

The Study of Genetic Diversity and Structure of Extremely Endangered *Manglietia longipedunculata* and Endangered *Manglietia insignis*

Qiaomei Qin

Guangdong Eco-Engineering Polytechnic

Yishan Yang

Guangxi Institute of Botany

Xiao Wei

Guangxi Institute of Botany

Lipo Pan

Guangxi Institute of Botany

Xianliang Zhu

South China Botanical Garden, CAS

Linfang Wu

Linfang Ecology Co., Ltd

Shunlai Liu

Nankunshan Provincial Nature Reserve of Guangdong Province

Qifeng Zhong

Nankunshan Provincial Nature Reserve of Guangdong Province

Wenchao Zhong

Nankunshan Provincial Nature Reserve of Guangdong Province

Jianmin Tang

18877384841@163.com

Guangxi Institute of Botany

Article

Keywords:

Posted Date: November 14th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3555362/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

To strengthen the protection of *Manglietia insignis* and *Manglietia longipedunculata* germplasm resources, the endangered mechanism and scientific protection strategies of these two species were investigated. A novel, effective, and flexible DNA sequencing library construction approach-Hyper-seq was used in this study to develop large-scale genomic single nucleotide polymorphism (SNP) and to evaluate the genetic diversity and population structure of *M. insignis* and *M. longipedunculata*. Genetic diversity analysis demonstrated that the genetic diversity of *M. insignis*, which expected heterozygosity (H_e) was 0.1435 and nucleotide diversity (π) was 0.1541. *M. longipedunculata*, lower than in *M. longipedunculata* (H_e = 0.1793, π = 0.1916) with narrow distribution. The genetic diversity of the two *ex-situ* conservation populations (H_e =0.1797, 0.1990, π = 0.1894, 0.2121) was higher than the wild populations (H_e =0.1591, π = 0.1733), suggesting that the artificial *ex-situ* conservation population can effectively protect its genetic diversity and the genetic breeding in future studies. The mean inbreeding coefficient (F_{IS}) in the *M. longipedunculata* and *M. insignis* populations were - 0.0141 and - 0.0176, respectively, indicating that the two populations were largely outcrossing in the natural state. The frequency of rare alleles in *M. insignis* (Tajima's D = 0.3436 > 0) and *M. longipedunculata* (Tajima's D = 0.2119 > 0) populations was low, which may be under balanced selection pressure or experienced population shrinkage. A high degree of genetic differentiation between the two species was detected, but moderate among the populations of the same species. The results of cluster analysis demonstrated a obvious separation between the two species with all individuals clustering according to species differences. It is worth noting that individuals 7, 8, and 9 in the ML-CW *M. insignis* population first clustered with *M. longipedunculata*. Furthermore, these individuals also contained a small genetic component from *M. longipedunculata*. Principal component and genetic relationship analysis also demonstrated that some *M. insignis* individuals exhibited an evident genetic relationship with *M. longipedunculata*. These populations can therefore be preferentially selected in the further cross-breeding experiments of the two species. The genetic diversity and structure of *M. insignis* and *M. longipedunculata* were compared in this study. Not only were the genetic characteristics of the two species explored for the first time, but the genetic relationship between the two species was also preliminarily explored. This study provides a scientific basis for the development of future protection strategies at the molecular level.

Introduction

Manglietia insignis is an evergreen tree widely distributed throughout Guizhou, Guangxi, Yunnan, and other provinces in China. This species is commonly found growing in broad-leaved forests in the form of individual plants or groups. However, pure forest or wild community with *Manglietia rosea* as the dominant population has not been found to date¹. *Manglietia insignis* trees are straight, tall, and exhibit rapid growth. The wood has the advantages of fine texture, luster, corrosion resistance, and easy processing. In addition, its flowers have a great aroma and bright color² and its roots, leaves, flowers, and fruits are commonly used for medicinal purposes³. *Manglietia insignis* has, therefore, become a popular garden tree species. However, due to its high ornamental value and economic value, *M. insignis* has

become a major target of deforestation due to its popularity. Combined with the poor regeneration ability of its natural populations, the number of wild populations of *M. insignis* has been decreasing annually. This species has become rare and endangered and has been listed as a national third-class protected plant in China ⁴. Similar to *M. insignis*, *Manglietia longipedunculata* is an evergreen tree in the Family Magnoliaceae. *Manglietia longipedunculata* was first discovered in 2003 by Professor Zeng Qingwen (Chinese Academy of Sciences) and his research team in Nankunshan Nature Reserve, Huizhou City, Guangdong Province, China ⁵. To date, only 11 wild plants have been identified in the Nankunshan Nature Reserve. No other populations of *M. longipedunculata* have been found in other parts of the world ⁶. This species is now listed in the red list of Magnoliaceae plants. Due to the lack of pollinators, the asynchronous development of pistil and stamen, and the extremely short stigma receptivity period, the propagation of this species is extremely challenging. In its natural state, *M. longipedunculata* only blooms and does not bear fruit. Therefore, to protect this rare species and its population size, researchers have rapidly performed a series of *in-situ* and *ex-situ* protection measures including artificial pollination and grafting ⁷. At present, *M. longipedunculata* has been successfully introduced to the Nankunshan Botanical Garden in Huizhou City and Dongguan Botanical Garden in Dongguan City (China).

Due to the limitation of population size, *M. insignis* research has been limited to resource investigation ^{8,9}, seedling breeding ^{10,11}, and medicinal ingredients ^{12,13}, all requiring further investigation. In addition to some research on resource investigation ¹⁴ and seedling breeding ⁸, there are few reports on *M. longipedunculata* at home and abroad. Therefore, to better implement effective protection strategies for this species, it is imperative to expand *M. longipedunculata* population size. Current research on the two species of *Manglietia* is still in its infancy. Research on the genetic diversity and structure of *M. insignis* and *M. longipedunculata* throughout the world is lacking. Investigating genetic diversity is therefore of great significance to the protection of species resources, specifically in rare and endangered plants. This study has aimed to effectively identify the environmental adaptability and evolutionary potential of this species, provide a scientific and effective basis for the formulation of protection measures for related species and the screening of core germplasm resources, and provide a solid material basis for resource utilization ¹⁵. Genetic diversity of species elucidation began from the initial simple phenotypic identification to the cell identification but now has progressively advanced to molecular markers. Compared with the previous marker techniques, molecular markers are highly polymorphic, fast, highly efficient, and are not affected by the developmental stage of the species or the external environment ¹⁶. With the continuous progression of molecular biology research, molecular marker technology has been developed to the third generation represented by single nucleotide polymorphisms (SNPs). SNPs are widely distributed within the genome, which commonly exhibits high polymorphisms, high genetic stability, and easy automatic detection, becoming the best option for population genetic analysis and genomics ¹⁷. SNP-based molecular marker technology has been widely used in genetic diversity studies of many species ^{18–20}. In addition, we used Hyper-seq method, which is a novel, effective, and flexible marker-assisted selection and genotyping approach. Compared with the traditional whole-genome sequencing (WGS), Hyper-seq approach has a certain gene region enrichment effect. This approach can

flexibly adjust the marker density by using different Hyper-seq primers based on the needs of different species and projects. And without additional enzyme digestion and ligation joints, through a special PCR method, a large number of samples can be built at the same time, and a large amount of genotype big data can be produced ²¹.

Theoretically, the genetic diversity of widely distributed species is higher than that of small and narrowly distributed species ²². Due to the high probability of inbreeding, unstable population structure, and genetic drift, small populations of plants will exhibit accelerated loss of genetic diversity. However, some recent studies have identified that the genetic diversity of small-scale populations with fewer individuals is higher than that of large-scale populations ^{23,24}. In addition, *M. insignis* is a relatively primitive species, occupying an important position in the study of the classification, distribution, and flora of the adjacent areas ²⁵. Therefore, this study investigated the genetic diversity of the widely distributed species *M. insignis* and the extremely small population of *M. longipedunculata*. The genetic diversity and structure differences were compared between these two species, to identify the cause for the decrease in the population size and the genetic relationship between the populations and present recommendations for the protection and restoration of the population. Furthermore, excellent germplasm resources were screened out to provide a theoretical scientific and material basis for future introduction and cultivation of this species.

Results

SNP calling based on the reference genome. A total of 430,379 raw reads were generated from 22 *M. insignis* and 36 *M. longipedunculata* samples. After filtering, 6,596 high-quality SNP loci were identified and evenly distributed on the 19 chromosomes (Fig. 1).

