The Complete Chloroplast Genome Sequence and Phylogenetic Analysis of Mulberry (Morus alba Lin.)

Guo Liangliang1, Shi Yisu1, Wu Mengmeng1, Michael Ackah1, Qiang Lin2 and Weiguo Zhao1*

1Key Laboratory of Silkworm and Mulberry Genetic Improvement, Ministry of Agriculture, School of Biology and Technology, Jiangsu University of Science and Technology, P.R. China
2Sericultural Research Institute, Guangxi Zhuang Autonomous Regin, China

*Corresponding author: Weiguo Zhao, Key Laboratory of Silkworm and Mulberry Genetic Improvement, Ministry of Agriculture, School of Biology and Technology, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, P.R. China

Introduction

The chloroplast (cp) is the photosynthetic organelle representing one of the most important organelles in green plants and algae [1]. The origin of chloroplasts can be dated back to about 390,000 years ago [2]. The mainstream view is the theory of endosymbiosis, which believes that plant chloroplasts originated from ancient cyanobacteria that were symbiotic in primitive eukaryotic cells [3]. In angiosperms, the chloroplast genome (cpDNA) is typically composed of a pair of inverted repeat regions (IRa and IRb), which are separated by a small single-copy (SSC) region and a large single-copy (LSC) region [4]. The contraction and expansion of IR region [5,6] determine the length of the chloroplast genome.

Most cp genomes are 120–220kb in length and contain 120–140 coding genes (Marcelo, Leila, Fraga, & Guerra, 2015; Méndez-Leyva et al.). The size of chloroplasts are similar in related specie [7]. The GC content in the chloroplast genome is generally low, about 37% on average, with AT tendency [8]. The chloroplast sequence is wildly used to phylogenetics, species identification, population genetics, and genetic engineering [9-11]. Recently, reports on phylogenetic analysis in rice [12], analysis clarify the taxonomic status of Capsicum L (Elmosallamy et al. 2019) and Korean ginseng [13]. The cpDNA sequence of Moraceae is incomplete and limited. The results of phylogenetic study on mulberry based on nuclear genome redefined the mulberry as eight species [14]. But the conclusions are not enough to reveal the complex systemic origin, evolutionary relationship, and phylogenetic studies of mulberry. DNA molecular markers like indels, SSRs and small inversions [15], ITS regions [16,17] have been used to study genetic and genome
diversity and phylogenetic analysis. In this study, the cpDNA sequences of *M. atropurpurea* and *M. multicaulis* were investigated, and a comparative analysis was performed between cultivated Morus and *M. mongolica*. The genome structure, gene order, repeat sequences, and phylogenetics were analyzed.

**Materials and Methods**

**Plant material, DNA Extraction, and Sequencing**

*M. atropurpurea* (Lunjiao40) and *M. multicaulis* (Husang32) fresh leaf were collected from National Mulberry Genebank Zhenjiang City, Jiangsu Province. The plants DNA were extracted using CTAB method for complete and high purity DNA to build DNA library following manufacture’s protocol. The DNA concentration was measured by Nanodrop instrument then qualified samples were sent to the Illumina NovaSeq for sequencing.

**Quality Control of Sequencing Data**

To improve the accuracy of the analysis, the Raw Reads were filtered again according to the following criteria:

1. Removal of the sequenced connectors and primer sequences in reads,
2. Reads with an average mass value less than Q5 were filtered out,
3. Reads with N number greater than 5 were removed. The quality reads after the above checks, called clean reads were subjected to subsequent analysis.

**Assembly and Annotation**

The SPAdes (3.13.0) software was used for the genome splicing [18]. The candidate sequence assembly was determined, annotate, and then GeSeq was used to draw the circular gene map.

**Analysis of Repeated Sequences**

Long repeats include three types: forward (P), palindrome (P) and tandem (T) repeats, which may promote chloroplast genome rearrangement and increase population genetic diversity. We use vmatch (http://www.vmatch.de/) The software (parameter: minimal repeat size 30bp) finds the scattered long repeat fragments in the chloroplast genome.

**Comparative Analysis of Chloroplast Genomes of Morus Species with Other Species**

The mVISTA online software in shuffle-LAGAN mode was applied to compare the complete chloroplast genomes of cultivated Morus species with four representatives. *M. multicaulis* (husang32) was used as a reference. Also, different families Arabidopsis thaliana (NC000923.1), Fragaria chiloensis (JN884816), Oryza sativa (NC-008155), Nicotiana tabacum (Z00044) were analysed together. The software (parameter: minimal repeat size 30bp) finds the scattered long repeat fragments in the chloroplast genome.

