

Multiple evidence shining a light on species delimitation in the tree genus *Catalpa* Scop. (*Bignoniaceae*) of China

Ying Liu¹, Wanting Ge¹, Huiling Yun², Xiaochi Yu¹, Jiangtao Zhang³, Minggang Zhang⁴, Junhui Wang¹ and Wenjun Ma^{1*}

(1. State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Catalpa National Innovation Alliance, Beijing 100091, China; 2. Gansu Xiaolongshan Forestry Science Research Institute, Tianshui 741020, China; 3. Henan Academy of Forestry, Zhengzhou 450008, China; 4. Guizhou Academy of Forestry, Guiyang 550005, China)

Abstract: *Catalpa* genus are distributed in South America, North America and East Asia. The existing taxonomy and phylogenetic studies of *Catalpa* focus on the evolution among intercontinental, and there are few studies on the field resources in China. In the latest taxonomic revision, the species of *Catalpa* in China were revised to *Catalpa ovata* and *Catalpa bungei*, and the rest were synonyms. However, the observation of greater morphological diversity than previously captured by the current taxonomic framework. Delimiting species requires multiple evidence, this research used integrative classification to explore species delimitation of *Catalpa* in China. On the basis of field investigation, *Catalpa* plants were pre-classified into six hypothetical species according to morphological and ecological species concepts. Subsequently, the DNA samples of six taxa were collected, the SNPs obtained by genotyping-by-sequencing (GBS) and chloroplast genome data were used to explore the species delimitation of *Catalpa*. The results showed that Chinese *Catalpa* contained at least three clade, and supported that *C. fargesii* was an independent species, not a synonym of *C. bungei*. In addition, there are two mystical taxa in the genus *Catalpa* in China, *C. jinsiquiu* and *C. huangxin*, which belong to different clade, and are not synonyms of existing species. The study found that there is a common gene introgression in Chinese *Catalpa* species, and the role of this phenomenon in the formation of species needs further research. This study provides a paradigm for the integrative taxonomy approach of plants and lays a foundation for the research and protection of *Catalpa*.

Key words: Species delimitation, Integrative taxonomy, *Catalpa* in China, Phylogeny

1. INTRODUCTION

The *Catalpa* Scop. (*Bignoniaceae* Juss., tribe Catalpeae DC. ex Meisn.) is a small tertiary relic genus that consists of over eight known species with disjunct distributions in Asia, North America and South America. They are divided into *C. sect. Catalpa* and *C. sect. Macrocatalpa* Griseb. Section *Macrocatalpa* is composed of four species of tropical, evergreen trees found only in the Greater Antilles (Gentry, 1992; Olesn and Kirkbride, 2017), including *Catalpa purpurea* Griseb., *Catalpa macrocarpa* (A. Rich.) Ekman, *Catalpa longissima* (Jacq.) Dum. Cours, and *Catalpa brevipes* Urb. (Olsen and Kirkbride, 2017; Paclt, 1952). Section *Catalpa* is composed of over four species of deciduous tall trees and is disjunctly distributed between temperate eastern Asia and eastern North America, including *Catalpa bignonioides* Walter, *Catalpa speciosa* Teas, *Catalpa bungei* C. A. Mey, *Catalpa ovata* G. Don, etc. (Dong *et al*, 2022; Olsen and Kirkbride, 2017; Li, 2008; Paclt, 1952). The well-known Chinese species *Catalpa fargesii* Bureau, *C. fargesii* f. *duclouxii* (Don) Gilmour (= *C. duclouxii* Dode), *Catalpa tibetica* Forrest are considered to be synonyms of *C. bungei* in the latest classification revision (Olsen and Kirkbride, 2017). All the species of Section *Catalpa* are high-grade timber species and bears beautiful inflorescences, have very important ornamental and economic value. Among them, *C. bignonioides* and *C. speciosa* are mainly distributed in North America, the remaining species such as *C. bungei* and *C. ovata* are mainly distributed in China. The taxonomy and phylogeny of Section *Catalpa* have been explored for decades, but are still subjects of controversy (Paclt, 1952; Pan, 1991; Li, 2008; Olsen and Kirkbride, 2010; 2017).

In the Section *Catalpa*, the classification of *C. ovata*, *C. speciosa* and *C. bignonioides* has been clarified, while the taxonomic status of other *Catalpa* taxa in China is inconclusive. The inflorescences of *C. ovata*, *C. speciosa* and *C. bignonioides* are significantly different from those of other Chinese *Catalpa* taxa such as *C. bungei*, *C. fargesii* and *C. duclouxii* (Fig.1). The former are thyrsoid and the latter are corymbose, racemose or paniculate. Many Molecular analyses (Olmstead, 2013) suggest that *Catalpa* sect. *Macrocatappa* probably originated in drier zones of the Greater Antilles (South America), and suggested there are two dispersal events. A dispersal to China involving an ancestor from the North America sect. *Macrocatappa* gave rise to the clade of sect. *Catalpa*. *C. bungei* and *C. ovata* subsequently evolved in China, and a second dispersal event from China to North America (NA) gave rise to the clade of *C. bignonioides* and *C. speciosa* (Olsen and Kirkbride, 2017; Dong *et al.*, 2022). However, the classification and species delimitation of *Catalpa* plants in China are controversial. For instance, Olsen and Kirkbride (2017) based on the observation of the specimens expressed the opinion that pubescence type and density cannot be used to separate the members of this species into specific or infraspecific taxa (2017), they treated *C. fargesii* and *C. duclouxii* as synonyms of *C. bungei*, *C. tibetica* as synonyms of *C. bignonioides*. Though, the Flora of China accepted *C. fargesii* and treated *C. duclouxii* as a form of *C. fargesii* (Zhang and Thawatchai, 1998). Through the morphological investigation of different volume samples, different researchers divided Chinese *Catalpa* species into 2-21 species or types (Paclt, 1952; Pan *et al.*, 1991; Elias and Newcombe, 1979; Wu *et al.*, 2008; Yongming *et al.*, 2013). In the past 20 years, researchers have used different molecular markers such as SSR, ISSR, SRAP and ITS and chloroplast gene fragments to analyze the phylogenetic and genetic diversity of some *Catalpa* plants in China (Li, 2008; Xin *et al.*, 2011; Xiulan, 2013; Hao, 2013; Ma *et al.*, 2020; Dong *et al.*, 2022; Li *et al.*, 2022).

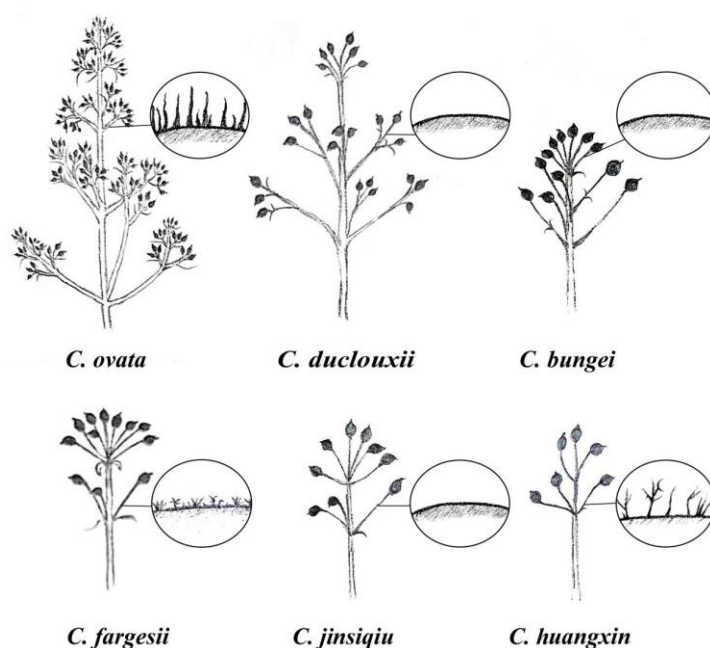


Fig. 1 Inflorescences and indumentum of six Chinese *Catalpa* taxa

As one of the most important taxonomic units in the biosphere, species is the basis of biological research (Coyne and Orr, 2004; Frankham *et al.*, 2012). Accurate delimitation of species is essential for the management and conservation of biodiversity (Agapow *et al.*, 2004; Hong, 2020). However, species is the most controversial definition or concept in biology (Mallet, 1995). Failure to recognize cryptic species diversity could lead to overestimates of population and range sizes, and thereby misinform the assessment of threat levels and compromise the prioritization of conservation resources (Bickford *et al.*, 2007). Conversely, over-splitting species might impair management through the exclusion of potentially valuable source stocks that could otherwise be used for relocations or genetic rescue efforts (Frankham *et al.*, 2012), furthermore, with limited funding for biological conservation, over-splitting dilutes and misdirects available resources (Agapow *et al.*, 2004; Moenga, 2018). Despite its

