The Chloroplast Genome Comparative Characteristic of Artificial Breeding Tree, a Case About Broussonetia Kazinoki × Broussonetia Papyifera

Wan zhang  
Central South University of Forestry and Technology

Guiyan Yang  
Northwest A&F University: Northwest Agriculture and Forestry University

Yunlin Zhao  
Central South University of Forestry and Technology

Zhenggang Xu (xuzhenggang@nwafu.edu.cn)  
Hunan Research Center of Engineering Technology for Utilization of Environmental and Resources Plant, Central South University of Forestry and Technology, 498 Shaoshan South Road, 410004 Changsha, Hunan Province, China;

Huimin Huang  
Central South University of Forestry and Technology

Jiakang Zhou  
Central South University of Forestry and Technology

Research article

Keywords: Broussonetia kazinoki, Broussonetia papyifera, Moraceae, phylogenetic relationship, maternal inheritance

DOI: https://doi.org/10.21203/rs.3.rs-87181/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: *Broussonetia kazinoki* × *Broussonetia papyifera* (ZJGS) is a hybrid species, which has a very complicated hybrid origin. Its excellent characteristics make it both ecological benefits and economically valuable.

Results: This study aims to further understand ZJGS and Moraceae taxa through the ZJGS chloroplast (cp) genome structure and the comparative with 12 closely related Moraceae species, especially the cross parent *B. kazinoki* and *B. papyifera*. The analyses show that ZJGS cp genome is 160,903 bp in length. Among the 13 Moraceae species, the cp genome length of seven *Broussonetia* species (ranges from 160,239 bp to 162,594 bp) is larger than that of six *Morus* species (ranges from 158,459 bp to 159,265 bp). However, the average GC content of *Broussonetia* species is lower than that of *Morus* species, which is 35.72% and 36.26%, respectively. Compared with Moraceae species, the ZJGS cp genome has a high degree of sequence similarity with *B. kazinoki* and *B. monoica*. In the comparison of repeated sequences, ZJGS and its maternal species *B. kazinoki* shows similar simple sequence repeats (SSRs) frequency. Among the 77 shared protein-coding genes (PCGs) in Moraceae species, the obvious positive selection of Ka/Ks ratios acted on *petD* and *rpl16* genes of *B. kazinoki* and *B. papyifera*, respectively, and the Ka/Ks ratio > 3. Phylogenetic analysis based on shared PCGs from 28 species shows that ZJGS is closely related to maternal *B. kazinoki*.

Conclusions: These findings provide data support for the origin of ZJGS hybridization, and provide genomic resources for future ZJGS resource development and molecular breeding.

Background

There are many hybridization phenomena in nature, which are regarded as effective methods to produce "positive species" [1–3]. Hybrid breeding is widely used in forestry production, specially in horticulture. At present, there are already many hybrid breeds used in agriculture and forestry production based on the heterosis. The hybrid aspen (*Populus tremula* × *Populus tremuloides*) has a higher yield than the parent species, and can be quickly regenerated from the root sucker [4–6]. Hybrids of *Abies* genus have the advantage of strong anti-pollution ability and resistance to pests and diseases [7, 8]. The heterosis is predicted by both positive gene contribution and environment. Hybridization is also an important way to produce gene diversity. Biologists have always focused on the role of hybridization in evolution. In-depth analyses of heterozygous genotypes showed that heterosis mainly came from the accumulation of a large number of rare superior alleles with positive dominance [9]. Lin et al. identified five candidate genes involved in high yield in super hybrid rice LYP9, among which heterozygous segments containing qSS7 and qHD8 showed superiority and contributed to heterosis [10]. In maize, the silent expression of Cell Number Regulator 1 gene (*CNRT*) can increase the size of plant and organ, and become a direct contributor to heterosis [11]. The nucleotide sequence of the plastid gene *rbcL* and the nuclear gene *PgiC* were used to study the reticulated evolution of the *Droopytes varia* complex in Japan and a haplotype not belong to any existing species was found. It was speculated that it came from an extinct species [12]. This indicates that hybrids may have higher fitness due to heterosis, and become an effective way to preserve their genetic diversity. However, above genetic information related to hybridization are mostly concentrated in nuclear genome. The chloroplast (cp) genome is the second largest genome after the nuclear genome, and has significant advantages in the study of phylogenetic evolution of species.

Chloroplast is a uniparentally inherited and stable organelle in green plants, who is often considered to be an informative resource for genetic diversity of plant [13–15]. The cp genome always consists of four conserved regions: a pair of inverted repeats (IRs), a small single-copy region (SSC) and a large single-copy region (LSC) [16–18]. Among these regions, the length of IRs regions is always diverse and determines the total length of cp genome [19, 20]. The cp genome contains a large number of functional genes, which can be divided into three categories: genes related to photosynthesis (Photosystem I: *psa*, Photosystem II: *psb*), genes related to gene expression (ribosomal RNA genes and transfer RNA genes, etc.), and other genes related to biosynthesis (ATP synthase gene, NADH dehydrogenase gene, etc.). It has been reported that *matK* and *ccsA* genes evolve faster than other gene groups [21, 22]. As the two largest genes, *ycf1* and *ycf2* located at the IR/SC junction and the IRs region, respectively, which proved to be useful for analyzing cp genome variation in higher plants [23]. As the development of sequencing technology, cp genome sequences are more and more available. The phylogenetic analyses based on the cp genome are achievable and revealed qualitative and valuable information to confirm the genetic relationship of different species [24, 25]. For instance, rice is an important cereal crop, so cultivating abundant varieties sounds vital for world’s population. The new cultivars are always evaluated by comparing to other ones on the affinity that according to the cp genome, and more than 20 species were categorized into 10 genome types [26]. To elucidate the evolution within neotropical-paleotropical bamboos, the cp genome of woody bamboo *Guadua angustifolia* was characterized and compared with other species, the results showed that the speciation events of extant species occurred during or after the Pliocene [27]. The complete cp genome sequence of *Morus cathayana* and *Morus multicaulis* were obtained and compared with other genus *Morus*, the results indicated that both natural selection and mutational bias have contributed to the codon bias [28].