**Phylogenetic Analysis**

The MEGA X software was used to determine the phylogenetic relationships between Morus species by the maximum likelihood (ML) and neighbor-joining (NJ) methods. Data on the cpDNA of Morus species are available, including those from NCBI *M. indica* (NC-008359), *M. mongolica* (KM491711), *M. notabilis* (KP939360).

**Results**

**Analysis of Sequencing Data and Quality Control**

An overview of the chloroplast sequencing reads derived from the *M. multicaulis* *M. atropurpurea* libraries is listed in Table 1. A total of 30417184 and 32296130 raw reads were obtained from *M. multicaulis* and *M. atropurpurea*, respectively. After quality control check on the raw reads, 3037110 and 32246870 clean reads were obtained from *M. multicaulis* and *M. atropurpurea* respectively. In addition, the Median Phred Quality Score as well as the Read Cycle in mulberry plants are shown in Figure 1.

![Figure 1: Raw data of M. multicaulis and M. atropurpurea.](image-url)
Genome Structural and Content

The cpDNA sequence was determined (Figures 2 & 3) In our studies, the length cpDNA of *M. multicaulis* is 158,817 bp, longer than *M. atropurpurea* 158,776 bp (Table 1). *M. multicaulis* with, a circular double-stranded DNA composed of two identical IR regions (25,551) bp, an LSC region (87,880 bp) and an SSC region (19,835 bp). *M. atropurpurea* (Lunjiao40) also has a typical quadripartite structure with 158,776 bp long, IRa and IRb with 25,667 bp in length separated by LSC region (87,670 bp) and SSC region (19,750 bp). The GC content of *M. multicaulis* chloroplast genome is 36.24%, LSC (33.90%) and SSC (29.24%) regions which are lower relatively lower, compared to the IR region (42.99%) (Table 2). The GC content of the chloroplast genome is very close, and no changes were found to occur in the IR region of the seven mulberry species selected. The cpDNA contains 113 functional genes, including 79 protein-coding genes, 30 tRNA genes and 4 rRNA genes. Pseudogenes and ORFs not included (Table 3). The cpDNA can be divided into three categories depends on the function. The first is related to self-replication containing 70 genes, the second is related to Photosynthesis containing 38 genes, and the third is related to fatty acids and amino acids biosynthesis. There was a total of 5 genes with unknown functions. The results is similar to that found in green plants [6].
Table 1: Rawdata of Morus.

| Material_             | length(bp) | Raw_Reads     | Raw_Bases(bp) | Clean_Reads | Clean_Bases(bp) | GC(%) | Q20(%) | Q30(%) |
|------------------------|------------|---------------|---------------|-------------|-----------------|-------|--------|--------|
| M. multicaulis         | 150        | 30,417,184    | 4,562,577,600 | 30,374,110  | 4,556,116,500   | 35.27 | 96.86  | 91.8   |
| M. atropurpurea        | 150        | 32,296,130    | 4,844,419,500 | 32,246,870  | 4,837,030,500   | 35.6  | 97.13  | 92.4   |

Table 2: Comparison of chloroplast genomes between seven Morus L species.

| Chloroplast feature | M.indica | M. notabilis | M. mongolica | M. atropurpurea | M. multicaulis |
|---------------------|----------|--------------|--------------|-----------------|---------------|
|                     | Yichuanhong | Lunjiao40 | Ribentiancheng | Husang32        |               |
| Genome size (bp)    | 158484   | 158680       | 158459       | 159113          | 158,776       |
| Total GC content (%)| 36.4     | 36.4         | 36.3         | 36.2            | 36.26         |
| genes (#)           | 133      | 129          | 133          | 130             | 130           |
| protein-coding genes (#) | 88     | 84           | 88           | 85              | 85            |
| LSC Length (bp)     | 87386    | 87470        | 87367        | 87824           | 87670         |
| percent (%)         | 55.14    | 55.12        | 55.14        | 55.19           | 55.15         |
| GC content (%)       | 34.1     | 34.1         | 34           | 33.9            | 33.9          |
| SSC length (bp)     | 19742    | 19776        | 19736        | 19875           | 19,750        |
| percent (%)         | 12.46    | 12.46        | 12.45        | 12.49           | 12.44         |
| GC content (%)       | 29.4     | 29.3         | 29.3         | 29.3            | 29.3          |
| IR length (bp)      | 25678    | 25717        | 25678        | 25707           | 25,667        |
| percent (%)         | 16.2     | 16.21        | 16.2         | 16.17           | 16.16         |
| GC content (%)       | 42.9     | 42.9         | 42.9         | 42.9            | 42.9          |

