Population study reveals genetic variation and introgression of four deciduous oaks at the junction between Taihang Mountain and Yanshan Mountain

Ziqi Pei
Beijing Forestry University

Qinsong Yang
Beijing Forestry University

Xining Chen
Beijing Forestry University

Yu Zong
Zhejiang Normal University

Jinjin Li
Beijing Forestry University

Xiong Yang
Beijing Forestry University

Chenrui Huo
Beijing Forestry University

Yixin Chen
Beijing Forestry University

Qian Zhang
Beijing Forestry University

Na Luo
Beijing Forestry University

Jialu Kang
Beijing Forestry University

Xiaoqian Meng
Beijing Forestry University

Wenjie Jiang
Beijing Forestry University

Yining Li
Beijing Forestry University

Huirong Zhou
Beijing Forestry University

Jiaxi Wang
Research Article

Keywords: Oak, population, genetic variation, SSR, DNA fingerprint

Posted Date: June 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1756058/v1

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

Oaks (*Quercus* spp.) are considered as model plants for studying plant evolution and natural gene introgression. However, interspecific hybridization often occurred in sympatric sibling specie, resulting in blurred interspecific boundaries and hindering the development of breeding. Beijing area is at the junction between Taihang Mountain and Yanshan Mountain. It is a key area for deciduous oaks native to China and an overlapping area of several oaks. It is urgently necessary to evaluate the genetic diversity and population structure of the 4 deciduous oak species to understand the degree of gene introgression and to screen out ideal breeding materials. In this study, we collected 11 populations of 4 oak species (*Q. variabilis*, *Q. mongolica*, *Q. dentata* and *Q. aliena*) in the junction between Taihang Mountain and Yanshan Mountain. By using the polymorphic SSR markers, we analyzed the genetic variation of the collected 400 individuals and investigated the population structure and found gene introgression events. *Q. variabilis* had a clearer genetic background as compared to the other three species and the artificial population of *Q. variabilis* might have been transplanted from other regions outside Beijing. *Q. mongolica* had a more frequent gene introgression with *Q. dentata* and *Q. aliena* in this area. In addition, our results provided DNA fingerprints of the sampled 40 individuals according 9 SSR markers. This research reported useful SSR data for analyzing the genetic variation of oak species native to China, laying the foundation of whole genome sequencing and conducting an oak germplasm nursery with clear genetic background.

Introduction

Oaks (*Quercus* spp.) are widespread all over the world. *Quercus* is a large genus in Fagaceae and oaks have been cultivated for a long history, especially in Asia, Europe and America, for their longevity and renewability (Plomion et al. 2018). There are over 450 *Quercus* species in the world and 51 in China. Oaks spread over south and north of China and play an important role of ecology and economy. In Beijing, the junction between Taihang Mountain and Yanshan Mountain, several deciduous oaks are natively distributed such as Chinese cork oak (*Q. variabilis*), Mongolian oak (*Q. mongolica*), *Q. dentata* and *Q. aliena*. The widespread of oaks has been described as “the evolutionary success”. One of the explanations of the success is propensity for hybridization, contributing to adaptive introgression (Kremer and Hipp 2020).

Oak species have been formalised into 2 subgenera: subgenus *Quercus*, comprising *Ponticae*, *Protobalanus*, *Virentes*, *Lobatae* and *Quercus*; and subgenus *Cerris*, encompassing sections *Ilex*, *Cerris* and *Cyclobalanopsis* (Liu et al. 2021; Hipp et al. 2020; Hubert et al. 2014). China, as a *Quercus* diversity centre, is home to 35–62 species. 4 native oak species are distributed in the same area, but they belong to different subgenus. *Q. variabilis* belongs to *Cerris* section East Asia Cerris with slender leaves, while *Q. mongolica*, *Q. dentata* and *Q. aliena* belong to *Quercus* section Roburoids with oval and irregular leaves (Hipp et al. 2020).

Studies have shown that interspecific hybridization and gene introgression are common among species of deciduous oaks, which is increasingly recognized as an important process across diverse lineages of
plants (Mir et al. 2006; Crowl et al. 2020). However, the introgression also makes it difficult to figure out the differences between two species with close relationship, leading to uncertainty when collecting the germplasm resources. Studies on gene introgression among *Quercus* species have been carried out for decades, and the phenomenon of hybridization and introgression has been clearly found in oak species widely distributed in Europe and North America, such as *Q. rubra*, *Q. suber*, *Q. alba* and *Q. robur* (López-Aljorna et al. 2007; Moran et al. 2012; Crowl et al. 2020). A study in Japan also demonstrate the interspecific hybridization of *Q. mongolica* and *Q. dentata* (Nagamitsu et al. 2020). Zeng et al. (Zeng et al. 2011) found that *Q. mongolica* and *Q. liaotungensis* had a high degree of gene introgression in Beijing area, with low degree of genetic differentiation between populations, leading to unclear species morphological characterization. However, the gene introgression among *Q. variabilis*, *Q. mongolica*, *Q. dentata* and *Q. aliena* have not been reported, and the genetic background of these oak species in Beijing area is still unknown. The ambiguity of genetic background also leads to the inaccuracy of some existing studies in representing a particular oak species.

