Genetics and Molecular Research

# Complete chloroplast genome sequence of cultivated *Morus* L. species

**Q.L. Li, J.Z. Guo, N. Yan and C.C. Li**

College of Forestry, Northwest A&F University, Yangling, China

Corresponding author: J.Z. Guo
E-mail: 18853881929@163.com

Genet. Mol. Res. 15 (4): gmr15048906
Received June 21, 2016
Accepted August 22, 2016
Published October 17, 2016
DOI http://dx.doi.org/10.4238/gmr15048906



**ABSTRACT.** The complete chloroplast genome (cpDNA) sequences of two cultivated species of *Morus* L. (*Morus atropurpurea* and *Morus multicaulis*) are reported and reconstructed in this study, and were compared with that of wild *Morus mongolica*. In *M. atropurpurea*, the circular genome is 159,113 bp in size and comprises two identical inverted repeat (IR) regions of 25,707 bp each, separated by a large single-copy (LSC) region of 87,824 bp and a small single-copy (SSC) region of 19,875 bp. The cpDNA sequence of *M. multicaulis* is longer than that of *M. atropurpurea* (159,154 bp), and consists of two IRs (25,678 bp), a LSC region (87,763 bp), and a SSC region (20,035 bp). Each cpDNA contains 112 unique genes including 78 protein-coding genes, 30 transfer RNA genes, and 4 ribosomal RNA genes, with a GC content of 36.2%. There were 83 simple sequence repeats (SSRs) with mononucleotides being the most common (60) and di-, tri-, tetra-, and hexanucleotides appearing less frequently in *M. atropurpurea*. *M. multicaulis* contains 81 SSRs containing 63 mononucleotide repeats. The genes and SSRs identified in this study may enhance understanding

of cpDNA evolution at both intra- and interspecific levels. MEGA 6.0 was used to construct a phylogenetic tree of 27 species, which revealed that *M. atropurpurea* and *M. multicaulis* are more related to their congeners than to others. The cpDNA of *M. atropurpurea* and *M. multicaulis* and its structural analysis are important for the chloroplast genome project, development of molecular markers for *Morus* species, and breeding of varieties.

**Key words:** *Morus atropurpurea*; *Morus multicaulis*; Phylogeny; Chloroplast genome; Complete sequences

## INTRODUCTION

The chloroplast (cp) is the photosynthetic organelle representing one of the most important organelles in green plants and algae. Its genome has proven to be useful for plant phylogenetics, species identification, population genetics, and genetic engineering (Nock et al., 2014). 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 (Jansen and Palmer, 1987; Wu et al., 2009).

The length of cpDNA ranges from 120 to 160 kb, owing to the loss and gain of introns (Delannoy et al., 2011), the expansion of the IR region (Dong et al., 2013; Zhang et al., 2013), and major structural rearrangements (Walker et al., 2014), which contain 110 to 130 genes (Huang et al., 2013). Chloroplasts are a valuable tool for use in phylogenetic studies because of their genes, which lack recombination and conversation (Ravi et al., 2008). To date, more than 1000 complete cpDNA sequences have been submitted to GenBank; however, the cpDNA sequence of Moraceae is incomplete. The cpDNA of the cultivated species *Morus atropurpurea* and *M. multicaulis* are described in detail in this study.

*Morus* L. is an economically significant crop belonging to the Moraceae family, which was once classified in the subclass Hamamelidae (Order: Urticales) (http://plants.usda.gov/), but has now been reclassified in the order Rosales in Fabidae (also known as Rosid I) according to some of its nuclear genes or cpDNA sequences (Zhang et al., 2011; Su et al., 2014). There are 68 species of mulberry, which are found mostly in Asia, mainly China and Japan, and continental America, and include cultivated (*M. atropurpurea* and *M. multicaulis*) and wild (*M. mongolica* and *M. notabilis*) species. This family is poorly represented in Africa and Europe and is virtually absent from Australia. Their leaves provide the sole source of food for the silkworm and their fruits are rich in nutrients and are beneficial to human health. Although there have been a few phylogenetic studies involving mulberry, these were restricted to only a few genes. A complete repertoire of genes would thus help us to establish the position of mulberry in the tree of life (Ravi et al., 2006). *M. atropurpurea* and *M. multicaulis* are native to China and are cultured in Shaanxi Province. 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.

## MATERIAL AND METHODS

### Plant material, sequences, assembly, and annotation

*M. atropurpurea* and *M. multicaulis* plants were collected from the mulberry field of Northwest A&F University. The DNeasy plant Minikit (Qiagen, Seoul, South Korea) was used to isolate total genomic DNA from 10 g fresh leaves and a UV-visible spectrophotometer was used to determine DNA concentration. High-quality DNA was sequenced using the Illumina Hiseq 2500 platform (Illumina Inc., San Diego, CA, USA).

The complete cpDNA sequence was assembled with MITOBIM V1.7 (Hahn et al., 2013) using default settings, with its congener *M. mongolica* (GenBank accession No. KM491711) as the reference sequence. Sequences were annotated in GENEIOUS R8 (Biomatters Ltd., Auckland, New Zealand) by aligning with that of *M. mongolica.* OGDraw was used to draw the circular gene map and ambiguous gaps or nucleotides were corrected manually (Lohse et al., 2007). Complete chloroplast genomes were submitted to GenBank under the following accession No. *M. atropurpurea*, KU355276 and *M. multicaulis*, KU355297.

### Comparative analysis of Rosales chloroplast genomes

The mVISTA online software in shuffle-LAGAN mode (Frazer et al., 2004) was applied to compare the complete chloroplast genomes of cultivated *Morus* species with four representatives of Rosales: *Prunus persica* (Rosaceae: NC-014697), *Pyrus pyrifolia* (NC-015996), *Fragaria vesca* subsp *vesca* (NC-015206), and *M. mongolica* (Moraceae), with the basal species *Nicotiana tabacum* L. (Solanaceae; Solanales; Z00044) used as the reference in the comparative analysis.

