#### Check for updates

#### **OPEN ACCESS**

EDITED BY Ibrar Ahmed, Alpha Genomics Private Limited, Pakistan

REVIEWED BY Gezahegn Girma, Donald Danforth Plant Science Center, United States Leonardo Alfredo Ornella, Cubiqfoods SL, Spain

\*CORRESPONDENCE Lili Luo, ⊠ luolili@ahzyygz.edu.cn

RECEIVED 14 September 2024 ACCEPTED 25 March 2025 PUBLISHED 08 April 2025

#### CITATION

Li W, Liu J, Wang S, Ma Y, Cui L, Yao Y, Sun K and Luo L (2025) Comparative analysis of chloroplast genomes in three Araceae species: genomic difference, genetic distance and species morphology association. *Front. Genet.* 16:1496262. doi: 10.3389/fgene.2025.1496262

#### COPYRIGHT

© 2025 Li, Liu, Wang, Ma, Cui, Yao, Sun and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Comparative analysis of chloroplast genomes in three Araceae species: genomic difference, genetic distance and species morphology association

Wengang Li<sup>1,2</sup>, Jingru Liu<sup>2</sup>, Siqin Wang<sup>2</sup>, Ying Ma<sup>2</sup>, Lulu Cui<sup>2</sup>, Yingxian Yao<sup>2</sup>, Ke Sun<sup>2</sup> and Lili Luo<sup>1\*</sup>

<sup>1</sup>Department of Pharmacy, Anhui College of Traditional Chinese Medicine, Wuhu, Anhui, China, <sup>2</sup>College of Life Sciences, Anhui Normal University, Wuhu, Anhui, China

Many species in the Araceae have extremely high medicinal value, while the chloroplast genome is relatively conserved, and the encoded and expressed bioactive substances are also abundant. Therefore, chloroplast genomes can serve as one of the basis for species evolution and are extremely important for individual material accumulation. To study the relationship between the chloroplast genome and morphology of target species, this study selected three Araceae species for chloroplast genome sequencing assembly, downloaded the complete chloroplast genomes sequences of another 11 Araceae species. Grouping based on genetic distance, we analyze the association between chloroplast genome structure and morphology. The results showed that there were significant differences in genome size among the three species, but Relative Synonymous Codon Usage (RSCU) exhibited high similarity; Based on the phylogenetic tree, these 14 species can be divided into three branches, with differences in genes such as rrn4, rrn5, rrn23, and trnN among species within each branch; Morphologically, the length of the male inflorescence in BranchIII is significantly greater than that in BranchII; There is a strong positive correlation between the length of the plant stem and the three parameters (Length of LSC, Length of SSC and Length of chloroplast genome) of the genome. This study conducted correlation research from the perspective of chloroplast genome and species morphology. On the one hand, the genetic distance and chloroplast genome structure differences between the target species were determined, and on the other hand, explored the correlation between chloroplast genome and species morphology, providing a theoretical basis for the study of phylogenetic relationships and morphology of Araceae species.

#### KEYWORDS

araceae, genetic distance, chloroplast genomes, morphology, species morphology

## **1** Introduction

# 1.1 Chloroplast genome structure and function

Chloroplasts are key organelles for photosynthesis, with the highest abundance of soluble proteins, membrane proteins, pigments, and lipids (Kirchhoff, 2019). The chloroplast genome size of most land plants is 120-160 kb (Wicke et al., 2011), containing inverted repeats (IR), a large single copy (LSC) region of 80-90 kb, and a small single copy (SSC) region of 10-20 kb (Wang and Lanfear, 2019). The IR region is highly conserved among different plants, with a length ranging from 20,000 to 25,000 bp (Morley et al., 2019), and within the IR region, intronic sequences are also considered to be highly conserved (Daniell et al., 2016). In most land plant individuals, two chloroplast structural haplotypes occur at equal frequencies. However, those species whose chloroplast genomes lack inverted repeats or have short inverted repeats have only one structural haplotype (Wang and Lanfear, 2019). Chloroplasts can dynamically adjust their energy conversion and metabolic performance according to plant metabolism and environmental demands. In this regard, they have developed different mechanisms to minimize and deal with toxic photochemical side reactions (reactive oxygen species, ROS), which are the inevitable cost of converting sunlight in an oxygenated atmosphere (Kirchhoff, 2019). In many angiosperms, cpDNA has a multibranched linear structure (Mower and Vickrey, 2018). However, the amount of cpDNA varies greatly during plant development (Oldenburg and Bendich, 2015). During evolution, many ancestral chloroplast genes have been transferred from cpDNA to the nucleus, a process known as endosymbiotic gene transfer (Ku et al., 2015).