Paper mulberry is a member of genus *Broussonetia* in Moraceae family. The tree displayed the features on fast-growing, easy breeding, resistant to pruning and widely distributing [29–31]. Vegetation survey in tailing’s area shows that the tree is an excellent native plant having a comprehensive absorption effect for various heavy metals [32]. Meanwhile, the leaves are an important source of pig feed and some chemical substances are also important resources for medicine [33]. Most importantly, as the name implies, because of its long fiber and ease of preparation, paper mulberry contributed to the invention of papermaking and then was also used in barkcloth [34]. Based on the multiple uses of paper mulberry, paper
mulberry has historically been an important economic tree species and even accompanied by human expansion [35]. At the same time, humans continue to select varieties of paper mulberry. Now, there is some confusion in the name of paper mulberry. Paper mulberry, *Broussonetia papyifera*, *B. kazinoki* × *B. papyifera* are all used to called the tree, while it may not the same tree [36]. In recent year, a hybrid variety for paper mulberry was bred [15, 37]. The breeding process is showed in Fig. 1 and in order to standardize the name, we called it as the hybrid paper mulberry (ZJGS) in the study. ZJGS is a hybrid between *B. kazinoki* and *B. papyifera*, whose maternal lineage is *B. kazinoki* and paternal is *B. papyifera*, respectively. With a long breeding process of more than ten years, the researchers first used modern breeding technology and then loaded the hybrid materials into the spacecraft and bred by space mutation. The mutant materials were optimized and cultivated after the spacecraft returned to the ground, and finally hybrid plants with excellent traits were obtained. However, in recent years, *B. kazinoki* was found to be probably a naturally occurring hybrid, with *B. monoica* as the female parent and *B. papyifera* as the male parent [15]. The complicated hybridization process has blurred the genetic background of paper mulberry and hindered the determination of the classification status of the tree and its utilization.

Due to the tree unclear genome is always large, a few forest tree genomes have been sequenced. Among the genetic resources of ZJGS, only unclear genome of *B. papyifera* was assembled [38]. Lacking of molecular genetic information about ZJGS may prevent optimal breeding of the tree and the exploration of adaptability mechanisms. In order to further determine the characteristics of the cp genome during the hybridization process, the cp genome of ZJGS was compared with 12 Moraceae species, including *B. papyifera* and *B. kazinoki*. The research is not only helpful for us to further determine the status of hybrid paper mulberry, but also for us to explore the role of chloroplasts in functional selection.

**Results**

**Chloroplast DNA (CpDNA) structure of ZJGS and comparative cp genomic analyses**

The cp genome of ZJGS is a circular molecule of 160,903 bp in length (Additional file 1: Table S1). 135 functional genes were identified in the ZJGS cp genome, including 81 PCGs, 46 tRNA genes, and eight rRNA genes (Table 1). Of the 81 PCGs, 77 unique genes were generated, of which four genes, *rps7*, *rpl23*, *ndhB*, and *ycf2*, are distributed in the IRs repeat regions (IRa/IRb). 15 tRNA genes and four rRNA genes are also repeated in the IRs regions. In the ZJGS cp genome, a total of eight different genes contain introns, even the longest intron is *ndhA* and has a length of 1,117 bp (Additional file 2: Table S2).
Table 1
List of genes encoded by ZJGS cp genome.

| Group of genes                        | Gene name                                                                 |
|---------------------------------------|---------------------------------------------------------------------------|
| Ribosomal RNA genes                   | rib16(*2), rib23(*2), rib4.5(*2), rib5(*2)                               |
| Transfer RNA genes                    | trnA-UGC(*4), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnF-MCAU, trnG-GCC, trnH-GUG, trnI-CAU(*2), trnI-GAU(*4), trnK-UUU, trnL-CAA(*2), trnL-UAU(*2), trnL-UAG, trnM-CAU(*2), trnN-GUX(*2), trnP-GGC, trnP-UGG, trnQ-UUG, trnR-ACG(*2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU(*2), trnT-UGU, trnV-GAC(*2), trnV-UAU(*2), trnW-CCA, trnY-GUA |
| Small subunit of ribosomal proteins   | rs11, rs12, rs14, rs15, rs16, rs18, rs19, rs2, rs3, rs4, rs7(*2), rs8     |
| Large subunit of ribosomal proteins  | rpl14, rpl16, rpl2(*2), rpl20, rpl23(*2), rpl32, rpl33, rpl36            |
| RNA polymerase subunits               | rpoA, rpoB, rpoC1(*2), rpoC2                                              |
| Large subunit of rubisco              | rbcL                                                                      |
| Photosystem I                         | psaA, psaB, psaC, psaL, psaJ                                              |
| Photosystem II                        | psbA, psbB, psbC, psbD, psbE, psbP1, psbH, psbl, psbJ, psbK, psbM, psbN, psbT, psbZ |
| NADH dehydrogenase                    | ndhA1, ndhB1(*2)1, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK  |
| Cytochrome b6/f complex               | petA, petB, petD, petG, petL, petN                                        |
| ATP synthase                          | atpA, atpB, atpE, atpF, atpH, atpI                                        |
| Conserved Open Reading Frames         | ycf1(*2), ycf2(*2), ycf3(*2), ycf4                                      |
| Maturase                              | matK                                                                      |
| Protease                              | clpP                                                                      |
| Subunit of Acetyl-CoA-carboxylase      | accD                                                                      |
| c-type cytochrome synthesis gene      | ccsA                                                                      |
| heme-binding protein                  | cemA                                                                      |

*2 means two gene copies and *4 means four gene copies; a gene containing a single intron; b gene containing two introns.