Simple sequence repeats (SSRs) were identified using the online software Wabsat and Gramene Ssrtool (in the chloroplast genome of *M. atropurpurea* and *M. multicaulis*). A total of 10, 5, 4, 3, 3, and 3 SSRs were identified for mono-, di-, tri-, tetra-, penta-, and hexanucleotides, respectively.

### Phylogenetic analysis

The MEGA 6.0 software was used to determine the phylogenetic relationships between *Morus* species by the maximum likelihood (ML) and neighbor-joining (NJ) methods (Tamura et al., 2013). Data on the cpDNA of *Morus* species are available, including those from *M. indica* (NC-008359), *M. mongolica* (KM491711), *M. notabilis* (KP939360), *M. multicaulis* (KU355297), and *M. atropurpurea* (KU355276), and *N. tabacum,* which was used as an outgroup. The likelihood bootstrap analysis of each branch was calculated with 1000 replications.

## RESULTS

### Genome content and organization

The cpDNA sequence of *M. multicaulis* was determined (Figure 1) and found to be 159,154-bp long, which is longer than that of the other congeners (Table 1). It comprises a circular double-stranded DNA structure composed of two identical IR regions (25,678 bp),

a LSC region (87,763 bp), and a SSC region (20,035 bp). The *M. atropurpurea* cpDNA was found to be 159,113 bp long (Figure 2) and was composed of a typical quadripartite structure containing a pair of IR regions of 25,707 bp each, which were separated by a SSC region (19,875 bp) and a LSC region (87,824 bp) (Table 1).

The GC content of the *M. multicaulis* chloroplast genome was found to be 36.2%, which is lower in LSC (33.9%) and SSC (29.3%) regions and higher in IR regions (42.9%). No changes were found to occur in the IR region of the five mulberry species. cpDNA contains 130 functional genes including eight rRNA genes, 37 tRNA genes, and 85 PCGs. Pseudogenes and ORFs were all non-coding.

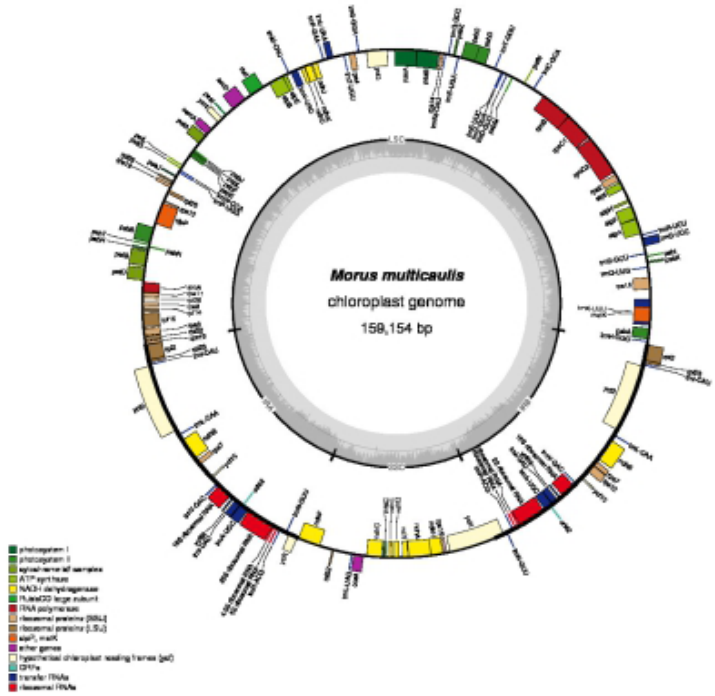


**Figure 1.** Gene map of the chloroplast genome of *Morus multicaulis*.

**Table 1.** Comparison of chloroplast genomes among five species of *Morus* L.

| Characteristics | *M. indica* | *M. mongolica* | *M. notabilis* | *M. atropurpurea* | *M. multicaulis* |
|---|---|---|---|---|---|
| Size (bp) | 158,484 | 158,459 | 158,680 | 159,113 | 159,154 |
| LSC length/percent/CG content | 87,386/55.14/34.1 | 87,367/55.14/34.0 | 87,470/55.12/34.1 | 87,761/55.18/33.9 | 87,763/55.15/33.9 |
| SSC length/percent/CG content | 19,742/12.46/29.4 | 19,736/12.45/29.3 | 19,776/12.46/29.3 | 19,875/12.50/29.3 | 20,035/12.59/29.3 |
| IR length/percent/CG content | 25,678/16.20/42.9 | 25,678/16.20/42.9 | 25,717/16.21/42.9 | 25,707/16.16/42.9 | 25,678/16.13/42.9 |
| GC content (%) | 36.4 | 36.3 | 36.4 | 36.2 | 36.2 |
| Number of genes | 128 | 133 | 129 | 130 | 130 |
| Protein-coding genes | 83 | 88 | 84 | 85 | 85 |

bp, base pairs.