#### 1.2 Medicinal value of araceae plants

Many species of the Araceae have high medicinal value. Pinellia pedatisecta (distributed in central and southern China, mainly growing in forests, valleys, and shaded areas (Abdullah et al., 2021).) has the effects of reducing phlegm, relieving nausea, dissolving lumps, and dissolving tumors. Recent pharmacological studies have shown that P. pedatisecta may be an effective anticancer immunomodulatory drug (Wang et al., 2021), P. pedatisecta contains heptacosterol, quercetin, sitosterol, and stigmasterol, which interact with the core targets AKT1, MAPK3, and ESR1. It has strong binding activity and can be used as an effective drug for the treatment of melanoma through the PI3K/Akt pathway (Wang et al., 2022). Arisaema erubescens (Widely distributed in southern China and Southeast Asian countries.) is a perennial herbaceous plant of the Araceae. In medicine, it is used as an important detoxifier for the treatment of several biological diseases (Li, 2007), Its main effects are to remove dampness and resolve phlegm, prevent convulsions, and promote the reduction of sclerosis and swelling (Yang et al., 2007), Its extracts show anticancer properties (Ducki et al., 1996). It has anticoagulant, gastric analgesic, sedative, antiemetic, anti-inflammatory, and antitumor activities (Dong et al., 2015). Arisaema heterophyllum [distributed in central and western tropical Africa, Angola. Growing on rocky ground along riverbanks or rivers, as well as in shaded areas of forests (Abdullah, et al., 2021)] rhizomes or tubers are used in traditional Chinese medicine to treat cancer and depression and relieve pain (Roshan et al., 2017). Studies have shown that *A. heterophyllum* extract can induce cell apoptosis and autophagy by inhibiting the PI3K/Akt pathway (Li-Xing et al., 2016).

As a relatively independent organelle, chloroplast genome plays an extremely important role in energy and material accumulation. Many species of Araceae have extremely high medicinal value. Therefore, this study extracted and sequenced the chloroplast genomes of three Araceae species, used the genetic distance between the three Araceae species as the basis for grouping, analyzed the correlation between phenotypic parameters and chloroplast genome parameters, and provided molecular evidence for the evolution and variation of Araceae species.

## 2 Materials and methods

### 2.1 Sample collection and DNA extraction

Fresh healthy leaf of three plants of the Araceae (*Pinellia pedatisecta, A. heterophyllum* and *A. erubescens*) were collected from the greenhouse of Anhui College of Traditional Chinese Medicine in Wuhu. Cetyl trimethyl ammonium bromide (CTAB) were used to extract whole-genomic DNA from 100 mg of fresh leaf tissue. Elute each extracted DNA in 125  $\mu$ L of elution buffer. DNA concentration and quality were determined using 1% gel electrophoresis and Nanodrop (ThermoScientific, Delaware, United States).