importance, species delimitation can be a challenging task (Fujita *et al.*, 2012; Sangster, 2014). The problem lies in the biological complexity and variability accompanying the speciation process, and how we define a species (De Queiroz, 2007; Schilthuizen, 2000). So far, there are nearly 30 (Wilkins, 2009) and even more definitions of species. The most commonly used species definitions are morphological species concept, biological species concept, ecological species concept, genetic species concept, evolutionary species concept and phylogenetic species concept, *et al.* Other definitions are derived from or synthesized from 2-3 species concepts (De Queiroz and Donoghue, 1988). The most basic assumption of species delimitation strategies is that “true species” will exhibit variation in traits of different biological dimensions (e.g. genetics, behaviour, morphology, ecology). This carries an additional level of complexity, as different traits along these biological dimensions could diverge at different times during the speciation process (De Queiroz, 2007). Meanwhile, recent speciation can produce complex species delimitation problems due to high levels of incomplete lineage sorting and introgression (Grummer *et al.*, 2018; Olave *et al.*, 2018). Furthermore, variation in traits may be the result of local adaptation or phenotypic plasticity and not be indicative of the speciation process (Frugone *et al.*, 2019; 2021). Indeed, previous research has emphasised that use of only a single line of evidence to delimit species may result in species delimitation errors (Edwards and Knowles, 2014). The discussion on the best methodology or source of information to delimit species had an important turn following De Queiroz’s (De Queiroz, 2007) proposal of a unified species concept, and Dayrat’s (Dayrat, 2005) and Will’s (Will *et al.*, 2005) proposals of integrative taxonomy. The field of taxonomy is moving toward a more integrative approach that takes account of different lines of evidence to delimit species. The rapid advance of genomic data acquisition, coupled with better understanding of speciation processes, has made species delimitation much more quantitative (Fujita *et al.*, 2012). Integrative taxonomy approach has been widely used in the delimitation of plant species (Wang *et al.*, 2022; Wu *et al.*, 2022; Clark, 2023). Nonetheless, reconciling differences between morphology and genetics can still be challenging in situations where they disagree. For example, species with deep divergences can remain morphologically indistinguishable due to static selective pressures or niche conservatism, resulting in morphologically cryptic species (Boumans *et al.*, 2007; Murphy *et al.*, 2016; Padial and la Riva, 2009; Welton *et al.*, 2013). At the other extreme, populations of the same species can display substantial morphological divergence due to local adaptation, phenotypic plasticity or other evolutionary mechanisms (Keogh *et al.*, 2005; Weitschat, 2016). Figuring out how to resolve discrepancies between genetics and morphology is difficult but necessary for taxonomic progress. Thus, integrative methods and comprehensive sampling are needed to relieve the problems with the delimitation of Chinese *Catalpa* species.

Based on the field investigation of wild *Catalpa* plants in most regions of China over the years by our research team. Our work suggested the *Catalpa* plants in China were divided into six taxa according to their stable morphological characteristics (Fig.1, 2, crown type, bark, flower, leaf, trichome type, fruit, *etc.*). These include *C. ovata* (most of China), *C. bungei* (northern, central China), *C. duclouxii* (southwestern China), *C. fargesii* (northwestern, central and eastern China), whose taxonomic status are widely discussed, and *C. jinsiquiu* (northern, central and eastern China), *C. huangxin* (southwestern China), which are not discussed. These six taxon occupy different distribution areas, different taxon have both sympatric and exotic distribution. The anthers of *C. ovata*, *C. bungei*, *C. duclouxii*, *C. fargesii* are all white, and the wood color are white to gray. In contrast, the anthers of *C. jinsiquiu* and *C. huangxin* are yellow, and the wood color are golden yellow, making them more popular in the market. Here, we used multiple evidence to evaluate the taxonomic treatment of Chinese *Catalpa*. The results showed that Chinese *Catalpa* contained at least three clades, and supported that *C. fargesii* was an independent species, not a synonym of *C. bungei*. The mysterious taxa *C. jinsiquiu* and *C. huangxin* belong to different evolutionary clades.



Fig. 2 Floral organs of *C. bungei* and *C. jinsiqui*. A is *C. bungei*, and B is *C. jinsiqui*. The differences in traits such as flower color, stamen spots, stripes in the floral tube and so on.

2. MATERIALS AND METHODS

2.1 Sampling, GBS library preparation and sequencing

We sampled six hypothetical species (as shown in Fig.3, Table S1), the samples were all mature trees grown naturally in the field. For DNA extraction, fresh leaves tissue were collected, immediately frozen in liquid nitrogen, and stored at -80°C , and a total of 74 samples were collected (Table S1). The complete chloroplast genome sequences of the six taxa measured in our previous research were deposited in GenBank with accession numbers OL628864 to OL628869 (Li, 2022). The DNA of the fresh leaves of 74 was extracted for GBS using a modified cetyltrimethylammonium bromide (CTAB) extraction method (Tel-Zur *et al.*, 1999). The genomic DNA was randomly disrupted into short DNA fragments by PstI/MspI enzyme, and then the flat ends were repaired. Then, the dA tail was connected at both ends of the DNA fragment and the sequencing adaptor was connected. The DNA fragment of the adapter was purified by AMPure XP magnetic beads, and the fragment in the range of 300-400 bp was selected for PCR amplification. The constructed library was purified, checked, and the DNA libraries were sequenced on the Illumina Hiseq X10 sequencing platform by Genedenovo Biotechnology Co., Ltd (Guangzhou, China).

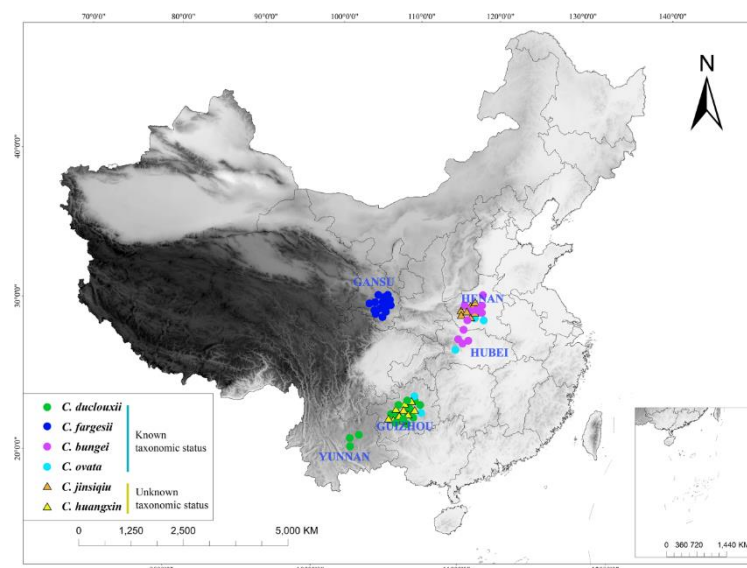


Fig. 3 Geographic distribution of 74 samples of six *Catalpa* hypothetical species in China

2.2 SNP calling

Quality control of the raw data was performed with the FastQC 0.11.9 application (Brown *et al.*, 2017). Adapter sequences and abnormal nucleotide bases at the 50 terminus were removed from the raw sequencing reads. Additionally, the low quality ends of the reads (sequence quality $<Q\ 20$) were trimmed and we removed reads that contained 10% Ns (undefined bases) and sequence lengths less than 25 bp after trimming. Without a reference sequence, SNP calling for each sample was performed using the Stacks V. 2.52 (Catchen *et al.*, 2013) pipeline to

build loci (ustacks), create a catalog of loci (cstacks), match samples back to the catalog (sstacks), transpose the data (tsv2bam), add paired-end reads to the analysis, call genotypes, and perform population genomics analysis (Catchen *et al.*, 2013). We used the *populations* programme in Stacks and Vcftools V. 4.1 (Danecek *et al.*, 2011) to filter loci. The obtained SNP data were filtered in depth with the following parameters: -min DP 2, -max-missing 0.7, -maf 0.05, and a distance between variants of 300 bp, and the remaining markers 3867 SNPs.