Table 3: Genos present in the chloroplast genome of M. atropurpurea and M. multicaulis.

| Function              | Gene group            | Gene_Name                                                                 |
|-----------------------|-----------------------|--------------------------------------------------------------------------|
| Self-replication      | ribosomal RNAs        | rrn16, rrn4.5, rrn5, rrn23                                              |
|                       | transfer RNAs         | trn-W-GAC, trn-T-UGU, trn-M-CAU, trn-T-GGU, trn-PUGG, trn-E-UUC, trn-R-AGG, trn-H-GUG, trn-S-GCU, trn-B-CAU, trn-L- UAA, trn-Q-UUG, trn-K-UUU, trn-D-GUC, trn-C-GCA, trn-S-GGA, trn-F-GAA, trn-V-UAC, trn-G-GCC, trn-L-UAG, trn-I-GAU, trn-R-UCU, trn-Y-GUA, trn-G- UCC, trn-N-GUU, trn-L-CAA, trn-I-CAU, trn-W-CCA, trn-S-UAGA |
|                       | ribosomal proteins (LSU) | rpl2, rpl33, rpl36, rpl22, rpl16, rpl14, rpl20, rpl32, rpl23            |
|                       | ribosomal proteins (SSU) | rps14, rps3, rps15, rps8, rps4, rps12, rps18, rps19, rps7, rps16, rps 11, rps2 |
|                       | RNA polymerase         | rpoB, rpoA, rpoC1, rpoC2                                                |
| Photosynthesis        | NADH dehydrogenase    | ndhA, ndhG, ndhL, ndhK, ndhB, ndhE, ndhF, ndhH, ndhI, ndhD, ndhC    |
|                       | photosystem I          | psaJ, psaA, psaC, psaL, psaB                                            |
|                       | photosystem II         | psbZ, psbB, psbE, psbEpsbC, psbK, psbA, psbH, psbD, psbM, psbT, psb8, psbL, psbI |
|                       | cytochrome b/f complex | petB, petD, petL, petN, petA, petG                                      |
|                       | ATP synthase           | atpA, atpI, atpB, atpF, atpH, atpE                                      |
|                       | RubisCO large subunit  | rbcL                                                                     |
|                       | Maturase               | matK                                                                     |
|                       | Protease               | clpP                                                                     |
|                       | other genes            | accD, cemA, pbf1, ccsA                                                  |
|                       | Unknown function       | hypothetical chloroplast reading                                        |
|                       | frames(ycf)            | ycf1(2), ycf3, ycf2, ycf4                                               |
Codon Usage

All 79 protein-coding genes in the cpDNA of *M. multicaulis* and *M. atropurpurea* were encoded by 60,765 codons (Table 4). Codon usages have an AT tendency. For *M. multicaulis* 63.76% of codons ending with A or T. Amino acid Leu has the most codons usage of 6,435 (10.59%) codons, followed by Ile with 5310. Ser with 4,464 and Glu with 4,464. These four amino acid codons represented one-third of the total. The least Cys had 699 codons (1.15%). AUU is the most frequent start codon (872), and UGA is the most common stop codon (11).

Table 4: Codon usage for *M. multicaulis* and *M. atropurpurea*.