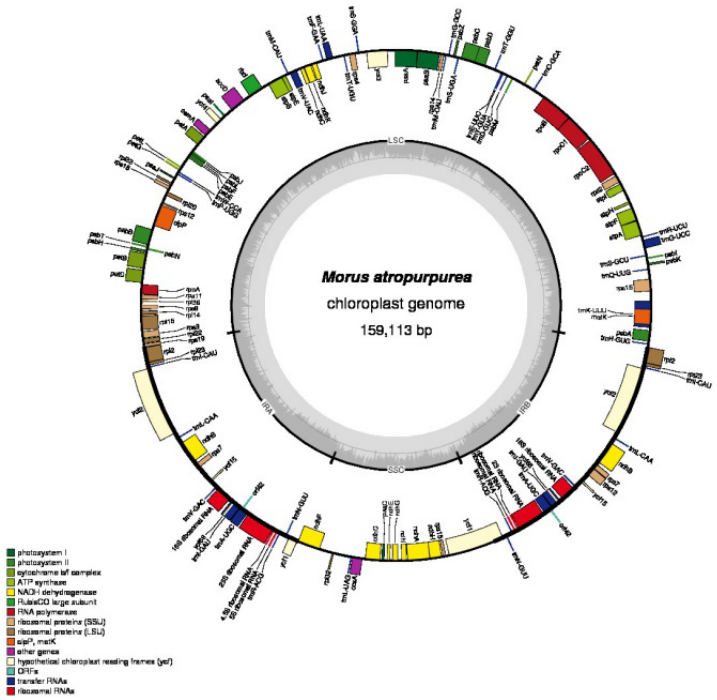


**Figure 2.** Gene map of the chloroplast genome of *Morus atropurpurea*. Genes shown on the side of the larger circle are transcribed clockwise. Inverted repeats (IRa and IRb) separate the genome into small and large single-copy regions.

Eighteen genes, including seven tRNA, seven PCGs, and all rRNA genes were duplicated in the IR regions. *M. multicaulis* contained 22 genes (eight tRNA, 12 PCGs, and two pseudogenes), with one intron, which is consistent with that found for *M. atropurpurea*, with the exception of two genes (ycf3 and clpP) that contain two introns (Table 2). Of the 22 genes, 10 are situated within the IR region (4 PCGs, 4 tRNAs, and 2 pseudogenes), 1 in the SSC region (ndhA), and 11 in the LSC region (7 PCGs and 4 tRNAs), and this study also found trnK-UUU has the largest intron that contains the protein-coding gene matK, which is similar to that found in green plants (Zhang et al., 2013).

## Codon usage

All 85 protein-coding genes in the cpDNA of *M. multicaulis* and *M. atropurpurea* were encoded by 53,051 and 53,037 codons, respectively (Table 3). Codon usage strongly reflects the AT tendency. For *M. atropurpurea*, 63.5% of codons end in A or T, with 73.5% stop codons ending in A or T. Leucine accounts for the highest codon usage (5624), followed by serine (4778), isoleucine (4731), and phenylalanine (3754). These four amino acids represent one third of the total codons. TAA is the most frequent stop codon found, accounting for 1268 uses, which is higher than that of TGA (1079) and nearly twice that of TAG (847) (Table 3). ATG (919) was the most common start codon, with the exception of GTG for rps19 and ACT for rps2.

**Table 2.** Genes present in the chloroplast genome of *Morus atropurpurea* and *M. multicaulis.*

| Function | Gene group | Gene name | | | |
|---|---|---|---|---|---|
| Self-replication | Ribosomal RNA genes | rrn4.5 (x2) | rrn5 (x2) | rrn16 (x2) | rrn23 (x2) |
| | Transfer RNA genes | trnA-UGC (x2)<br>trnF-GAA<br>trnH-GUG<br>trnL-CAA (x2)<br>trnN-GUU (x2)<br>trnR-UCU<br>trnT-GGU<br>trnW-CCA | trnC-GCA<br>trnfM-CAU<br>trnI-CAU (x2)<br>trnL-UAA<br>trnP-UGG<br>trnS-GCU<br>trnT-UGU<br>trnY-GUA | trnD-GUC<br>trnG-GCC<br>trnI-GAU (x2)<br>trnL-UAG<br>trnQ-UUG<br>trnS-GGA<br>trnV-GAC (x2) | trnE-UUC<br>trnG-UCC<br>trnK-UUU<br>trnM-CAU<br>trnR-ACG (x2)<br>trnS-UGA<br>trnV-UAC |
| | Small subunit of ribosome | rps2<br>rps8<br>rps15 | rps3<br>rps11<br>rps16* | rps4<br>rps12 (x2)<br>rps18 | rps7 (x2)<br>rps14<br>rps19 |
| | Lange subunit of ribosome | rpl2* (x2)<br>rpl22<br>rpl36 | rpl14<br>rpl23 (x2) | rpl16*<br>rpl32 | rpl20<br>rpl33 |
| | RNA polymerase subunits | rpoA | rpoB | rpoC1 * | rpoC2 |
| Photosynthesis | NADH dehydrogenase | ndhA*<br>ndhE<br>ndhI | ndhB* (x2)<br>ndhF<br>ndhJ | ndhC<br>ndhG<br>ndhK | ndhD<br>ndhH |
| | Photosystem I | psaA<br>psaJ | psaB | psaC | psaI |
| | Photosystem II | psbA<br>psbE<br>psbJ<br>psbN | psbB<br>psbF<br>psbK<br>psbT | psbC<br>psbH<br>psbL<br>psbZ | psbD<br>psbI<br>psbM |
| | | petA<br>petL | petB*<br>petN | petD* | petG |
| | ATP synthase | atpA<br>atpH | atpB<br>atpI | atpE | atpF* |
| | Large subunit of rubisco | rbcL | | | |
| Other genes | Maturase | matK | | | |
| | Protease | ClpP* | | | |
| | Envelope membrane protein | cemA | | | |
| | Subunit of acetyl-CoA-carboxylase | accD | | | |
| | C-type cytochrome synthesis | ccsA | | | |
| | Component of TIC complex | yf1 (x2) | | | |
| Unknown function | Hypothetical chloroplast reading frames | ycf2 (x2)<br>ycf68* (x2) | ycf3* | ycf4 | ycf15 (x2) |
| | ORFs | orf42 (x2) | | | |

Asterisks indicate genes containing one or more introns.

