Low genetic diversity in the endangered *Taxus yunnanensis* following a population bottleneck, a low effective population size and increased inbreeding

Y. C. Miao, Z. J. Zhang and J. R. Su

Research Institute of Resource Insects, Chinese Academy of Forest (CAF), Kunming, 650233, Yunnan, China

*Corresponding author: J.R. Su, Phone: +86-871-63860017, E-mail: jianrongsu@vip.sina.com

Abstract

*Taxus yunnanensis*, which is an endangered tree that is considered valuable because it contains the effective natural anticancer metabolite taxol and heteropolysaccharides, has long suffered from severe habitat fragmentation. In this study, the levels of genetic diversity in two populations of 136 individuals were analyzed based on eleven polymorphic microsatellite loci. Our results suggested that these two populations were characterized by low genetic diversity ($N_E = 2.303/2.557; H_O = 0.168/0.142; H_E = 0.453/0.517$), a population bottleneck, a low effective population size ($N_e = 7/9$), a high level of inbreeding ($F_{IS} = 0.596/0.702$), and a weak, but significant spatial genetic structure ($S_p = 0.001, b = -0.001^*$). Habitat fragmentation, seed shadow overlap and limited seed and pollen dispersal and potential selfing may have contributed to the observed genetic structure. The results of the present study will enable development of practical conservation measures to effectively conserve the valuable genetic resources of this endangered plant.

**Keywords**: bottleneck, effective population size ($N_e$), inbreeding, spatial genetic structure (SGS), *Taxus yunnanensis*

Introduction

The genetic diversity of a species is important to its current persistence and long-time evolutionary potential (Anderson et al., 2011). The levels of genetic variation of a species are mainly determined by ecological, demographic, and life history (Nevo et al., 1984). However, recent human-induced habitat fragmentation has had a severely adverse impact on genetic diversity, resulting in decreased population sizes (Martin et al., 2014), reduced effective population sizes ($N_e$) (Manoel et al., 2012), increased inbreeding (Vu et al., 2016) and the occurrence of strong spatial genetic structure (SGS; Sjölund and Jump, 2015), which can severely reduce within-population genetic diversity and adaptive ability under changing environmental conditions. Thus, understanding the levels of genetic diversity is crucial to developing practical and effective conservation strategies for endangered species.

*Taxus yunnanensis* is one of six *Taxus* species in China (Zhang et al., 2012) that is mainly distributed in the southwest region (Fu et al., 1999). This species is a patchily distributed, understory woody gymnosperm that has an extremely high medicinal value because of the presence of taxol and heteropolysaccharides, which are natural anti-cancer substances that are effective for treatment of ovarian, breast and liver cancers (Yin et al., 2010; Hai et al., 2014). In China, a number of natural *T. yunnanensis* populations have been led to smaller and more isolated populations because of natural habitat fragmentation caused by excessive anthropogenic logging for taxol extraction (Miao et al., 2014). Such fragmentation is likely to increase the negative effects of biparental inbreeding; however, it has been identified as an endangered plant on the national red list of China (State Forestry Bureau, 1999).

In the genus *Taxus*, sexual formations are complex, including the exclusive monoecious *T. canadensis* (Wilson et al., 1996), pure dioecious *T. wallchiana* (Vu et al., 2016) and *T. contorta* (Poudel et al., 2014) and mixed *T. brevifolia* (Hogg et al., 1996) and *T. baccata* (Di Cosmo, 2005). Studies of genetic
diversity have shown that many exclusive and mixed Taxus species have an especially high level of inbreeding, with FIS values of 0.226 to 0.472 (El-Kassaby and Yanchuk, 1994; Chung et al., 1999; Myking et al., 2009; Dubreuil et al., 2010; Zhang and Zhou, 2013; Poudel et al., 2014; Vu et al., 2016). T. yunnanensis has long been believed to be a strictly dioecious tree; however, coosexuality has recently been confirmed in this species (Wang et al., 2008).

