Insertion |
| ycf1 | Am: a | 1 | Insertion |
| ycf1 | Am: atc | 3 | Insertion |
| ycf1 | Am: attattcattattcattatt | 15 | Insertion |
| ycf1 | Ab: g | 1 | Insertion |
| ycf1 | Am: g | 1 | Insertion |
| ycf1 | Ab: a | 1 | Insertion |
| ycf1 | Am: a | 1 | Insertion |
| ycf1 | Ab: c | 1 | Insertion |
| ycf1 | Am: c | 1 | Insertion |
| ycf1 | Ab: c | 1 | Insertion |
| ycf1 | Am: tc | 3 | Insertion |
| ycf1 | Ab: gg | 2 | Insertion |
| ycf1 | Am: gg | 2 | Insertion |
| ycf1 | Ab: e | 1 | Insertion |
| ycf1 | Am: e | 1 | Insertion |
| ycf1 | Ab: cg | 2 | Insertion |
| ycf1 | Am: ga | 2 | Insertion |
| ycf1 | Ab: ttaaaaagg | 9 | Deletion |
| ycf1 | Am: ttaaaaagg | 9 | Deletion |
| ycf1 | Ab: aacccgggg | 10 | Insertion |
| ycf1 | Am: ttgttccagttgttccgg | 18 | Insertion |
| ycf1 | Am: caa | 3 | Insertion |

Table 6. Comparisons of InDels of protein coding cp genes among *A. bulbifer*, *A. konjac* and *A. muelleri*. Ab Amorphophallus bulbifer cp genome, Ak Amorphophallus konjac cp genome, Am Amorphophallus muelleri cp genome.
Data Availability Statement
All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

References
1. Niu, Y. et al. The germplasm resources of Amorphophallus rivieri Durieu: a review. *Southwest Horticulture* 33, 22–24 (2005).
2. Cescutti, P., Campa, C., Delben, F. & Rizzo, R. Structure of the oligomers obtained by enzymatic hydrolysis of the glucosamann produced by the plant Amorphophallus konjac. *Carbohydrate Research* 337, 2505–2511 (2002).
3. Hu, J. B., Liu, J., Yan, H. B. & Xie, C. H. Histological observations of morphogenesis in petiole derived callus of Amorphophallus rivieri Durieu in vitro. *Plant Cell Reports* 24, 642–648 (2005).
4. Zhong, L. et al. High embryogenic ability and regeneration from floral axis of Amorphophallus konjac (Araceae). *Open Life Sciences* 12, 34–41 (2017).
5. Claudel, C. et al. Large-scale phylogenetic analysis of Amorphophallus (Araceae) derived from nuclear and plastid sequences reveals new subgeneric delineation. *Botanical Journal of the Linnean Society* 184, 32–45 (2017).
6. Diao, Y. et al. De Novo Transcriptome and Small RNA Analyses of Two Amorphophallus Species. *PLoS One* 9, e95428, https://doi.org/10.1371/journal.pone.0095428 (2014).
7. Neuhau, H. & Emes, M. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* 51(51), 111–140, https://doi.org/10.1146/annurev.plant.1151.1141.1111 (2000).
8. Wicke, S., Schnewein, G. M., Depumahl, C. W., Kai, F. M. & Quandt, D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76, 273–297 (2011).
9. Provan et al. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* 16, 142–147 (2001).
10. Ravi, V., Khurana, J., Akhilish, K. T. & Khurana, P. An update on chloroplast genomes. *Plant Systematics & Evolution* 271, 101–122 (2008).
11. Hui, H., Chao, S., Yuan, L., Mao, S. Y. & Gao, L. Z. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evolutionary Biology* 14, 151 (2014).
12. Bedbrook, J. R. & Bogorad, L. Endonuclease recognition sites mapped on Zea mays chloroplast DNA. *Proc. Natl. Acad. Sci. USA* 73, 4309 (1976).
13. Shinozaki, K. et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. The *EMBO Journal* 5, 2043–2049 (1986).
14. Ohyama, K. et al. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. *Nature* 322, 572, https://doi.org/10.1038/322572a0 (1986).