2.3 Phylogenetic analysis

We construct phylogenetic trees of the samples in six *Catalpa* hypothetical species of China using *C. ovata* as an outgroup based on the nuclear SNPs data and the chloroplast genomic sequences. For the SNPs data, phylogenetic trees were inferred using the neighbor-joining (NJ) method implemented in MEGA V.7.0 in Linux (Kumar *et al.*, 2016), and the bootstrap was set to 1,000 times to assess the branch reliability. For the chloroplast genome data of six *Catalpa* hypothetical species (OL628864 to OL628869) of China and *C. speciosa* of NA (MT319818), phylogenetic trees were inferred using maximum likelihood (ML) methods. The ML analysis was performed using RAxML-HPG BlackBox v.8.1.24 (Stamatakis, 2014) with 1000 bootstrap, and the (GTR)+ G+ I model was used in ML phylogenetic analyses.

2.4 Population genetic structure analysis

The composition of population structure in the 74 accessions belonging to the six *Catalpa* hypothetical species was determined by using ADMIXTURE based on filtered SNPs. ADMIXTURE adopts the likelihood model embedded in STRUCTURE and runs considerably faster than other tools to estimate ancestry (Alexander *et al.*, 2009). First, the vcf format datasets were converted into Plink ped format using the --plink option in software VCFtools. Then, Plink program (Purcell *et al.*, 2007) was used to convert and filter ped format into bed format. To calculate individual ancestries, we performed a maximum likelihood (ML) model-based clustering analysis using ADMIXTURE V 1.3.0 (Alexander *et al.*, 2009) based on SNPs that are in HWE at $P > 0.05$. We tested K values from 1 to 20 and used the lowest cross-validation error (CV error) estimates to determine the optimal K value across 10 independent runs (Alexander *et al.*, 2009; Evanno *et al.*, 2005).

2.5 Principal component analysis

Principal component analysis (PCA) is a pure mathematics method that can select a small number of important variables from multiple related variables. PCA of SNPs obtained from the 74 samples was performed using GCTA v1.92.4 (Yang *et al.*, 2011) software for the clustering of the main components. Then, based on the out.eigenvec file, the results is represented with R.

2.6 Discriminant analysis of principal components

In order to complement and verification the previous analysis results, We performed a discriminant analysis of principal components (DAPC) to determine the number of discrete clusters implemented in the R package Adegenet package v.2.1.1 (Jombart, 2008; Jombart *et al.*, 2010; Jombart and Ahmed, 2011) using R version 3.4.1. Compared with other widely used programs such as STRUCTURE, this analysis does not assume Hardy-Weinberg or linkage equilibrium, and determines the number of principal components (PCs) to be retained through hierarchical cross-validation. We employed the find.clusters function, using a sequential k-means clustering algorithm to determine the optimal number of groups, and the Bayesian Information Criterion (BIC) to select the best model (Jombart, 2008; Jombart and Ahmed, 2011). The number of retained PCs was chosen using the optim.a.score command to prevent overfitting (Feng, 2020).

2.7 ABBA-ABBA analysis

In the previous analysis, we found that there were suspected hybrid individuals in the sample, and we removed the SNPs of these individuals (A11, A01, A05, A09, A10, A14, A07, B04, B11, D03, D10, D12, D13, HX046), remaining 60 samples represent six hypothetical species to analysis the genetic introgression among species. We employed the ABBA-BABA test (also known as the D -statistics, Green *et al.*, 2010; Durand *et al.*, 2011) to test for the possibility of admixture among the six *Catalpa* hypothetical species. Dsuite has a fast C++ implementation, allowing genome-scale calculations of the D and related statistics across all combinations of tens or hundreds of populations or species directly from a VCF file (Malinsky *et al.*, 2020). The statistics applied to biallelic SNPs across quartets (populations, taxa, species) on a typical tree topology (((P1, P2), P3), O), where O is the outgroup,

served to define the ancestral allele (A), while the derived alleles were denoted by (B). Site patterns were ordered as follows: BBAA (P1 and P2 sharing the derived allele), ABBA (P2 and P3 sharing the derived allele) and BABA (P1 and P3 sharing the derived alleles). Under the null model (no gene flow), ABBA and BABA patterns were expected to occur with equal frequencies, while a significant deviation from that suggested possible introgression between P3 and either P1 or P2 (Malinsky *et al.*, 2020; Wang *et al.*, 2020; Tonatiuh *et al.*, 2021). D value was expected to be 0 if no gene flow had occurred and 1 if introgression existed between P2 and P3. We used Dsuite (Malinsky *et al.*, 2020) to estimate ABBA-BABA statistic to detect gene introgression between the six taxon. We inferred the D -statistic through all combinations of trios (samples trios) inside each recovered major clade as implemented in Dtrios (Malinsky *et al.*, 2020). To assess whether D differed significantly from zero (null model), Dtrios calculated Z-scores and associated p -values.

2.8 Treemix analysis

Further analysis were carried out to infer the genetic relationship, genetic differentiation and genetic introgression among the hypothetical species. We used the Stacks to assess the weighted pairwise F_{ST} based on 60 samples of six hypothetical species. The Treemix software uses genome-wide allele frequency data to infer patterns of differentiation and mixing of populations. TreeMix analysis is also used to analyze gene introgression between species when species are considered as highly genetically differentiated populations (Gagnaire *et al.*, 2018). We determined genetic introgression between species using TREEMIX v1.13 (Pickrell *et al.*, 2012). Migration events (–m) were sequentially added to the tree.

3. RESULTS

3.1 Phylogenetic relationship

The phylogenetic tree was constructed using the NJ method (Fig.4) based on the nuclear SNPs. Through the NJ tree, it can be intuitively found that the samples of the six Chinese *Catalpa* hypothetical species were clustered together. Among them, five *C. ovata* samples were clustered into one clade and the genetic distance was far from other groups, the tree strongly support the monophyly of *C. ovata*. The other Chinese *Catalpa* complex (containing *C. bungei*, *C. fargesii*, *C. duclouxii*, *C. jinsiquiu* and *C. huangxin*) forms a clade, and the tree strength supports its monophyly. The samples of *C. fargesii* and *C. jinsiquiu* (sample B and D) form a robust clade, and in this clade, most of *C. fargesii* (B11 to B13) are form a clade, and they are sisters related with ten *C. jinsiquiu* individuals, while B04 should be the sample of *C. jinsiquiu* that is regarded as *C. fargesii*, gathered together with the sample of *C. jinsiquiu*; The samples of *C. bungei*, *C. duclouxii* and *C. huangxin* form another robust clade. *C. duclouxii* and *C. huangxin* form a clade, and the sample of *C. huangxin* form a robust clade, and the genetic distance with *C. duclouxii* (DQ269 to DQ009) is far. Phylogenetic tree based on chloroplast genome also strongly support the monophyly of Chinese *Catalpa* complex (Fig.5B). The difference is that in the nuclear gene phylogenetic tree, *C. fargesii* and *C. jinsiquiu* formed a monophyletic clade, but it is not in the chloroplast genome phylogenetic tree. Although both *C. huangxin* and *C. jinsiquiu* have the characteristics of yellow heartwood, yellow pollen, and thick flower tube stripes, phylogenetic trees based on SNPs of the nuclear genome and chloroplast genome revealed that the *C. huangxin* and *C. jinsiquiu* populations were recognized as two different taxon.

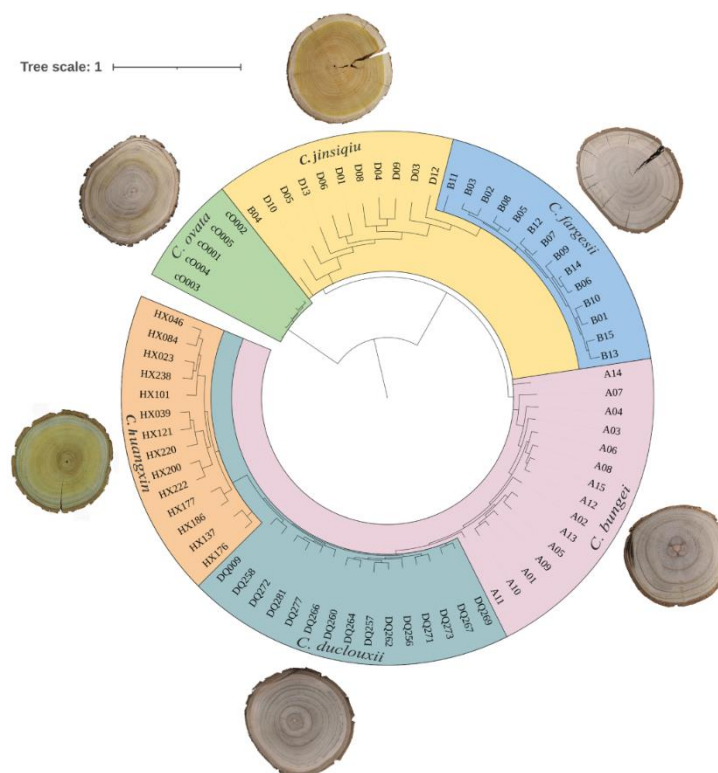


Fig. 4 Phylogenetic relationships of six *Catalpa* hypothetical species of China inferred from Neighbor-Joining (NJ) analyses based on SNPs of the nuclear genome. The periphery of the tree represents the taxonomic group, and the interior represents the individual number. The outer section of the tree is a cross-section of the mature trees of each species, and the color of the heartwood of *C. jinsiquiu* and *C. huangxin* are yellow.