Genetic diversity. The genetic diversity of *M. insignis* and *M. longipedunculata* was calculated using 430,379 SNPs (Tables 1 and 2 and Fig. 2). Among the five populations, ML-NKC H_0 was the largest, followed by ML-DG and ML-NKW, and the smallest was in MI-CW. The H_E of ML-NKC was the largest, followed by ML-DG and ML-NKW, and the smallest was in MI-CW. The ranking of the H_E in each population was consistent with the H_0 . The average H_0 value of the two populations of *M. insignis* was 0.1649, which was higher than that of H_E . The average H_0 value of the three populations of *M. longipedunculata* was 0.2024, which was also higher than that of H_E . The mean value of π (0.1916) in the three populations of *M. longipedunculata* was greater than in *M. insignis*. The genetic diversity was evaluated using H_E and π as follows: *M. longipedunculata* (ML-NKC > ML-DG > ML-NKW) > *M. insignis* (MI-YZL > MI-CW). Between the three *M. longipedunculata* populations, the genetic diversity of the two *ex-situ* populations was higher than that of the wild population. These results demonstrate the significance of *ex-situ* conservation for maintaining and improving the genetic diversity of *M. longipedunculata*.

The F_{IS} coefficient was used to assess the inbreeding level within the populations. Among the two *M. insignis* populations, the F_{IS} coefficient of one population (MI-CW, $F_{IS}=0.0226$) was greater than 0, while

the other (MI-YZL, $F_{IS}=-0.0507$) was less than 0. Among the three *M. longipedunculata* populations, only ML-DG ($F_{IS}=0.0080$) F_{IS} was greater than 0, while the other two populations (ML-NKC and ML-NKW) were less than 0. This suggests that selfing or inbreeding between individuals may be occurring in the MI-CW population of *M. insignis* and the ML-DG population of *M. longipedunculata*, demonstrating the loss of heterozygosity phenomenon. The results of Tajima's D calculation of *M. insignis* and *M. longipedunculata* were positive at 0.3039–0.3833 and 0.0440–0.3513, respectively (Fig. 2). These results suggest that the rare allele frequency in these two populations was low and may be under balanced selection pressure or experiencing population shrinkage. The genetic differentiation (F_{st}) of the population was preliminarily established according to the genetic differentiation coefficient between the populations (Table 2). The genetic distance between *M. insignis* and *M. longipedunculata* populations was large, as was the degree of differentiation. The F_{st} of the *M. insignis* MI-CW population and *M. longipedunculata* populations was lower than that of MI-YZL and *M. longipedunculata* populations, suggesting that *M. insignis* MI-CW population was closer to *M. longipedunculata* than the MI-YZL population.

Table 1
Genetic diversity of five *M. insignis* and *M. longipedunculata* populations.

Pop ID	Private SNP	H_0	H_E	π	F_{IS}	Tajima's D
MI-CW	119	0.1485	0.1421	0.1554	0.0226	0.3039
MI-YZL	365	0.1812	0.1426	0.1528	-0.0507	0.3833
Mean value		0.1649	0.1424	0.1541	-0.0141	0.3436
ML-DG	18	0.1890	0.1797	0.1894	0.0080	0.2404
ML-NKC	30	0.2385	0.1990	0.2121	-0.0495	0.3513
ML-NKW	23	0.1826	0.1591	0.1733	-0.0112	0.0440
Mean value		0.2024	0.1793	0.1916	-0.0176	0.2119

Table 2
Coefficient of genetic differentiation among *M. insignis* and *M. longipedunculata* populations.

	MI-CW	MI-YZL	ML-DG	ML-NKC	ML-NKW
MI-CW		0.141	0.168	0.170	0.189
MI-YZL			0.234	0.236	0.251
ML-DG				0.051	0.065
ML-NKC					0.073

Population structure. The most direct manifestation of genetic diversity is the level of genetic variation. However, a single individual of a species has a limited lifespan and can be susceptible to external factors. A continuous population composed of individuals is therefore a unit of genetic evolution. The specific distribution pattern of these types of populations in nature is the genetic structure of the population. The mixed analysis results demonstrated that $K = 2$ was the best clustering result among all the samples of *M. insignis* and *M. longipedunculata* (Fig. 3). When $K = 2$, the CV error value was the lowest, and the two species were obviously separated (Fig. 4). Since the two *M. longipedunculata ex-situ* populations were derived from wild populations (ML-NKW), some individuals were not separated according to population differences but exhibited some individual confounding phenomena. All individuals within the two populations were separated according to population differences. It is worth noting that some individuals (ML-CW-2,3,7,8,9) in the *M. insignis* ML-CW population were mixed, suggesting that a small number of genetic components of *M. longipedunculata* were mixed in the ML-CW *M. insignis* population. Furthermore, the *M. insignis* ML-CW-7, 8, and 9 individuals were the first to be clustered with *M. longipedunculata* in the phylogenetic tree. This demonstrates that *M. longipedunculata* may have a certain relationship with *M. insignis* in its genetic evolution.

When $K = 3$, all *M. longipedunculata* individuals still clustered into one group (Fig. 4). While the individuals of the two populations were clustered into one group after the separation of *M. longipedunculata* and *M. insignis*, two of the populations did not show a common genetic component. Two individuals (ML-CW-7 and 8) from the *M. insignis* ML-CW population were also mixed with *M. longipedunculata* genetic components. When $K = 4$, most *M. longipedunculata* individuals were divided into three categories according to different populations (Fig. 4). However, a small number of individuals was separated from the population and clustered with other populations. A few *M. longipedunculata* individuals (ML-NKC-5, 6, and ML-NKW-4) exhibited same genetic components of other *M. longipedunculata* populations, but all same genetic components within the population. All *M. insignis* individuals clustered into one group and individuals (ML-CW-2, 3, 7, 8, and 9) in the ML-CW population were mixed with *M. longipedunculata*. Furthermore, ML-CW-7 and 8 individuals contained mixed genetic information from the three *M. longipedunculata* populations.

Principal component analysis (PCA) and genetic relationship analysis. According to the two-dimensional PCA diagram (Fig. 5), the contribution rates of the first and the second principal components were 31.15 and 7.34%, respectively. The position distance in the two-dimensional map demonstrated the distance of the genetic relationship between the germplasm resources of the *M. insignis* and *M. longipedunculata* populations. The results showed that *M. longipedunculata* and *M. insignis* were obviously separated (Fig. 5). The two *M. insignis* populations exhibited considerable separation, with the MI-YZL in the first quadrant and the MI-CW in the third and fourth quadrants, likely due to the geographical distance between the two populations. The three *M. longipedunculata* populations exhibited a small degree of separation and close genetic distance, and all individuals were mixed, which may be related to all individuals originating from the same population. Furthermore, PCA results demonstrated that some ML-CW individuals tended to move closer to *M. longipedunculata*, suggesting the presence of a genetic

evolutionary relationship between the two populations, consistent with the genetic structure analysis results.

The genetic relationship analysis of all *M. insignis* and *M. longipedunculata* individuals demonstrated that *M. insignis* was largely divided into two categories according to the genetic relationship (Fig. 6). Most individuals in the MI-CW population were clustered into one group, as were all individuals in the MI-YZL population. In addition, the MI-CW-7, 8, and 9 individuals in the MI-CW population exhibited considerable separation from most of the MI-CW individuals, suggesting significant differences in the genetic relationship between the two populations. Furthermore, the three MI-CW-7, 8, and 9 individuals demonstrated an evident genetic relationship. In the *M. longipedunculata* genetic relationship analysis, most individuals of the three populations were generally divided into three categories according to the differences in populations. Some individuals however were mixed into other populations. Overall, the genetic distance between all individuals of the three *M. longipedunculata* populations was not significantly different. The genetic differentiation was not obvious when compared with *M. insignis*. This was consistent with the PCA results.

Discussion

Comparison of genetic diversity between *M. insignis* and *M. longipedunculata*. Genetic diversity is the basis of species adaptability and evolutionary potential²⁶. Generally, higher genetic diversity of the species results in greater adaptability to the environment, decreased influence of the external environment, and increased evolutionary potential. On the contrary, the evolutionary potential is reduced with lower genetic diversity of the species, increasing the vulnerability to environmental factors²⁷. In this study, a large-scale genomic analysis of SNP loci was performed for the first time to evaluate the genetic diversity of *M. insignis* and *M. longipedunculata*. The results showed that the genetic diversity of *M. insignis* ($\pi = 0.1541$) was lower than that of *M. longipedunculata* ($\pi = 0.1916$). On one hand, this phenomenon may be due to *M. insignis* ecology and habitat has undergone great changes in recent years despite having accumulated rich genetic variation over the long-term evolution process. In addition, due to excessive anthropogenic deforestation and natural disasters, its distribution has become fragmented with the individual numbers continually decreasing, leading to a decrease in the genetic diversity of its population. On the other hand, due to the continuous habitat fragmentation of *M. insignis*, its original wide and continuous distribution has been broken up. This reduces the species adaptability, causing some populations to completely disappear, increasing the inbreeding rate in the residual population, and resulting in the disappearance of some alleles, which in turn leads to the continuous reduction of *M. insignis* genetic diversity. In addition, this study only collected samples from two wild populations of *M. insignis*. The low sampling rate or one-sidedness may also cause the low results. Although *M. longipedunculata* population was small, high genetic diversity was still maintained. A previous study demonstrated that the high genetic diversity of small populations is dependent on frequent gene flow²⁸. Similar phenomena have been reported in extremely small populations of *Camellia dongxingensis*²⁹, Chinese wild apricot³⁰, and other species.