DNA molecular markers especially SSR (Simple Sequence Repeat) and SNP (Single Nucleotide Polymorphism) markers have been developed for many years and have been widely used in recognition of species or cultivars (Lupini et al. 2019; Zurn et al. 2020). In coconut, 74 polymorphic SSR markers were identified by genotyping-by-sequencing (GBS) analysis of 40 coconut accessions and different traits could be divided into several groups, which was consistent with the amplicons of SSR markers (Riangwong et al. 2020). Recently, SSR markers also have been used in the identification of tree peony varieties and 18 SSR markers were found to be associated with the floral scent traits (Luo et al. 2021). In oaks, although SSR markers have been used to characterize the populations worldwide for many years (Ramos-Ortiz et al. 2016; An et al. 2017; Yang et al. 2017; Burkardt et al. 2020), there were few studies focusing on large amount of oak genetic resources at the junction between Taihang Mountain and Yanshan Mountain, a specific region that near the north border for *Q. variabilis* and *Q. aliena* natural distribution and near the south border for *Q. mongolica* natural distribution in China. Using SSR markers, we can generate the DNA fingerprint of the individuals, which can provide an exclusive identity information for each individual.

DNA fingerprints can be used to reveal and compare the relationship among individuals within a species or different species, and can be effectively applied to genetic analysis and breeding (Nybom 1991) (Hilde Nybom 2014). In pea, DNA fingerprint and ISSR marker analysis indicated that pea distributed in two places had a tendency of differentiation (Stavridou et al. 2020). Thus, germplasm from different regions were collected by authors for breeding and genetic improvement. Meanwhile, modern fingerprinting is crucial for varietal protection and germplasm characterization even for protecting plant breeders’ rights. 5 microsatellite markers were used to distinguish different varieties of rice, ultimately solving the problem of mixing of 2 rice cultivars (Ganopoulos et al. 2011). As there is no oak cultivar native to China, introgression is very important for plant breeding. It is urgent to analyze the genetic variation and introgression of native oaks in China, especially in Beijing area, fingerprinting the Chinese oak germplasm resources, thus we can conduct the oak germplasm nursery with clear genetic background.
Here, we collected 400 individuals of four oak species, including individuals both in natural and artificial forests in the junction between Taihang Mountain and Yanshan Mountain. By using 9 SSR markers that were previously developed with high polymorphism, we analyzed the genetic variation among 4 oak species (Q. variabilis, Q. aliena, Q. dentata and Q. mongolica) and the genetic distance within one species from different area. Giving fingerprint of these oak individuals, we provided valuable data for further germplasm resource collecting and laid the foundation of the construction of oak germplasm resource nursery using oaks with clear genetic background, providing experimental basis for subsequent molecular breeding and other molecular biological and genetic studies of oaks native to China.

Materials And Methods

Sample distribution and plant materials

Briefly, in this study, we chose 6 representative regions for sampling in Beijing, which have abundant oak resources (see Supplementary Table S1 for details). Q. variabilis, Q. aliena, Q. dentata and Q. mongolica were studied in this work. Sampling sites are mostly forest parks rich in forest resources. Except for the Q. variabilis and Q. aliena populations in Jiufeng, other populations are natural populations. These areas are the north border of the North China Plain and belong to warm temperate semi-humid continental monsoon climate zone, with 4 distinct seasons and abundant rainfall.

According to the earlier investigation on the spot, all individuals in good growth condition were labeled and numbered. After photographing and recording the position, young leaves or dormant buds were collected. All the samples were frozen in liquid nitrogen and stored at −80 °C until used for DNA extraction and isolation. Numbers of samples in each region are shown in Supplementary Table S1.

DNA isolation and SSR amplification

Genomic DNA of 400 samples was extracted using Plant DNA Isolation Mini Kit (Vazyme, China), according to the user manual and modified cetyltrimethyl ammonium bromide method was also used as previously described (Yue et al. 2014; Wang et al. 2020). The DNA was eluted into 50 μL Tris-EDTA buffer as stock solution. Subsequently the quality and the concentration of the genomic DNA were quantified using 1% agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) respectively. The genomic DNA was diluted and separated to 20 ng/μL aliquots and stored at -20°C for polymerase chain reactions (PCR) amplification. Nine polymorphic SSR primer pairs were selected as previously reported (Kampfer et al. 1998), namely QrZAG96, QrZAG102, QrZAG112, QrZAG7, Qden03011, Qden03021, Qden03032, Qden05011 and Qden05031. The information of primers were listed in Supplementary Table S2.

The nine pairs of verified SSR primers with M13 (-21) tail at their 5’ end were used for PCR amplification together with M13 (-21) tailed fluorescent primers (M13-FAM, M13-ROX, M13-HEX) as previously described (Schuelke 2000; Wang et al. 2020). PCRs were conducted in 15μL reaction volumes containing 1 μL of genomic DNA (20 ng/μL), 7.5μL of 2×Rapid Taq Master Mix (Vazyme), 0.4 μL of fluorescent
primer, 0.1 μL of forward SSR primer, 0.5 μL of reverse primer and 5.5 μL of ddH$_2$O. PCR procedures were as follows: 95 ºC (5 min), 32 cycles at 95 ºC (30 s) / 55-59 ºC (30 s) (depending on the Tm value of SSR primers) / 72 ºC (30 s), followed by 8 cycles 95 ºC (30 s) / 53 ºC (30 s) / 72 ºC (15 s) and a final extension at 72 ºC for 5 min. PCR products were sent to Ruibo BioTech (Beijing, China) for capillary fluorescence electrophoresis detection to read the length of the PCR products.