| Condon | AA | *M. multicaulis* | *M. atropurpurea* | Condon | AA | *M. multicaulis* | *M. atropurpurea* |
|--------|----|-----------------|-----------------|--------|----|-----------------|-----------------|
|        |    | Number          | RSCU            |        |    | Number          | RSCU            |
| UAA    | *  | 27              | 1.69            | UUG    | *  | 27              | 1.69            |
| UAG    | 10 | 0.62            | 10              | AAC    | 11 | 0.69            | 11              |
| UGA    | *  | 11              | 0.69            | UAC    | 11 | 0.69            | 11              |
| GCU    | A  | 503             | 1.86            | CCU    | P  | 317             | 1.54            |
| GCC    | A  | 163             | 0.6             | CCC    | P  | 146             | 0.71            |
| GCA    | A  | 301             | 1.11            | CCA    | P  | 239             | 1.16            |
| GCC    | A  | 116             | 0.43            | CCG    | P  | 120             | 0.58            |
| GCG    | C  | 177             | 1.52            | CAU    | Q  | 579             | 1.56            |
| GGU    | G  | 56              | 0.48            | CUG    | Q  | 164             | 0.44            |
| GGC    | G  | 678             | 1.61            | CGA    | R  | 267             | 1.35            |
| GAT    | D  | 163             | 0.39            | CGG    | R  | 77              | 0.39            |
| GAA    | E  | 862             | 1.54            | CAG    | R  | 260             | 1.32            |
| GAG    | E  | 259             | 0.46            | CGG    | R  | 77              | 0.39            |
| UUU    | F  | 784             | 1.37            | AGA    | R  | 375             | 1.9             |
| UUC    | F  | 362             | 0.63            | AGG    | R  | 130             | 0.66            |
| GGU    | G  | 469             | 1.36            | UCU    | S  | 418             | 1.69            |
| GGC    | G  | 148             | 0.43            | UCC    | S  | 236             | 0.95            |
| GGA    | G  | 549             | 1.59            | UCA    | S  | 300             | 1.21            |
| GGG    | G  | 212             | 0.62            | UCG    | S  | 130             | 0.52            |
| CAU    | H  | 380             | 1.54            | AGU    | S  | 309             | 1.25            |
| CAC    | H  | 113             | 0.46            | AGC    | S  | 95              | 0.38            |
| AUU    | I  | 516             | 0.56            | ACC    | T  | 187             | 0.76            |
| AAC    | I  | 260             | 0.48            | TTT    | T  | 116             | 0.43            |
| AAA    | K  | 433             | 1.21            | TTA    | T  | 143             | 0.53            |
| AAG    | K  | 128             | 0.36            | GUC    | V  | 362             | 1               |
| UUC    | L  | 701             | 1.96            | GUA    | V  | 629             | 1.62            |
| UUG    | L  | 460             | 1.29            | GUU    | V  | 147             | 0.38            |
| CUU    | L  | 285             | 0.81            | UAU    | Y  | 147             | 0.38            |
| CGG    | L  | 134             | 0.37            | UAC    | Y  | 147             | 0.38            |

Repeat Sequences Analysis

Long repeats include three types: forward (P), palindromic (P) and tandem (T) repeats, which may promote chloroplast genome rearrangement and increase population genetic diversity. The *M. atropurpurea* with the most extended repetitive sequence located in 133,098 with 25,678 bp, which is 127 bp longer than that of *M. multicaulis* and the shortest is 31 bp, which is 8 bp shorter than that of *M. multicaulis* (Tables 5 & 6). This is a phenomenon of genome rearrangement and may be related to genetic diversity.
### Table 5: *M. multicaulis* Long repeat sequence analysis.

| Length | Position | Type | Length | Position | Distance_value | E-value | Score_value | Percent_identity |
|--------|----------|------|--------|----------|----------------|---------|-------------|------------------|
| 39     | 45935    | F    | 39     | 101,936  | 0              | 2.35E-14| 78          | 100              |
| 25551  | 87880    | P    | 25551  | 133,266  | 0              | 0.00E+00| 51,102      | 100              |
| 39     | 45935    | P    | 39     | 144,722  | 0              | 2.35E-14| 78          | 100              |
| 30     | 125482   | P    | 30     | 125,482  | 0              | 6.15E-09| 60          | 100              |
| 127    | 0        | P    | 127    | 87,753   | 0              | 2.45E-67| 254         | 100              |
| 30     | 9104     | P    | 30     | 47,631   | 0              | 6.15E-09| 60          | 100              |

### Table 6: *M.atropurpurea* Long repeat sequence analysis.

| Length | Position  | Type | Length | Position  | Distance_value | E-value | Score_value | Percent_identity |
|--------|-----------|------|--------|-----------|----------------|---------|-------------|------------------|
| 39     | 45866     | F    | 39     | 101,853   | 0              | 2.35E-14| 78          | 100              |
| 31     | 33278     | F    | 31     | 33,302    | 0              | 1.54E-09| 62          | 100              |
| 25678  | 87670     | P    | 25,678 | 133,098   | 0              | 0.00E+00| 51,356      | 100              |
| 39     | 45866     | P    | 39     | 144,554   | 0              | 2.35E-14| 78          | 100              |
| 30     | 125314    | P    | 30     | 125,314   | 0              | 6.15E-09| 60          | 100              |
| 30     | 8984      | P    | 30     | 47,562    | 0              | 6.15E-09| 60          | 100              |