## Comparison with other Rosales chloroplast genomes

The cp genomes of *M. multicaulis* and *M. atropurpurea* contained 83 and 81 SSRs, respectively, of at least 10 bp in size (Table 4). A total of 60, 8, 3, 10, and 2 mono-, di-, tri-, tetra-, and pentanucleotide repeats were found in the *M. atropurpurea* chloroplast genome. All mononucleotides and 17 other SSRs were comprised of T and A nucleotides, with a high AT content (92.2%). Of the 83 SSRs, 23 were located within gene-coding regions and 60 were located within intergenic spacers. SSRs were rarer in protein-coding genes than in non-coding regions (Rajendrakumar et al., 2007). Fifty-two loci were identical between *M. atropurpurea* and *M. multicaulis*, 31 were unique, and three were not found (Table 4).

The borders of the two inverted repeats (IRa and IRb) with the LSC and SSC regions play an important role in the expansion and contraction of the chloroplast genome (Goulding et al., 1996). It is believed that the locations of the SSC/IR and LSC/IR junctions are markers of chloroplast genome evolution (Zhang et al., 2013). The IR junction among the potential impact of these changes in the cp genome of *Morus* was compared.

**Table 3.** Codon usage in *Morus multicaulis* and *M. atropurpurea.*

| Codon | Amino acid | *M. atropurpurea* | *M. multicaulis* | Codon | Amino acid | *M. atropurpurea* | *M. multicaulis* |
|---|---|---|---|---|---|---|---|
| GGG | Gly (G) | 542 | 494 | TGG | Trp (W) | 648 | 684 |
| GGA | Gly (G) | 740 | 759 | TGA | stop | 1079 | 1032 |
| GGT | Gly (G) | 551 | 599 | TGT | Cys (C) | 711 | 725 |
| GGC | Gly (G) | 332 | 350 | TGC | Cys (C) | 436 | 435 |
| GAG | Glu (E) | 599 | 550 | TAG | stop | 847 | 786 |
| GAA | Glu (E) | 1245 | 1368 | TAA | stop | 1268 | 1306 |
| GAT | Asp (D) | 1022 | 1064 | TAT | Try (Y) | 1524 | 1624 |
| GAC | Asp (D) | 412 | 425 | TAC | Try (Y) | 730 | 690 |
| GTG | Val (V) | 448 | 418 | TTG | Leu (L) | 1083 | 1073 |
| GTA | Val (V) | 748 | 728 | TTA | Leu (L) | 1359 | 1250 |
| GTT | Val (V) | 797 | 792 | TTT | Phe (F) | 2369 | 2343 |
| GTC | Val (V) | 432 | 430 | TTC | Phe (F) | 1385 | 1471 |
| GCG | Ala (A) | 228 | 249 | TCG | Ser (S) | 586 | 578 |
| GCA | Ala (A) | 431 | 430 | TCA | Ser (S) | 1017 | 979 |
| GCT | Ala (A) | 463 | 511 | TCT | Ser (S) | 1193 | 1273 |
| GCC | Ala (A) | 328 | 321 | TCC | Ser (S) | 858 | 864 |
| AGG | Arg (R) | 632 | 596 | CGG | Arg (R) | 366 | 350 |
| AGA | Arg (R) | 1036 | 1044 | CGA | Arg (R) | 564 | 596 |
| AGT | Ser (S) | 654 | 718 | CGT | Arg (R) | 383 | 363 |
| AGC | Ser (S) | 470 | 478 | CGC | Arg (R) | 244 | 236 |
| AAG | Lys (K) | 1050 | 1039 | CAG | Gln (Q) | 462 | 440 |
| AAA | Lys (K) | 2206 | 2280 | CAA | Gln (Q) | 1067 | 1013 |
| AAT | Asn (N) | 1924 | 1883 | CAT | His (H) | 907 | 945 |
| AAC | Asn (N) | 802 | 728 | CAC | His (H) | 391 | 362 |
| ATG | Met (M) | 919 | 855 | CTG | Leu (L) | 505 | 489 |
| ATA | Ile (I) | 1700 | 1729 | CTA | Leu (L) | 859 | 799 |
| ATT | Ile (I) | 1945 | 1965 | CTT | Leu (L) | 1172 | 1065 |
| ATC | Ile (I) | 1086 | 1083 | CTC | Leu (L) | 646 | 581 |
| ACG | Thr (T) | 332 | 399 | CCG | Pro (P) | 378 | 400 |
| ACA | Thr (T) | 733 | 689 | CCA | Pro (P) | 723 | 738 |
| ACT | Thr (T) | 683 | 690 | CCT | Pro (P) | 616 | 730 |
| ACC | Thr (T) | 593 | 587 | CCC | Pro (P) | 578 | 580 |

Four complete chloroplast genome sequences of Rosales and the sister group Cucurbitales were selected, namely, *M. atropurpurea*, *M. multicaulis*, *M. mongolica*, *M. notabilis*, *Rosa odorata* var. *gigantea*, *Cucumis melo* subsp *melo*, and *Corynocarpus laevigatus*. The IR boundaries of cpDNA from *M. atropurpurea* and *M. multicaulis* were very similar (Figure 3). The IRb-SSC junction was found to be located at the ndhF gene, and the ndhF and ycf1 genes overlapped by 32 bp in *C. melo* subsp *melo*. The IRa-SSC was located in ycf1, resulting in the formation of a ycf1 pseudogene. The boundary of the LSC/IR was located within the rps19 gene, also resulting in the formation of an rps19 pseudogene, which is consistent with the findings of a previous study (Nazareno et al., 2015).