**Data analysis**

**Characteristics and genetic diversity of SSR primers**

GeneMarker 2.2.0 software was used to read the polymorphic loci and obtain the genotypes of all samples. Linkage disequilibrium analysis was tested for all locus pairs in each population by randomization using Arlequin 3.1 (Excoffier et al. 2005). Polymorphism parameters of 9 pairs of SSR markers were calculated by GenAlEx6.5 (Peakall and Smouse 2012), respectively by the number of alleles (Na), effective alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho), expected heterozygosity (He), Fixation Index (F) and so on.

**Genetic differentiation among oak populations and optimal population classification**

The analysis of the population structure was assessed in STRUCTURE 2.3.4 (Pritchard et al. 2000) using a Bayesian clustering approach setting parameters with a burn-in period of 100,000 iterations and 100,000 MCMC iterations after burn-in. In this study, we used “admixture model” and “allele frequencies correlated” during the modeling process. This approach revealed genetic structure by assigning individuals or predefined groups to $K$ clusters. Different $K$ values that ranged from 1 to 10 were used to infer the number of clusters for 10 replicate runs. $\Delta K (\Delta K=\text{mean (|L'(K)|)/sd[L(K)]})$ (EVAANNO, REGNAUT and GOUDET 2005, 2611-2620) and InPPK were showed in STRUCTURE Harvester (Earl and vonHoldt 2011) using the results of STRUCTURE.

Different results could be produced in the same conditions by different 10 replicate cluster analyses of the same data. Thus, CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) and R 4.2.0 were used to average these 10 replicate results and visualize the outcomes separately. Proportion of gene pools distributed in 10 oaks populations were plotted on maps in the form of pie charts. The clustering tree of both individuals (NJ) and populations (UPGMA) were made using PowerMarker 3.25 (Liu and Muse 2005). Trees were then coloured and modified using Adobe Illustrator CC 2018.

**Genetic diversity of 11 oak population and PCoA analysis**

Na, Ne, Ho, He and fixation index (F), Fst and gene flow (Nm) among 11 populations were calculated by GenAlEx6.5. The software was used to conduct the PCoA of the four oak species and analyze the length range of the amplified alleles among different species. According to the length range of the amplified alleles, histogram was showed by using SigmaPlot 12.5.
Results

Polymorphism of the chosen SSR primers

As previous studies have validated that the 9 pairs of SSR markers have high polymorphism and stability (Wang 2012), we chose these SSR markers to identify the genetic background of the four oak species (Q. variabilis, Q. mongolica, Q. dentata, and Q. aliena). Meanwhile, the results of LD analysis showed 9 locus are in linkage equilibrium. First, we collected 400 individual samples (including 136 of Q. variabilis, 113 of Q. mongolica, 77 of Q. dentata, and 74 of Q. aliena) from 6 regions of Beijing (see Supplementary Table S1 for detail). We used these 400 samples to validate the polymorphism of the 9 SSR markers. The 9 SSR markers amplified 215 alleles in total, with an average of 23.9 alleles per locus (Table 1). Each SSR marker at least amplified 13 alleles for all the individuals. The observed heterozygosity (Ho) ranged from 0.45 to 0.82, with an average of 0.61. For all the observed loci, the Ho value was lower than the expected (He), which was ranged from 0.81 to 0.96 with an average of 0.87 (Table 1).

| Locus      | Na  | Ne   | Ho  | He  | I   | F   |
|------------|-----|------|-----|-----|-----|-----|
| QrZAG96    | 27  | 7.47 | 0.64| 0.87| 2.45| 0.26|
| QrZAG102   | 38  | 22.37| 0.59| 0.96| 3.28| 0.38|
| QrZAG112   | 23  | 7.44 | 0.65| 0.87| 2.36| 0.25|
| QrZAG7     | 24  | 12.02| 0.82| 0.92| 2.71| 0.10|
| Qden03011  | 13  | 7.75 | 0.72| 0.87| 2.18| 0.17|
| Qden03032  | 20  | 7.30 | 0.39| 0.86| 2.34| 0.55|
| Qden03021  | 13  | 5.31 | 0.47| 0.81| 1.83| 0.43|
| Qden05011  | 24  | 8.77 | 0.76| 0.89| 2.41| 0.14|
| Qden05031  | 33  | 5.74 | 0.45| 0.83| 2.35| 0.46|
| Mean       | 23.9| 9.35 | 0.61| 0.87| 2.43| 0.30|

Genetic Variation Within Oak Species And Populations

We further analyzed the genetic variation of populations of different oak species from different region by using SSR amplification results. Overall, Q. variabilis has the lowest Ho and He value among the four species while Q. mongolica has the highest Ho and He value (Table 2). In Q. variabilis, the number of alleles per locus (Na) ranged from 6.56 to 10.89 among different populations in different region and the average Na is 8.61. Q. variabilis populations on Dayang Mountain and Zhoukoudian have lower Ho as
compared to those on Jiufeng and Shangfang Mountain, with the highest inbreeding coefficient ($F = 0.26$ and 0.31).