#### 2.2 Library construction

We used 55 µL (20 ng/µL) of DNA for library preparation based on the protocol of Illumina TruSeq kits in the Pires lab at the University of Missouri, Columbia following the TruSeq DNA Sample Preparation Guide protocol (Illumina, Inc., 2010), except where noted. The DNA was sheared by sonication of 15-24 min using a Bioruptor (Diagenode, Inc., New Jersey, United States). X-tracta disposable gel extraction tools United States Scientific, Ocala, Florida, United States) were used to perform gel extractions for size-selection of samples between 200-400 bp. The gel extractions were purified with the Gel Extraction kit (Qiagen) for the end repair, adenylation of 3' ends, ligation, and enrichment steps. The sheared DNA was visualised for size selection by running on 1% gel for 1 h at 120 V. Before sequencing, we first used ABI 3730xl DNA analyzer to determine the number and size of DNA fragments. Only qualified samples were used for the next step of sequencing. Qualified DNA libraries were sequenced using Illumina HiSeq 2000 (Illumina, Inc., San Diego, California) with a single-end read length of 100 bp. After sequencing, the read quality of each sequenced sample was assessed using FastQC (Andrews, 2017), and then RawData was output using MultiQC (Ewels et al., 2016) for subsequent assembly.

#### 2.3 Genome assembly and annotation

We used the default parameters of Fast-Plast v. 1.2.2 (McKain and Wilson, 2017) and NOVOPlasty (Dierckxsens et al., 2017) for chloroplast assembly. In this process, Trimmomatic v. 0.36 (Bolger et al., 2014) was first used to trim the adapters and quality of the original Illumina reads, and then Bowtie2 v. 2.2.9 (Langmead and Salzberg, 2012) was used to adjust the parameters to "verysensitivelocal" to trim the trimmed reads. Mapping with GenBank. Mapped reads were assembled using SPAdes v. 3.9.0 (Bankevich et al., 2012) and K - mers were assembled by selecting the "assembler only" option based on the size of the reads. Chloroplast genomes, in circular form, were annotated using GeSeq with BLAT search of 85% identify for annotation of proteincoding genes, rRNAs and tRNAs. The annotations of tRNAs were further confirmed by using tRNAscan-SE v. 2.0.3 (Lowe and Chan, 2016) and ARAGORN v. 2.38 (Laslett and Canback, 2004) by selection of default parameters of chloroplast genome. Blast was used to search for homologous genes in the chloroplast genome to further confirm the positions of the start codon and stop codon. Finally, using Geneious Prime to generate a table of annotated genes, submit the sequencing results and annotation results to NCBI (National Center for Biotechnology Information), and then using Geneious Prime to create a fully annotated chloroplast circle diagram (Lohse et al., 2007).

# 2.4 Data collection and comparison between plant chloroplast genomes

To compare the genomic features of the three Araceae species, we used Geneious R8.1 to determine codon bias and amino acid usage frequency (Kearse, et al., 2012). IRscope was used to visualize the contraction and expansion of IR regions at chloroplast genome junctions (Amiryousefi, et al., 2018), mVISIT was used to focus on gene rearrangements (Darling et al., 2004), and we obtained the genome sequences from NCBI. Complete chloroplasts of 11 other Araceae species [Arisaema amurense (NC085264), Arisaema franchetianum (MN046885), Arisaema flavum (NC062735), Arisaema ringens (NC044118), Arisaema decipiens (NC064687), (MW338731), Arisaema nepenthoides Arisaema bockii (MZ380241), Arisaema prazeri (NC072165), P. peltata (NC052862), P. ternata (NC027681), P. cordata (MT863558)], and a orchid species as an outgroup [Bletilla sinensis (ON243844)] were downloaded from NCBI. A phylogenetic tree was then constructed using iqtree (1,000 iterations, 1,000 replicates) with the Neighbor-Joining (NJ) Algorithm (Nguyen et al., 2015). In order to compare the phenotypic characteristics of Araceae plants, we collected six indicators including tuber diameter, petiole length, inflorescence stalk length, tube diameter, female inflorescence length and male inflorescence length according to Flora of China (Li, 2007) to analyze the phenotypic characteristics of plants and grouped them according to their evolutionary relationships. Using the Psych package to calculate the correlation coefficient between genome size and plant phenotype, and drawing the correlation heatmap by ComplexHeatmap. Each block is represented by a different color to indicate different correlations, and coefficients are marked within the block. Significance is represented by \*, where \* represents 0.01 < P < 0.05, and not significate with P > 0.05.