mVISTA (Frazer et al., 2004) was used to compare sequence identity between the six cpDNAs, referring to the annotation of the *N. tabacum* cpDNA (Figure 4). Although some divergent regions were found, Rosales cpDNAs were found to be rather conserved through the complete aligned than their non-coding regions. For *M. atropurpurea*, *M. mongolica* was the closer relative, followed by *M. multicaulis*, *M. notabilis*, *P. pyrifolia*, *Prunus kansuensis*, *F. vesca* subsp *vesca*, and *N. tabacum.*

The complete chloroplast genomes of Rosales clades were used to construct the phylogenetic tree in MEGA6.0 via the ML (Figure 5) and NJ methods (Figure 6). The two methods group the *Morus* species together. The ML and NJ methods grouped M. *atropurpurea* and *M. mongolica* together. However, we cannot conclude that *M. atropurpurea* and *M.*

*mongolica* have a close genetic relationship, because the cp genomes of other *Morus* species were not sequenced. Moreover, further research into *Morus* species is needed in order to reach a conclusion.

**Table 4.** Distribution of SSR loci in the *Morus atropurpurea* (M.A) and *M. multicaulis* (M.M) chloroplast genomes.

| Repeat unit | Length(bp) | Number of SSRs | | Position in the chloroplast genome (gene name) | |
|---|---|---|---|---|---|
| | | M.A | M.M | M.A | M.M |
| A | 10 | 8 | 10 | 3997 (trnK-UUU); 5100; 5998 (rps16); 29085; 49740; 68673; 68688; 114237 (ndhF) | 2142 (trnK-UUU); 3980 (trnK-UUU); 5079; 5977 (rps16); 29067; 49740; 68616; 68631; 114154 (ndhF); 116262 |
| | 11 | 4 | 3 | 53953; 62875; 87528; 116346 | 9589; 62837; 87467 |
| | 12 | 2 | 3 | 13603 (atpF); 84635 | 4830; 53982; 85376 (rpl16) |
| | 13 | | 1 | | 13596 (atpF) |
| | 14 | 1 | 1 | 128093 | 128163 |
| | 15 | 2 | 1 | 9583; 74234 (clpP) | 74160 (clpP) |
| | 16 | 1 | 1 | 9002 | 8990 |
| | 17 | 1 | | 4846 | |
| T | 10 | 18 | 20 | 5279; 9801; 24375 (rpoC1); 30690; 30956; 54024; 54921;5 7117 (atpB); 58017 (rbC1); 62648; 66988; 68800; 70966; 74032; 116849; 122289; 130417 (ycf1); 132174 (ycf1) | 66; 5258; 8582; 9802; 68743; 70892; 73958 (clpP); 83130; 14098; 14919; 24357 (rpoC1); 30672; 30938; 54024; 57098 (atpB); 62610; 66927; 116784; 130487 (ycf1); 132244 (ycf1) |
| | 11 | 7 | 6 | 127; 526; 8593; 59604; 74750; 78755 (petB); 131276 (ycf1) | 513, 34264; 69552; 78684 (petB); 122351; 131346 (ycf1) |
| | 12 | 9 | 5 | 12711; 27635 (rpoB); 34289; 37832; 57588; 68549; 69620; 72545; 85868 (rpl16) | 27617 (rpoB); 57549; 59565; 72471; 85809 (rpl16) |
| | 13 | 3 | 5 | 9225; 13293 (atpF); 128515 | 12703; 13286 (atpF); 68491; 81352; 128585 |
| | 14 | 1 | 5 | 63903 | 9213; 51829; 63865; 74676 (clpP); 86927 |
| | 16 | 1 | | 81423 | |
| | 17 | 2 | 1 | 49438; 84685 | 49475 |
| | 19 | | 1 | | 116631 |
| AT | 10 | 2 | 1 | 68872; 115739 (ndhF) | 11566 (ndhF) |
| | 12 | 1 | 3 | 10814 | 5522; 118643; 118871 |
| TA | 12 | 4 | 1 | 5543; 21253 (rpoC2); 118731; 118839 | 21234 (rpoC2) |
| TC | 10 | 1 | 1 | 64630 (cemA) | 645927(cemA) |
| TAT | 12 | 1 | | 49786 | |
| TTC | 12 | 1 | 1 | 70983 | 70909 |
| AAT | 12 | 1 | 1 | 128481 | 128565 |
| ATTT | 12 | 1 | 1 | 14192 | 62140 |
| | 16 | | 1 | | 14187 |
| AAAT | 12 | 2 | 2 | 24069 (rpoC1); 46696 (ycf3) | 24056 (rpoC1); 46731 (ycf3) |
| TATT | 12 | 1 | 1 | 24406 (rpoC1) | 24388 (rpoC1) |
| ATTA | 12 | 2 | 2 | 34075; 116528 | 33980; 116443 |
| TTTA | 12 | 1 | | 62179 | |
| TCTT | 12 | 1 | 1 | 111648 | 111575 |
| TTAT | 12 | 1 | | 117879 | |
| AAAG | 12 | 1 | 1 | 135264 | 135331 |
| AAGGA | 15 | 1 | 1 | 14037 (atpF) | 14021 (atpF) |
| ATTTC | 15 | 1 | | 24509 (rpoC1) | |

## DISCUSSION

In recent years, researchers have used cpDNA for the study of plant evolution, along with the published chloroplast data available in the NCBI database (Drew et al., 2014). In our study, we prudentially selected cpDNAs of different taxa from the NCBI database that were potentially published. Additionally, long-branch attraction will mislead to a wrong phylogenetic tree.