Many studies have found that the ability of some rare and endemic species to maintain high genetic diversity is closely related to the evolutionary history of species, the maintenance of genetic diversity of glacial refuge population, the regional characteristics of distribution area, and the ecological habits of species including the mating system³¹. Therefore, the potential effects of these factors on the genetic diversity of *M. insignis* and *M. longipedunculata* were comprehensively analyzed in the current study. The Tajima's D values of the five populations of these two species were positive, suggesting that the frequency of rare alleles in these two species was low and the existing populations may be under balanced selection pressure or have already experienced population shrinkage. The inbreeding ($F_{IS}>0$) was detected in the *M. insignis* MI-CW population, while in the *M. longipedunculata* population, the wild population (ML-NKW) specifically was dominated by outcrossing ($F_{IS}<0$). In the case of limited pollinators, selfing is one of the most effective strategies for species to maintain or increase population size³². Although the outcrossing-based mating system of *M. longipedunculata* limits the rapid population expansion, it ensures the high genetic diversity of the population and reduces the risk of species extinction due to any future changes in the habitat. Despite the small size of the wild *M. longipedunculata* population, selfing was not detected. Therefore, it is likely that in natural habitat, there may be self-incompatibility and partial asexual reproduction in *M. longipedunculata*, eventually leading to heterozygous and homozygous deletions within the population^{33,34}. However, further research is required via experimental analyses such as artificial pollination.

When the raw habitat of species is destroyed, the measures of ex-situ migration and protection are called 'ex-situ conservation'³⁵. However, due to the lack of complete sampling, unclear resources, and repeated sampling during the implementation of the ex-situ conservation plan, the isolated populations established by the ex-situ conservation often have the risk of loss or reduction in genetic diversity³⁶. The genetic diversity and population structure information of ex-situ populations is one of the most important reference indices to evaluate the success of ex-situ conservation. If the genetic diversity of the ex-situ population is higher than or generally consistent with the genetic diversity of the wild population, the protection measures are considered successful³⁷. The genetic diversity ($\pi = 0.1894, 0.2121$) in this study of the two ex-situ conservation populations of *M. longipedunculata* was higher than that of its wild population ($\pi = 0.1733$). Therefore, the two artificial ex-situ conservation populations of ML-DG and ML-NKC can effectively protect the genetic diversity of *M. longipedunculata* species and fully represent the genetic diversity of its wild populations. The ex-situ conservation of *M. longipedunculata* individuals largely originated from the seeds collected from the sexual reproduction of wild populations (ML-NKW). In addition, the mating strategy of the wild population based on outcrossing ensures frequent gene exchange. This may be one of the reasons why the genetic diversity of ex-situ conservation populations was higher than in the wild populations. Therefore, the ex-situ population can be used as the key application material in future genetic breeding work. In this study, the average H_e of *M. insignis* and *M. longipedunculata* was 0.1424 and 0.1793, respectively. This value was significantly lower than in other Magnoliaceae species, such as *Manglietia conifera* ($H_e = 0.66$)³⁸, *Kmeria septentrionalis* ($H_e = 0.69$)³⁹,

indicating that the genetic diversity of these two *Manglietia* species was low and the genetic basis was narrow, which may be related to the small natural distribution range of these two plant species.

Comparison of genetic structure and differentiation between *M. insignis* and *M. longipedunculata*. The investigation of the genetic structure of species is very important for the protection of species and the formulation of protection strategies^{40, 41}. The *M. longipedunculata* and *M. insignis* clustering map demonstrated a few individual confounding phenomena between the two populations. However, due to the geographical distance between the two populations, there may be two reasons for this phenomenon. On the one hand, a population of *M. insignis* was located near *M. longipedunculata* and there is gene exchange between the two groups. On the other hand, combined with phylogenetic tree clustering analysis, *M. longipedunculata* may be derived from the genetic differentiation of some individuals in *M. insignis*. This hypothesis is also supported by PCA and genetic relationship analysis between individuals. These two analyses demonstrate that some individuals in the MI-CW population of *M. insignis* tend to gradually approach *M. longipedunculata*. Therefore, in future breeding studies, these species with close genetic relationships can be first selected to improve the success rate of the study.

In addition, the F_{st} was also used to measure the distance of genetic relationships or the degree of genetic differentiation⁴². The genetic differentiation coefficients of the two populations of *M. insignis* and the three populations of *M. longipedunculata* were different, and the value of MI-CW was lower, suggesting that the population was closely related to *M. longipedunculata*. Previous studies have demonstrated an absence of differentiation among subgroups in populations with F_{st} values between 0 and 0.05. If the F_{st} value was 0.05 to 0.15, it was moderately differentiated. High differentiation was considered if the F_{st} value was between 0.15 and 0.25⁴³. The results of this study showed that there was moderate differentiation between the two populations ($F_{st} = 0.141$), similar to a high degree of differentiation. This may be related to the geographical distance between the two populations. The three *M. longipedunculata* populations ($F_{st} = 0.051, 0.065, 0.073$) also belonged to moderate differentiation, with almost no differentiation between the two ex-situ conservation populations. The degree of genetic differentiation between the wild and ex-situ conservation populations was higher than that of the two ex-situ conservation populations. The degree of differentiation between ex-situ conservation and wild populations and between ex-situ conservation populations will gradually increase³⁵. Furthermore, the adaptability of ex-situ conservation species to the native habitat will gradually be lost, eventually delaying the subsequent wild regression experiments of the species⁴⁴. Therefore, to avoid this putative risk, an increase in the gene exchange between the ex-situ conservation and the wild populations is necessary via artificial pollination, systematic updates of the ex-situ conservation population gene diversity, and finally ensuring that the ex-situ population maintains the adaptability to the native environment.

Protection recommendations. The results of this study demonstrated that the genetic diversity of *M. insignis* was low. In addition to ensuring that the population was able to self, the number of individuals was also an important factor due to the continuous reduction of anthropogenic activities. It is, therefore, necessary to improve the investigation of the plant resources and appropriately protect the native

environment of the large-scale population and the *ex-situ* conservation of the small-scale population. The protection of species does not only provide protection for a single group but for different groups or different genotype groups. In regard to multi-population species, the screening of excellent germplasm resources should not only improve the protection and management efficiency of germplasm resources but also reduce costs⁴⁵. However, the genetic information of only two *M. insignis* populations was investigated. The sampling range in future studies should be expanded to accurately screen out excellent germplasm resources. In addition, a one-way mixing phenomenon was identified between the two species. These two species should avoid proximity in *ex-situ* conservation to ensure the purity of genetic information for the species. Due to the geographical distance between the two species, this gene mixing phenomenon can not only explain the possible evolutionary relationship between the two species but also more importantly, shows that there may be an *M. longipedunculata* population in Caowang Mountain, Leye County, Baise City (China). Therefore, the species investigation in this area should be accelerated and the corresponding protection should be rapidly established. The inability to rapidly expand the population is due to the self-pollination mating strategy. Therefore, the primary task should be to protect the wild seedlings *in situ* or *ex-situ* and expand the population size based on maintaining the excellent characteristics of the female parent via tissue culture, cuttings, and other asexual reproduction methods, to provide a solid material basis for future introduction and breeding.

Two *ex-situ* conservation populations with high genetic diversity can therefore be utilized for breeding improved varieties. Although the *ex-situ* conservation duration of *M. longipedunculata* was relatively short, the degree of differentiation between *M. longipedunculata* and wild populations has gradually increased. Therefore, to ensure the adaptability of the species to the native habitat and the success of the subsequent introduction and regression experiments, it is necessary to identify the genetic diversity of the *ex-situ* conservation population and increase the gene exchange between the *ex-situ* and the wild populations in time via artificial pollination and other breeding methods.