3.2 Population genetic structure

The genetic structure of the population showed that the value of *CV error* was small and the classification was better when the number of clusters was 4, 5, 6 and 7, with $K = 5$ minimizing the *CV error* (Fig.S1). When $K = 4$ (Fig.5A), *C. ovata*, *C. huangxin*, *C. fargesii* were all single ancestral clusters, *C. bungei* and *C. duclouxii* belong to a cluster from an ancestral. The samples of *C. jinsiquiu* are clustered together as a multi-ancestor population. When $K = 5$, the mixed population of *C. bungei* and *C. duclouxii* was divided into two different clusters, and six *C. bungei* samples contained almost 50% of the ancestral genetic information of *C. duclouxii*. *C. jinsiquiu* is a multi-ancestor cluster, which contains most of the genetic information of the ancestors of *C. fargesii* and a small part of the genetic information of the ancestors of *C. huangxin* and some individuals contain a small amount of genetic information of the ancestors of *C. bungei*.

3.3 Principal component analysis

Principal component analysis (PCA) is a pure mathematical method that can select a small number of variables with high weights from multiple related variables. The results of principal component analysis are shown in Fig.5C. The five samples of *C. ovata* were clustered together, which were far away from the samples of other clusters. Except for two samples, most of the samples of *C. bungei* and *C. duclouxii* were clustered together. Except for two samples, the samples of *C. fargesii* were clustered together. Most of samples of *C. jinsiquiu* were independently clustered together, which was relatively close to the population of *C. fargesii*. The samples of *C. huangxin* were clustered together and separated from other groups. Principal component analysis including all six taxa revealed five separate, tightly clustered groups according to species boundaries (Fig.5C), and the genetic distance between *C. bungei* and *C. duclouxii* was relatively close. Most of the sample of groups were clustered together, and a small number of samples were in the middle of each group, suggesting the existence of hybridization or introgression.

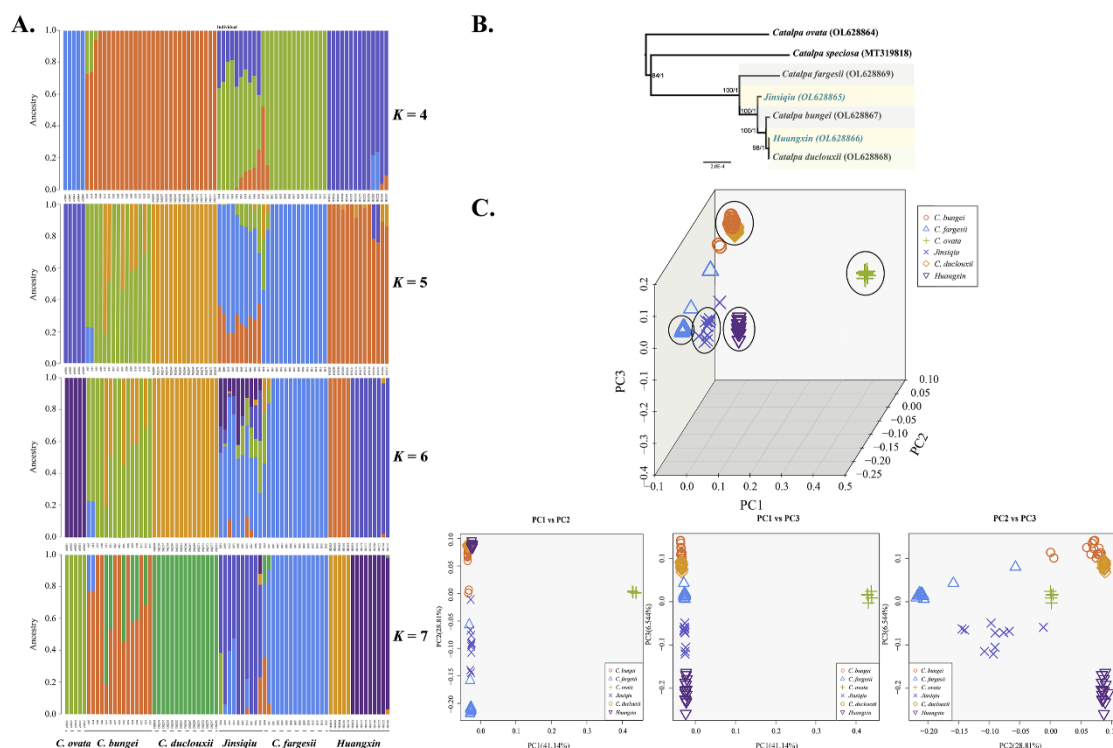


Fig. 5 Genetic structure and phylogenetic tree of 6 hypothetical species

A. Population Genetic structure based on SNPs data of 74 samples of 6 hypothetical species ($K=4$ to 7).

B. Maximum Likelihood (ML) analyses based on the whole chloroplast genome of seven species. The taxon marked by dark blue italic bold fonts are not found in the existing classification research, but they are significantly different from the existing taxa in morphology and distribution. **C. Principal component analysis of 74 samples of six Chinese *Catalpa* hypothetical species based on SNPs.**

3.4 Discriminate analysis of principal components (DAPC)

The results showed that the Bayesian Information Criterion (BIC) values (Fig.6A) were lower when the clustering number was five and six, indicating that 74 samples were more reasonable when clustered into five or six taxon, and the best cluster number is five. The results of discriminant analysis in Fig.6 showed that 74 samples were divided into 5 clusters, *C. ovata*, *C. huangxin*, *C. fargesii*, *C. jinsiqui*, *C. bungei* and *C. duclouxii* mixed groups respectively. Fig.6 C shows the clustering of DAPC of all samples. When the PCA axis is 75, the cumulative variance is close to 100%, which proves that the results are reliable. The clustering results showed that the samples of *C. huangxin* were clustered into one group (yellow dots), the samples of *C. fargesii* were clustered into one group (pink dots), the samples of *C. ovata* were clustered into one group (green dots), the samples of *C. jinsiqui* were clustered into one group (orange dots), and the samples of *C. bungei* and *C. duclouxii* were clustered into one group (blue-purple dots). DAPC shows that the *Huang* and *C. jinsiqui* were recognized as two different species, and also shows the possible genetic relationship between species.

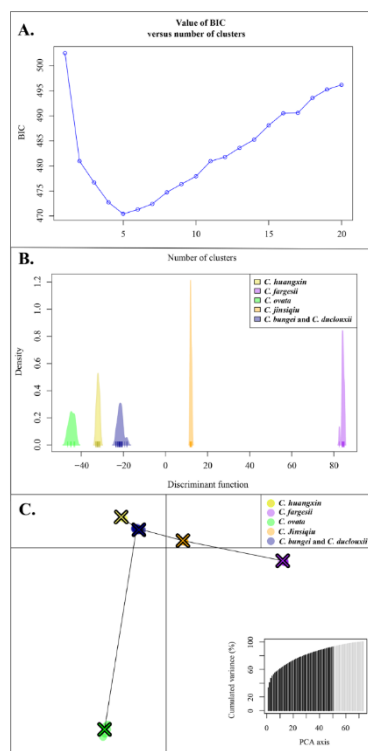


Fig. 6 Discriminate analysis of principal components (DAPC) results for the 74 samples of six Chinese *Catalpa* hypothetical species based on SNPs by adegenet.

3.5 ABBA-ABAB statistic

Population genetic structure analysis is limited in the inference of demographic history and gene flow (Lawson, 2018). The ABBA-BABA model (*Patterson's D* statistic) provides a new way to analyze interspecies introgression. This method was originally proposed by Green *et al.* (2010) to detect introgression between modern human and Neanderthal populations (Durand *et al.*, 2011; Green *et al.*, 2010). We removed the suspected hybrid individuals inferred from the above analysis. Based on the SNP data of 60 samples, the genetic introgression of six *Catalpa* taxa was analyzed by Dsuite software. In this research, the *D*-statistic uses Z-score (> 3) and *p*-value to reflect the significance of introgression. The result (Table 1) shows that there are five species combinations with significant genetic introgression in 20 ternary combinations composed of six taxon. The results showed that there were significant genetic introgression between *C. bungei* and *C. fargesii*, *C. bungei* and *C. jinsiquiu*, *C. fargesii* and *C. jinsiquiu*, *C. jinsiquiu* and *C. huangxin*.