### Table 2

#### Genetic variation of different oak populations

| Species         | Populations | Na   | Ne   | Ho   | He   | $F$  |
|-----------------|-------------|------|------|------|------|------|
| *Q. variabilis*  | JF-Qv       | 6.89 | 4.19 | 0.59 | 0.66 | 0.04 |
|                 | ZK-Qv       | 10.11| 4.92 | 0.50 | 0.71 | 0.31 |
|                 | SF-Qv       | 6.56 | 4.59 | 0.56 | 0.67 | 0.18 |
|                 | DY-Qv       | 10.89| 4.64 | 0.49 | 0.69 | 0.26 |
|                 | Qv-mean     | 8.61 | 4.59 | 0.54 | 0.68 | 0.20 |
| *Q. aliena*     | JF-Qa       | 10.11| 5.90 | 0.59 | 0.73 | 0.19 |
|                 | SF-Qa       | 12.11| 6.49 | 0.61 | 0.76 | 0.21 |
|                 | Qa-mean     | 11.11| 6.19 | 0.60 | 0.75 | 0.20 |
| *Q. dentata*    | SF-Qd       | 6.22 | 4.77 | 0.61 | 0.76 | 0.18 |
|                 | ZK-Qd       | 3.33 | 2.78 | 0.52 | 0.59 | 0.07 |
|                 | DY-Qd       | 13.11| 6.19 | 0.69 | 0.81 | 0.14 |
|                 | Qd-mean     | 7.56 | 4.58 | 0.61 | 0.72 | 0.13 |
| *Q. mongolica*  | YM-Qm       | 16.89| 8.28 | 0.67 | 0.86 | 0.22 |
|                 | BH-Qm       | 12.67| 8.15 | 0.74 | 0.86 | 0.14 |
|                 | Qm-mean     | 14.78| 8.21 | 0.70 | 0.86 | 0.18 |

In SF-Qa population, Na, Ne, Ho and He values are higher than those in JF-Qa population while they have similar sample amount (35 and 39 in SF and JF respectively, Supplementary Table S1). And the inbreeding coefficient of SF-Qa population is lower than that in JF-Qa population (Table 2), suggesting the higher genetic diversity of *Q. aliena* populations in Shangfang Mountain where distributed natural populations. For *Q. mongolica*, BH-Qm population has higher Ho value and much lower inbreeding coefficient as compared to YM-Qm population, indicating that *Q. mongolica* in Baihua Mountain has abundant genetic diversity.

Then we calculated the genetic differentiation coefficient ($F_{st}$) and gene flow (Nm) among populations within the same oak species in different region. The results showed that *Q. dentata* had the highest genetic differentiation coefficient and lowest gene flow among populations from Shangfang Mountain, Dayang Mountain and Zhoukoudian (Table 3). *Q. mongolica* had the lowest $F_{st}$ (0.01), indicating that only 1% variation happened between YM-Qm population and BH-Qm population.
### Table 3
Genetic variation among populations within the same oak species in different region

| Populations | Fst  | Nm    |
|-------------|------|-------|
| JF-Qv       | SF-Qv| DY-Qv | ZK-Qv | 0.04 | 6.16 |
| JF-Qa       | SF-Qa| -     | -     | 0.02 | 12.25|
| -           | SF-Qd| DY-Qd | ZK-Qd | 0.08 | 2.80 |
| YM-Qm       | BH-Qm| -     | -     | 0.01 | 17.61|

### Population Structure And Gene Introgression

A neighbor-joining tree was conducted to show the genetic relationship among all the observed individuals. The results showed that *Q. mongolica* might be the first evolved species among these species and *Q. aliena* and *Q. dentata* evolved separately. Most of the individuals of *Q. variabilis* had the longest genetic distance with the other three species. Notably, a few individuals of *Q. variabilis* were grouped with *Q. dentata* and *Q. mongolica* (Fig. 1), indicating that possible gene introgression has happened among species. Meanwhile, *Q. dentata*, *Q. aliena*, and *Q. mongolica* had more individuals grouped with each other, which suggested that gene introgression happened more frequently among these three species, confirming that these three species had closer relationship as compared to *Q. variabilis*.

The population structure of 400 oak individuals was also confirmed by principal coordinates analysis (PCoA). We first used all the samples to do the PCoA analysis and the results showed that the first coordinate separated *Q. variabilis* and the other three species while the other three species were grouped together (Fig. 2a). In addition, PCoA analysis by using the samples for each species in different regions could not generate the significant group (Supplementary Fig. S1), the results show that there was no distinct genetic differentiation among the different populations from different regions in Beijing area for each oak species.

The 400 oak individuals were evaluated for population stratification. The assumed number of clusters was set from $K = 1$ to 10. The Bayesian analysis by STRUCUTRE Harvester showed that the individuals were suitable to be divided into four groups ($K = 4$) (Supplementary Fig. S2). The STRUCTURE results showed that *Q. variabilis* (dark blue) has a relative clear genetic background but still a few individuals had gene introgression with *Q. dentata* (cyan) (Fig. 2b). This result was consistent with PCoA analysis and phylogenetic tree. *Q. mongolica* had abundant genetic diversity and had gene introgression individuals from *Q. aliena* and *Q. dentata*, indicating that *Q. mongolica*, had a closer genetic relationship with *Q. aliena* and *Q. dentata* as compared to *Q. variabilis* as we could also see the population structure when $K = 2$ and $K = 3$ (Supplementary Fig. S3).
The regions where we collected samples are located on the edge of the North China Plain (Fig. 3), and they are near the borders of natural distribution of _Q. mongolica_ and _Q. aliena_. Using the data processed from STRUCTURE, we drew the pies on the map of Beijing area to show the genetic background of all the populations. Southwest of Beijing had abundant species diversity of deciduous oaks, while northeast of Beijing we mainly collected _Q. mongolica_ (Fig. 3).

We then used the Unweighted Pair-Group Method and the Arithmetic (UPGMA) algorithm to build the phylogenetic tree of populations of 4 species from different regions in Beijing area. Populations from one species were grouped together and _Q. mongolica_ had a closer genetic relationship with _Q. aliena_ than that with _Q. dentata_ (Fig. 4). Notably, JF-Qv population, which was an artificial population, had a further genetic distance with other three _Q. variabilis_ populations, indicating that the individuals might transplanted from other area outside Beijing.