## **3** Results

# 3.1 Comparison of genomes after species sequencing and assembly

The chloroplast genome sizes of P. pedatisecta, A. heterophyllum, and A. erubescens differed greatly (168,262 bp, 170,584 bp, and 150,394 bp, respectively) (Figure 1). Table 1 shows that there were great differences in the total number of genes obtained from the chloroplast genomes of the three species. The lowest total number of genes was 98 (P. pedatisecta) and the highest was 147 (A. heterophyllum). However, the number of rRNA genes in the three species was 8. The number of tRNAs in both plants was the same, which was much larger than that in P. pedatisecta. The protein-coding genes in P. pedatisecta were the highest, at 99, while A. erubescens had the lowest, at only 83 genes. In addition, the GC content of the three plants was also very similar (42% (P. pedatisecta), 41% (A. heterophyllum), and 43% (A. erubescens)). Relative synonymous codon usage showed that codons with A or T rather than C or G at the 3'end had a higher coding rate. The RSCU of codons with A/T at the 3'end was mostly  $\geq 1$ , while the RSCU of codons with C or G at the 3'end was mostly  $\leq 1$ . Amino acid frequency analysis showed that leucine and isoleucine had the highest frequencies, while cysteine was a rare amino acid. Overall, we found that the RSCU of the three Araceae species had high similarity (Figure 2).

# 3.2 Determination of phylogenetic relationships and grouping of araceae

To determine the evolutionary relationship of these three species in the Araceae, we selected the chloroplast whole genome sequences of 11 species from Arisaema and Pinellia, with B. sinensis as the outgroup, and constructed a phylogenetic tree based on the NJ Algorithm (Figure 3). The results indicate that P. pedatisetta and other species of Pinellia cluster on the same large branch, while A. heterophyllum and A. erubescens cluster on the same large branch as Arisaema. Strangely, in plant taxonomy, the A. nepenthoides, which belongs to Arisaema, forms a separate branch and seems to have significant differences in chloroplast genome variation compared to Arisaema species. Therefore, based on the tree diagram results, we artificially divided them into three different branches, namely, Branchl (A.nepenthoides), Branchll (P. peltata, P. cordata, P. pedatisecta, P. ternata), and BranchIII (A.prazeri, A. heterophyllum, A. ringens, A. amurense, A. bockii, A. decipiens, A. franchetianum, Α. *flavum*), for the convenience of subsequent analysis.

# 3.3 Comparison of chloroplast genomes in araceae

In order to analyze the length differences and gene location differences in different regions, we used the IRscope online tool to analyze the IR contraction and expansion phenomena in the chloroplast genomes of different branching species (Figure 4). Divide the 14 species into three branches based on the results of



#### FIGURE 1

Schematic map of overall features of the chloroplast genome of Arisaema heterophyllum, Arisaema erubescens and P. pedatisecta. The map contains four circles. From the center going outward, the first circle shows the GC skew (green means less than 0, red means greater than 0). The next circle shows the GC content (red represents greater than the average GC content of the genome, and green represents less than the average GC content of the genome). The third circle shows the inverted repeat sequence. The fourth circle shows the gene location (mRNA is red, tRNA is green, rRNA is blue)