Conclusions

In this study, Hyper-seq method was used to sequence a total of 58 samples from two wild populations of *M. insignis* as well as one wild population and two *ex-situ* conservation populations of *M. longipedunculata*. A total of 6,596 high-quality SNP loci were detected. The genetic diversity, population genetic structure, and phylogenetic tree of the two populations were constructed. The results showed that the genetic diversity of the widely distributed *M. insignis* was lower than that of the narrowly distributed *M. longipedunculata*. This may be related to the significant anthropogenic destruction of *M. insignis* habitat. In addition, although the *M. longipedunculata* population was small, high genetic diversity was still maintained. The results of this study demonstrate that the individuals were primarily outcrossing, which may be the reason for this phenomenon. The genetic diversity of the two *ex-situ* conservation populations was higher than that of their wild populations, suggesting that the *ex-situ* conservation can effectively protect the genetic diversity of *M. longipedunculata* and this population can be used for future breeding. In addition, the results of the population genetic structure of the two populations demonstrated

mixed genetic components between the two populations. This material could be preferred for future cross-breeding testing. Since only two populations of *M. insignis* were selected for genetic diversity investigation, the results were not comprehensive enough. Therefore, the sampling range should be expanded in the future to screen out high-quality germplasm resources and accelerate the protection of these species.

Methods

Plant materials. Leaf samples, which are all healthy and mature were collected from 22 *M. insignis* individuals from two populations and 36 *M. longipedunculata* individuals from three populations (Fig. 7 and Table 3). Five leaves were collected from each plant. Among the three populations of *M. longipedunculata*, both ML-DG and ML-NKC were *ex-situ* conservation populations, originating from ML-NKW. Due to the small size of some populations, the total number of samples with the sampling interval was balanced in order to collect as many possible representative samples. The interval between plant individuals sampled was more than ten meters. The samples were all collected from healthy and mature individuals. Fresh leaf samples were collected and quickly dried in discolored silica gel. The materials used in this study were collected and identified by associate researcher Tang Jianmin (Guangxi Botanical Research Institute, China). One sample of each population was preserved in the Herbarium of Guangxi Botanical Research Institute (China). The plant materials collected in this study are in compliance with Chinese and international guidelines and laws, including the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. One plant sample was collected from each population of *M. insignis* and *M. longipedunculata*., and all of them were preserved in the South China Botanical Garden, Guangdong, China. The specimen numbers were MI-CW-20221024, MI-YZL-20221025, ML-DG-20221105, ML-NKC-20221106, ML-NKW-20221107, respectively.

Table 3. Sampling information of *M. insignis* and *M. longipedunculata*.

Population	Location	Type	Latitude	Longitude	No. of samples
MI-CW	Caowang Mountain, Leye County, Guangxi, China	Wild population	24°43'46 N	106°21'21 E	9
MI-YZL	Yinzulao Mountain, Ziyuan County, Guangxi, China	Wild population	26°15'19 N	110°33'19 E	13
ML-DG	Dongguan Botanical Garden, Guangdong, China	Ex-situ population	22°51'16 N	113°47'26 E	14
ML-NKC	Nankunshan Botanical Garden, Huizhou, Guangdong, China	Ex-situ population	23°38'40 N	113°52'47 E	12
ML-NKW	Nankun Mountain, Huizhou County, Guangdong, China	Wild population	23°36'60 N	113°50'60 E	10

Hyper-seq library construction and sequencing. E.Z.N.A. Tissue DNA Kit (Omega Bio-Tek, USA) was used to extract DNA from 22 *M. insignis* and 36 *M. longipedunculata* leaf samples according to the manufacturer's instructions. The DNA quality was assessed using 1% agarose gel electrophoresis and a Nanodrop 2000 (Thermo Fisher, USA) spectrophotometer. The extracted DNA was quantified using Qubit 3.0 (Thermo Fisher, USA) to ensure that each sample met the following criteria: total mass > 3 µg, concentration > 30 ng/µL, and OD₂₆₀/OD₂₈₀ = 1.80-1.20. The barcode adaptor was then ligated and the fragment size selection and PCR amplification were performed to obtain a Hyper-seq library²¹. The constructed library was sent for 150-bp paired-end sequencing on the Illumina Novaseq 6000 platform (Illumina, San Diego, USA).

SNP calling. FastQC⁴⁶ was used to evaluate the quality of the original data. Trimmomatic v0.36 (set to the default parameters)⁴⁷ was used to filter the adaptor sequences, cleavage sites (the first six base pairs of the reads), and low-quality reads (Q20 < 20, length < 36 bp) from the raw sequence data⁴⁸. The clean reads were then mapped to the *Magnolia hypoleuca* (Magnoliaceae) reference genome⁴⁹ using BWA v0.7.12⁵⁰. The rm-pcr-duplicates parameter of the gstacks module in the Stacks v.2.59 software⁵¹ was used to remove the PCR repeats and perform SNP calling. Vcftools v0.1.11⁵² was used to filter SNPs, removing sites with a population loss rate of more than 40%, maf > 0.05, and a sequencing depth of less than 3.

Population genetics analysis. Based on the selected high-quality SNP data, Stacks v2.59 was used to calculate genetic diversity parameters, including observed heterozygosity (H_0), expected heterozygosity (H_E), nucleotide diversity (π), and inbreeding coefficient (F_{IS})⁵³. The Tajima's D value of the site was calculated using Vcftools, with the sliding window size of the operation set to 500 kb.

Phylogenetic and population structure analyses. A phylogenetic tree between individuals was constructed for 22 *M. insignis* and 36 *M. longipedunculata* using the maximum likelihood (ML) method in IQtree v2.0⁵⁴, with *M. insignis* set as the outgroup. Bayesian clustering was performed on all individuals using Admixture v1.3.0⁵⁵ with clustering values (K) set from 1 to 9. The optimal K value was determined based on the minimum cross-validation (CV) error.

Principal component analysis (PCA) and genetic relationship analysis. PCA and kinship analysis among individuals were performed using MingPCACluster, a new simple and efficient software for population VCF file⁵⁶.

Declarations

Acknowledgments

This work was supported by the Special Funds for Natural Resource Affairs in Guangdong Province in 2023, the National Key Research and Development Program (No.2022YFF1300700), Chinese Academy of Sciences 'Light of West China' Program (2022) and Guangxi Forestry Science and Technology Promotion Demonstration Project (2023LYKJ03 and [2022]GT23). This publication is partially based on work at Guangxi Key Laboratory of Plant Functional Substances and Resources Sustainable Utilization of Guangxi Institute of Botany.

Author Contributions

Conceptualization, X.W.; methodology, X.L.; validation, Q.M.; formal analysis, Y.Y. and X.L.; investigation, L.W., S.L., Q.F., W.Z. (sampling), L.Z. and J.T. (sequencing), X.L. and L.P. (data analysis); writing—original draft preparation, Y, Y.; writing—review and editing, Y, Y.; visualization, X.W. and J, T; supervision, J.T.; funding acquisition, X.W. All co-authors have reviewed and approved the final version of the manuscript for publication. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests

Correspondence and requests for materials should be addressed to J.T.

Data Availability

Raw sequencing data are available at NCBI with the SRA accession number of PRJNA1037033 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA1037033?reviewer=pjnht5rev90an34aft2ojnjfti>).