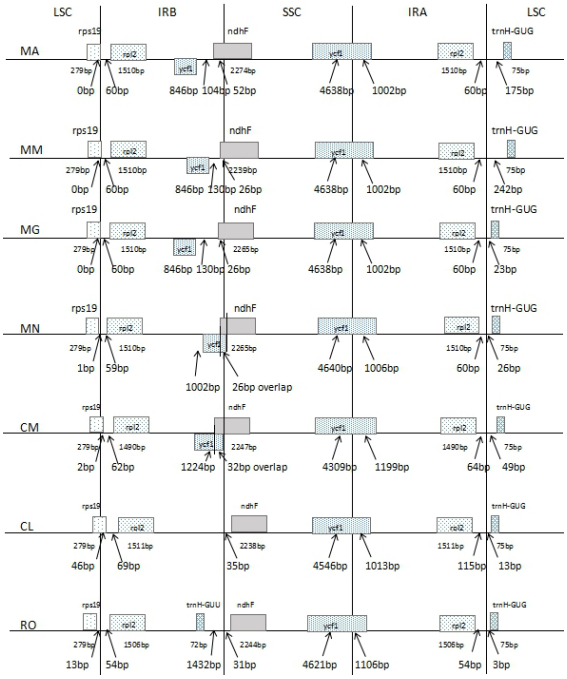


**Figure 3.** Comparison of the junction between IR and SC regions among Rosales and its sister group. MA: *Morus atropurpurea*; MM: *M. multicaulis*; MG: *M. mongolica*; MN: *M. notabilis*; CM: *Cucumis melo* subsp *melo*; CL: *Corynocarpus laevigatus*; RO: *Rosa odorata* var. *gigantea.*

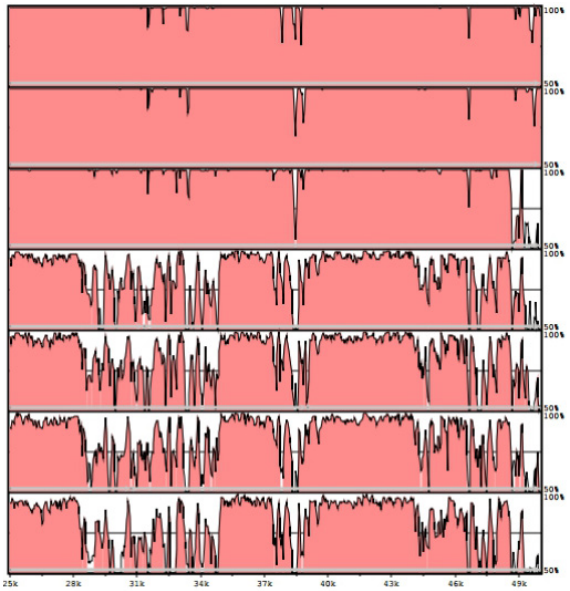


**Figure 4.** Y-scale represents identity, ranging from 50 to 100%. Genomes are arranged according to the number of conserved bases relative to Rosales.

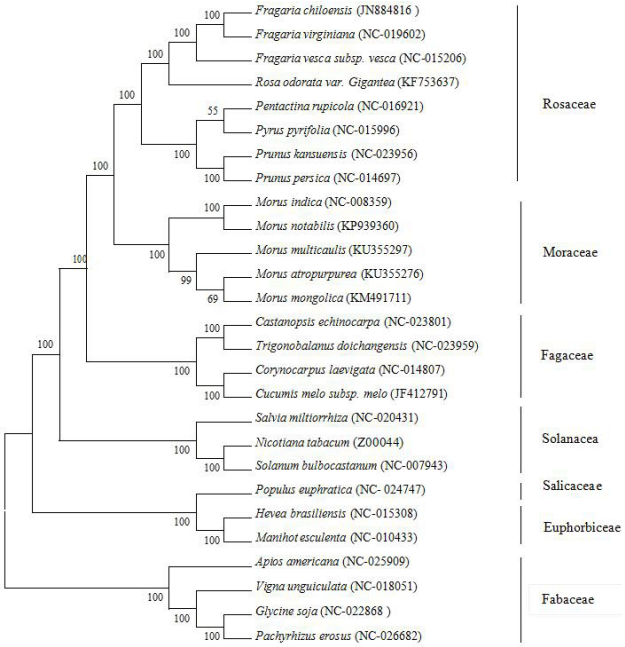


**Figure 5.** Phylogenetic analysis of *Morus* species using the complete chloroplast genome by the ML method.

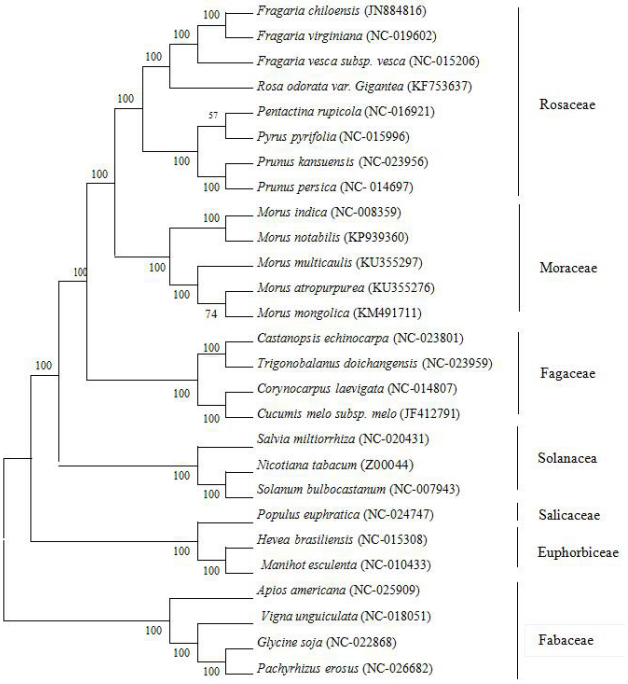


**Figure 6.** Phylogenetic analysis of *Morus* species using the complete chloroplast genome by the NJ method. *Nicotiana tabacum* is included as the outgroup to root the tree.