Although it is limited and only based on isoenzymes, present genetic knowledge regarding T. yunnanensis indicates that it is on the verge of extinction (Chen et al., 2001). To better understand the levels of genetic diversity in this species, bottleneck, SGS, Ne and inbreeding analyses within two populations were conducted using microsatellite markers. Practical conservation strategies for this endangered species were then proposed based on the results.

Material and Methods

Study populations and sampling

Two natural T. yunnanensis populations, TY1 (n = 79 in a 79x69 m area; 26°39’10.08” N, 99°36’7.2” E) and TY2 (n = 57 in a 132x97 m area; 26°27’18.36” N, 99°14’27.24” E) separated by approximately 42 km and located in Lanping County, Yunnan Province, China were selected for the study (see Table S1). The tree diameter at breast height (DBH) and tree height ranged from 0 to 46.2 cm and 12.7 to 90.4 cm, with a mean of 11.1 cm and 44.7 cm, and from 0 to 20 m and 3 to 20 m, with a mean of 7.4 m and 10.1 m in TY1 (four seedlings in TY1) and TY2, respectively. According to the regression equation of the relationship between tree height and tree age Y = 1.09937X0.46538 (Y = tree age, X = tree height; Wang et al., 2006), the tree range from 0 to 509.6 y and 8.6 to 509.6 y, with a mean of 92.6 y and 144.7 y in TY1 and TY2, respectively. Based on the observation of an individual needing 15 years to grow to maturity (Wang et al., 2006), the tree age ranged from 0 to 509.6 y and 8.6 to 509.6 y, with a mean of 92.6 y and 144.7 y in TY1 and TY2, respectively. The relative spatial locations between individuals within each population are shown in Figure 1. Fresh leaves of each individual were collected in separate plastic bags containing silica gel and stored at room temperature prior to DNA extraction.

DNA extraction and microsatellite genotyping

Extraction of genomic DNA, primer sequences of the 11 polymorphic microsatellites (Miao et al., 2008), polymerase chain reaction (PCR) conditions, and microsatellite genotyping methods were performed essentially as described by Miao et al. (2014).

Data analyses

Increased linkage disequilibrium (LD) between loci can cause a reduction in Ne because of selection at linked sites (Charlesworth and Wright, 2001); thus LD tests between all pairs of loci were performed using FSTAT v 2.9.3 (Goudet, 1995) to evaluate the independence of loci.

Genetic variation within each population

INEST v 2.1 was used to calculate the observed heterozygosity (H0), expected heterozygosity (H), inbreeding coefficient (FIS) and null frequency (FN) of each locus and the mean FIS of each population with the Markov Chain Monte Carlo (MCMC) based on 5x105 cycles, 1x104 thinning and 5x105 burn-in periods. POPGENE v 1.31 was used to calculate the effective number of alleles (Ne) of each locus. The BOTTLENECK v 1.2.02 (Piry et al., 1999) program was used to infer the presence of demographic expansion/ contraction in each population using the infinite allele model (IAM). The same program was also used to assess the significance of homozygosity excess with 1x105 iterations. Finally, the program LDNE (Waples and Do, 2008) was employed to calculate the numbers of Ne, and their confidence intervals (CI) for these two populations.

Structure analyses of each population

To investigate the individual genetic relationship, viz., the numbers of family structures within each population, the program STRUCTURE v 2.2 (Pritchard et al., 2000) was used. This software calculates the proportion of membership of each individual to the inferred K clusters (K = number of potential groups). The most likely value of K was the smallest value of K that captured the major structure [maximum value of Ln Prob of Data (LnP(D))] in the data. For both populations, the K range was set from one to 12 (because K = 8 captured the maximum structure and a range of K of one to 12 clearly showed the data trend). Each simulation consisted of 5x106 MCMC iterations following a burn-in period of 5x105 iterations. Ten independent repeats of each K were carried out to quantify the mean amount of LnP(D).