**Dna Fingerprints Of Oak Species In Beijing Area**

As there was no oak germplasm nursery with clear genetic background in China, here we provided the fingerprints of the observed 400 oak individuals. According to above results, 10 trees were selected for each species to facilitate future studies. The first five are the single genetic background, and the last five are complex. The DNA fingerprints of these 40 individuals were represented with numbers that showed the allele length and letters that indicated the amplified SSR alleles (Supplementary Table S3). The length of amplified alleles varied among different species. Qden03021 had a large range of amplicons in _Q. variabilis_ as compared to the other species while Qden05031 had a large range of amplicons in _Q. mongolica_ (Supplementary Fig. S4), suggesting the different polymorphism of different alleles in different oak species.

**Discussion**

Oaks are very important species in the world, not only for their ecologic and economic value but also for their research value. Oaks are proved to be the successfully evolved species in the world with extensive radiation and expansion (Kremer and Hipp 2020) and they are ideal materials to study gene introgression and natural hybridization of woody plants (Curtu et al. 2009; An et al. 2017; Crowl et al. 2020). Hipp et al. (Hipp et al. 2020) have used genotyping-by-sequencing and SNP markers to generate a large-scale data, demonstrating frequent gene diversity and introgression of oak species in _Quercus_ genus. Here we conducted fine-scale study to reveal the genetic relationship of oak species at the junction between Taihang Mountain and Yanshan Mountain, taking a detail look at the genetic diversity within and among species, and also observed the gene introgression among these oak species. We collected 400 individuals of oak species in the area, which is the north border of North China Plain. In this area, several oak species are naturally distributed including _Q. variabilis, Q. aliena, Q. dentata_ and _Q. mongolica_, which are species native to China. As there is no improved cultivar for any of these oak species native to China, it is
meaningful to detect the genetic variation of these individuals, preparing for germplasm nursery using accessions with clear genetic background.

Previous studies have screened some SSR markers with high polymorphism for oak species (Kampfer et al. 1998; Guo 2018). Here, we used these SSR markers and further confirmed the polymorphism of the picked 9 markers. For the individuals of the 4 oak species in Beijing area, these SSR markers exhibit high polymorphism and 215 alleles were detected in total and 23.9 in average (Table 1), much higher than the studies in *Quercus petraea* (Lupini et al. 2019) and similar with the studies in *Q. mongolica* and *Q. liaotungensis* which collected 1166 individuals (Zeng et al. 2011). This result indicates that these SSR markers are efficient for oak species characterization (Guo 2018). Range size shows some relationship to heterozygosity and allelic richness (Spence et al. 2021), in that *Q. mongolica* had the highest heterozygosity (He = 0.86) and allelic richness (Na = 14.78), suggesting that *Q. mongolica* had the most widespread distribution among the 4 species. Naturally, the phylogenetically close relationship between *Q. dentata* and *Q. aliena* with *Q. mongolica* could be well explained. Whether they originated and evolved in *Q. Mongolica* remains to be verified. Both total alleles and expected heterozygosity (He) were lower than the studies in *Q. variabilis* on a national scale (Shi et al. 2017), speculating that it might due to artificial population. Interestingly, the inbreeding coefficient of the artificially planted *Q. variabilis* population (JF-Qv) was 0.04 (Table 2), suggesting that the individuals probably came from random mating populations. Meanwhile, the UPGMA analysis showed that JF-Qv artificial population had a phylogenetically distant relationship with other *Q. variabilis* populations in this area (Fig. 4), indicating that the trees in Jiufeng might be transplanted from other area outside Beijing. Thus, the sampled individuals might come from different provenance. In contrast, the other artificial population, *Q. aliena* population in Jiufeng (JF-Qa), had a strong gene flow (Nm = 12.25) with the natural population in Shangfang Mountain (SF-Qa) (Table 3), suggesting that some individuals in JF-Qa population might be transplanted from Shangfang Mountain. Gene flow determines the genetic structure and survival potential of future populations of a species (Hamrick et al. 1992), previous study reported a lower gene flow (Nm = 3.648) among the population of *Q. variabilis* than our study (Shi et al. 2017), which objectively reflected *Q. variabilis* distributed in the junction between Taihang Mountain and Yanshan Mountain had more survival potential. In addition, SF-Qa natural population has more abundant genetic diversity than JF-Qa artificial population (Table 2), further proving that JF-Qa might be generated from the same provenance.