TABLE 1	Comparison	of the	general	information	of	three	species	of	Araceae
			-						

Species names	Genome	PCGs	rRNA	tRNA	Total genes
Pinellia pedatisecta	168,262	85	8	5	98
Arisaema heterophyllum	170,584	99	8	40	147
Arisaema erubescens	150,394	83	8	40	131

the tree diagram. Branchl contains one species (A. nepenthoides) with a chloroplast genomes length of 166,390 bp. A. nepenthoides lacks the rpl22 and rps19 genes in the LSC region, while the majority of rpl23 regions have LSC regions and lack ndhF and trnl genes. Branchll contains four species, with chloroplast genomes lengths ranging from 164013bp to 168262bp. Compared with the other three species, P. cordata contains more small gene, such as rrn4,

rrn5, rrn23, trnN. Only A. prazeri contains the ndhF gene, spanning the IR and SSC regions. BranchIII contains eight species, with chloroplast genomes lengths ranging from 150,403 bp to 75,537 bp. Compared to the other seven species, chloroplast genomes of A. ringens do not contain the ndhF gene, A. erubescens does not contain the rpl2 gene, and only A. flavum contains the *ycf2* gene.



# 3.4 Analysis of phenotypic parameters of araceae plants

In order to analyze the differences in plant morphology of different branches, we collected six indicators, including tuber diameter, petiole length, inflorescence stalk length, tube diameter, female inflorescence length and male inflorescence length, based on the records of 14 species in Flora of China. The results showed (Figure 5) that the tuber diameter of the 14 species ranged from 2 to 7 cm, the petiole length ranged from 15 to 90 cm, the inflorescence stalk length ranged from 0.8 to 8 cm, the female inflorescence length ranged from 0.5 to

3.8 cm, and the male inflorescence length ranged from 0.6 to 4 cm. The male inflorescence length of BranchIII is significantly higher than that of BranchII, but there is no significant difference in the other five morphological parameters between branches.

# 3.5 Analysis of association between plant phenotype and chloroplast genome

To analyze the correlation between chloroplast genome and species phenotypic data, we calculated the correlation coefficients between chloroplast genome parameters (full length, LSC length,



#### FIGURE 3

A Neighbor-Joining tree based on the entire chloroplast genome sequence of 15 species in the Araceae. The number above the branch indicates the branch length, the blue circle represents the Bootstraps, and the red font represents our sample.



SSC length and IR length) and plant phenotypes (tuber diameter, petiole length, inflorescence stalk length, tube diameter, female inflorescence length and male inflorescence length) and displayed

them in a heat map (Figure 6). The results showed that plant tube length showed a strong positive correlation with three genome parameters, namely LSC length (R = 0.6), chloroplast genome



length (R = 0.58) and SSC length (R = 0.4). SSC length showed a strong positive correlation with tube diameter (R = 0.48). IR length showed a strong negative correlation with tube length (R = -0.37) and tuber diameter (R = -0.5). LSC length showed a strong correlation with petiole length (R = -0.34) and inflorescence stalk length (R = -0.43). The full length of chloroplast genes also showed a strong correlation with petiole length (R = -0.326) and inflorescence stalk length (R = -0.38).

## 4 Discussion

# 4.1 Chloroplast gene loss and changes in species survival strategies

In this study, Illumina sequencing technology was used to assemble the chloroplast genomes of three Araceae species and conduct comparative analysis. All methods were carried out in accordance with relevant guidelines and regulations. According to the results of the phylogenetic tree, we believe that among the branches divided, the one with the largest difference in chloroplast genome structure is *A. nepenthoides*, the one with the largest difference in Branchll is *A. prazeri*, and the one with the largest difference in BranchIII is A. ringens. In BranchIII, the main difference between A. nepenthoides and the other three species is the loss of rpl22 gene and rps19 gene. In plants, the transfer of plastid genes to nuclear genes is a very common phenomenon, such as infA (Millen et al., 2001) of rosids, rpl22 (Gantt et al., 1991) of Pisum, rpl32 (Cusack and Wolfe, 2007) of some Salicaceae, and rpoA (Sugiura et al., 2003) of mosses. The loss of rpl22 gene in A. nepenthoides may be due to transfer to nuclear genes. However, there is no literature showing that the transfer of this gene has any effect on the evolution and functional expression of this species. The ndh genes are homologous to components encoding mitochondrial and bacterial respiratory complex I (NADH dehydrogenase) (Ohyama et al., 1986; Shinozaki et al., 1986). Chloroplasts are evolutionarily derived from primitive endosymbionts in host cells, cyanobacteria (Martin et al., 2002), Many genes from cyanobacteria were gradually transferred to the nucleus of the host cell, and the engulfed cyanobacteria evolved into only partially autonomous chloroplasts. Most chloroplast proteins are encoded in nuclear DNA, and only a few chloroplast proteins (about 100) are encoded in genes retained in plastid DNA; among them are ndh genes. Although these genes have been lost in most algal lineages, they are conserved in the plastid DNA of Streptophyta and derived land plants. This suggests