References

1. Gao, Z. Q. *Manglietia insignis* and its cultivation techniques. *Northern Horticulture*.1, 160-161 (2007).

2. Li, M., Ma, H.C. & Li, F.X. Study of Photosynthesis Characteristics of *Manglietia insignis* at Seedling Stage. *Journal of West China Forestry Science*. 1, 42-45 (2004).
3. Peng, C. S. Breeding technology of improved varieties of *Manglietia insignis*. *Forest By-Product and Specialty in China*. 1, 52-54 (2022).
4. Chen, Y. J., Deng, L. X., Chen J. Y.; Chen, Z. P. & Tan, H. M. A Study on the Annual Growth Dynamics of *Manglietia insignis* Seedlings from Different Provenances. *Chinese Agricultural Science Bulletin*. 29, 8-14 (2013).
5. Zeng, Q. W. & Law, Y. *Manglietia longipedunculata* (Magnoliaceae), a new species from Guangdong, China. *Annales Botanici Fennici*. 41 (2004).
6. Xie, C., Lin F., Liu, D. M., Wen, X. Y., & Zhong, W.C. Conservation of Rare and Endangered Species *Manglietia longipedunculata* (Magnoliaceae). *Proceedings of the second international symposium on the family Magnoliaceae*. Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, PR China Graduate University of the Chinese Academy of Sciences, Beijing 100049, PR China; Key Laboratory.
7. Ren, H. et al. The use of grafted seedlings increases the success of conservation translocations of *Manglietia longipedunculata* (Magnoliaceae), a Critically Endangered tree. *Oryx*. 50, 437-445. DOI: 10.1017/S0030605315000423 (2015).
8. Zhu H. Vegetation geography of evergreen broad-leaved forests in Yunnan, southwestern China. *Chinese Journal of Plant Ecology*. 45, 224-241 (2021).
9. Zou T.C. et al. Species diversity conservation and utilization of Guizhou rare and endangered spermatophyte. *Guihaia*. 41, 1699-1717 (2021).
10. Yu, Y. P., Liang, D. L., Liu K. C., Wang, L. H. & Huang, R. F. Effects of Chilling Stress on Physiological Characteristics of *Manglietia florida* Seedlings. *Northern Horticulture*. 1, 69-71 (2013).
11. Guan, L. W., Kuang, D. L. Yang, W. L. & Liu, Z. Y. Effects of Cutting Roots during Different Periods on Transplanting Survival in *Manglietia insignis*. *Southwest China Journal of Agricultural Sciences*. 24, 707-711 (2011).
12. Shang, S. Z. et al. Two new neolignans from *Manglietia insignis*. *Natural Products and Bioprospecting*. 2, 227-230. DOI: 10.1007/s13659-012-0063-7 (2012).
13. Shang S. Z. et al. Bioactive phenolics and terpenoids from *Manglietia insignis*. *Fitoterapia*. 84, 58-63. DOI: 10.1016/j.fitote.2012.10.010 (2013).
14. Zhang Y. K. Investigation and Evaluation of Wild Ornamental Plants in Nankun Mountain Nature Reserve in Guangdong Province. Ph.D. Thesis, South China Agricultural University, Guangdong, China, 2016.
15. Eduardo, C., Marie Stéphanie, S., Paul, G. & Fernández P. J. A. Genetic structure in peripheral Western European populations of the endangered species *Cochlearia pyrenaica* (Brassicaceae). *Plant Systematics and Evolution*. 297, 75-85. DOI: 10.1007/s00606-011-0500-9 (2011).

16. Liu, J. X., Shan, J. P. & Wang, P. Genetic Diversity Analysis of Sunflowers Resources Based on SNP Markers. *Institute of Crops. Molecular Plant Breeding*. 1, 1-18 (2023).
17. Wang, X., Lu, B., Shao, L., Ali, A. & Sun, F. Genome-wide SNPs reveal fine-scale population structure of *Laodelphax striatellus* in China using double-digest restriction site-associated DNA sequencing *Genomics*. 114, 110329-110329. DOI: 10.22541/au.159141074.44080801 (2020).
18. Phoebe, M. et al. Genetic Diversity and Population Structure of Sorghum (*Sorghum bicolor* (L.) Moench) Landraces Using DArTseq-Derived Single-Nucleotide Polymorphism (SNP) Markers. *Journal of molecular evolution*. 91, 552-561. DOI: 10.1007/S00239-023-10108-1 (2023).
19. Liu, C. G. et al. Genetic Diversity of Tea Plant (*Camellia sinensis* (L.) Kuntze) Germplasm Resources in Wuyi Mountain of China Based on Single Nucleotide Polymorphism (SNP) Markers. *Horticulturae*. 8, 932. DOI: 10.3390/HORTICULTURAE8100932 (2022)
20. Yukio, N., Hiroaki, T., Sayoko, N., Naofumi, H., Atsushi, J. N. & Shinji, F. Genetic diversity of loquat (*Eriobotrya japonica*) revealed using RAD-Seq SNP markers. *Scientific Reports*. 12, 10200 (2022).
21. Zou, M., Xia, Z. Hyper-seq: A novel, effective, and flexible marker-assisted selection and genotyping approach. *Innovation*. 3, 100254 (2022). DOI: 10.1016/J.XINN.2022.100254.
22. Kaijund, K. & Jaaska, V. No loss of genetic diversity in small and isolated populations of *Medicago sativa* subsp. *Falcata*. *Biochemical Systematics and Ecology*. 38, 510-520. DOI: 10.1016/j.bse.2010.05.007 (2010).
23. Li W. J.; Su, Z. H. & Yang, L. Genetic diversity of the critically endangered *Ferula sinkiangensis* K. M. Shen (Apiaceae) and the implications for conservation. *Turkish Journal of Botany*. 44, 145-152. DOI: 10.3906/bot-1906-2 (2020).
24. Schou, M. F., Loeschche, V., Bechsgaard, J., Christian, S. & Torsten, N. K. Unexpected high genetic diversity in small populations suggests maintenance by associative overdominance, *Molecular Ecology*. 26, 6510-6523. DOI: 10.1111/mec.14262 (2017).
25. Wang, C. M., Zhuang, Z. Y., Li, W. D. & Li, W. B. Study on Seedling Reproduction Technique of *Manglietia insignis*(Wall.). *Modern Agricultural Science and Technology*. 6, 193-195 (2010).
26. Du, L. M., Yang, L. W., Lin, X. J., Huang, G. N., Li, X. B. & Guo, X. Genetic Diversity Analysis of *Ficus hirta* Vahl Based on SNP Molecular Marker. *Molecular Plant Breeding* 21, 5709-5719 (2023).
27. Nybom, H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* 13, 1143-1155. DOI: 10.1111/j.1365-294X.2004.02141.x (2004).
28. Cai, M., Wen, Y., Uchiyama, K., Onuma, Y. & Tsumura, Y. Population genetic diversity and structure of ancient tree populations of *Cryptomeria japonica* var. *Sinensis* based on RAD-seq data. *Forests*. 11, 1192. DOI: 10.3390/f11111192 (2020).
29. Tang, J. M. et al. Genetic diversity and mating system of *camellia tunghinensis* Chang extremely small population. *Fresenius Environmental Bulletin*. 29, 6508-6517 (2020).
30. Wang, Z. et al. High-level genetic diversity and complex population structure of Siberian apricot (*Prunus sibirica* L.) in China as revealed by nuclear SSR markers. *PloS one*. 9, e87381. DOI: 10.1371/journal.pone.0087381 (2014).

31. Eduardo, S. L., Márcia, G., Camila, M. Z., & Fernanda B. Genetic diversity and population structure of *Vriesea reitzii* (Bromeliaceae), a species from the Southern Brazilian Highlands. *Genetics and Molecular Biology*. 41, 308–317. DOI: 10.1590/1678-4685-GMB-2017-0062 (2018).
32. Kalisz, S., Vogler, D. W. & Hanley, K. M. Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. *Nature*. 430, 884–887. DOI: 10.1038/nature02776 (2004).
33. Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I. & Mariette, S. Heterozygote excess in a self-incompatible and partially clonal forest tree species: *Prunus avium* L. *Mol Ecol*. 15, 2109–2118. DOI: 10.1111/j.1365-294X.2006.02926.x (2006).
34. Wang, B., Yan, L.L., Zhang, Y. & Song, L.H. Genetic diversity analysis of four wild Jujube populations in Ningxia based on SNP molecular markers. *Journal of Anhui Agricultural University*. 49, 432–437 (2022).
35. Cibrian-Jaramillo, A. et al. What is the Conservation Value of a Plant in a Botanic Garden? Using Indicators to Improve Management of Ex Situ Collections. *The Botanical Review*. 79, 559–577. DOI: 10.1007/s12229-013-9120-0 (2013).
36. Wei, X. Z. & Jiang, M. X. Meta-analysis of genetic representativeness of plant populations under ex-situ conservation in contrast to wild source populations. *Conservation Biology*. 35, 12–23. DOI: 10.1111/cobi.13617 (2021).
37. Su, Z. H., Bryce, R. B., Zhuo, L., & Jiang, X. L. Divergent population genetic structure of the endangered *Helianthemum* (Cistaceae) and its implication to conservation in northwestern China. *Frontiers in Plant Science*. 7, 2010. DOI: 10.3389/fpls.2016.02010 (2017).
38. Wen, S. N., Zhong, C. L., Jiang, Q. B., Zhang, Y., Chen, Z. & Chen, Y. Genetic Diversity Analysis of *Manglietia conifera* Dandy Germplasm by SSR Markers. *Molecular Plant Breeding*. 15, 3788–3797 (2017).
39. Lin, Y. F. The population genetic structure and gene flow of *Kmeria septentrionalis* assessed by microsatellite markers. Ph.D. Thesis, Guangxi Normal University, Ph.D. Thesis, Guilin, China, 2012.
40. Pan, W., Sun, J., Yuan, Q., Zhang, L., Deng, K. & Li, Y. Analysis of genetic diversity and structure in different provenances of *liriodendron* by RAD-seq technique. *Scientia Silvae Sinicae*. 58, 74–81 (2022).
41. Zhu, X. L. et al. Genome-wide diversity evaluation and core germplasm extraction in ex situ conservation: A case of golden *Camellia tunghinensis*. *Evolutionary Applications*. 16, 1519–1530. DOI: 10.1111/eva.13584 (2023).
42. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: Defining, estimating and interpreting. *Nature Reviews Genetics*. 10, 639–650 (2009).
43. Wright, S. Evolution and the genetics of populations. A treatise in four volumes. Volume 4. Variability within and among natural populations. *Journal of Biosocial Science*. 4, 253–256 (1972).
44. Enßlin, A., Sandner, T. M. & Matthies, D. Consequences of ex-situ cultivation of plants: Genetic diversity, fitness and adaptation of the monocarpic *Cynoglossum officinale* L. in botanic gardens. *Biological Conservation*. 144, 272–278. DOI: 10.1016/j.biocon.2010.09.001 (2011).