### Comparison with Other Species Chloroplast Genomes

We compared with others 6 species and two exogenous Pyrus pyrifolia and Zea mays based on the complete cp genomic sequences (Figure 3). The results showed that the IR boundary of Morus chloroplast genome is slightly different. The rp12 gene at the junction of LSC/IRb 67bp in LSC and the rest are in IRb and trnH gene which is 150bp away from the IRa/LSC boundary of *M. multicaulis*. Pyrus pyrifolia is 101 bp away from the IRa/LSC boundary. The junction IRb/SSC 52bp of ndhF gene is located at IRb, and the rest is located at SSC of *M. alba*. Ycfl is located in IRa/SSC boundary, resulting in the formation of ycfl pseudogene. The rps19 gene is located at the LSC/IR boundary, which resulted in the formation of a rps19 pseudogene. These results are similar to previous study findings [19]. We analyzed the chloroplast genome sequences of 7 mulberry species and *Arabidopsis thaliana, Fragaria chiloensis, Oryza sativa, Nicotiana tabacum* using mVISTA software (Figure 4). Results showed that the chloroplast genome within the genus Morus was relatively similar. The regions with higher variation were distributed in the LSC and SSC regions, and the IR region was more conservative.

![Figure 4: Percent identity plot for comparison of seven Morus L. chloroplast genomes using m VISTA.](image-url)
Phylogenetic Analysis

Using MEGA-X (10.2.2) software through the ML (Figure 5) and NJ methods (Figure 6) cluster analysis based on the complete chloroplast genome sequences. *M. multicaulis* (Husang32) and *M. atropurpurea* (Lunjiao40) are grouped together it’s means they are closely related and diverged from the *M. indica* and *M. notabilis* earlier.

![Figure 5: Phylogenetic analysis of the Morus L. using complete chloroplast genome by the ML method.](image)

![Figure 6: Phylogenetic analysis of *Morus* species using the complete chloroplast genome by the NJ method.](image)
Discussion

In this study, we collected two cultivated species of Morus L. (M. atropurpurea and M. multicaulis) assembled and annotated the cp genome and performed extensive analyses based on the complete cp genome sequences and amino acid sequences of the annotated genes. The Morus L cp genome is circular contain LSC, IR, SSC. compared with wild species of M. mongolica, M. notabilis and M. indica in the genome sequence, providing more detailed information for phylogenetic studies. The results show that the chloroplast genome length of 7 mulberry species ranges from 158,459 to 159,154 bp. M. atropurpurea and M. multicaulis with two species different accession also have slight differences. “Husang32” and “Ribentiancheng” of M. multicaulis, with “Husang32” 377bp being shorter than the “Ribentiancheng” (Table 1). This result agrees with the gene sequence used by Li Qiaoli [20] on M. multicaulis. Also, M. atropurpurea “Lunjiao40” is 337bp shorter than “Yichuanhong” (Table 2). There were few differences in the length of the IR and SSC regions of the cpDNA from seven species, but most of the differences mainly focus on the LSC region. The results indicated that these species are closely related and confirmed by the phylogenetic tree analysis.

IR expansion/contraction studies reveal the considerable difference even in the same family [13]. It is believed that the borders of the IR region (IRa and IRb) with the LSC and SSC regions, some expansion and contraction may play an important role in evolution [21]. Therefore, it is of great interest to compare IR/SC junctions in different varieties. Based on this, we use 6 mulberry species along with other members, complete chloroplast genome sequences. The results showed that rpl2 gene located in the junction of LSC/IRb 67bp in LSC and the rest are in IRb and trnH gene is 150 bp away from the IRa/LSC boundary of M. multicaulis. With ndhF located in the junction of IRb/SSC 52bp at IRb, the rest is located at SSC of M. alba. Based on IR boundary, it could be seen that M. atropurpurea and M. multicaulis are closely related and together also closely related to M. mongolica and M. notabilis. This result agrees with the finding by Li Qiaoli [20].