Table 1 Species with significant gene introgression in ABBA-BABA statistic

Species 2	Species 3	<i>D</i> -statistic	Z-score	<i>p</i> -value
<i>C. bungei</i>	<i>C. fargesii</i>	0.355758	4.35591	1.33E-05
		0.45399	5.40993	6.31E-08
<i>C. bungei</i>	<i>C. jinsiquiu</i>	0.28094	3.71361	0.00020432
<i>C. fargesii</i>	<i>C. jinsiquiu</i>	0.372269	5.67318	1.40E-08
<i>C. jinsiquiu</i>	<i>C. huangxin</i>	0.379065	5.53026	3.20E-08

3.6 Treemix analysis

Treemix is able to estimate population splits and migration by first building a tree model of the populations and then subsequently adding migration events (or edges) between populations that poorly fit the tree model (Pickrell *et al.*, 2012), which has been used to infer gene flow between populations of many species including plants, reptiles, mammals, and numerous others. In order to further explore the introgression of different taxa, we further

filtered the SNPs database (-min DP 2, -max-missing 0.8, -maf 0.05) and retain 2212 SNPs performed Treemix analysis. We reconstructed the ML tree of the six taxa using TreeMix to address the population history relationships and identify pairs of populations related to each other independent of that captured by the tree (Fig.7). The ML tree without migration events inferred from TreeMix analysis divided the six taxa into three clusters (Fig.S2), was similar to the population structuring patterns identified from phylogenetic tree analysis, PCA, and genetic structure analysis. When potential migration edges (i.e., migration events between branches) were added to the ML tree, strong migration events were detected between the clusters (Fig.7). The calculated m value is from 0 to 7. When $m=2$, the model fits the covariance between the actual populations well. The ML Tree shows that there are two genetic introgression events, which are the genetic migration of the ancestor of *C. huangxin* to *C. jinsiquiu* and the genetic migration of *C. fargesii* to *C. bungei*. The results of R package OptM (Fitak, 2021) analysis showed that the migration edges were the optimal number of migration edges when the Δm value was the largest. The results showed that when Δm was one, it was the largest (Fig.S3). The results showed that there was introgression from *C. huangxin* to *C. jinsiquiu*, which may be related to the formation of *C. jinsiquiu*, which was consistent with the results of ABBA-BABA analysis (Table 1).

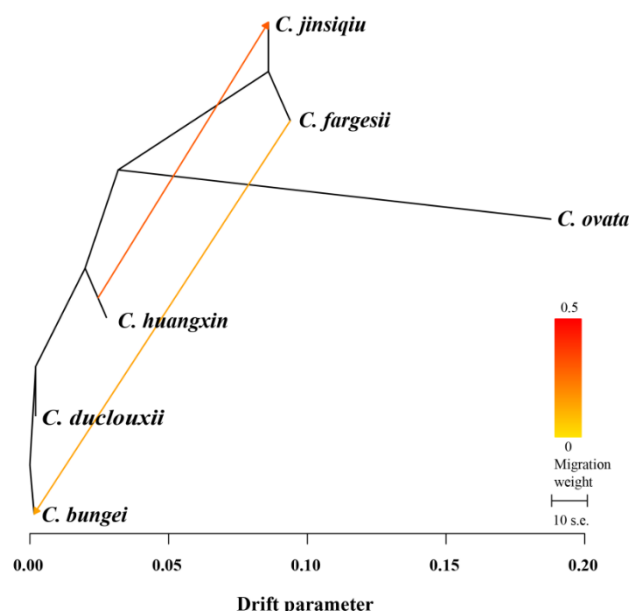


Fig. 7 ML tree of six taxon with inferred migration edges

4. DISCUSSION

4.1 Insight into species concepts and *Catalpa* species delimitation in China

The most common and widely accepted species concept are the biological species concept (BSC, Mayr, 1942, 1963) and phylogenetic species concept (PSC, Eldredge and Cracraft, 1980). According to the BSC, species are defined as a group of individuals that can mate with each other or potentially mate with each other, and there is reproductive isolation between other similar populations. According to this species concept, interspecific and intraspecific variation of sexual reproduction can be effectively distinguished, but it is difficult to apply in practice, because many species do not have strict reproductive isolation. With the development of molecular technology, genetic markers have been gradually applied to species division, and the concept of phylogenetic species has emerged. This species concept holds that species are the smallest monophyletic group with a common ancestor, emphasizing the history of biological phylogeny and is accepted by many taxonomists. However, as more and more markers and even genomic-level genetic information are used for phylogenetic reconstruction, it has been found that gene trees based on limited gene markers may not represent real species relationships. However, in practical applications, this concept of species also exists: 1. It is difficult to find a fixed genetic difference to distinguish species and populations; 2. Due to the incomplete lineage sorting (ILS) based on limited genetic markers, the gene tree obtained may be different from the actual population divergence order; 3. Due to selection pressure and random drift, the same evolutionary history may leave different traces on different genes, especially genes with different

genetic patterns. The conflict between phenotype and heredity, the conflict between different genetic markers and the objective existence of hybridization phenomenon, the difficulty and complexity of traditional species identification are increased (Petit and Excoffier, 2009; Carstens, 2013). With the development of molecular technology and the continuous optimization of analytical methods, the integrative taxonomy approach has emerged as the times require. It refers to a method of species division that comprehensively considers evidence from multiple sources such as molecules, morphology, ecology, and geographical distribution (Padial *et al.*, 2010). This method combines evidence from different sources (Schlick-Steiner *et al.*, 2010), by making objective species assumptions in a non-prior situation, and using a variety of data to verify the rationality of the assumptions, effectively increasing the rigor of species division. It is considered to be the most reliable method for identifying species (Puillandre, 2012). At present, it is widely used in animals and is gradually applied to plants.

We have studied the *Catalpa* plants in China for many years. It is well known that the confusion of Chinese *Catalpa* species classification has brought negative impacts on basic and applied research. Through the observation of the morphology, phenology, geographical distribution and common garden planting experiment of the *Catalpa* germplasm in China, combined with the previous research results based on some groups, the *Catalpa* plants in China were divided into six hypothetical species (Wu *et al.*, 2008; Ma *et al.*, 2020). Among them, the morphology of *C. ovata* is the most different from other species (e.g. different inflorescence, flower color, flower size, leaf shape), and the molecular evidence also proves that the genetic distance between *C. ovata* with other species is the farthest. In addition to *C. ovata*, the existing taxa include *C. bungei*, *C. fargesii* and *C. duclouxii*. *C. duclouxii* has been widely considered to be a hairless variety of *C. fargesii*. Until Olsen and Kirkbride (2017) observed a large number of specimens, and suggested that pubescence type and density cannot be used to separate the members of this species into specific or infraspecific taxa, they treated *C. fargesii* and *C. duclouxii* as synonyms of *C. bungei*. Dong (2022) conducted a molecular study on the specimens of these three taxa. After analysis, it was considered that *C. bungei* may be a hybrid of *C. duclouxii* and *C. fargesii*, and supported Olsen and Kirkbride's proposal to classify the three taxa as synonyms. In our study, *C. bungei*, *C. fargesii* and *C. duclouxii* occupy different geographical distributions, and have significant differences in tree type, stem shape, inflorescence, wood color and adaptability. Of course, their floral anatomical structures have more similarities. At the same time, in our survey, we also found that there are a large number of hybrid offspring (intermediate type) in the overlapping region or transition region of taxa distribution. In the samples of our study, there are also some suspected hybrid individuals, which proves the universality of hybridization and introgression between different taxa. We observed the specimens of *Catalpa* species in the specimen bank (e.g. Museum national d'Histoire naturelle (P), Royal Botanic Garden Edinburgh (E), PE, HUH, NY, KUN), and found that there were a large number of hybrid offspring specimens and misidentifications in the specimen bank, and there was also a complex genetic background sample in the study of Dong (2022). Therefore, we should abandon the simple dependence on specimens in the classification and evolution of *Catalpa* species in China, and turn to field investigation and sampling and common garden planting experiments. There are also two unclassified taxa in our research, *C. jinsiquiu* and *C. huangxin*.