The phylogenetic tree combining with UPGMA analysis and the STRUCTURE analysis showed that *Q. variabilis* is evolved separately from the other three species (Figs. 1, 2 & 4), which is consistent with previous studies using SNP markers and limited individuals within one species (Hipp et al. 2020). As Hipp et al. (Hipp et al. 2020) reported, *Q. mongolica*, *Q. aliena* and *Q. dentata* belong to *Quercus* section *Roburoids*, while *Q. variabilis* belongs to *Cerris* section *East Asia Cerris* together with *Q. accutisima* and *Q. dentata* evolves separately from *Q. mongolica* and *Q. aliena*. The introgression of *Q. variabilis* and *Q. accutisima* have already been confirmed by using high-quality genomic resources (Fu et al. 2022). Our results also demonstrate that *Q. mongolica* and *Q. aliena* have a closer genetic relationship as compared to *Q. dentata* (Figs. 1, 2 & 4). The gene flow between BH-Qm and YM-Qm reached 17.61, much higher
than the other 3 species. Considering the geographical distance between Baihua Mountain and Yunmeng Mountain, the two *Q. mongolica* populations might be evolved together and separated in recent time. As many previous studies have demonstrated that gene introgression happens frequently in *Quercus* genus (Moran et al. 2012; An et al. 2017; Crowl et al. 2020; Manos and Hipp 2021), it could not be ignored when analyzing the oak population in Beijing area. The STRUCTURE analysis and the phylogenetic tree showed that gene introgression individuals were grouped with corresponding species and *Q. mongolica* has the most individuals that have gene introgression with *Q. dentata* and *Q. aliena* (Figs. 1 & 2). *Q. variabilis* has the least gene introgression individuals, suggesting that *Q. variabilis* has a phylogenetically distant relationship with the other three species. Presumably due to the lower successful rate of interspecific hybridization with the other three species in natural conditions. This result provides guidance for distant hybridization breeding of *Q. variabilis*, which should take the hybridization compatibility into consideration. In addition, as previously reported that *Q. variabilis* has distinct flowering time compared to the other 3 species (Liu 2020), which might also cause the less chance for interspecific hybridization (Schermer et al. 2020). Previous study on *Q. dentata* and *Q. aliena* showed that the male flowers of *Q. dentata* the female flowers of *Q. aliena* in this area had a certain overlap (Liu et al. 2018), suggesting these 2 oaks had frequent outcrossing.

The junction between Taihang Mountain and Yanshan Mountain is not only near the north border of the natural distribution of *Q. variabilis* and *Q. aliena* but also near the south border of the natural distribution of *Q. mongolica*. *Q. mongolica* distributed on the mountains with an altitude over 600 m. The north part of this distribution area is high while the south part is low topographically with warm temperate subhumid continental monsoon climate. Considering the pollen dispersal of oak species and the northwest wind direction in spring in the junction between Taihang Mountain and Yanshan Mountain, it could be explained that southwest of Beijing had abundant species diversity of deciduous oaks, which may also due to the warmer climate (Fig. 3). Fine-scale spatial genetic structure of the oak species could be further studied in Beijing area to further identify the interspecific hybrids and their adaption to the environment (Valbuena-Carabana et al. 2007; Curtu et al. 2009; Curtu et al. 2014). During our investigation on the spot, we observed the leaf shape of *Q. dentata* were similar to *Q. mongolica* (Supplementary Fig. S5), which were quite different from the leaf of *Q. variabilis*. Further studies could use SNP markers to link the morphological traits or environmental adaptions (eg. temperature, drought, salt and etc.) and the genotype and more oak species native to China should be identified by molecular markers. With the clear identified genetic background, we could take a use of these genetic resources for breeding.

DNA fingerprints could give the putative cultivar an identity, which is very important for intellectual property protection. Here we selected 10 individuals for each tree species and provided their DNA fingerprints through STRUCTURE, PCoA and phylogenetic tree analysis (Supplementary Table S3). Further study could be done to get the link between DNA fingerprints and their morphological traits by using more SSR markers or SNP markers like studies in some other species (Riangwong et al. 2020; Luo et al. 2021), making it clearer to focus on some specific trait breeding.
In conclusion, we detected the genetic variation of natural and artificial population of four oak species native to China and confirmed the genetic relationship and gene introgression of *Q. variabilis*, *Q. mongolica*, *Q. aliena* and *Q. dentata*. Our work laid the foundation for collecting oak germplasm and screened some individuals with clear genetic background for whole genome sequencing and cross breeding.

**Declarations**

All authors declare that they have no competing interest of this work.

**Acknowledgements**

We thank Dr. Yan Wang for providing useful suggestions on SSR data analysis and manuscript preparing. This work was supported by the Fundamental Research Funds for the Central Universities (2021ZY11) and the China Postdoctoral Science Foundation (2021M700451).

**Data Archiving Statement**

All data generated or analysed during this study are included in this published article [and its supplementary information files].

**Author Contributions**

Q. Y. and G. L. designed the experiments. Z. P. conducted most of the experiment, analyzed the data. Z.P. and Q.Y. wrote the manuscript. Z. P., X. C. and J. L. conducted DNA extraction and SSR PCR experiment. Y. Z. analyzed the data and revised the manuscript. Z. P., Q. Y., X. C., J. L., X. Y., C. H., Y. C., Q. Z., N. L., J. K., X. M., W. J., Y. L., H. Z. and J. W. collected the samples. Y. L and G. L supervised the project and revised the manuscript. Z. P. and Q. Y. contributed equally to this work.

**References**

1. An M, Deng M, Zheng SS, Jiang XL, Song YG (2017) Introgression Threatens the Genetic Diversity of Quercus austrocochinchinensis (Fagaceae), an Endangered Oak: A Case Inferred by Molecular Markers. Front Plant Sci 8:229. doi:10.3389/fpls.2017.00229

2. Burkardt K, Pettenkofer T, Ammer C, Gailing O, Leinemann L, Seidel D, Vor T (2020) Influence of heterozygosity and competition on morphological tree characteristics of Quercus rubra L.: a new single-tree based approach. New Forests 52 (4):679–695. doi:10.1007/s11056-020-09814-1

3. Crowl AA, Manos PS, McVay JD, Lemmon AR, Lemmon EM, Hipp AL (2020) Uncovering the genomic signature of ancient introgression between white oak lineages (Quercus). New Phytol 226 (4):1158–1170. doi:10.1111/nph.15842