Usually, cpDNA also tends to AT [7,22]. The AT content of the four regions of the seven species (Table 2) showed The IR region has the lowest with an average of only 57.00%, LSC region and the SSC region are 66.02% and 70.67%, M. atropurpurea, and M. multicaulis were slightly higher than other groups. The average of 4 species was about 63.78%, SC regions with a high AT content harbour more variation SSR polymorphisms between M. multicaulis and M. atropurpurea are all involved A or T mutations. The lowest AT content in the IR region maybe because it contains all the rRNA, which is very conservative, and the GC content is high and relatively stable. The rpl21 gene is known to only exist in the plastomes of ferns and bryophytes [23]. This result shows that Morus contains two pseudogenes ycf1 and rpsl9. SSR as an important molecular marker used for studying population genetics [24,25]. Which may suggest their usefulness in future evolutionary studies on Morus species. With the Phylogenetic analysis, seven species of Morus gathered together. Both M. atropurpurea and M. multicaulis gather together, which is the closest. This result is similar to Li Qiaoli [20]. In the present research, cpDNA sequences of Morus species are minimal. Our experimental results do not represent the final classification of the Morus system. Therefore, the evolutionary relationship of Morus needs further research to draw more accurate conclusions. Reported here enhance genome information on Morus and contribute to the study of germplasm diversity. These data represent a valuable Phylogenetic analysis source for future studies on Morus populations diversity. Genetic diversity within the chloroplast is considered effective tool for understanding population genetic structure and species evolution. We assemble the full chloroplast DNA sequences of two cultivated species of M. atropurpurea and M. multicaulis and compared with other six wild Morus species. These sequences enabled us to identify evolution divergence time of Morus and can be used for further research of genetic diversity, genetic structure, and genome evolution history of Morus [26,27].

Acknowledgement

This work was supported by China Agriculture Research System of MOF and MARA (CARS-18-ZJ0207), National Key R&D Program of China, key projects of international scientific and technological innovation cooperation (2021YFE0111100), Guangxi innovation-driven development project (AA19182012-2), Zhenjiang Science and Technology support project (GJ2021015).

Author Contribution Statement

GL: conceived and designed the project, assembled the genomes, analyzed the data, and wrote the original manuscript; SY: designed the experiments, analyze portions of data, revised and edited the manuscript; WM, MA, , LQ: Collected leaf samples and extracted chloroplast DNA. ZW: supervised the experiment and edited the manuscript. All authors contributed to the editing of the final manuscript.

Data Availability Statement

The whole chloroplast genome data have been deposited at NCBI with the following accession numbers: MW548981 and MW548982. The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

MW548981(https://www.ncbi.nlm.nih.gov/nuccore/MW548981)
MW548982(https://www.ncbi.nlm.nih.gov/nuccore/MW548982)
Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

References

1. Daniell H, HT Chan, Pasoreck EK (2016) Vaccination via Chloroplast Genetics: Affordable Protein Drugs for the Prevention and Treatment of Inherited or Infectious Human Diseases. Annu Rev Genet 50: 595-618.

2. Hohmann N, Wolf EM, Rigault P, WB Zhou, Kiefer M, et al. (2018) Ginkgo biloba’s footprint of dynamic Pleistocene history dates back only 390,000 years ago. BMC Genomics 19: 299.

3. Howe CJ, Barbrook AC, Koumandou VL, Nisbet RER, Symington HA, et al. (2003) Evolution of the chloroplast genome. Philos Trans R Soc Lond B Biol Sci 358: 99-106.

4. Daniell H LS-B, Greavich J, Sasaki C, Quesada-Vargas T, Guda C, et al. (2006) Complete chloroplast genome sequences of Solanum bulbocastanum, Solanum lyropersicum and comparative analyses with other Solanaceae genomes. Theor Appl Genet 112: 1503.

5. W Dong, X Chao, C Tao, S Zhou, Nicolas SJ (2013) Complete Chloroplast Genome of Sedum sarmentosum and Chloroplast Genome Evolution in Saxifragales. Plos One 8: e80508.

6. HY Zhang, C Li, HM Miao, SJ Xiong (2013a) Insights from the Complete Chloroplast Genome into the Evolution of Sesamum indicum L. Plos One 8: e80508.

7. JX Liu (2015) De novo Assembly and Gene Annotation of the Chloroplast Genomes of Lycium barbarum and Lycium ruthenicum. Dissertation, Beijing Institute of Genomics Chinese Academy of Sciences.