As early as more than a thousand years ago, there were records of the classification of *Catalpa* plants according to different wood colors and textures (Jia, 544; Li, 1578), referring to the yellow and white wood of *Catalpa* species. However, in the existing taxonomic studies, such taxa are not involved. In our many years of field investigation, we found that there are two taxa with yellow wood, and the anthers of these two taxa are yellow, and the stripes in the flower tube are thicker than those of *C. bungei*, *C. fargesii* and *C. duclouxii* (Fig.2). However, there are obvious differences in morphology and geographical distribution between *C. jinsiquiu* and *C. huangxin* (Fig.8). A large number of basic traits have been studied and popularized for these two taxa (Xin *et al.*, 2011). Our results are not completely consistent, but the overall results show that: 1. Among the six taxa in China, the genetic differentiation index (F_{ST}) between *C. duclouxii* and *C. bungei* is 0.119479465 (Table S2), all the analysis results showed that the relationship between *C. duclouxii* and *C. bungei* was closer (Fig.4, 5, 7), which is consistent with the results of Dong (2022), but our results show that the relationship between the complex of *C. duclouxii* and *C. bungei* is far from *C. fargesii* (F_{ST} were 0.557576384 and 0.572014464, respectively), which is consistent with the results of Li (2008), so we do not support the classification of *C. bungei*, *C. duclouxii* and *C. fargesii* as the synonyms, at least *C. fargesii*

is another species; 2. The application of *C. jinsiquiu* is always referred to as a local ‘variant’ of *C. bungei*, but in the investigation of the Chinese *Catalpa* resources by various research teams, it was found that the distribution range of *C. jinsiquiu* is far more than that of *C. bungei*, and has a series of morphological characteristics that are significantly different from *C. bungei*. Followed by molecular genetic research (Ma, 2020; Li, 2022) further proved that the genetic relationship between *C. jinsiquiu* and *C. bungei* is much farther than that between *C. jinsiquiu* and *C. fargesii* (F_{ST} were 0.325643201 and 0.283541907, respectively), and that it is an important group that needs to clarify its taxonomic status and phylogenetic relationship. The results of this study show that *C. jinsiquiu* is not a type of *C. bungei*, Although the *C. jinsiquiu* may originated in the *C. fargesii* (Fig.4, 5, 6, 7), it has been developed on a large scale in China, and occupies different ecological niches with *C. fargesii*. Therefore, we believe that it comes from *C. fargesii* but forms a stable and widely distributed species with strong resistance; 3. *C. huangxin* and *C. jinsiquiu* have some similar morphological characteristics (e.g. yellow wood and yellow anther, Fig.2, Fig.4), but not the same taxa (Fig.8). *C. huangxin* should be a narrow-range species with the exotic distribution of *C. jinsiquiu*.

Therefore, through morphological classification, distribution range investigation and molecular-based phylogenetic studies, it is shown that there are at least five *Catalpa* species in China.



Fig. 8 Morphological and distribution characteristics of *C. huangxin* and *C. jinsiquiu*. *C. huangxin* with long branched hairs and long simple hairs in sepals, pedicels, petioles and veins, while *C. jinsiquiu* is glabrous.

4.2 Inconsistency in the phylogenetic study of *Catalpa*

Previous nrDNA, nrDNA ITS and plastome gene phylogenetic analyses (Li, 2008; Dong, 2022) strongly support the monophyly of clade of genus *Catalpa* (except *C. jinsiquiu* and *C. huangxin*). Four sect. *Macrocatappa* species form a robust clade and so do the sect. *Catalpa* species in the research. Their monophyly also gets support from morphology. For example, leaves are elliptic and evergreen in sect. *Macrocatappa* but broadly ovate and deciduous in sect. *Catalpa* (Paclt, 1952; Li, 2008). Previous research support that the first dispersal event occurred from southwestern North America (NA) and Mexico into the Greater Antilles giving rise to the ancestor of the sect. *Macrocatappa*. The second dispersal event also occurred from southwestern NA and Mexico but led to central and northern NA, subsequently reaching China through the Bering land bridge and reaching Europe through the North Atlantic land bridge. The third dispersal event took place in the Miocene from China to NA and gave rise to a clade composed of *C. bignonioides* and *C. speciosa* (Olmstead, 2013; Olsen and Kirkbride, 2017; Dong, 2022). However, the existing phylogenetic studies of sect. *Catalpa* only based on part of the genetic information, involving part taxa

and fewer samples, and did not cover all *Catalpa* taxa in China, and the results showed that there have plenty of gene trees discordance and plastid-nuclear discordance (Li, 2008; Dong, 2021; Li, 2022). For example, Li (2008) used nrDNA ITS and chloroplast gene *ndhF* to estimated interspecific relationships of *Catalpa*, found *C. ovata* is more closely related to NA species than to other China species and pointed *C. duclouxii* forms a clade that is sister to the clade containing *C. bungei* and *C. fargesii*, he didn't support the close relationship of *C. fargesii* and *C. duclouxii* that is suggested by Rehder (1914) and Gilmour (1936), even with plastid-nuclear trees discordance in the study. The genetic distance based on the plastome data supports a closer relationship between *C. ovata* and the three eastern Asian species, while the nuclear data supports a closer relationship between *C. ovata* and the two NA species. The genetic distance based on the plastome data supports a closer relationship between *C. ovata* and the three eastern Asian species, while the nuclear data supports a closer relationship between *C. ovata* and the two NA species. Dong (2022) used chloroplast and nuclear genome (ITS and 5759 single-copy nuclear genes) data to explore the phylogenetic relationship of *Catalpa* species infer the evolutionary history of *Catalpa* species, and indicated that current interspecific gene flow (between *C. bungei*, *C. duclouxii*, and *C. fargesii*) and incomplete lineage sorting have played important roles in the evolutionary history of *Catalpa*. The phylogenomic analysis of the genus supports the monophyly of three Asian species (*C. bungei*, *C. duclouxii*, and *C. fargesii*), and suggested that the relationships among these three species quite complex. Similarly, there are gene trees discordance and plastid-nuclear trees discordance in the study. We previously used the chloroplast genome to explore the phylogenetic relationship of *Catalpa* in China (the species involved the same as this research, Li, 2022). The phylogenetic relationship shows that *C. ovata* and *C. speciosa* each form a clade, and have a close relationship with each other, and the others five Chinese *Catalpa* taxa form a clade, and supports the monophyly of these five taxa. The above research shows that there is inconsistency in the position of *C. ovata*.

This study mainly focused on *Catalpa* plants in China. The results showed that the NJ tree based on nuclear genome SNP divided *Catalpa* into three clade (Fig.4). *C. ovata* formed a clade, *C. fargesii* and *C. jinsiqui* formed a clade, *C. bungei*, *C. duclouxii* and *C. huangxin* formed a clade, and these three clades were monophyletic. In the ML tree constructed based on the chloroplast genome, the Chinese *Catalpa* is divided into four monophyletic clade (Fig.5B), of which *C. ovata* formed a clade, *C. fargesii* formed a clade, *C. jinsiqui* formed a clade, *C. bungei*, *C. duclouxii* and *C. huangxin* formed a clade. Although the phylogenetic tree is not completely consistent, it suggests that there are at least three monophyletic clades in Chinese *Catalpa*, which highlights the problems in the classification and phylogenetic research of *Catalpa* in China.

Although *C. jinsiqui* and *C. huangxin* have some similar traits and are different from other taxa (e.g. similar deep purple thick stripes in the flower tube, large deep purple spots in the flower tube (>1 mm), yellow pollen, yellow heartwood). However, the two taxa belong to different phylogenetic clade and the distribution areas do not overlap. They are exotic species with stable morphology, but these similar characteristics make us very curious about how the two species originated and what is the relationship between them? Genetic introgression analysis showed that there was a genetic introgression from *C. huangxin* to *C. jinsiqui* (Fig.7, Fig.S2, S3, Table 1), which may have an important relationship with the speciation of *C. jinsiqui*, but when were the two species formed? When does the genetic introgression occur? What region of the genome does introgression occur and what effect it has on species maintenance need to be further explored.

5. CONCLUSION

In view of the inconvenience caused by the confusion of the classification of *Catalpa* plants in China to basic and applied research, we conducted morphological pre-classification through field investigation, and explored the distribution region of the preliminary taxa, through these, we obtained six taxa that conform to the concept of morphological and ecological species. In order to demonstrate whether the existing Chinese *Catalpa* taxonomic system is reasonable and verify our pre-classification, phylogenetics of the six hypothetical species of Chinese *Catalpa* inferred from GBS data and the chloroplast genome we previously sequenced. Although the results of different analyses are not completely consistent, the following points are generally shown: 1. Phylogenetic studies of the six hypothetical species showed that the six hypothetical species were divided into three to four clades, which proved that the *Catalpa* species in China were not only *C. ovata* and *C. bungei*; 2. The data analysis of the

controversial taxa (*C. bungei*, *C. fargesii* and *C. duclouxii*) showed that *C. duclouxii* had a close genetic relationship with *C. bungei* (belonged to the same phylogenetic clade), not a variant of *C. fargesii*. However, *C. fargesii* belonged to another phylogenetic clade with *C. duclouxii* and *C. bungei*. Although there are hybridization and introgression among the three species, they are not the synonyms; 3. The two unknown taxa of *C. jinsiquiu* and *C. huangxin* are not only different from other taxa in morphology and distribution, but also separated from other taxa in the results of genetic analysis, indicating that the taxonomic system of *Catalpa* in China needs to be further revised and improved. The results point to that *C. jinsiquiu* and *C. huangxin* are separate taxa, which are consistent with the differences in morphology and ecological characteristics; 4. Hybridization and introgression are common in *Catalpa*, which are also common in other plants, insects and animals; 5. The gene introgression between *C. jinsiquiu* and *C. bungei*, *C. fargesii* and *C. huangxin* showed the complexity of its genetic origin, and its wide distribution, stable morphological characteristics and strong adaptability showed that it was a stable taxonomic taxa with complex sources.