4. Curtu AL, Craciunesc I, Enescu CM, Vidalis A, Sofletea N (2014) Fine-scale spatial genetic structure in a multi-oak-species (*Quercus* spp.) forest. iForest - Biogeosciences and Forestry 8 (3).
5. Curtu AL, Gailing O, Finkeldey R (2009) Patterns of contemporary hybridization inferred from paternity analysis in a four-oak-species forest. BMC Evol Biol 9:284. doi:10.1186/1471-2148-9-284

6. Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4 (2):359–361. doi:10.1007/s12686-011-9548-7

7. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

8. Fu R, Zhu Y, Liu Y, Feng Y, Lu RS, Li Y, Li P, Kremer A, Lascoux M, Chen J (2022) Genome-wide analyses of introgression between two sympatric Asian oak species. Nat Ecol Evol. doi:10.1038/s41559-022-01754-7

9. Ganopoulos I, Argiriou A, Tsaftaris A (2011) Adulterations in Basmati rice detected quantitatively by combined use of microsatellite and fragrance typing with High Resolution Melting (HRM) analysis. Food Chem 129 (2):652–659. doi:10.1016/j.foodchem.2011.04.109

10. Guo B (2018) Construction of SSR fingerprint and research of genetic structure in relative Quercus species. Journal of Beijing Forestry University 40 (5):10–18. doi:10.13332/j.1000-1522.20170444

11. Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New Forests 6:95–124

12. Hilde Nybom KWA BR (2014) DNA fingerprinting in botany: past, present, future. Investigative Genetics 5 (1)

13. Hipp AL, Manos PS, Hahn M, Avishai M, Bodenes C, Cavender-Bares J, Crowl AA, Deng M, Denk T, Fitz-Gibbon S, Gailing O, Gonzalez-Elizondo MS, Gonzalez-Rodriguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodriguez-Correa H, Schulze ED, Simeone MC, Sork VL, Valencia-Avalos S (2020) Genomic landscape of the global oak phylogeny. New Phytol 226 (4):1198–1212. doi:10.1111/nph.16162

14. Hubert F, Grimm GW, Jousselin E, Berry V, Franc A, Kremer A (2014) Multiple nuclear genes stabilize the phylogenetic backbone of the genus Quercus. Systematics and Biodiversity 12 (4):405–423. doi:10.1080/14772000.2014.941037

15. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23 (14):1801–1806. doi:10.1093/bioinformatics/btm233

16. Kampfer S, Lexer C, Glossl J, Steinkellner H (1998) Characterization of (GA)n microsatellite loci from Quercus robur. Hereditas 129:183–186

17. Kremer A, Hipp AL (2020) Oaks: an evolutionary success story. New Phytol 226 (4):987–1011. doi:10.1111/nph.16274

18. Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21 (9):2128–2129. doi:10.1093/bioinformatics/bti282
19. Liu P (2020) Studies on hybridization and seed dispersal of three Quercus species in Beijing area., Beijing Forestry University.

20. Liu X, Chang E, Liu J, Jiang Z (2021) Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China. Journal of Forestry Research 32 (5):2203–2218. doi:10.1007/s11676-020-01288-3

21. Liu Y, Li Y, Song J, Zhang R, Yan Y, Wang Y, Du FK (2018) Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: Quercus dentata Thunberg and Quercus aliena Blume (Fagaceae). Annals of Forest Science 75 (4). doi:10.1007/s13595-018-0770-2

22. López-Aljorna A, Bueno MA, Aguinagalde I, Martín JP (2007) Fingerprinting and genetic variability in cork oak (Quercus suber L.) elite trees using ISSR and SSR markers. Annals of Forest Science 64 (7):773–779. doi:10.1051/forest:2007057

23. Luo L, Yang Y, Zhao H, Leng P, Hu Z, Wu J, Zhang K (2021) Development of EST-SSR markers and association analysis of floral scent in tree peony. Scientia Horticulturae 289. doi:10.1016/j.scienta.2021.110409

24. Lupini A, Aci MM, Maucer A, Luzzi G, Bagnato S, Menguzzato G, Mercati F, Sunseri F (2019) Genetic diversity in old populations of sessile oak from Calabria assessed by nuclear and chloroplast SSR. Journal of Mountain Science 16 (5):1111–1120. doi:10.1007/s11629-018-5335-1

25. Manos PS, Hipp AL (2021) An Updated Infrageneric Classification of the North American Oaks (Quercus Subgenus Quercus): Review of the Contribution of Phylogenomic Data to Biogeography and Species Diversity. Forests 12 (6). doi:10.3390/f12060786

26. Mir C, Toumi L, Jarne P, Sarda V, Di Giusto F, Lumaret R (2006) Endemic North African Quercus afares Pomel originates from hybridisation between two genetically very distant oak species (Q. suber L. and Q. canariensis Willd.): evidence from nuclear and cytoplasmic markers. Heredity (Edinb) 96 (2):175–184. doi:10.1038/sj.hdy.6800782

27. Moran EV, Willis J, Clark JS (2012) Genetic evidence for hybridization in red oaks (Quercus sect. Lobatae, Fagaceae). Am J Bot 99 (1):92–100. doi:10.3732/ajb.1100023

28. Nagamitsu T, Uchiyama K, Izuno A, Shimizu H, Nakanishi A (2020) Environment-dependent introgression from Quercus dentata to a coastal ecotype of Quercus mongolica var. crispula in northern Japan. New Phytol 226 (4):1018–1028. doi:10.1111/nph.16131

29. Nybom H (1991) Applications of DNA Fingerprinting in Plant Breeding. In: DNA Fingerprinting: Approaches and Applications. pp 294–311

30. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28 (19):2537–2539. doi:10.1093/bioinformatics/bts460

31. Plomion C, Aury JM, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonne N, Labadie K, Le Provost G, Lesur I, Bartholome J, Faivre-Rampant P, Kohler A, Leple JC, Chantret N, Chen J, Diewart A, Alaeitabar T, Barbe V, Belser C, Berges H, Bodenes C, Bogeat-Triboulot MB, Bouffaud ML, Brachi B, Chancerel E, Cohen D, Couloux A, Da Silva C, Dossat C, Ehrenmann F, Gaspin C, Grima-Pettenati J,
32. Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959

33. Ramos-Ortiz S, Oyama K, Rodríguez-Correa H, González-Rodríguez A (2016) Geographic structure of genetic and phenotypic variation in the hybrid zone between Quercus affinis and Q. laurina in Mexico. Plant Species Biology 31 (3):219–232. doi:10.1111/1442-1984.12109

34. Riangwong K, Wanchana S, Aesomnuk W, Saensuk C, Nubankoh P, Ruanjaichon V, Kraithong T, Toojinda T, Vanavichit A, Arikit S (2020) Mining and validation of novel genotyping-by-sequencing (GBS)-based simple sequence repeats (SSRs) and their application for the estimation of the genetic diversity and population structure of coconuts (Cocos nucifera L.) in Thailand. Hortic Res 7:156. doi:10.1038/s41438-020-00374-1

35. Schermer E, Bel-Venner MC, Gaillard JM, Dray S, Boulanger V, Le Ronce I, Oliver G, Chune I, Delzon S, Venner S (2020) Flower phenology as a disruptor of the fruiting dynamics in temperate oak species. New Phytol 225 (3):1181–1192. doi:10.1111/nph.16224

36. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18:233–234

37. Shi X, Wen Q, Cao M, Guo X, Xu L-a (2017) Genetic Diversity and Structure of Natural Quercus variabilis Population in China as Revealed by Microsatellites Markers. Forests 8 (12). doi:10.3390/f8120495

38. Spence ES, Fant JB, Gailing O, Griffith MP, Havens K, Hipp AL, Kadav P, Kramer A, Thompson P, Toppila R, Westwood M, Wood J, Zumwalde BA, Hoban S (2021) Comparing Genetic Diversity in Three Threatened Oaks. Forests 12 (5). doi:10.3390/f12050561

39. Stavridou E, Lagiotis G, Karapetsi L, Osathanunkul M, Madesis P (2020) DNA Fingerprinting and Species Identification Uncovers the Genetic Diversity of Katsouni Pea in the Greek Islands Amorgos and Schinoussa. Plants (Basel) 9 (4). doi:10.3390/plants9040479

40. Valbuena-Carabana M, Gonzalez-Martinez SC, Hardy OJ, Gil L (2007) Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. Mol Ecol 16 (6):1207–1219. doi:10.1111/j.1365-294X.2007.03231.x

41. Wang Y (2012) Interspecific hybridization of Quercus dentata, Q.mongolica-Q. wutaishansea based on SSR markers. SHANDONG UNIVERSITY,

42. Wang Y, Jia HM, Shen YT, Zhao HB, Yang QS, Zhu CQ, Sun DL, Wang CY, Zhou CC, Jiao Y, Chai CY, Yan LJ, Li XW, Jia HJ, Gao ZS (2020) Construction of an anchoring SSR marker genetic linkage map and detection of a sex-linked region in two dioecious populations of red bayberry. Hortic Res 7 (1):53. doi:10.1038/s41438-020-0276-6
43. Yang J, Vazquez L, Chen X, Li H, Zhang H, Liu Z, Zhao G (2017) Development of Chloroplast and Nuclear DNA Markers for Chinese Oaks (Quercus Subgenus Quercus) and Assessment of Their Utility as DNA Barcodes. Front Plant Sci 8:816. doi:10.3389/fpls.2017.00816

44. Yue XY, Liu GQ, Zong Y, Teng YW, Cai DY (2014) Development of genic SSR markers from transcriptome sequencing of pear buds. J Zhejiang Univ Sci B 15 (4):303–312. doi:10.1631/jzus.B1300240

45. Zeng YF, Liao WJ, Petit RJ, Zhang DY (2011) Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. Mol Ecol 20 (23):4995–5011. doi:10.1111/j.1365-294X.2011.05354.x

46. Zurn JD, Nyberg A, Montanari S, Postman J, Neale D, Bassil N (2020) A new SSR fingerprinting set and its comparison to existing SSR-and SNP-based genotyping platforms to manage Pyrus germplasm resources. Tree Genetics & Genomes 16 (5). doi:10.1007/s11295-020-01467-7

**Figures**
Figure 1

Neighbor-joining tree of 400 oak individuals from 4 species

Different colors indicate different species. Length of lines show the genetic distance. The scale bar represents 0.1 genetic distance.
Figure 2

PCoA and population structure for all the 400 individuals

(A) PCoA groups of four oak species. Different colors indicate different species. (B) Population structure based on Bayesian clustering approaches with K = 4 by STRUCTURE software. Each bar represent a
Figure 3

Population distribution map and genetic background of all samples

Different colors indicate the different genetic clusters in STRUCTURE analysis.
Figure 4

The UPGMA analysis of populations from different regions

Eleven populations mentioned above were evaluated. The scale bar represents 0.05 genetic variation. The abbreviations of the population names were listed in Table S1.

Supplementary Files

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

- SupplementaryInformation.docx