In the Moraceae species of comparison, the cp genome length of all seven Broussonetia (ranges from 160,239 bp to 162,594 bp) is higher than that of the six Morus (ranges from 158,459 bp to 159,265 bp), but the total GC content is relatively lower in the genus Broussonetia, the average value is 35.72%, which in Morus is 36.26%. (Table 2). From the perspective of cp genome length, the species that are closer to ZJGS are B. papyifera, B. kazinoki, B. monoica and B. kaempferi, and the GC content of B. kazinoki is most similar to ZJGS. The length ratio of the IRs regions in ZJGS cp genome is the lowest among all the comparative species, which is 16.04%. In addition, the ZJGS cp genome contains the largest number of genes, while the number of genes that direct synthetic proteins is the least, only 81.
CpDNA variation and conservation of ZJGS

In order to detect the global sequence variability of the ZJGS cp genome, the sequence of ZJGS was used as a reference to compare with the other Moraceae species (Fig. 2). It is found that the ZJGS cp genomic sequence has the highest similarity with B. kazinoki and B. monoica by comparison. Furthermore, the sequence of its paternal B. papyrifera exhibits higher divergence. These highly different regions include matK, rps16, atpF, rpoC2, rpoC1, ycf3, ctpP, ndhF, ccsA, ndhA and ycf1. It has become apparent to us that these genes are good sources for interspecies phylogenetic analysis and evolutionary studies. There are also some genes that are relatively conserved among these species, such as atpH, petN, psbE, rpl2, psbH, psbM and rps7. As a whole, the non-coding sequence of the Moraceae plants is more divergent than the coding sequence, and the sequence variation in the IRs regions is smaller than that in the LSC and SSC regions, which also shows that the IRs regions are more conservative than the LSC and SSC regions.

The BRIG image shows that the cp genome sequence of the Moraceae species is conserved (Fig. 3). We use the ZJGS cp genome as the reference sequence, which is distributed in the innermost circle of the concentric circle, and the outermost circle is the PCGs of ZJGS. Here, it can be clearly seen that the cp genome of the six innermost Broussonetia genus species is more closed than the other Morus species in the outer circle compared to the reference cp genome sequence, indicating that the cp genome sequence of Broussonetia species have high similarity with ZJGS. In addition, from the figure we can also intuitively see that the most conservative sequence is located in the IRs regions, and the GC content is also the highest among all species.

Through the previous study of sequence diversity, we know that the IRs regions are the most conserved regions in the cp genome. Most studies have shown that the contraction and expansion of the IRs regions determine the length of the cp genome, and this change directly affects the length and distribution of genes at the boundary of this region [23, 39–41]. Intuitive results show that the gene composition in IR/SC junction of ZJGS cp genome is somewhat different from that of several other species (Fig. 4). Among them, the most obvious is that on the left of the IRb/LSC region is the gene rpl23, while the remaining species are rpl2. Same as B. monoica, B. kaempferi, B. luzonica and B. kurzii, in IRa/SC region, ycf1 is replaced by trnN in ZJGS cp genome. However, the genetic constitution in the two parent species of ZJGS at the IR/SC junction is similar. In terms of gene length, the rpl2 gene (401 bp) in the ZJGS cp genome is significantly shortened in the IRa region and is more than three times shorter than other species (1,509 bp), which is related to the IRa boundary contraction.

Repeat sequence comparison

In this study, the SSRs of ZJGS and 12 Moraceae species cp genome were detected (Fig. 5). Similarly, all cp genomes of Moraceae plants have mononucleotides, dinucleotides, trinucleotides, tetranucleotides and pentanucleotides. While, hexanucleotide repeats are only detected in B. papyrifera and B. kurzii. Moreover, the number of SSRs in Broussonetia species is higher than that of Morus species, except for B. papyrifera. On the
whole, ZJGS and its maternal species *B. kazinoki* have the most similar performance among different types of SSRs (Table 3). Mainly reflected in the following SSRs: C/G, AAT/ATT, AAG/CTT, AAAT/ATTT, AGAT/ATCT, ATCG/ATCG, AAATT/AATTT, AATAG/ATTCT, AATAT/ATATT, and AAATGT/ACATTT, which shows the characteristics of maternal inheritance of ZJGS cpDNA. In order to obtain more genetic regulation information, the long repeat sequences of 13 Moraceae plants were further analyzed (Fig. 6A). Four types of repeats were detected in the cp genome of *M. atropurpurea* and *Broussonetia* species except for *B. luzonica*. And the number of palindromic repeat is the largest, then the type of forward repeat and reverse repeat, the least is the complementary repeat. The forward repeats, palindromic repeats and complementary repeats in the ZJGS cp genome all contain the most repeting elements. Repetitive sequences of 20–30 bp in length reach 50% or more (Fig. 6B-E). It can be clearly seen that the long repeat sequences of the ZJGS cp genome is dominant in this interval, and these repeat elements can be used as important genetic resources for the genetics and phylogenetic of the ZJGS population.
| SSR      | Species     | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | total |
|----------|-------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-------|
| A/T      | ZJGS        |   |   |   |   |   |   |   |   | 15| 18| 20| 5  | 3  | 6  | 1  | -   | 68   |
|          | B. papyifera|   |   |   |   |   |   |   |   | 27| 11| 2  | 6  | 2  | 4  | 1  | 1   | 54   |
|          | B. kazinoki |   |   |   |   |   |   |   |   |18 | 17|19  | 3  | 7  | 3  | -  | -   | 67   |
| C/G      | ZJGS        |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 0    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 1    |
| AG/CT    | ZJGS        |   |   |   |   |   |   |   |   |   | 20 | 1  | -  | -  | -  | -  | -   | 21   |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 21   |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 21   |
| AT/AT    | ZJGS        |   |   |   |   |   |   |   |   |   | 47 | 7  | 2  | 2  | -  | 1  | -   | 59   |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | 38 | 12 | 2  | 1  | -  | -   | 53   |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | 20 | 1  | -  | -  | -  | -  | -   | 21   |
| AAT/ATT  | ZJGS        |   |   |   |   |   |   |   |   |   | 4  | -  | -  | -  | -  | -  | -   | 4    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | 2  | -  | -  | -  | -  | -  | -   | 2    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | 4  | -  | -  | -  | -  | -  | -   | 4    |
| AAG/CTT  | ZJGS        |   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 0    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 0    |
| AAAT/ATT  | ZJGS        |   |   |   |   |   |   |   |   |   | 4  | -  | -  | -  | -  | -  | -   | 4    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | 2  | -  | -  | -  | -  | -  | -   | 6    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | 4  | -  | -  | -  | -  | -  | -   | 4    |
| AATT/AATT | ZJGS        |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 0    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 0    |
| AGAT/ATCT | ZJGS        |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 1    |
| ATCG/ATCG | ZJGS        |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 0    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
| AAAGG/CCTT | ZJGS        |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
| AAATT/AATTT | ZJGS        |   |   |   |   |   |   |   |   |   | 2  | -  | -  | -  | -  | -  | -   | 2    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | -  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. kazinoki |   |   |   |   |   |   |   |   |   | 2  | -  | -  | -  | -  | -  | -   | 2    |
| AATAG/ATTCT | ZJGS        |   |   |   |   |   |   |   |   |   | 1  | -  | -  | -  | -  | -  | -   | 1    |
|          | B. papyifera|   |   |   |   |   |   |   |   |   | 2  | -  | -  | -  | -  | -  | -   | 2    |
The form of sequence evolution: Ka/Ks ratio