The purpose of this study is to prove the genetic differences and phylogenetic relationships of the hypothetical species based on geographical distribution and morphological classification by molecular evidence, and to propose and preliminarily discuss the classification of *Catalpa* species in China, so a small amount of sampling was carried out. In the following research, we will further explore the species delimitation of *Catalpa* in China at the genome-wide level, taxonomic revision, and explore the gene introgression region of the genome and species maintenance of Chinese *Catalpa* species.

References

- Agapow P M, Bininda-Emonds O R P, Crandall K A, *et al.*, 2004. The impact of species concept on biodiversity studies. *Q. Rev. Biol.* 79(2), 161-179. <https://doi.org/10.1086/383542>
- Alexander D H, Novembre J, Lange K, 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9), 1655-1664. <https://doi.org/10.1101/gr.094052.109>
- Bickford D, Lohman D J, Sodhi N S, *et al.*, 2007. Cryptic species as a window on diversity and conservation. *Trends ecol evol.* 22(3), 148-155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Boumans L, Vieites D R, Glaw F, Vences, *et al.*, 2007. Geographical patterns of deep mitochondrial differentiation in widespread Malagasy reptiles. *Mol Phylogenet Evol.* 45, 822-839. <https://doi.org/10.1016/j.ympev.2007.05.028>
- Brown J, Pirrung M, Mccue L A, 2017. FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. <https://doi.org/10.1093/bioinformatics/btx373>.
- Carstens B C, Pelletier T A, Reid N M, *et al.*, 2013. How to fail at species delimitation. *Mol Ecol.* 22, 4369-4383. <https://doi.org/10.1111/mec.12413>
- Catchen J, Hohenlohe P A, Bassham S, *et al.*, 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 22(11), 3124-3140. <https://doi.org/10.1111/mec.12354>
- Clark L G, 2023. Integrating genomic and morphological data into bamboo taxonomic and evolutionary studies. *Plant Divers.* 45(2), 123-124. <https://doi.org/10.1016/j.pld.2023.02.002>.
- Coyne J A and Orr H A, 2004. *Speciation*. Sinauer Associates, Sunderland, UK.
- Dayrat B. 2005. Towards integrative taxonomy. *Biol J Linn Soc.* 85, 407-417. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- De Queiroz K and Donoghue MJ, 1988. Phylogenetic systematics and the species problem. *Cladistics*, 4, 317-338. <https://doi.org/10.1111/j.1096-0031.1988.tb00518.x>
- De Queiroz K, 2007. Species concepts and species delimitation. *Syst Biol.* 56, 879-886. <https://doi.org/10.1080/10635150701701083>
- Danecek P, Auton A, Abecasis G, *et al.*, 2011. The variant call format and VCFtools. *Bioinformatics.* 27, 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Dong W P, Liu Y L, Li E Z, *et al.*, 2022. Phylogenomics and biogeography of *Catalpa* (Bignoniaceae) reveal incomplete lineage sorting and three dispersal events. *Mol Phylogenet Evol.* 166:107330. <https://doi.org/10.1016/j.ympev.2021.107330>
- Durand E Y, Patterson N, Reich D, *et al.*, 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol.* 28, 2239-2252. <https://doi.org/10.1093/molbev/msr048>
- Edwards D L and Knowles L L. 2014. Species detection and individual assignment in species delimitation: can integrative data

- increase efficacy? P Roy Soc B-Biol Sci, 281(1777), 20132765. <https://doi.org/10.1098/rspb.2013.2765>
- Eldredge N and Cracraft J, 1980. Phylogenetic Analysis and the Evolutionary Process, New York: Columbia University Press.
- Elias T S and Newcombe L F, 1979. Foliar nectaries and glandular trichomes in *Catalpa* (bignoniaceae). Acta Botanica Sinica, 21(3), 215-227.
- Evanno G S, Regnaut S J, Goudet J, 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 14(8), 2611-2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Feng L, Ruhsam M, Yi-Han Wang, *et al.*, 2020. Using demographic model selection to untangle allopatric divergence and diversification mechanisms in the rheum palmatum complex in the eastern asiatic region. Mol Ecol. 29. <https://doi.org/10.1111/mec.15448>
- Frankham R, Ballou J D, Dudash M R, *et al.*, 2012. Implications of different species concepts for conserving biodiversity. Biol Conserv. 153, 25-31. <https://doi.org/10.1016/j.biocon.2012.04.034>
- Fitak R R, 2021. Optm: estimating the optimal number of migration edges on population trees using treemix. Biology Methods and Protocols. 1–6. <https://doi.org/10.1093/biomethods/bpab017>
- Frugone M J, López M E, Segovia N I, *et al.*, 2019. More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. Mol Phylogenet Evol. 139, 106563. <https://doi.org/10.1016/j.ympev.2019.106563>
- Frugone M J, Cole T L, López M E, *et al.*, 2021. Taxonomy based on limited genomic markers may underestimate species diversity of rockhopper penguins and threaten their conservation. Divers Distrib. 27, 2277–2296. <https://doi.org/10.1111/ddi.13399>
- Fujita M K, Leaché A D, Burbrink F T, *et al.*, 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol Evol. 27(9), 480-488. <https://doi.org/10.1016/j.tree.2012.04.012>
- Gagnaire P A, Lamy J B, Cornette F, *et al.*, 2018. Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. Genome Biol Evol. 10, 2518-2534. <https://doi.org/10.1093/gbe/evy194>
- Gentry A H, 1992. Bignoniaceae—part II (tribe Tecomeae). Flora Neotropica Monograph 25(2): 1–370.
- Gilmour J S L, 1936. *Catalpa fargesii* forma *duclouxii*. Curtis's Botanical Magazine. 159, 9458.
- Green R E, Krause J, Briggs A W, *et al.*, 2010. A draft sequence of the neandertal genome. Science. 328, 710-722. <https://doi.org/10.1126/science.1188021>
- Green R E, Krause J, Briggs A W, *et al.*, 2018. Genome-wide Admixture is Common across the Heliconius Radiation. bioRxiv. <https://doi.org/10.1101/414201>
- Grummer J A, Morando M M, Avila L J, *et al.*, 2018. Phylogenomic evidence for a recent and rapid radiation of lizards in the Patagonian *Liolaemus fitzingerii* species group. Mol Phylogenet Evol. 243-254. <https://doi.org/10.1016/j.ympev.2018.03.023>
- Hao Mingzhuo, 2013. Analysis of Genetic Diversity and Distribution of *Catalpa bungei* and *Catalpa fargesii* in China. Nanjing: Nanjing Forestry University.
- Hong D Y, 2020. Gen-morph species concept—A new and integrative species concept for outbreeding organisms. J. Syst. Evol. 58, 725-742. <https://doi.org/10.1111/jse.12660>
- Jia Sixie, 544. Qimin Yaoshu.
- Jombart T, 2008. adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics. 24, 1403-1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart T, Devillard S, Balloux F, 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. <https://doi.org/10.1186/1471-2156-11-94>
- Jombart T, Ahmed I, 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics. 27, 3070-3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Keogh J S, Scott I A W, Hayes C, 2005. Rapid and repeated origin of insular gigantism and dwarfism in Australian tiger snakes. Evolution. 59, 226–233. <https://doi.org/10.1111/j.0014-3820.2005.tb00909.x>
- Kumar S, Stecher G, Tamura K, 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33, 1870-1874. <https://doi.org/10.1093/molbev/msw054>
- Lawson D J, Dorp L V, Falush D, 2018. A tutorial on how not to over-interpret structure and admixture bar plots. Nat Commun. <https://doi.org/10.1101/066431>
- Li F, Liu Y, Wang J, *et al.*, 2022. Comparative Analysis of Chloroplast Genome Structure and Phylogenetic Relationships Among Six Taxa Within the Genus *Catalpa* (Bignoniaceae). Front Genet. 13, 845619. <https://doi.org/10.3389/fgene.2022.845619>
- Li J H. 2008. Phylogeny of *Catalpa* (Bignoniaceae) inferred from sequences of chloroplast *ndhF* and nuclear ribosomal dna. J Syst Evol. 46(3), 341-348. <https://doi.org/10.3724/SP.J.1002.2008.08025>