Figure 1

Spatial distribution of Taxus yunnanensis samples collected from TY1 (a) and TY2 (b) populations.
Spatial genetic structure of each population
To examine the SGS within each population, the kinship coefficient \( F_s \) (Loiselle et al., 1995) was calculated for both populations using SPAGEDI v 1.3 (Hardy and Vekemans, 2002). For each population, 10 distance classes with equal numbers of sample pairs were defined. To visualize the SGS, the multilocus \( F_s \) for each distance class was plotted against the corresponding physical distance. Moreover, to quantify the extent of SGS in these two populations, the regression slopes \( b \) (Curtu et al., 2015), Sp (Vekemans and Hardy, 2004) and significance of both multilocus \( F_s \) values per distance class and \( b \) were also estimated using the SPAGEDI program. Finally, isolation by distance was estimated using GENEAIEX v 6.0 (Peakall and Smouse, 2006) by determining the correlation between \( F_s \) and the corresponding spatial geographic distances for the TY1 and TY2 populations.

Results
Genetic variation within each population
LD tests revealed 55 possible pairwise comparisons between the eleven loci for both populations that did not show significant linkage disequilibrium \( (P > 0.05) \). The genetic variation of both populations is shown in Table 1. The \( N_e, H_o, H_e \) and \( F_s \) ranged from 1.000 to 5.280, 0 to 0.797, 0 to 0.817 and 0.0006 to 0.0025, with mean values of 2.303, 0.168, 0.453 and 0.0018 per locus for TY1, while these values ranged from 1.036 to 6.315, 0 to 0.947, 0.035 to 0.848 and 0.0006 to 0.0024, with a mean of 2.557, 0.142, 0.517 and 0.0016 per locus for TY2. The inbreeding coefficients were relatively high (up to 1) for most loci, while the mean \( F_s \) reached up to 0.596 and 0.702 for TY1 and TY2, respectively. The effective population sizes were 7 and 9, with the bounds of CI varying between 4.4 and 10.1, and between 6.5 and 12.3, and estimates of \( Ne/N \) being 0.09 and 0.16 for TY1 and TY2, respectively. Bottleneck tests showed a significant size reduction in both populations \( (P < 0.01) \).

Structure analyses for each population
Structure analyses clearly showed that eight family structures existed in each population (Figure 2). The largest increase break based on the average \( LnP(D) \) of 10 repeats for \( K \) was 8, with a maximum value of \( LnP(D) \) = -932.12 and -791.00 for TY1 and TY2, respectively. The corresponding bar plots for \( K = 8 \) for both populations are shown in Figure 2.

Spatial genetic structure of each population
Spatial autocorrelation analyses revealed a weak, but significant genetic structure within both populations \( (b = -0.001; \text{Sp} = 0.001) \) (Figure 3). Moreover, the multilocus correlograms showed that significant SGS was apparent when the individual distance was less than 23 m and 50 m for the TY1 and TY2 populations, respectively. This was also supported by the structure analyses which identified eight kinship families within each population (Figure 2). In addition, significant and negative linear decreases in the pairwise kinship coefficient \( F_s \) with the geographical distance (Figure 4) were detected in TY1 \( (y = -0.001x + 0.0261; P = 0.006) \) and TY2 \( (y = -0.0011x + 0.0547; P = 0.000) \) populations, indicating that individuals physically close

Table 1
Genetic indices of two natural Taxus yunnanensis populations, TY1 and TY2