15. Zhang, Y. et al. The Complete Chloroplast Genome Sequences of Five Epipedium Species: Lights into Phylogenetic and Taxonomic Analyses. *Frontiers in plant science* 7, 506, https://doi.org/10.3389/fpls.2016.00506 (2016).
16. Aawadhi, P., Ahmad, I., Gandhi, S. G. & Bedi, Y. S. Development of chloroplast microsatellite markers for phylogenetic analysis in Brassicaceae. *Acta Biol Hung* 63, 463–473, https://doi.org/10.1556/ABiol.63.2012.4.3 (2012).
17. Barbosa, T. et al. Molecular phylogenetics of New Caledonian Diospyros (Ebenaceae) using plastid and nuclear markers. *Molecular Phylogenetics & Evolution* 69, 740 (2013).
18. Danniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome biology* 17, 134, https://doi.org/10.1186/s13059-016-1004-2 (2016).
19. Ahmed, I. et al. Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of Colocasia esculenta (L.) Schott (Araceae) and closely related taxa. *Molecular Ecology Resources* 13, 929 (2013).
20. Kim, K. J. & Lee, H. L. Complete chloroplast genome sequences from Korean ginseng (Panax schinseng Poiret) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 11, 247–261, https://doi.org/10.1093/dnares/110.1094.1247 (2004).
21. Roman, L. T. et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41, https://doi.org/10.1186/1471-2105-4-41 (2003).
22. Kuang, D. Y. et al. Complete chloroplast genome sequence of Magnolia kwangsiensis (Magnoliales): implication for DNA barcoding and population genetics. *Genome* 54, 663 (2011).
23. Chen, J. et al. The complete chloroplast genome sequence of the relict woody plant Metasequoia glyptostroboides Hu et Cheng. *Frontiers in plant science* 6, 447, https://doi.org/10.3389/fpls.2015.00447 (2015).
24. Shahid, M. M. et al. The complete nucleotide sequence of wild rice (Oryza nivara) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. *Gene* 340, 133–139 (2004).
25. Chung, H. J. et al. The complete chloroplast genome sequences of Solanum tuberosum and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Reports* 25, 1369–1379 (2006).
26. Dong, W. et al. A chloroplast genomic strategy for designing tax on specific DNA mini-barcodes: a case study on ginsengs. *BMC Genet.* 15, 138, https://doi.org/10.1186/1471-2156-15-41 (2014).
27. Song, Y. et al. Comparative analysis of complete chloroplast genome sequences of two tropical trees Machilus yunnanensis and Machilus balansae in the family Lauraceae. *Frontiers in plant science* 6, 662, https://doi.org/10.3389/fpls.2015.00662 (2015).
28. Dong, W., Liu, J., Yu, J., Wang, L. & Zhou, S. Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. *PLoS One* 7, e35071 (2012).
29. Sarkinen, T. & George, M. Predicting Plastid Marker Variation: Can Complete Plastid Genomes from Closely Related Species Help? *PLoS One* 8, e62266 (2013).
30. Matariesa, D. C., Trevor, H. & Susanne, B. Chloroplast DNA markers (cpSSRs, SNPs) for Miscanthus, Saccharum and related grasses (Poaceae, Poaceae). *Molecular Breeding* 26, 539–544 (2010).
31. McCaauley, D. E. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution* 10, 198–202 (1995).
32. McGrath, S., Hodkinson, T. R. & Barth, S. Extremely high cytoplasmic diversity in natural and breeding populations of Lolium (Poaceae). *Heredity* 99, 531–544 (2007).
33. Peter, J. N., Gounaris, K., Coomber, S. A. & Hunter, C. N. pbG is not a photosystem two gene but may be an ndh gene. *Journal of Biological Chemistry* 264, 14129–14135 (1989).
34. Whelan, J., Young, S. & Day, D. A. Cloning of ndhK from soybean chloroplasts using antibodies raised to mitochondrial complex I. *Plant molecular biology* 20, 887–895, https://doi.org/10.1007/BF00027160 (1992).