- Li Shizhen. 1578. Compendium of Materia Medica.
- Ma Q G, Wang Z J, Xu H M, *et al.*, 2020. AFLP Analysis of *Catalpa bungei* Superior Clones. Forest Research. <https://doi.org/s10.13275/j.cnki.lykxyj.2020.05.018>
- Malinsky M, Matschiner M, Svardal H, 2020. Dsuite-Fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 21, 584–595. <https://doi.org/10.1111/1755-0998.13265>
- Mallet J, 1995. A species definition for the modern synthesis. Trends Ecol Evol, 10, 294-299. [https://doi.org/10.1016/0169-5347\(95\)90031-4](https://doi.org/10.1016/0169-5347(95)90031-4)
- Mayr E, 1942. Systematics and the Origin of Species, Columbia University Press, New York.
- Mayr E, 1963. Animal Species and Evolution. Harvard University Press Cambridge, Massachusetts.
- Moenga S M, Bedada G, Greenlon A, *et al.*, 2018. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. Nat Commun. 9(1), 649. <https://doi.org/10.1038/s41467-018-02867-z>
- Murphy J C, Jowers M J, Lehtinen R M, *et al.*, 2016. Cryptic, sympatric diversity in tegu lizards of the *Tupinambis teguixin* group (Squamata, Sauria, Teiidae) and the description of three new species. PLoS ONE. 11, e0158542. <https://doi.org/10.1371/journal.pone.0158542>
- Olave M, Avila L J, Sites J W, *et al.*, 2018. Hybridization could be a common phenomenon within the highly diverse lizard genus *Liolaemus*. J Evolution Biol. 31, 893-903. <https://doi.org/10.1111/jeb.13273>
- Olsen R T and Kirkbride J H, 2010. Manchurian *Catalpa*: *Catalpa bungei*. Arnoldia. 68:75-76.
- Olsen R T and Kirkbride J H, 2017. Taxonomic revision of the genus *Catalpa* (bignoniaceae). Brittonia. 69(6), 1-35. <https://doi.org/10.1007/s12228-017-9471-7>
- Olmstead R T, 2013. Phylogeny and biogeography in solanaceae, verbenaceae and bignoniaceae: a comparison of continental and intercontinental diversification patterns. Bot Journal Linn Soc. 171(1), 80-102. <https://doi.org/10.1111/j.1095-8339.2012.01306.x>
- Paclet, J. 1952. Synopsis of the Genus *Catalpa* (Bignoniaceae) III. Candollea. 13, 241–285.
- Padial J M, la Riva De I, 2009. Integrative taxonomy reveals cryptic Amazonian species of *Pristimantis* (Anura: Strabomantidae). Zool J Linn Soc-Lond. 155(1), 97-122. <https://doi.org/10.1111/j.1096-3642.2008.00424.x>
- Padial J M, Miralles A, la Riva De I, Vences M. 2010. The integrative future of taxonomy. Front in Zool. 7, 16. <https://doi.org/10.1186/1742-9994-7-16>
- Pan Q K, Kang P S, Guo M. 1991. *Catalpa*. China Forestry Publishing House, Beijing.
- Petit R J and Excoffier L. 2009. Gene flow and species delimitation. Trends Ecol Evol. 24, 386-389. <https://doi.org/10.1016/j.tree.2009.02.011>
- Pickrell J K, Pritchard J K, Tang H. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8(11), e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Puillandre N, Lambert A, Brouillet S, Achaz G, 2012. ABGD, automatic barcode gap discovery for primary species delimitation. Mol Ecol. 21, 1864-1877. <https://doi.org/10.1111/j.1365-294X.2011.05239.x>
- Purcell S, Neale B, Todd-Brown K, *et al.*, 2007. Plink: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559-575. <https://doi.org/10.1086/519795>
- Rehder A, 1914. *Catalpa*. In: L. H. Bailey (ed.), The standard cyclopedia of horticulture. 2, 684-685.
- Sangster G, 2014. The application of species criteria in avian taxonomy and its implications for the debate over species concepts. Biological Reviews of the Cambridge Philosophical Society. 89(1), 199-214. <https://doi.org/10.1111/brv.12051>
- Schilthuizen M, 2000. Dualism and conflicts in understanding speciation. BioEssays. 22(12), 1134-1141. [https://doi.org/10.1002/1521-1878\(200012\)22:12<1134::AID-BIES11>3.0.CO;2-5](https://doi.org/10.1002/1521-1878(200012)22:12<1134::AID-BIES11>3.0.CO;2-5)
- Schlick-Steiner B C, Steiner F M, Seifert B, *et al.*, 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annu Rev Entomol. 55,421-38. <https://doi.org/10.1146/annurev-ento-112408-085432>
- Stamatakis A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Tel-Zur N, Abbo S, Myslabodski D, Mizrahi Y. 1999. Modified ctab procedure for DNA isolation from epiphytic cacti of the genera *hylocereus* and *selenicereus* (cactaceae). Plant Mol Biol Rep. 17(3), 249-254. <https://doi.org/10.1023/A:1007656315275>
- Tonatiuh Ramírez-Reyes, Flores-Villela O A, Piero D, *et al.*, 2021. Genomic assessment of the *Phyllodactylus tuberculatus* complex (Reptilia: phyllodactylidae) in America. Zool Scr. (3), 1-14. <https://doi.org/10.1111/zsc.12492>
- Wang N, Yuan Y, Wang H, *et al.*, 2020b. Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. Sci Rep. 10, 16308. <https://doi.org/10.1038/s41598-020-73321-8>
- Wang M Z, Fan X K, Zhang Y H, *et al.*, 2022. Phylogenomics and integrative taxonomy reveal two new species of *Amana* (Liliaceae), Plant Divers. <https://doi.org/10.1016/j.pld.2022.03.001>

- Weitschat W. 2016. Intraspecific variation of *Svalbardiceras spitzbergensis* (Frebold) from the Early Triassic (Spathian) of Spitsbergen. POLAR RES. 27, 292–297. <https://doi.org/10.1111/j.1751-8369.2008.00041.x>
- Welton L J, Siler C D, Oaks J R, Diesmos A C, Brown R M. 2013. Multilocus phylogeny and Bayesian estimates of species boundaries reveal hidden evolutionary relationships and cryptic diversity in Southeast Asian monitor lizards. Mol Ecol. 22, 3495-3510. <https://doi.org/10.1111/mec.12324>
- Wilkins J S. 2009. Species: A History of the Idea. UC Press, Berkeley.
- Will K W, Mishler B D, Wheeler Q D, 2005. The Perils of DNA Barcoding and the Need for Integrative Taxonomy. Syst. Biol. 54, 844-851. <https://doi.org/10.1080/10635150500354878>
- Wu C L, Hao M Z, Peng F R, Du X H, Zhu Z X. 2008. Comparative investigation of the growth and physiology characteristics for different *Catalpa* spp.saplings. Journal of Nanjing Forestry University (Natural Sciences Edition). <https://doi.org/10.3969/j.issn.1000-2006.2008.06.028>
- Wu J F, Jia D R, Liu R J, *et al.*, 2022. Multiple lines of evidence supports the two varieties of *Halenia elliptica* (Gentianaceae) as two species. Plant Divers. 44(03), 290-299. <https://doi.org/10.1016/j.pld.2021.09.004>
- Xin S, Ya L I, Ru-Tong Y, Qing W, Gan Y, 2011. Issr analysis of genetic diversity of *Catalpa bungei* C. A. Mey germplasm resources in China. Jiangsu Journal of Agricultural Sciences.
- Xiulan Li, 2013. Population Genetics diversity of *Catalpa bungei* C.A.Mey and its Relationship with Some Other *Catalpa* Species. Yichang: China Three Gorges University.
- Yang J, Weedon M N, Purcell S, *et al.*, 2011. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 19(7), 807-812. <https://doi.org/10.1038/ejhg.2011.39>
- Yongming Y, Junhui W, Wenjun M, *et al.* 2013. Characteristic difference of catalpa scop species in vitro culture. *Journal of Northeast Forestry University*, 41(12), 17-20.
- Zhang Z Y and Thawatchai Santisuk, 1998. Flora of China, 18, 215-216. Science Press, Beijing.