Ka/Ks ratio represents the ratio between non-synonymous substitution (Ka) and synonymous substitution (Ks) at a particular site, which is used to infer the direction and magnitude of natural selection acting on PCGs. A ratio > 1 means positive selection; < 1 means purifying selection; and = 1 indicates no selection [42, 43]. According to previous study, except for the most genes with faster evolution rate, the frequency of synonymous nucleotide substitutions is higher than non-synonymous substitutions due to the purifying selection [40]. In this study, the cp genome sequence of ZJGS was used as reference, the Ka/Ks ratio of Moraceae species was presented (Fig. 7, Additional file 3: Table S3). The Ka/Ks ratio of such genes: atpH, pabC, petG, petN, psaC, psbE-psbI, psbL-psbT, rps7 and rps16 is 0, which reflects the strong purifying options and also indicates that these genes are highly conserved. On the contrary, most of the NADH genes (ndhA-ndhK) have higher Ka/Ks ratio, and in the comparison of all species, the Ka/Ks ratio of atpF, ndhB and rpoC1 is > 1, these three genes are strongly positively selected. Except for the above three genes, clpP, ndhA, ndhD, petB and petD genes are also positively selected compared with maternal B. kazinoki of ZJGS. While compared with paternal B. papyrifera, ndhA, ndhD, ndhF, petB, rpl16, rpoA and ycf1 genes were positively selected. Among them, petD and rpl16 have the largest Ka/Ks ratio, and both are > 3. These two genes are highly evolved and can be used for subsequent gene identification.

Phylogenetic analysis

In the phylogenetic tree based on 77 shared PCGs, ZJGS and its maternal B. kazinoki formed a single clade with high bootstrap support (99%) through three different methods (Fig. 8). All species are divided into four branches according to their evolutionary history. The first branch is the family of Moraceae, which was divided into three genera, namely, Broussonetia, Ficus and Morus. Here, Broussonetia and Ficus species are highly clustered with bootstrap values ≥ 99%. In addition, ZJGS, B. kazinoki and B. monoica formed a well-supported clade, as a natural hybrid, B. kazinoki also has the closest phylogenetic relationship with its maternal B. monoica. The second branch is Cannabaceae includes C. sativa and H. lupulus, which is closely to Moraceae. Studies have shown that both Cannabaceae and Moraceae belong to the Urticalean Rosid clade [44, 45]. The remaining two families, Rosaceae and Fagaceae, belong to Rosanae. In general, the classification results provide important information for the phylogenetic of ZJGS, as well as confirm the source of ZJGS hybridization.

Discussion

In this study, we determined and analyzed the cp genome of ZJGS (a hybrid plant) and compared it with the Moraceae species for structural comparison and sequence alignment. From the structure composition of the cp genome, it can be seen that the seven Broussonetia species and the six Morus species are grouped independently and each of them have similar compositions. Among the seven Broussonetia species, total GC content of ZJGS cp genome sequence is the closest to the maternal B. kazinoki, shows high species affinity [41], which is also reflected in the comparison of genome sequence diversity and SSRs. In addition, the sequence comparison analysis of ZJGS cp genome and other 12 Moraceae plants at the genomic level reflects the high sequence conservation of the IRs regions, so the IRs regions changes can be used as a marker of species evolution [46, 47]. At the boundary of the IRs regions, the changes in the length of the rpl2, rps19, tmnN and ycf1 reflect the contraction and expansion of the IRs regions. In the evolution of higher plants, all kinds of repetitive elements have a common feature, which tends to show "co-evolution", which plays a crucial role in the genome sequence diversity and gene rearrangement [48, 49]. Studies have shown that the number of repeats is related to the degree of genome rearrangement [50–52]. Here, long repeat sequences in ZJGS cp genome are the largest, and the number of 20–30 bp repeats may promote the rearrangement of ZJGS [53], these repeats can serve as good molecular markers for species evolution, and play an important role in the variation of cp genome sequence.

The breeding process of ZJGS is long and complicated. It has undergone many cross-breeding and space mutation to form excellent plants with stable inheritance. And its maternal B. kazinoki has experienced controversy [15, 54], therefore, it is necessary to compare ZJGS and its hybrid parent systematically. Unlike the previous study [36], the cp genome of ZJGS was analyzed in detail. Through comparison, it is found that the total
length and total GC content of ZJGS cp genome sequence (160,903 bp and 35.74%, respectively) are closest to its maternal *B. kazinoki* (160,841 bp and 35.73%, respectively), while its paternal *B. papyrifera* is 160,239 bp and 35.83%, respectively. This reflects a high degree of species affinity in the history of system evolution [41, 55]. Similarly, in the sequence diversity comparison, we also found that the cp genome sequence of ZJGS and maternal *B. kazinoki* remains highly consisent. This is related to sequence conservation and provides a basis for the species evolution of ZJGS [40, 56]. The construction of the phylogenetic tree shows us the most intuitive results, closely related species are grouped into the same branch. *And ZJGS is closest to maternal B. kazinoki*, which strongly supports the characteristics of cp maternal inheritance [57, 58].