<table>
<thead>
<tr>
<th>Locus</th>
<th>( N_e )</th>
<th>( H_o )</th>
<th>( H_e )</th>
<th>( F_s )</th>
<th>( F_w )</th>
<th>( N_e )</th>
<th>( H_o )</th>
<th>( F_s )</th>
<th>( F_w )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TY05</td>
<td>2.265</td>
<td>0</td>
<td>0.566</td>
<td>1</td>
<td>0.0019</td>
<td>2.080</td>
<td>0</td>
<td>0.655</td>
<td>1</td>
</tr>
<tr>
<td>TY08</td>
<td>1.568</td>
<td>0</td>
<td>0.367</td>
<td>1</td>
<td>0.0024</td>
<td>1.407</td>
<td>0</td>
<td>0.294</td>
<td>1</td>
</tr>
<tr>
<td>TY12</td>
<td>1.640</td>
<td>0</td>
<td>0.395</td>
<td>1</td>
<td>0.0018</td>
<td>1.949</td>
<td>0</td>
<td>0.496</td>
<td>1</td>
</tr>
<tr>
<td>TY16</td>
<td>1.591</td>
<td>0.038</td>
<td>0.377</td>
<td>0.899</td>
<td>0.0019</td>
<td>1.693</td>
<td>0</td>
<td>0.417</td>
<td>1</td>
</tr>
<tr>
<td>TY24</td>
<td>1.026</td>
<td>0</td>
<td>0.025</td>
<td>1</td>
<td>0.0014</td>
<td>1.036</td>
<td>0</td>
<td>0.035</td>
<td>1</td>
</tr>
<tr>
<td>TY27</td>
<td>1.733</td>
<td>0</td>
<td>0.428</td>
<td>1</td>
<td>0.0025</td>
<td>2.714</td>
<td>0</td>
<td>0.643</td>
<td>1</td>
</tr>
<tr>
<td>TY29</td>
<td>2.522</td>
<td>0</td>
<td>0.611</td>
<td>1</td>
<td>0.0025</td>
<td>2.940</td>
<td>0</td>
<td>0.672</td>
<td>1</td>
</tr>
<tr>
<td>TY44</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0014</td>
<td>1.407</td>
<td>0</td>
<td>0.294</td>
<td>1</td>
</tr>
<tr>
<td>TS07</td>
<td>3.806</td>
<td>1</td>
<td>0.744</td>
<td>-0.344</td>
<td>0.0007</td>
<td>6.315</td>
<td>0.947</td>
<td>0.848</td>
<td>-0.117</td>
</tr>
<tr>
<td>TB01</td>
<td>2.844</td>
<td>0.013</td>
<td>0.651</td>
<td>0.981</td>
<td>0.0021</td>
<td>2.798</td>
<td>0</td>
<td>0.676</td>
<td>1</td>
</tr>
<tr>
<td>TW01</td>
<td>5.280</td>
<td>0.797</td>
<td>0.817</td>
<td>0.024</td>
<td>0.0006</td>
<td>2.883</td>
<td>0.614</td>
<td>0.659</td>
<td>0.069</td>
</tr>
<tr>
<td>Average</td>
<td>2.303</td>
<td>0.168</td>
<td>0.453</td>
<td>0.596</td>
<td>0.0018</td>
<td>2.557</td>
<td>0.142</td>
<td>0.517</td>
<td>0.702</td>
</tr>
</tbody>
</table>

\( N_e \), effective number of alleles; \( H_o \), observed heterozygosity; \( H_e \), expected heterozygosity; \( F_s \), inbreeding coefficients; \( F_w \), null allele frequencies; \( Ne \), effective population sizes and their confidence intervals in parenthesis; \( Ne/N \), ratio of \( Ne \) to the number of census population sizes; P-value is significance of no excess of heterozygote; \( *P < 0.01 \).
...to each other were more genetically similar than those separated by greater distances.

Discussion

In this study, the two natural *Taxus yunnanensis* populations showed a lower level of genetic diversity than other congener *Taxus* species ($N_e = 2.303/2.557$; $H_o = 0.168/0.142$; $H_e = 0.453/0.517$, Table 1). Evidently, the $N_e$ values were lower than those reported for *T. baccata* in Poland ($N_e = 6.800$; Chybicki et al., 2011) and *T. wallichiana var. mairei* ($N_e = 2.678$; Zhang and Zhou, 2013) from China. The $H_o$ values were also lower than those reported for *T. baccata* in the northeastern Iberian Peninsula ($H_o = 0.487$; Dubreuil et al., 2010) and from Poland ($H_o = 0.541$; Poudel et al., 2014) and *T. wallichiana var. mairei* from China ($H_o = 0.538$; Zhang and Zhou, 2013).