45. Miao, L., Wang, S., Zou, M., Li, J., Kong, L. & Yv, X. Review of the studies on Core collection for horticultural crops. *Journal of Plant Genetic Resources*. 17, 791–800 (2016).
46. Andrews, S. Fastqc: A quality control tool for high throughput sequence data. 2010.
47. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30, 2114–2120. DOI: 10.1093/bioinformatics/btu170 (2014).
48. Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 30, 2114–2120. DOI: 10.1093/bioinformatics/btu170 (2014).
49. Zhou, L. J. et al. The genome of *Magnolia hypoleuca* provides a new insight into cold tolerance and the evolutionary position of magnoliids. *Frontiers in Plant Science*. 14, 1108701-1108701. DOI: 10.3389/FPLS.2023.1108701 (2023).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 25, 1754-1760. DOI: 10.1093/bioinformatics/btp324 (2009).
51. Rochette, N. C., Rivera-Colón, A.G., Catchen, J.M. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol.Ecol.* 28, 4737–4754. DOI: 10.1111/mec.15253 (2019).
52. Danecek, P. et al. The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. DOI: 10.1093/bioinformatics/btr330 (2011).
53. Rochette, N. C., Rivera-Colón, A. G. & Catchen, J. M. Stacks 2: Analytical methods for paired-end sequencing improve RADseqbased population genomics. *Molecular Ecology*. 28, 4737-4754. DOI: 10.1111/mec.15253 (2019).
54. Minh, B.Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. DOI: 10.1093/molbev/msaa015 (2020).
55. Alexander, D. H., Novembre, J. & Lange, K. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* 19, 1655–1664. DOI: 10.1161/01.ATV.0000137190.63214.c5 (2009).
56. Zhu, X. L., Zou, R., Tang, J. M., Deng, L. L. & Wei, X. Genetic diversity variation during the natural regeneration of *Vatica guangxiensis*, an endangered tree species with extremely small populations. *Global Ecology and Conservation* 1, 42. DOI: 10.1016/J.GECCO. 2023. E02400 (2023).