As we have known before, when Ka/Ks > 1, the gene is strongly positive selection [42, 43]. We can screen some genes according to Ka/Ks ratio and then carry out functional studies, which have been commonly applied to the field of molecular evolution [59, 60]. In our research, three genes of *atpF, ndhB* and *rpoC1* show a complete positive selection. In previous studies, some genes have been reported with a faster evolution rate, including *ycf1, ycf2, accD, clpP, ndhA, rbcL, matK, ccsA* and *cema* [22, 61]. Compared with two cross parents, the same genes (*atpF, ndhA, ndhB, ndhD, petB* and *rpoC1*) with positive selection were screened out. Most of the NADH genes (*ndhA-ndhK*) have a higher Ka/Ks ratio than photosynthesis genes (Photosystem I: *psaA, psaB, psaC, psaL, psaJ*, Photosystem II: *psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbM, psbN, psbT, psbZ*), which is similar to the previous research [55, 62], shows that photosynthesis genes have strong purification options. NADH genes have higher activity during cellular senescence and oxidative stress [63, 64], and early hybrid selection experiments showed that most proteins synthesized by chloroplasts in the early stages of aging are NDH polypeptides [65]. It can be seen that these genes under positive selection belong to the advantageous genes of hybrid plants, and may be the key to speciation.

**Conclusions**

This study is based on the cp genome of ZJGS, a hybrid species, whose structure and composition are the same as most angiosperms, and the IRs regions are highly conserved. The detailed comparison of 13 closely related Moraceae plants revealed the evolutionary characteristics of ZJGS cp genome: (1) The comparison of total GC content, sequence diversity, and SSRs all indicate that the cp genome of ZJGS has significant maternal inheritance; (2) The high sequence variability of the LSC region, the SSC region, and the intergenic regions, and the contraction and expansion of the IRs regions bounder leads to the changes in gene length; (3) Compared with Morus species, the *Broussonetia* species have a longer cp genome length, higher total GC content and SSRs number. And among all the *Broussonetia* plants, ZJGS is closest to maternal *B. kazinoki*. (4) Ka/Ks ratio reveals positive selection of genes such as *atpF, ndhB* and *rpoC1*, and NADH genes. Phylogenetic analysis supporting the close relationship between ZJGS and its hybrid parents (the maternal *B. kazinoki* and the paternal *B. papyrifera*). And it strongly supports the genetic relationship between ZJGS and maternal *B. kazinoki*. Our results provide a basis for overcoming phylogenetic problems at the species level, and also provide important genomic resources for the functional utilization of ZJGS.

**Methods**

**Comparative analyses of cp genomes structure**

Our research group first reported the complete cp genome of ZJGS and the sequence was submitted to Genebank with the accession number of MF496038. The more experiment details can be got from the announcement [36]. At the same time, 12 Moraceae plants, including six *Broussonetia* species and six *Morus* species were selected to compare cp genomes structure with ZJGS. The cp genome information of 13 Moraceae plants was downloaded from NCBI database, and the GenBank accession numbers were listed in Additional file 4: Table S4. Firstly, the characteristics of cp genomes, such as GC content, length of IRs regions and number of genes were compared manually. Secondly, mVista in Shuffle-LAGAN mode and Blast Ring Image Generator (BRIG) were employed to compare cp genomes of the above selected 13 plants and ZJGS was set as the reference [66, 67]. To explore the evolutionary event of ZJGS, the details of IRs junction regions were detected and displayed, for the IRs regions were considered as the most conserved regions and these junctions were regarded as an index of cp genome evolution.

**Simple sequence repeats (SSRs) and long repeat sequences analysis**

SSRs, which is one of the most widely used molecular marker systems in plant genetic breeding and revealed more polymorphism [68–70]. The SSRs within the cp genome of each species were predicted by MISA with the parameters were set as: ≥ 10 for mono-nucleotides, ≥ 4 for di-nucleotides, ≥ 3 for tetra-nucleotides, penta-nucleotide and hexa-nucleotide. The number and details of SSRs were showed. Long repeat sequences, including forward match, reverse match, palindromic match and complementary match, within the cp genome, were identified by REPuter [71]. The related settings were showed as follows: (1) 90% or greater sequence identity; (2) a minimal repeat size of 20 bp. The graphs were all plotted using software Simplot 12.5.

**Molecular evolution analysis**

The Ka/Ks ratio is a good indicator of selective pressure at the sequence level, which can calculate selective pressure within protein coding regions. Here, the Ka/Ks ratio of 77 shared protein-coding genes (PCGs) of 12 Moraceae plants cp genomes were compared with ZJGS. Firstly, the multiple nucleotide sequences of the homologous genes that code for proteins were aligned by MEGA 7 [72], and then the Ka/Ks ratio was calculated using DnaSP v5 [73].
Phylogenetic analysis

In order to clarify the phylogenetic relationship of ZJGS, 27 related species were downloaded from the NCBI for phylogenetic analysis (the GenBank accession numbers were listed in Additional file 4: Table S4). The software MEGA 7 was used for multiple sequences alignment and evolutionary analysis [72]. Three methods were used to infer the evolutionary history and construct phylogenetic tree, including neighbor-joining method (NJ), maximum likelihood method (ML) and maximum parsimony method (MP). The evolutionary distances of NJ method were calculated using the Poisson correction method [74] and were in the units of the quantity of amino acid substitutions per site. In the ML method, initial tree was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using JTT model, and then selecting the topology with superior log likelihood value. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches.