It is not clear why wind-pollinated plants such as *T. yunnanensis* exhibit such high inbreeding (mean $F_{IS} = 0.596/0.702$) within populations. Indeed, six *Taxus* species have been shown to have particularly high levels of inbreeding to date, namely, *T. baccata* ($F_{IS} = 0.226/0.448$; Myking et al., 2009; Dubreuil et al., 2010), *T. chinensis* ($F_{IS} = 0.250$; Vu et al., 2016), *T. wallichiana* ($F_{IS} = 0.270/0.400$; Zhang and Zhou, 2013; Vu et al., 2016), *T. contorta* ($F_{IS} = 0.418$; Poudel et al., 2014), *T. brevifolia* ($F_{IS} = 0.472$; El-Kassaby and Yanchuk, 1994) and *T. cuspidata* ($F_{IS} = 0.229$; Chung et al., 1999). Theoretically, high inbreeding could be explained by the presence of null alleles, biparental inbreeding, low $N_e$ and/or selfing. Null allele frequency analyses have excluded null allele impact on the high inbreeding because of the extremely low frequency of 0.0018 and 0.0016 per locus in TY1 and TY2. In our opinion, the biparental inbreeding, low $N_e$ and potential selfing may have all contributed to the unexpected

Figure 2

Eight subgroups within the *Taxus yunnanensis* populations, TY1 (a) and TY2 (b), identified by structure analyses.
inbreeding. Specifically, biparental inbreeding due to limited seed and pollen dispersal could contribute to the high inbreeding. A recent study showed that severe habitat fragmentation in *T. yunnanensis* has sharply reduced the population size and increased the level of inbreeding (Miao et al., 2014). In this study, weak but significantly positive $F_{ij}$ values were found to extend up to 23 m and 50 m in both bottlenecked populations, which is consistent with a mechanism of limited seed and pollen dispersal combined with seed shadow overlap that occurs at high adult density (see Study populations and sampling) and a potential male bias within each population (Chybicki et al., 2016). Structure analyses also indicated that eight kinship families existed within each population, suggesting that mating has not been at random within these two populations and that consanguineous matings occurred between genetically related individuals. Limited seed dispersal can be caused by the following two scenarios: when seeds of *T. yunnanensis* mature, they drop directly beneath a maternal tree by gravity, which could lead to clumping of relatives in the vicinity of the mother parent and result in spatial kin-structured patches. Moreover, it is generally believed that birds disperse seeds of this species; however, there is no direct indication that seeds would be especially attractive to birds. Indeed, it is more likely that rodents would have greater interest in seeds of this understory plant than birds; therefore, rodent caches would contribute to the construction of observed kinship families. Thus, it is necessary to determine the proportions of seeds removed by birds versus rodents in a future study (Wilson et al., 1996). Limited pollen flow may also contribute to the occurrence of biparental inbreeding. Although no data regarding the spatial scale of pollen dispersal in natural *T. yunnanensis* were available, long-distance pollen movement was not expected. This is because the species is an understory tree that occurs in areas in which the wind velocity is greatly reduced, resulting in pollen flow being highly spatially restricted. These findings are indirectly supported by investigations of the pollen dispersal scale of *T. canadensis* (Allison, 1990), which revealed that a majority of the pollen dropped to the ground within a few meters of the plant due to low wind-pollination efficiency in the forest. It is also important to consider the effects of territorial pollinators on high inbreeding (Franceschinelli and Bawa, 2000) because of changes in ecological factors. Additionally, the extremely small $N_e$ may have contributed to the emergence of strong inbreeding. This is because the degree of inbreeding was inversely proportional to the $N_e$ as shown by the equation $N_e = N/1 + F_{IS}$ (Pollak, 1987). Past intensive felling has undoubtedly reduced the pollen donors and seed-producing females, increasing the relative individual mating. In the present study, the low number of $N_e$ of 7 and 9 for TY1 and TY2, respectively, likely contributed to the high inbreeding. Finally, potential selfing may also contribute to the presence of high inbreeding. Cosexual individuals of *T. yunnanensis* have been discovered (Wang et al., 2008), despite being listed as exclusively dioecious, which may be a response to ecology constraints of limited pollen availability (Wilson et al., 1996). Considering the understory properties, the wind might be less efficient; accordingly, it is
possible that the monoecious individuals are becoming favored. Once selfing occurs, the inbreeding would increase. Estimates of Ne are fundamental to understanding population structure, genetic conservation and evolutionary potential for endangered species such as T. yunnanensis. In the present study, the numbers of Ne were found to be very small for both populations (7 for TY1 and 9 for TY2). These values were significantly lower than 50, which is considered the threshold at which species are vulnerable to the immediate effects of inbreeding depression (Mace and Lande, 1991). Moreover, the Ne to N ratios of these two populations were only 0.09 and 0.16, respectively, which were both close to that of the 0.11 averaged across 102 animal and plant species (Frankham, 1995), but below the Mace-Lande’ endangerment criteria of 0.2 (Mace and Lande, 1991). Consequently, the rate of loss of genetic variation of endangered T. yunnanensis may be more rapid than previously believed. There are two possible explanations for the Ne value being below the census population numbers in these two populations. One is the unequal sex ratios. Although studies conducted by Litkowiec et al. (2015) revealed that the biased sex ratios of 32-68% females between T. baccata populations had no effect on the assessment of genetic diversity, the potential extremely low percentage of females within a T. yunnanensis population would consistently support the results of a study by Frankham (1995) in which unequal sex ratios were found to greatly reduce Ne to below the actual population size. However, this was difficult for us to find micro- and macro-strobili to enable accurate determination of the sex for many trees in these two populations because they were too high to reach (>10.0 m, see Table S1). Nevertheless, the potential overwhelming male-biased sex ratio occurrence due to environmental stress (Ortiz et al., 2002) in both populations was indirectly confirmed by the results observed in other natural populations. For instance, Wang et al. (2006) reported that the average sex ratio of species to males was 1:3. Moreover, Li et al. (2005) detected smaller ratios of females to males in two natural Lijiang populations, with ratios of 1:34 and 0:14, respectively. The other possible explanation is variance in fecundity. Wang et al. (2006) reported that the seed set of a T. yunnanensis tree was positively correlated with the sunlight intensity, varying from 7 kg to 20 g. This species is usually distributed in the undergrowth of coniferous and broad-leaved mixed forests (Miao et al., 2014). In general, individuals with marginal growth tended to produce more seeds than those with central growth. Thus, different individuals sharing different sunlight intensities would not have the same progeny production.