Figures

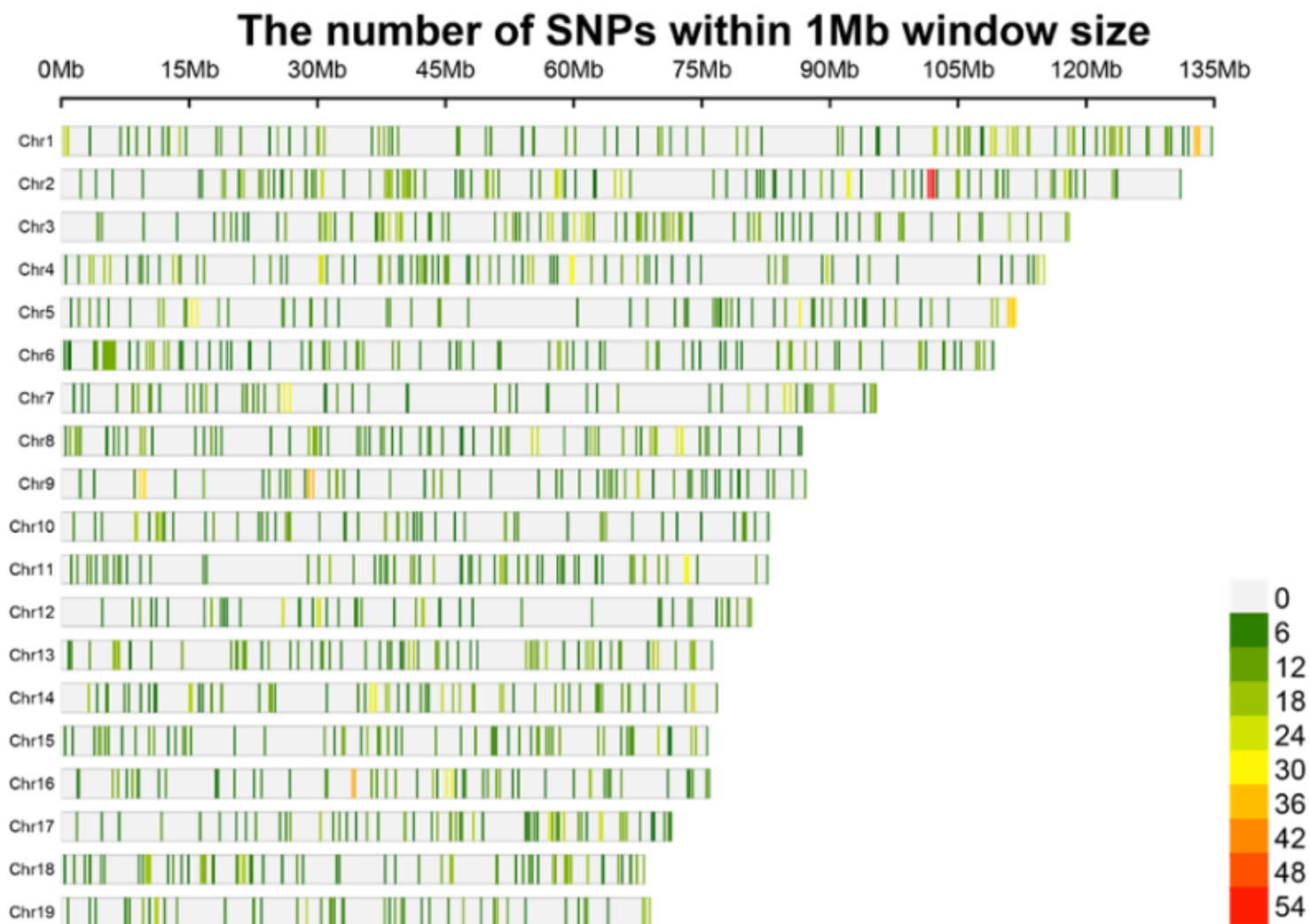


Figure 1

The distribution of 6,596 SNPs on 19 chromosomes

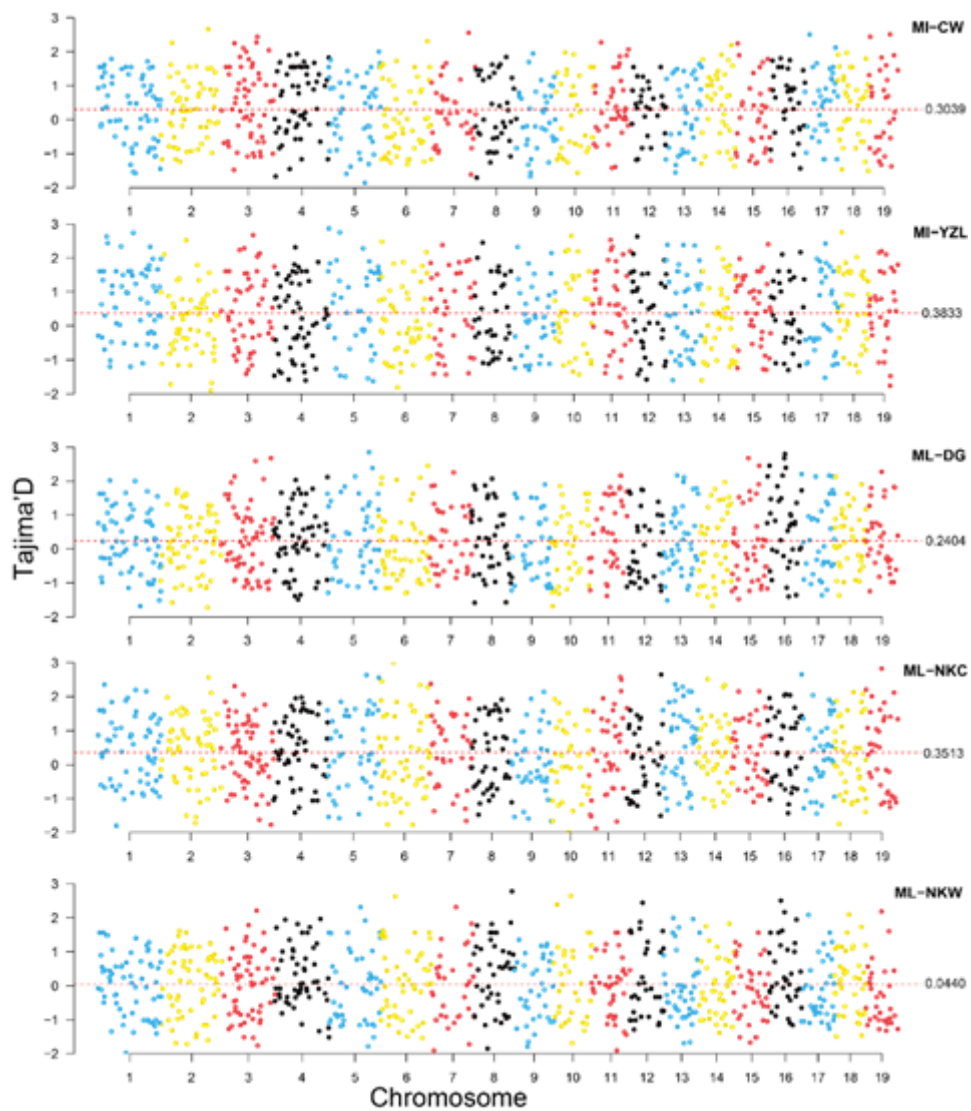


Figure 2

The distribution of Tajima's D values in five *M. insignis* and *M. longipedunculata* populations

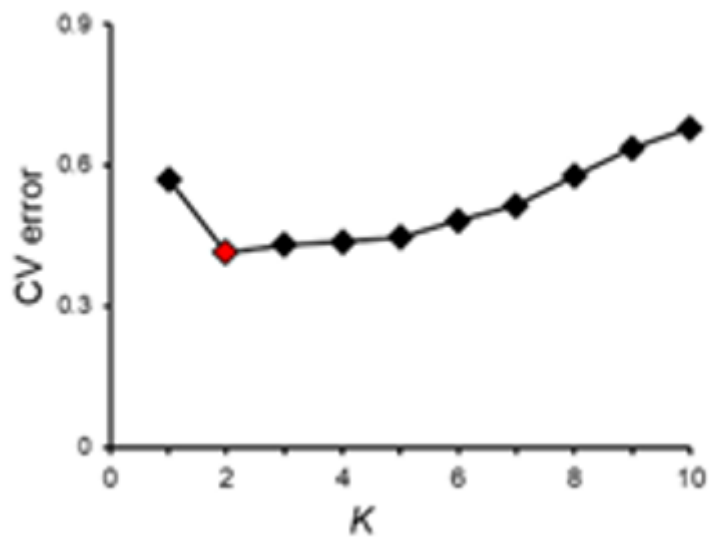


Figure 3

CV error distribution for K from 1 to 10. K with the smallest CV value is marked in red

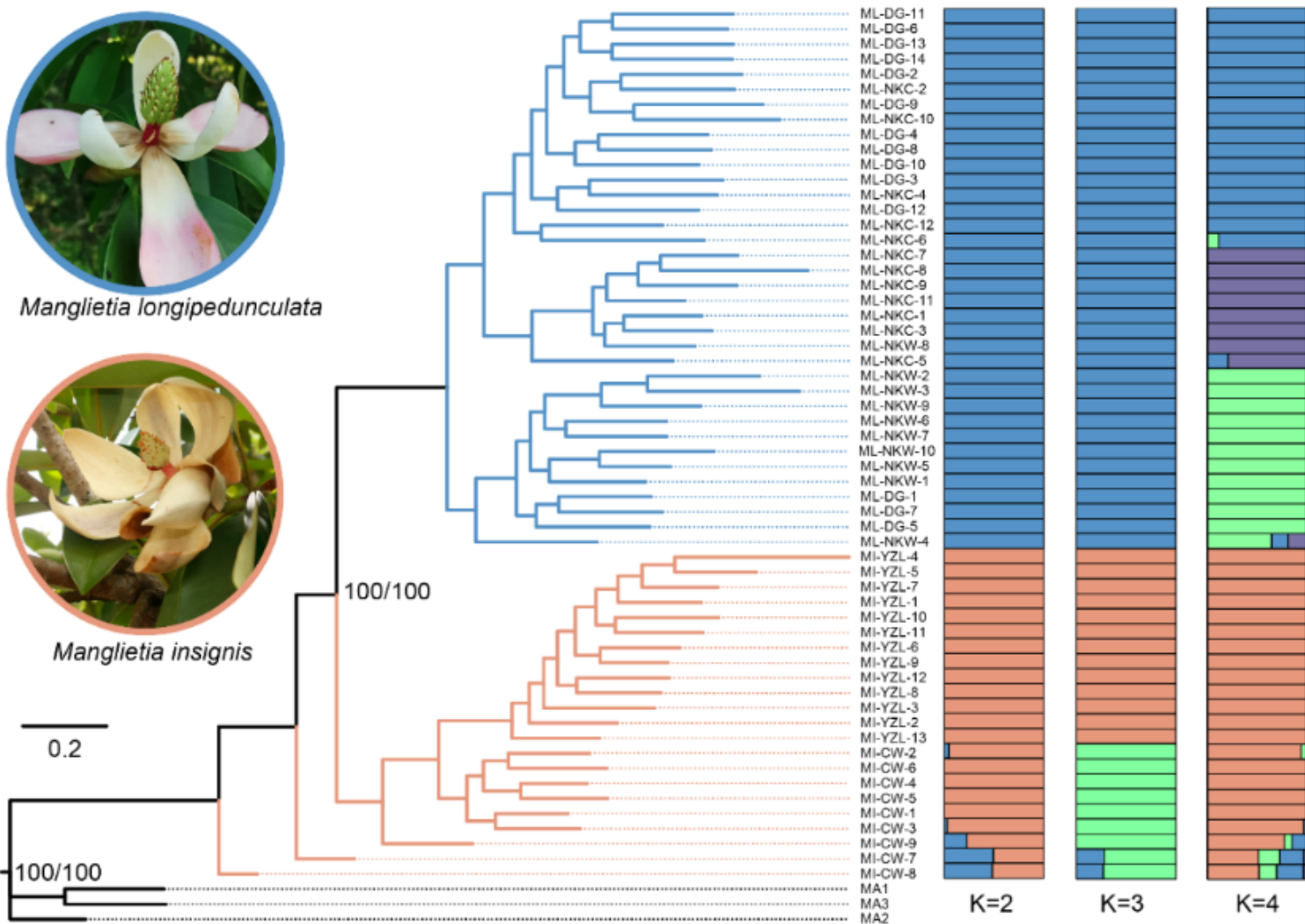


Figure 4

The maximum likelihood phylogenetic tree for 22 *M. insignis* and 36 *M. longipedunculata*.