Research has shown that *M. mongolica* and *M. indica* are wild species of *Morus* L. (Yang and Yoder, 1999), whereas, *M. atropurpurea* and *M. multicaulis* are cultivated species of *Morus* L. In the present study, the complete chloroplast genome sequence of *M. atropurpurea* was determined and compared with that of *M. multicaulis* and the wild species of *Morus*. The genome sequence, the size of the LSC, IR, SSC, and the CG content, among other variables, were analyzed providing detailed information for phylogenetic studies of the chloroplast. The results revealed that the size of the *M. atropurpurea* chloroplast genome is 159,113 bp, which is 41 bp shorter than that of *M. multicaulis* and 654 bp longer than that of *M. mongolica*. Moreover, there were few differences in the length of the IR and SSC regions of the cpDNA from all five species, with differences accounted for by the LSC region. Analysis of the results also indicated that these species are closely related and this will be confirmed by construction of the phylogenetic tree.

The expansion and contraction of IR are common evolutionary events in plants (Liu et al., 2013). It is believed that the locations of SSC/IR and LSC/IR junctions are markers of chloroplast genome evolution (Zhang et al., 2013). Here, we compared the positions of the IR/SC boundary in six complete cpDNA sequences. The IR boundaries of *Morus* species followed the same pattern in terms of the order of genes and structure, except for the IRb/SSC and IRa/LSC boundaries. In the IRb/SSC junction, 52 bp of the ndhF gene was located in IRb, with the rest located in the SSC in *M. atropurpurea*; this differed in *Morus* species. In the IRa/LSC boundary, the trnH-GUG gene was 175-bp away from the boundary of IRa/LSC in *M. atropurpurea,* 242-bp away in *M. multicaulis,* and 23-bp away in *M. mongolica.* The IR boundary showed that *M. atropurpurea* and *M. multicaulis* are closely related, and have a closer genetic relationship to *M. mongolica* than to *M. notabilis*. Studies based on IR/SC junction regions and other variable regions from different *Morus* species would be of great help in systematics. In addition, the information generated from such studies would be useful for taxonomic analyses of other species of *Morus*, other genera within Moraceae, and other families within the same subclass. The cpDNA sequence of cultivated *Morus* described in the present study will contribute to further studies on molecular breeding, phylogenetics, and genetic engineering.

Most cpDNAs are AT rich (AT content above 60%), have conserved regions with lower AT contents, and have unevenly distributed AT contents (Cai et al., 2006). cpDNA from *M. atropurpurea* and *M. multicaulis* exhibited the same features, and the AT content in the whole cpDNA, SSC, LSC and IR regions was 63.8, 70.7, 66.1, and 57.1%, respectively, with no changes observed between the two mulberry species (Table 1). Similarly, regions with a high AT content harbor more variation, such as hypervariable regions and SSRs. SSR polymorphisms between *M. multicaulis* and *M. atropurpurea* all involved A or T mutations. These phenomena indicate that a positive correlation exists between sequence divergence and AT content, and that there is a bias toward A and T changes over G and C changes in plant cpDNAs.

The rpl21 gene is only present in the plastomes of ferns and bryophytes (Steane, 2005) and the infA gene is known to have been transferred to the nucleus and lost from almost all known rosid plastomes (Millen et al., 2001). The *Morus* plastome also contains two pseudogenes, ycf15 and ycf68. ycf15 is not believed to be a protein-coding gene (Schmitz-Linneweber et al., 2002). The ycf15 gene fragment indicates that it is a remnant of an ancestral functional gene. The deletion observed in the ycf68 gene, which causes the frame-shift, does not appear to have been a sequencing artifact, as the coverage and read quality in the concerned region were high.

The SSRs identified in *M. atropurpurea*, serving as important molecular markers, can be applied to further population genetics studies (Katti et al., 2001; Shaw et al., 2007). We

identified 83 and 81 SSRs in the *M. multicaulis* and *M. atropurpurea* cp genomes, respectively. Due to their variability at inter- and intrapopulation levels, these SSRs may be useful in evolutionary studies. Future research should focus on the validity of SSRs in phylogenetic and ecological studies of *Morus*. Data on the SSRs of *Morus* are available and were used in the present study. We found that the numbers of SSRs in the complete cpDNAs of different *Morus* species were almost identical. A number of SSRs were located within the same gene (Nguyen et al., 2015). For example, dinucleotides were observed in rpoC2, cemA, and ndhF, and trinucleotides were observed in the non-coding region. Moreover, three mononucleotides were observed in the ycf1 gene and two mono-, two tetra-, and one pentanucleotide SSRs were found in the rpoC1 gene. SSRs distributed in coding genes between *M. atropurpurea* and *M. multicaulis* were similar, containing atpF, ycf1, cemA, atpB, rpoC2, and ndhF, which was consistent with the findings of Kong and Yang (2016).

The nucleotide sequence and structure of the complete chloroplast genomes of *M. multicaulis* and *M. atropurpurea*, and the sequence differences between *Morus* species and other species presented in this study will contribute to future evolution and ecological studies.

The cpDNA sequences of *Morus* species, including *M. mongolica*, *M. indica*, and *M. notabilis*, have been reported; however, data on the cpDNA of cultivated *Morus* species are limited. The complete cpDNA sequences of *M. atropurpurea* and *M. multicaulis* reported here enhance genome information on *Morus* and contribute to the study of germplasm diversity. These data represent a valuable source of markers for future studies on *Morus* populations. Moreover, the complete cp genome sequence also provides data on functional protein variability within the chloroplast.

## Conflicts of interest

The authors declare no conflict of interest.