In summary, the two natural T. yunnanensis populations investigated in this study were characterized by low genetic diversity following a population bottlenecks, weak but significant SGS, extremely low effective population size and high inbreeding. To effectively conserve the genetic diversity of this species to maintain the adaptive potential under changing environmental conditions, in situ and ex situ conservation should be implemented (Miao et al., 2015), especially for the actual breeding individuals (Ne). Moreover, increasing the Ne can not only promote the quantity of expected heterozygosity (HR), but also decrease the degree of inbreeding within a population based on the equation \( Ht/Ho = \left[1 - l/(2Ne)\right] = 1 - F \) given by Falconer (1989). Thus, we should try to increase the number of actual breeding individuals with different genetic representations to maximize the genetic variation within a population. Greater attention should also be paid to seed dormancy mechanisms. Miao et al. (2015) found that individuals established by seed propagation had greater genetic diversity and contained more unique alleles than those that were vegetatively propagated. Similarly, Yang et al. (2011) reported that the former tended to have greater numbers of prosperous roots, straighter trunks and stronger disease resistance than the latter. However, the seed germination of T. yunnanensis has been a challenge because of its deep dormancy, which leads to a small number of seeds germinating only after experiencing dormancy in soil for about 15 months to eliminate dormant factors (Bian et al., 2015). The most important step is to understand which component(s) of T. yunnanensis seeds, testa, endosperms and embryos, separately or jointly, contribute to seed dormancy. Once this information is obtained, the seed germination rates will likely be greatly improved (Blumenthal et al. 1986). Finally, deliberate augmentation of the gene flow within and between populations is also essential to this small, isolated and inbred species (Frankham, 2015).

Acknowledgements
This work was supported by the National Natural Science Foundation of China (grant number 31470617).

References