that *ndh* genes have some advantages in the adaptation from aquatic to terrestrial environments (Sabater, 2021). Angiosperms that have lost their plastid ndh genes survive in mild and moderate stress environments or adopt heterotrophic or carnivorous metabolism to compensate for their low or absent photosynthetic efficiency. As biological evolution changes over time, different environments put angiosperms without *ndh* at risk of permanent extinction, just as the endangered golden plants may have done (Sun et al., 2020). In this study, many species in the Pinellia lacked the ndhf gene, and a few species in the Araceae lacked the *ndfh* gene. The lack of *ndf* genes can lead to differences in the way plants survive, which may indicate that the species of the Pinellia and Araceae have different living environments, and this difference is determined by genes. Araceae has higher photosynthesis efficiency than species in the Pinellia.

# 4.2 Relationship between genome structure and plant phenotypic differences

Genome size has important implications for the adaptive evolution of species, which influence phenotype through gene expression (Bennett and Smith, 1976) and may further influence cross-compatibility between genotypes. In plants, DNA amount can be related to several characteristics such as minimum generation time and ecological behavior (Eaton et al., 2004). In current research, people pay more attention to the correlation between nuclear gene expression and plant phenotype, while there are relatively few studies on the species phenotype of chloroplast genome. Studies have shown that there are differences in growth habits and leaf color between wild-type and cultivated individuals of Poa pratensis, and there are also significant differences in chloroplast genome length (Raggi et al., 2015). Buckwheat species differ mainly in height, leaf shape, seeds and sowing type. There are differences in chloroplast genome length and number of SSRs between these four species (Wang et al., 2017); differences in morphological characteristics such as style, stigma, stigma, male scales, seed shape and integument decoration of Cuscuta reproductive organs, chloroplast genome analysis results showed significant differences in chloroplast genome length and gene order between the two plants (Park et al., 2019). This study collected 6 phenotypic data including tuber diameter and petiole length of 14 species of Araceae in Flora of China, and conducted correlation analysis with chloroplast genome parameters. The data showed that there is a strong relationship between chloroplast gene parameters and plant phenotype. Although this study does not provide direct evidence of the correlation between chloroplast genome and plant phenotypes, it provides an idea and lays the foundation for subsequent research.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, No. OR772808.

## **Ethics statement**

Ethical approval for the study was obtained from the Ethical Committee of Anhui College of Traditional Chinese Medicine.

### Author contributions

WL: Data curation, Formal Analysis, Writing – original draft. JL: Investigation, Software, Writing – original draft. SW: Methodology, Software, Writing – review and editing. YM: Formal Analysis, Software, Validation, Writing – review and editing. LC: Investigation, Supervision, Visualization, Writing – review and editing. YY: Investigation, Software, Writing – review and editing. KS: Writing – review and editing. LL: Funding acquisition, Resources, Writing – review and editing.