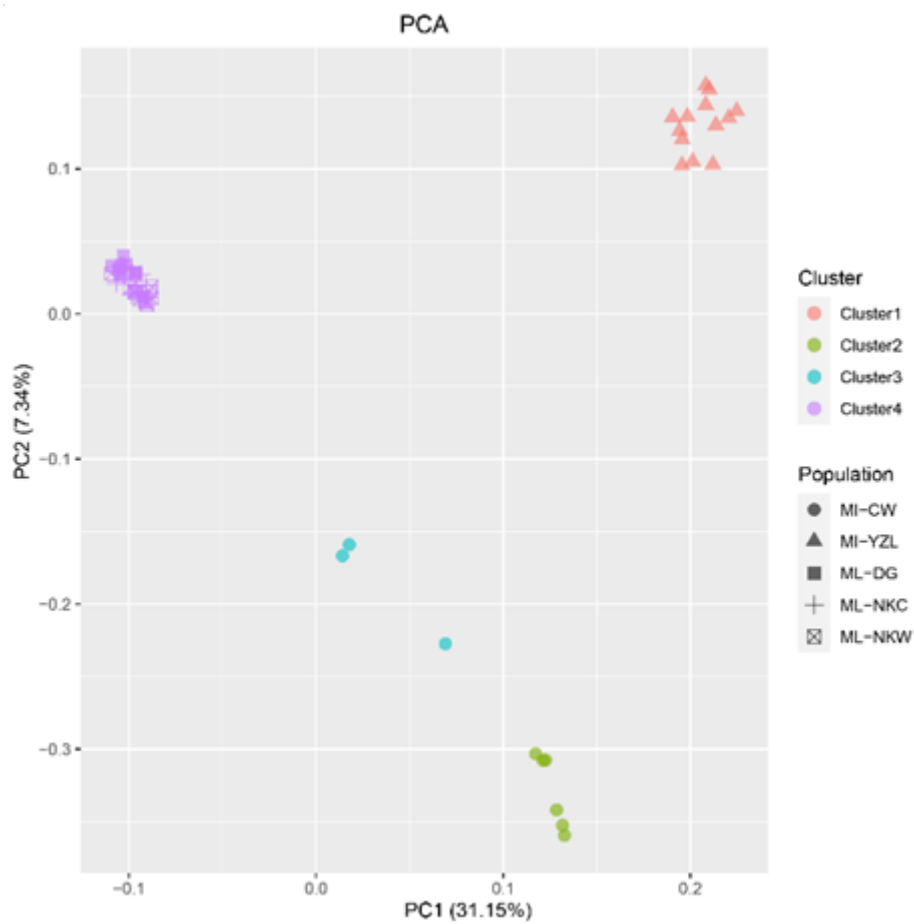


Figure 5

PCA between *M. insignis* and *M. longipedunculata* individuals

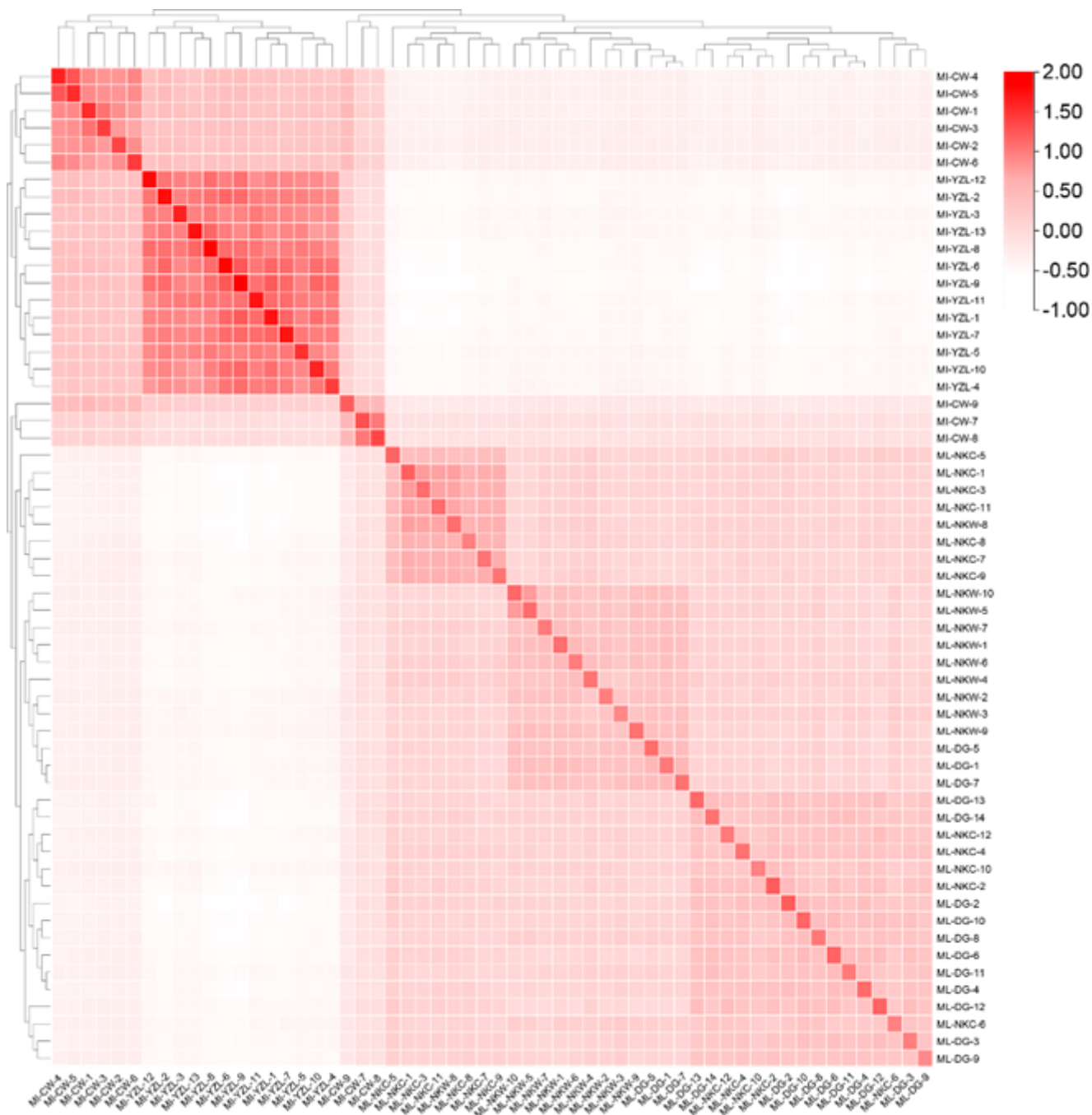


Figure 6

The genetic relationship between *M. insignis* and *M. longipedunculata* individuals.

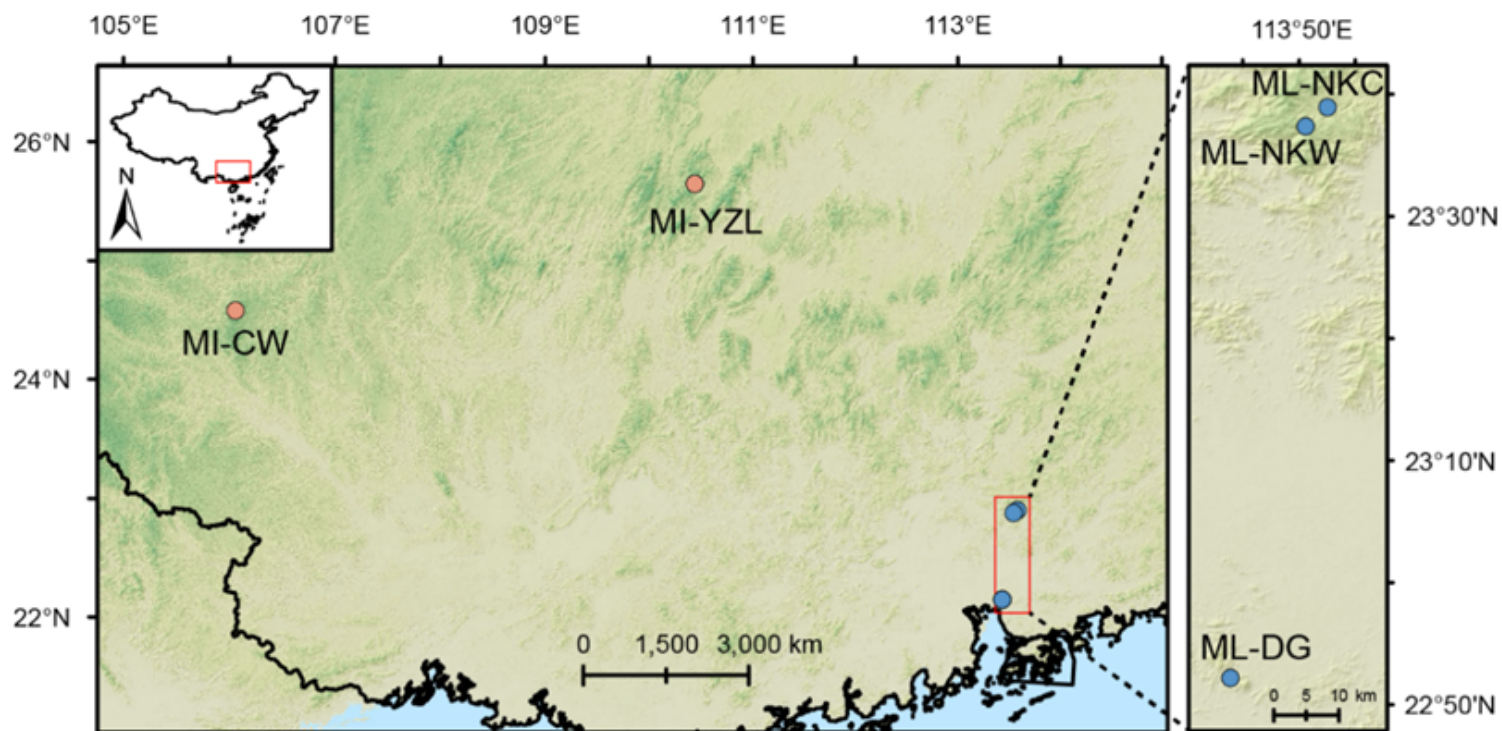


Figure 7

Sampling sites for *M. insignis* and *M. longipedunculata*.