8. Daniell H, Wurdack KJ, Kanagaraj A, Lee SB, Sasaki C, et al. (2008) The complete nucleotide sequence of the cassava (Manihota esculenta) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. Theor Appl Genet 116: 723-737.

9. Asaf S, Khan AL, Khan MA, Waqas M, SM Kang, et al. (2017) Chloroplast genomes of Arabidopsis halleri ssp. gemmifera and Arabidopsis lyrata ssp. petraea: Structures and comparative analysis. Sci Rep 7: 7556.

10. S Xue, T Shi, WJ Luo, XP Ni, Lqbal S, et al. (2019) Comparative analysis of the complete chloroplast genome among Prunus mume, P. armeniaca, and P. salicina. Hort Res 6: 1-9.

11. YD Gao, Harris A, XI He (2015) Morphological and ecological divergence of Lilium and Nomocharis within the Hengduan Mountains and Qinghai-Tibetan Plateau may result from habitat specialization and hybridization. BMC Evol Biol 15: 147-168.

12. L Cheng, Nam J, SH Chu, Rungnapa P, Park YJ (2019) Signatures of differential selection in chloroplast genome between japonica and indica. Rice 12: 65.

13. KJ-Joong K, Hea-Lim L (2004) Complete Chloroplast Genome Sequences from Korean Ginseng (Panax schinseng Nees) and Comparative Analysis of Sequence Evolution among 17 Vascular Plants. DNA Research 4: 247-261.

14. QW Zeng, HY Chen, C Zhang, MJ Han, TLI, et al. (2015) Definition of Eight Mulberry Species in the Genus Morus by Internal Transcribed Spacer-Based Phylogeny. Plos One 10: e0135411.

15. Sancho R, Cantalapiedra CP, López-Alvarez D, Gordon SP, Vogel JP, et al. (2017) Comparative plastome genomics and phylogenomics of Brachypodium: flowering time signatures, introgression and recombination in recently diverged ecotypes. New Phytol 218: 1631-1644.

16. Ghosh JS, Bhattacharya S, Pal A (2017) Molecular phylogeny of 21 tropical bamboo species reconstructed by integrating non-coding internal transcribed spacers (ITS1 and 2) sequences and their consensus secondary structure. Genetica 145: 319-333.

17. B Xu, XM Zeng, XF Gao, DP Jin, LB Zhang (2017) ITS non-concerted evolution and rampant hybridization in the legume genus Lepidium (Fabaceae). Sci Rep 7: 44057.

18. Anton Bankovic, Sergey Nurk, Dmitry Antipov, Alexey A Gurevich, Mikhail Dvorkin, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19: 455-477.

19. GonçAlves NA, Monica C, Garcez LLI. (2015) Complete Chloroplast Genome of Tanaecium tetragonolobum: The First Bignoniaceae Plastome. Plos One 10: e0129950.

20. QL Li, JZ Guo, N Yan, CC Li (2016) Complete chloroplast genome sequence of cultivated Morus L. species. Genet Mol Res 15.

21. HY Zhang, C Li, HM Miao, SJ Xiong (2013b) Insights from the Complete Chloroplast Genome into the Evolution of Sesamum indicum L. Plos One 8: e80508.

22. Zhengqiu Cai, Cynthia Penaflor, Jennifer V Kuehl, James Leebens-Mack, John E Carlson, et al. (2006) Complete plastid genome sequences of Drimys, Lirioidendron , and Piper : implications for the phylogenetic relationships of magnolids. BMC Evol Biol 16: 77.

23. Steane DA (2005) Complete Nucleotide Sequence of the Chloroplast Genome from the Tasmanian Blue Gum, Eucalyptus globulus (Myrtaceae). DNA Res 12: 215-220.

24. Katti MV, Ranjekar PK, Gupta VS (2001) Differential Distribution of Simple Sequence Repeats in Eukaryotic Genome Sequences. Mol Biol Evol 18: 1161-1168.

25. Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94: 275-288.

26. Magdy M, Ou L, Yu H, Rong Chen, Yuhong Zhou, et al. (2019) Pan-plastome approach empowers the assessment of genetic variation in cultivated Capsicum species. Hortic Res 6: 108.

27. Naqui S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, et al. (2019) Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. Science 365: eaaw7317.