Abbreviations

B. papyrifera
Broussonetia papyifera; bp:Base pairs; BRIG:Blast Ring Image Generator; cp:Chloroplast; CNR1:Cell Number Regulator 1 gene; DOGMA:Dual Organellar GenoMe Annotator; F. carica:Ficus carica; IRs:inverted repeats; Ka:the number of nonsynonymous substitutions per non-synonymous site; Ks:the number of synonymous substitutions per synonymous site; LSC:large single copy; ML:maximum likelihood method; MP:maximum parsimony method; M. atropurpurea:Morus atropurpurea; NJ:neighbor-joining method; NAD(P)H:Nicotinamide adenine dinucleotide (reduced); NDH:NAD(P)H dehydrogenase; PCGs:protein-coding genes; SSC:small single copy; rRNA:Ribosomal RNA; SSR:Simple sequence repeats; SPR:Subtree-Pruning-Regrafting ; tRNA:Transfer Ribonucleic Acid; ZJGS:Broussonetia kazinoki × Broussonetia papyifera.

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable.

Availability of data and materials
All datasets supporting the conclusions of this article are included within the article (and its Additional files). All sequences can be downloaded from NCBI Genebank (https://www.ncbi.nlm.nih.gov/) with the accession number which has been mentioned in the article.

Competing interests
All the authors declare that they have no conflict of interest.

Funding
This research was funded by National Natural Science Foundation of China (31700332), Hunan Provinicial Natural Science Foundation of China (2019JJ50027), Forestry science and technology project of Hunan Province (XKL201731), Science and Technology Project of Hunan Province (No: 2016TP2007, 2016TP1014, 2017TP2006), Postdoctoral Science Foundation project of Shaanxi Province (2016BSHEDZZ117) and Major Science and Technology Program of Hunan province (2017NK1014).

Authors' Contributions
YG and ZY are the lead investigator and conceived of the study. XZ and ZW wrote the paper and analyzed all the data. HH and ZJ is the lead investigator for the genome sequence. XZ and YG designed and fund for the current study. All authors have read and commented on the article.

Acknowledgements
Not applicable

References

1. Gowda M, Kling C, Würschum T, Liu W, Maurer HP, Hahn V, et al. Hybrid Breeding in Durum Wheat: Heterosis and Combining Ability. Crop Sci. 2010;50:2224–30.
2. Pucher A, Sy O, Sanogo MD, Angarawai II, Zangre R, Ouedraogo M, et al. Combining ability patterns among West African pearl millet landraces and prospects for pearl millet hybrid breeding. Field Crops Research. 2016;195:9–20.

3. Tompkins DM, Mitchell RA, Bryant DM. Hybridization increases measures of innate and cell-mediated immunity in an endangered bird species. J Anim Ecol. 2006;75:559–64.

4. Ryter L. Nutrient content in stems of hybrid aspen as affected by tree age and tree size, and nutrient removal with harvest. Biomass Bioenerg. 2002;23:13–25.

5. Ryter L, Stener L-G. Productivity and thinning effects in hybrid aspen (Populus tremula L. × P tremuloides Michx.) stands in southern Sweden. Forestry: An International Journal of Forest Research. 2005;78:285–95.

6. Liesebach M, Wuehlsch G, Muhs H.J. Aspen for short-rotation coppice plantations on agricultural sites in Germany: Effects of spacing and rotation time on growth and biomass production of aspen progenies. Forest Ecology Management. 1999;121:25–39.

7. Kobliha J, Stejskal J. Recent fir hybridization research in the light of Czech-American cooperation. JOURNAL OF FOREST SCIENCE. 2009;55:162–70.

8. Kobliha J, Stejskal J, Lstiburek M, Typta J, Jakubuv P. Testing of hybrid progenies and various species of genus Abies for forestry, decorating horticulture and christmas tree production. Acta Scientiarum Polonorum-Hortorum Cultus. 2013;12:85–94.

9. Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, et al. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nature communications. 2015;6:6258.

10. Lin T, Zhou C, Chen G, Yu J, Wu W, Ge Y, et al. Heterosis-associated genes confer high yield in super hybrid rice. Theor Appl Genet. 2020.

11. Guo M, Rupe MA, Dieter JA, Zou J, Spielbauer D, Duncan KE, et al. Cell Number Regulator1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis. Plant Cell. 2010;22:1057–73.

12. Hori K, Tono A, Fujimoto K, Kato J, Ebihara A, Watano Y, et al. Reticulate evolution in the apogamous Dryopteris varia complex (Dryopteridaceae, subg. Erythrovariae, sect. Variae) and its related sexual species in Japan. Journal of plant research. 2014;127:661–84.

13. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. Curr Biol. 2015;25:1375–80.

14. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Reviews in Genetics. 2004;5:123–35.

15. Won H. Test of the hybrid origin of Broussonetia × kazinoki (Moraceae) in Korea using molecular markers. Korean Journal of Plant Taxonomy. 2019;49:282–93.

16. Wang M, Cui L, Feng K, Deng P, Du X, Wan F, et al. Comparative Analysis of Asteraceae Chloroplast Genomes: Structural Organization, RNA Editing and Evolution. Plant Molecular Biology Report. 2015;33:1526–38.

17. Peng S, Yang G, Liu C, Yu Z, Zhai M. The complete chloroplast genome of the Juglans regia (Juglandales: Juglandaceae). Mitochondrial DNA. 2015;41:1–2.

18. Li C, Zhao Y, Xu Z, Yang G, Peng J, Peng X. Initial Characterization of the Chloroplast Genome of Vicia sepium, an Important Wild Resource Plant, and Related Inferences About Its Evolution. Frontiers in genetics. 2020;11:73.

19. Amaral-Zettler LA, Dragone NB, Schell J, Slikas B, Murphy LG, Morrall CE, et al. Comparative mitochondrial and chloroplast genomics of a genetically distinct form of Sargassum contributing to recent “Golden Tides” in the Western Atlantic. Ecology Evolution. 2017;7:516–25.