### References

Abdullah, Henriquez, C. L., Croat, T. B., Poczai, P., and Ahmed, I. (2021). Mutational dynamics of aroid chloroplast genomes II. *Front. Genet.* 11, 610838. doi:10.3389/fgene. 2020.610838

Amiryousefi, A., Hyvönen, J., and Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi:10.1093/bioinformatics/bty220

Andrews, S. (2017). FastQC: a quality control tool for high throughput sequence data.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to singlecell sequencing. *J. Comput. Biol.* 19, 455–477. doi:10.1089/cmb.2012.0021

Bennett, M. D., and Smith, J. (1976). Nuclear DNA amounts in angiosperms. Philosophical Trans. R. Soc. Lond. B, Biol. Sci. 274, 227–274. doi:10.1098/rstb.1976.0044

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Cusack, B. P., and Wolfe, K. H. (2007). When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet.* 23, 270–272. doi:10.1016/j.tig.2007.03.010

Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134–229. doi:10.1186/s13059-016-1004-2

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi:10.1101/gr.2289704

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic acids Res.* 45, e18. doi:10.1093/nar/gkw955

Dong, G., Wang, J., Guo, P., Wei, D., Yang, M., and Kong, L. (2015). Toxicity assessment of Arisaematis Rhizoma in rats by a 1 H NMR-based metabolomics approach. *Mol. Biosyst.* 11, 407–417. doi:10.1039/c4mb00583j

Ducki, S., Hadfield, J. A., Zhang, X., Lawrence, N. J., and McGown, A. T. (1996). Isolation of aurantiamide acetate from Arisaema erubescens. *Planta medica*. 62, 277–278. doi:10.1055/s-2006-957878

Eaton, T., Curley, J., Williamson, R., and Jung, G. (2004). Determination of the level of variation in polyploidy among Kentucky bluegrass cultivars by means of flow cytometry. *Crop Sci.* 44, 2168–2174. doi:10.2135/cropsci2004.2168

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi:10.1093/bioinformatics/btw354

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by funding from Anhui Provincical Department of Education, China (KJ 2019A1086, KJ2021ZD0148 and gxgnfx2020136).

## **Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Gantt, J. S., Baldauf, S. L., Calie, P. J., Weeden, N. F., and Palmer, J. D. (1991). Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO J.* 10, 3073–3078. doi:10.1002/j.1460-2075. 1991.tb07859.x

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi:10.1093/bioinformatics/bts199

Kirchhoff, H. (2019). Chloroplast ultrastructure in plants. *New Phytol.* 223, 565–574. doi:10.1111/nph.15730

Ku, C., Nelson-Sathi, S., Roettger, M., Garg, S., Hazkani-Covo, E., and Martin, W. F. (2015). Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. *Proc. Natl. Acad. Sci.* 112, 10139–10146. doi:10.1073/pnas. 1421385112

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. methods 9, 357–359. doi:10.1038/nmeth.1923

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic acids Res.* 32, 11–16. doi:10.1093/nar/gkh152

Li, J. (2007). Flora of China. Harv. Pap. Bot. 13, 301-302.

Li-Xing, F., Peng, S., Tian, M., Liu, M., Liu, W., Yao, S., et al. (2016). Agglutinin isolated from Arisema heterophyllum Blume induces apoptosis and autophagy in A549 cells through inhibiting PI3K/Akt pathway and inducing ER stress. *Chin. J. Nat. Med.* 14, 856–864. doi:10.1016/S1875-5364(16)30102-9

Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi:10.1007/s00294-007-0161-y

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic acids Res.* 44, W54–W57. doi:10. 1093/nar/gkw413

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., et al. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci.* 99, 12246–12251. doi:10.1073/pnas.182432999

McKain, M. R., and Wilson, M. (2017). Fast-Plast: rapid *de novo* assembly and finishing for whole chloroplast genomes. Available online at: https://github.com/mrmckain/Fast-Plast.