35. Arizmendi, J. M., Runswick, M. J., Skel, J. M. & Walker, J. E. Nadh - ubiquinone oxidoreductase from bovine heart mitochondria - a 4th nuclear encoded subunit with a homolog encoded in chloroplast genomes. *FEBS Lett.* 301, 237–242, https://doi.org/10.1016/S0014-5793(92)80248-F (1992).
36. Felix, G., Guo, W., Gubbels, E. A., Katie, H. A. & Mower, J. P. Complete plastid genomes from Ophioglossum limbosum and Platymelum nudum, and Equisetum hystremerveal an ancestral land plant genome structure and resolve the position of Equisetales among monophyletes. *BMC Evolutionary Biology* 13, 8 (2013).
37. Yao, X. et al. The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. PLoS One 10, e0129347 (2015).
38. Choi, K. S., Chung, M. G. & Park, S. The Complete Chloroplast Genome Sequences of Three Veronicaeae Species (Plantaginaceae): Comparative Analysis and Highly Divergent Regions. Frontiers in plant science 7, 355, https://doi.org/10.3389/fpls.2016.00355 (2016).
39. Wang, Y. et al. Complete Chloroplast Genome Sequence of Aquilaria sinensis (Lour.) Gilg and Evolution Analysis within the Malvales Order. Front Plant Sci 7, 280, https://doi.org/10.3389/fpls.2016.00280 (2016).
40. Pritchard, A. E., Venuti, S. E., Ghalambor, M. A., Sable, C. L. & Cummings, D. J. An unusual region of Paramedium mitochondrial DNA containing chloroplast-like genes. Gene 78, 121–134 (1989).
41. Gao, Y., Yin, S., Yang, H., Wu, L. & Yan, Y. Genetic diversity and phylogenetic relationships of seven Amorphophallus species in southwestern China revealed by chloroplast DNA sequences. Mitochondrial DNA Part A, 1–8, https://doi.org/10.1080/24701394.2017.1350855 (2017).
42. Sedayu, A., Eurlings, M., Gravendeel, B. & Hetterscheid, W. Morphological character evolution of Amorphophallus (Araceae) based on a combined phylogenetic analysis of trnL, rbcL, and LEAFY second intron sequences. Vol. 51 (2010).
43. Ghislain, A. R., Pawar, K. D., Yadav, S. R., Bapat, V. A. & Jadhav, J. P. Reconstruction of molecular phylogeny of closely related Amorphophallus species of India using plastid DNA marker and fingerprinting approaches. Physiology and Molecular Biology of Plants 23, 155–167, https://doi.org/10.1007/s12298-016-0400-0 (2017).
44. Grob, G. B. J., Gravendeel, B. & Eurlings, M. C. M. Potential phylogenetic utility of the nuclear FLORICAULA/LEAFY second intron: comparison with three chloroplast DNA regions in Amorphophallus (Araceae). Molecular Phylogenetics and Evolution 30, 13–23, https://doi.org/10.1016/S1055-7903(03)00183-0 (2004).
45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114 (2014).
46. Andrews, S. FastQC: A quality control tool for high throughput sequence data. Reference Source (2010).
47. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18, 821–829 (2008).
48. Wyman, S. K., Jansen, R. K. & Boore, J. L. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252 (2004).
49. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402, https://doi.org/10.1093/nar/25.17.3389 (1997).
50. Lohse, M., Drechsel, O. & Bock, R. Organellar Genome DRAW (OGDRAW):a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr.Genet. 52, 267–274, https://doi.org/10.1007/s00294-0007-00161-y (2007).

Acknowledgements
This work was supported by grants from the Basic Scientific Research Business Special Funds of Wuhan University (2042016kf1106).

Author Contributions
Lingling Zhao designed and performed the experiments and drafted the manuscript. Erxi Liu, Chaozhu Yang and Jiangdong Liu processed some of the data. Ying Diao, Nunung Harijati, Zhongli Hu and Surong Jin revised the manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-37456-z.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019