20. Ye N, Wang X, Li J, Bi C, Xu Y, Wu D, et al. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant Salix suchowensis. PeerJ. 2017;5:e3148.

21. Jose CC, Roberto A, Victoria I, Javier T, Manuel T, Joaquin D. A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus Citrus. Molecular Biology Evolution. 2015;32:2015–35.

22. Salamin N, Dong W, Xu C, Cheng T, Zhou S. Complete Chloroplast Genome of Sedum sarmentosum and Chloroplast Genome Evolution in Saxifragales. Plos One. 2013;8:e77965.

23. Zhang H, Li C, Miao H, Xiong S. Insights from the complete chloroplast genome into the evolution of Sesamum indicum L. Plos One. 2013;8:e80508.

24. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci USA. 2010;107:4623–8.

25. Nguyen PAT, Kim JS, Kim J. The complete chloroplast genome of colchicine plants (Colchicum autumnale L. and Gloriosa superba L.) and its application for identifying the genus. Planta. 2015;242:223–37.

26. Asaf S, Waqas M, Khan AL, Khan MA, Kang S, Imran QM, et al. The Complete Chloroplast Genome of Wild Rice (Oryza minuta) and Its Comparison to Related Species. Front Plant Sci. 2017;8:304.

27. Kuo C, Wu M, Lan S, Cai B, Chen S, Chen H, et al. The Complete Chloroplast Genome of Guadua angustifolia and Comparative Analyses of Neotropical-Paleotropical Bamboos. Plos One. 2015;10:e0143792.
28. Kong WQ, Yang JH. The complete chloroplast genome sequence of Morus cathayana and Morus multicaulis, and comparative analysis within genus Morus L. PeerJ. 2017;5:e3037.
29. Peng X, Wu Q, Teng L, Tang F, Pi Z, Shen S. Transcriptional regulation of the paper mulberry under cold stress as revealed by a comprehensive analysis of transcription factors. BMC plant biology. 2015;15:1–14.
30. Yan J, Wu PS, Du HZ, Zhang QE. First Report of Black Spot Caused by Colletotrichum gloeosporioides on Paper Mulberry in China. Plant Dis. 2011;95:880-.
31. Morgan EC, Overholt WA. Wildland Weeds: Paper Mulberry. Entomology & Nematology. 2016.
32. Zhao XP, Liu J, Xia X, Chu J, Wei Y, Shi S, et al. The evaluation of heavy metal accumulation and application of a comprehensive bioconcentration index for woody species on contaminated sites in Hunan, China. Environ Sci Pollut Res Int. 2014;21:5076–85.
33. Ryu HW, Curtis-Long MJ, Jung S, Jeong JY, Kim DS, Kang KY, et al. Anticholinesterase potential of flavonols from paper mulberry (Broussonetia papyrifera) and their kinetic studies. Food Chem. 2012;132:1244–50.
34. Moncada X, Payacán C, Arriaza F, Lobos S, Seelenfreund D, Seelenfreund A. DNA extraction and amplification from contemporary polynesian bark-cloth. Plos One. 2013;8:e56549.
35. Chang C-S, Liu H-L, Moncada X, Seelenfreund A, Seelenfreund D, Chung K-F. A holistic picture of Austronesian migrations revealed by phylogeography of Pacific paper mulberry. Proceedings of the National Academy of Sciences. 2015;112:13537-42.
36. Xu ZG, Yang GY, Dong M, Wu L, Zhang W, Zhao YL. The complete chloroplast genome of an economic and ecological plant, paper mulberry (Broussonetia kazinoki × Broussonetia papyrifera). Mitochondrial DNA Part B. 2018;3:28–9.
37. Xianjun P, Linhong T, Xiaoman W, Yucheng W, Shihua S. De Novo assembly of expressed transcripts and global transcriptomic analysis from seedlings of the paper mulberry (Broussonetia kazinoki × Broussonetia papyrifera). Plos One. 2014;9:e97487.
38. Peng XJ, Liu H, Chen PL, Tang F, Hu YM, Wang FF, et al. A Chromosome-Scale Genome Assembly of Paper Mulberry (Broussonetia papyrifera) Provides New Insights into Its Forage and Papermaking Usage. Molecular Plant. 2019:1–17.
39. Ravi V, Khurana JP, Tyagi AK, Khurana P. The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. Tree Genomics. 2006;3:49–59.
40. Ivanova Z, Sablok G, Daskalova E, Zahmanova G, Apostolova E, Yahubyans G, et al. Chloroplast genome analysis of resurrection tertiary relict Haberlea rhodopensis highlights genes important for desiccation stress response. Front Plant Sci. 2017;8:204.
41. Guo L, Guo S, Xu J, He L, Carlson JE, Hou X. Phylogenetic analysis based on chloroplast genome uncover evolutionary relationship of all the nine species and six cultivars of tree peony. Ind Crops Prod. 2020;153:112567.
42. Hurst LD. The K a/ K s ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18:486–7.
43. Yang ZH, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 2000;15:496–503.
44. Leme FM, Borella PH, Marinho CR, Teixeira SP. Expanding the laticifer knowledge in Cannabaceae: distribution, morphology, origin, and latex composition. Protoplasma. 2020:1–17.
45. Sytsma KJ, Morawetz J, Pires JC, Nepokroeff M, Conti E, Zhirua M, et al. Urticalean rosids: circumscription, rosid ancestry, and phylogenetics based on rbcL, trnL-F, and ndhF sequences. Am J Bot. 2002;89:1531–46.
46. Wang RJ, Cheng CL, Chang CC, Chaw SM. Dynamics and evolution of the inverted repeat/single copy junctions in the chloroplast genomes of monocots. Bmc Evolutionary Biology. 2008;8:1–14.
47. Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol. 2014;14:151.
48. Dover G, Coen E. Springcleaning ribosomal DNA: a model for multigene evolution? Nature. 1981;290:731–2.
49. Hanson RE, Zhao XP, Islam-Faridi MN, Paterson AH, Zwick MS, Crane CF, et al. Evolution of interspersed repetitive elements in Gossypium (Malvaceae). Am J Bot. 1998;85:1364–8.
50. Xue S, Shi T, Luo W, Ni X, Iqbal S, Ni Z, et al. Comparative analysis of the complete chloroplast genome among Prunus mume, P. armeniaca, and P. salicina. Horticulture Research. 2019;6:1–13.
51. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 2011;28:583–600.
52. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of Jasminum and Menodora (Oleaceae) are due to multiple, overlapping inversions. Mol Biol Evol. 2007;24:1161–80.
53. Peng XJ, Kuo W-H, Hsu Y-H, Li Y-H, Rubite RR, Xu W-B. Molecular recircumscription of Broussonetia (Moraceae) and the identity and taxonomic status of B. kaempferi var. australis. Botanical Studies. 2017;58:1–12.
55. Choi KS, Park S. The complete chloroplast genome sequence of Aster spathulifolius (Asteraceae); genomic features and relationship with Asteraceae. Gene. 2015;572:214–21.
56. Yin K, Zhang Y, Li Y, Du FK. Different Natural Selection Pressures on theatpFGene in Evergreen Sclerophyllous and Deciduous Oak Species: Evidence from Comparative Analysis of the Complete Chloroplast Genome of Quercus aquifolioides with Other Oak Species. Int J Mol Sci. 2018;19:1–15.
57. Corriveau JL, Coleman AW. Rapid screening method to detect potential biparental inheritance of plastid dna and results for over 200 angiosperm species. Am J Bot. 1988;75:1443–58.
58. Quan Z, Yang L, Sodmergen. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. Plant Cell Physiology. 2003;44:941–51.
59. Tae-Kun S, Hirohisa K. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. Syst Biol. 2008;57:367–77.
60. Yoshihiro M, Yukiko Y, Yasunari O, Koichiro T. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. Molecular Biology Evolution. 2002;19:2084–91.
61. Stephan G, Xi W, Herrmann RG, Uwe R, Klaus M, Georg H, et al. The complete nucleotide sequences of the 5 genetically distinct plastid genomes of Oenothera, subsection Oenothera: II. A microevolutionary view using bioinformatics and formal genetic data. Molecular Biology Evolution. 2008;25:2019–30.
62. Yang Y, Tao Z, Dong D, Jia Y, Li F, Zhao G. Comparative Analysis of the Complete Chloroplast Genomes of Five Quercus Species. Front Plant Sci. 2016;7:959.
63. Martin M, Casano LM, Sabater B. Identification of the Product of ndhA Gene as a Thylakoid Protein Synthesized in Response to Photooxidative Treatment. Plant Cell Physiology. 1996;37:293.
64. Peng L, Yamamoto H, Shikanai T. Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex II. BBA - Bioenergetics. 2011;1807:945–53.
65. Vera A, Tomás R, Martin M, Sabater B. Apparent expression of small single copy cpDNA region in senescent chloroplasts. Plant Sci. 1990;72:63–7.
66. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, et al. Glocal alignment: finding rearrangements during alignment. Bioinformatics. 2003;19:54–62.
67. Alikhan NF, Petty NK, Zakour NLB, Beatson SAA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. Bmc Genomics. 2011;12:402.
68. Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1996;1:215–22.
69. Suo z, Li W, Jin X, Zhang HA, New Nuclear DNA. Marker Revealing Both Microsatellite Variations and Single Nucleotide Polymorphic Loci: A Case Study on Classification of Cultivars in Lagerstroemia indica L. Journal of Microbial Biochemical Technology. 2016;8:266–71.
70. M U, P V, S M. Assessment of genetic diversity of rice (Oryza sativa) cultivars using ssr markers. African Journal of Biotechnology. 2014;13:3547–52.
71. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.
72. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.
73. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.
74. Zuckerkandl E, Pauling L. Evolutionary Divergence and Convergence in Proteins. Evolving Genes Proteins. 1965;97:97–166.