Millen, R. S., Olmstead, R. G., Adams, K. L., Palmer, J. D., Lao, N. T., Heggie, L., et al. (2001). Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13, 645–658. doi:10.1105/tpc.13.3.645

Morley, S. A., Peralta-Castro, A., Brieba, L. G., Miller, J., Ong, K. L., Ridge, P. G., et al. (2019). *Arabidopsis thaliana* organelles mimic the T7 phage DNA replisome with specific interactions between Twinkle protein and DNA polymerases Pol1A and Pol1B. *BMC plant Biol.* 19, 241–316. doi:10.1186/s12870-019-1854-3

Mower, J. P., and Vickrey, T. L. (2018). Structural diversity among plastid genomes of land plants. Adv. botanical Res. 85, 263-292. doi:10.1016/bs.abr.2017.11.013

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi:10.1093/molbev/msu300

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., et al. (1986). Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. *Nature* 322, 572–574. doi:10.1038/322572a0

Oldenburg, D. J., and Bendich, A. J. (2015). DNA maintenance in plastids and mitochondria of plants. Front. Plant Sci. 6, 883. doi:10.3389/fpls.2015.00883

Park, I., Song, J.-H., Yang, S., Kim, W. J., Choi, G., and Moon, B. C. (2019). Cuscuta species identification based on the morphology of reproductive organs and complete chloroplast genome sequences. *Int. J. Mol. Sci.* 20, 2726. doi:10.3390/ijms20112726

Raggi, L., Bitocchi, E., Russi, L., Marconi, G., Sharbel, T. F., Veronesi, F., et al. (2015). Understanding genetic diversity and population structure of a Poa pratensis worldwide collection through morphological, nuclear and chloroplast diversity analysis. *PLoS One* 10, e0124709. doi:10.1371/journal.pone.0124709

Roshan, R., Ahmed, S., and Hasan, M. M. (2017). Arisaema jacquemontii Blume (Araceae): a review of medicinal uses, phytochemistry and pharmacology. *J. Pharmacogn. Phytochemistry* 6, 429–432.

Sabater, B. (2021). On the edge of dispensability, the chloroplast ndh genes. *Int. J. Mol. Sci.* 22, 12505. doi:10.3390/ijms222212505

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., et al. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5, 2043–2049. doi:10.1002/j.1460-2075. 1986.tb04464.x

Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C., and Sugita, M. (2003). Complete chloroplast DNA sequence of the moss Physcomitrella patens: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. *Nucleic Acids Res.* 31, 5324–5331. doi:10.1093/nar/gkg726

Sun, Y., Deng, T., Zhang, A., Moore, M. J., Landis, J. B., Lin, N., et al. (2020). Genome sequencing of the endangered Kingdonia uniflora (Circaeasteraceae, Ranunculales) reveals potential mechanisms of evolutionary specialization. *IScience* 23, 101124. doi:10. 1016/j.isci.2020.101124

Wang, C.-L., Ding, M.-Q., Zou, C.-Y., Zhu, X. M., Tang, Y., Zhou, M.-L., et al. (2017). Comparative analysis of four buckwheat species based on morphology and complete chloroplast genome sequences. *Sci. Rep.* 7, 6514. doi:10.1038/s41598-017-06638-6

Wang, F., Bai, J., Li, F., Liu, J., Wang, Y., et al. (2022). Investigation of the mechanism of the anti-cancer effects of *Astragalus propinquus* Schischkin and Pinellia pedatisecta Schott (A&P) on melanoma via network pharmacology and experimental verification. *Front. Pharmacol.* 13, 895738. doi:10.3389/fphar.2022.895738

Wang, K., Chen, Q., Shao, Y., Yin, S., Liu, C., Liu, Y., et al. (2021). Anticancer activities of TCM and their active components against tumor metastasis. *Biomed. and Pharmacother.* 133, 111044. doi:10.1016/j.biopha.2020.111044

Wang, W., and Lanfear, R. (2019). Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biol. Evol.* 11, 3372–3381. doi:10. 1093/gbe/evz256

Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi:10.1007/s11103-011-9762-4

Yang, N., Liu, W., Huo, X., Gao, Y., and Liu, J. (2007). Determination of chemical constituents of the volatile oil from Arisaema erubescens (Wall.) Schott. *Biotechnology* 17, 52–54.