## REFERENCES

Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, et al. (2006). Complete plastid genome sequences of *Drimys, Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.* 6: 77. http://dx.doi.org/10.1186/1471-2148-6-77

Delannoy E, Fujii S, Colas des Francs-Small C, Brundrett M, et al. (2011). Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* 28: 2077-2086. http://dx.doi.org/10.1093/molbev/msr028

Dong W, Xu C, Cheng T and Zhou S (2013). Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. *PLoS One* 8: e77965. http://dx.doi.org/10.1371/journal.pone.0077965

Drew BT, Ruhfel BR, Smith SA, Moore MJ, et al. (2014). Another look at the root of the angiosperms reveals a familiar tale. *Syst. Biol.* 63: 368-382. http://dx.doi.org/10.1093/sysbio/syt108

Frazer KA, Pachter L, Poliakov A, Rubin EM, et al. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32: W273-9. http://dx.doi.org/10.1093/nar/gkh458

Goulding SE, Olmstead RG, Morden CW and Wolfe KH (1996). Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* 252: 195-206. http://dx.doi.org/10.1007/BF02173220

Hahn C, Bachmann L and Chevreux B (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Res.* 41: e129. http://dx.doi.org/10.1093/nar/gkt371

Huang YY, Matzke AJM and Matzke M (2013). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (Cocos nucifera). *PLoS One* 8: e74736. http://dx.doi.org/10.1371/journal.pone.0074736

Jansen RK and Palmer JD (1987). A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Acad. Sci. USA* 84: 5818-5822. http://dx.doi.org/10.1073/pnas.84.16.5818

Katti MV, Ranjekar PK and Gupta VS (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.* 18: 1161-1167. http://dx.doi.org/10.1093/oxfordjournals.molbev.a003903

Kong W and Yang J (2016). The complete chloroplast genome sequence of *Morus mongolica* and a comparative analysis within the Fabidae clade. *Curr. Genet.* 62: 165-172. http://dx.doi.org/10.1007/s00294-015-0507-9

Liu Y, Huo N, Dong L, Wang Y, et al. (2013). Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS One* 8: e57533. http://dx.doi.org/10.1371/journal.pone.0057533

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. http://dx.doi.org/10.1007/s00294-007-0161-y

Millen RS, Olmstead RG, Adams KL, Palmer JD, 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. http://dx.doi.org/10.1105/tpc.13.3.645

Nazareno AG, Carlsen M and Lohmann LG (2015). Complete chloroplast genome of *Tanaecium tetragonolobum*: the first Bignoniaceae plastome. *PLoS One* 10: e0129930. http://dx.doi.org/10.1371/journal.pone.0129930

Nock CJ, Baten A and King GJ (2014). Complete chloroplast genome of *Macadamia integrifolia* confirms the position of the Gondwanan early-diverging eudicot family Proteaceae. *BMC Genomics* 15 (Suppl 9): S13. http://dx.doi.org/10.1186/1471-2164-15-S9-S13

Nguyen PA, Kim JS and Kim JH (2015). The complete chloroplast genome of colchicine plants (*Colchicum autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus. *Planta* 242: 223-237. http://dx.doi.org/10.1007/s00425-015-2303-7

Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K, et al. (2007). Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics* 23: 1-4. http://dx.doi.org/10.1093/bioinformatics/btl547

Ravi V, Khurana JP, Tyagi AK and Khurana P (2006). The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. *Tree Genet. Genomes* 3: 49-59. http://dx.doi.org/10.1007/s11295-006-0051-3

Ravi V, Khurana JP, Tyagi AK and Khurana P (2008). An update on chloroplast genome. *Plant Syst. Evol.* 271: 101-122. http://dx.doi.org/10.1007/s00606-007-0608-0

Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, et al. (2002). The plastid chromosome of Atropa belladonna and its comparison with that of Nicotiana tabacum: the role of RNA editing in generating divergence in the process of plant speciation. *Mol. Biol. Evol.* 19: 1602-1612. http://dx.doi.org/10.1093/oxfordjournals.molbev.a004222

Shaw J, Lickey EB, Schilling EE and 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. http://dx.doi.org/10.3732/ajb.94.3.275

Steane DA (2005). Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Res.* 12: 215-220. http://dx.doi.org/10.1093/dnares/dsi006

Su HJ, Hogenhout SA, Al-Sadi AM and Kuo CH (2014). Complete chloroplast genome sequence of Omani lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. *PLoS One* 9: e113049. http://dx.doi.org/10.1371/journal.pone.0113049

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729. http://dx.doi.org/10.1093/molbev/mst197

Walker JF, Zanis MJ and Emery NC (2014). Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am. J. Bot.* 101: 722-729. http://dx.doi.org/10.3732/ajb.1400049

Wu CS, Lai YT, Lin CP, Wang YN, et al. (2009). Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol. Phylogenet. Evol.* 52: 115-124. http://dx.doi.org/10.1016/j.ympev.2008.12.026

Yang Z and Yoder AD (1999). Estimation of the transition/transversion rate bias and species sampling. *J. Mol. Evol.* 48: 274-283. http://dx.doi.org/10.1007/PL00006470

Zerega NJ, Clement WL, Datwyler SL and Weiblen GD (2005). Biogeography and divergence times in the mulberry family (Moraceae). *Mol. Phylogenet. Evol.* 37: 402-416. http://dx.doi.org/10.1016/j.ympev.2005.07.004

Zhang H, Li C, Miao H and Xiong S (2013). Insights from the complete chloroplast genome into the evolution of *Sesamum indicum* L. *PLoS One* 8: e80508. http://dx.doi.org/10.1371/journal.pone.0080508

Zhang SD, Soltis DE, Yang Y, Li DZ, et al. (2011). Multi-gene analysis provides a well-supported phylogeny of Rosales. *Mol. Phylogenet. Evol.* 60: 21-28. http://dx.doi.org/10.1016/j.ympev.2011.04.008