Figures
Figure 1
Schematic diagram of ZJGS breeding. Note: The *Broussonetia monoica* ♀ was named as *Broussonetia kaempferi* in flora of China. The *Broussonetia kazinoki* ♂ was also called as *B. × hanjiana* in Korea and *B. × kazinoki* in Japan.
Figure 2

Comparison of sequence similarity among 13 Moraceae plants chloroplast genomes using ZJGS as reference. Arrows indicate the direction of the gene, where exons are shown in blue, tRNA or rRNA genes are shown in green, and CNS represent non-coding sequences, indicated in red.
Figure 3

Similarity between the central reference genome (ZJGS) and other Moraceae plants cp genome sequences. The color represents a BLAST match for a specific percentage identity.
Figure 4

Comparison of ZJGS cp genome with the IR/SC borders of other 12 Moraceae species.
Figure 5

Simple sequence repeats (SSRs) in cp genomes of ZJGS and 12 Moraceae plants.
Figure 6

Long repeat sequences in the cp genomes of ZJGS and 12 Moraceae species. (A) The four repetitive sequences; (B) the number of the forward repeat by length; (C) the number of the reverse repeat by length; (D) the number of the palindromic repeat by length; (E) the number of the complementary repeat by length.
Figure 7

The Ka/Ks ratios of 77 shared PCGs in 12 Moraceae cp genomes were compared with ZJGS.
Figure 8

Phylogenetic tree was constructed for ZJGS and 27 related species using different methods based on the 77 shared PCGs. Each data set was used with three different methods: neighbor-joining (NJ), maximum likelihood (ML) and maximum parsimony (MP). The numbers next to the branches are the bootstrap values of NJ, ML, and MP, respectively. The black dot represents the position of ZJGS (MF496038) in the phylogenetic tree.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile4.xlsx
- Additionalfile3.xlsx
- Additionalfile2.xlsx
- Additionalfile1.xlsx