Recent demographic histories of temperate deciduous trees inferred from microsatellite markers

Yu Cao  
Beijing Normal University

Da-Yong Zhang  
Beijing Normal University

Yan-Fei Zeng  
Chinese Academy of Forestry

Wei-Ning Bai  (baiwn@bnu.edu.cn)  
Beijing Normal University  https://orcid.org/0000-0002-6289-9459

Research article

Keywords: Asian butternuts, demographic inference, effective population size, MIGRAINE, oaks, VarEff

DOI: https://doi.org/10.21203/rs.3.rs-119096/v1

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

Background

Accurate inference of demographic histories of temperate tree species can aid our understanding of current climate change as a driver of evolution. Microsatellites are more suitable for reflecting recent historical events due to their high mutation rates. However, most programs analyze microsatellite data following a strict stepwise mutation model (SMM), which could cause false detection of population shrinkage when microsatellite mutation is not according with SMM.

Results

This study aims to reconstruct the recent demographic histories of five cool-temperate tree species, *Quercus mongolica*, *Q. liaotungensis*, *Juglans cathayensis*, *J. mandshurica* and *J. ailantifolia*, in eastern Asia by using 19 microsatellite markers and two methods considering generalized stepwise mutation model (GSM) (MIGRAINE and VarEff). Both types of software revealed that all populations experienced expansions after the Last Glacial Maximum (LGM). In particular, *J. cathayensis* experienced a more serious bottleneck in its history than the other species, leading to a smaller effective population of ancestors, while *Q. mongolica* showed only a moderate increase in population size and remained stable after the expansion. In addition, the point estimates of the multistep mutation proportion in the generalized stepwise mutation model ($\rho_{GSM}$) in all populations were between 0.50 and 0.65, which indicates that when inferring population demographic history of the above forest species using microsatellite molecular markers, it is better to assume a GSM rather than a SMM.

Conclusions

This study provides the first direct evidence that five cool-temperate tree species in East Asia have experienced expansions after the LGM using microsatellites data. Moreover, since the mutation model of microsatellite have a vital influence on demographic inference, combining multiple software programs such as MIGRAINE and VarEff can effectively reduce unnecessary errors caused by inappropriate model selection and prior setting.

Background

Understanding the demographic history of populations and species is a central issue in evolutionary ecology. At present, it is widely believed that repeated Quaternary glacial cycles since the Cenozoic, especially Last Glacial Maximum (LGM, 20–18 kya [1]), have strongly influenced the distribution and population dynamics of many temperate tree species in the northern hemisphere [2–4] [5–8]. In recent years, whole-genome sequencing coupled with new methods, such as Sequentially Markovian Coalescent (SMC) based methods, e.g., Pairwise Sequentially Markovian Coalescent (PSMC) model [9] and multiple sequentially Markovian coalescent (MSMC) model [10], and SFS based method, e.g., stairway plot [11], contribute to illustrate the pattern of population demography. However, SMC methods cannot efficiently reveal the continuous changes of population size within the last ten thousand years [12]. SFS method can't provide an accurate estimation if the sample size is limited, which may be common for whole-genome resequencing projects of non model organism [13]. Thus, predicting recent population dynamics accurately is still a major challenge for evolutionary biologist, which needs some molecular markers with much more polymorphism. Microsatellites may be a good choice [14] because of their high mutation rate [15, 16] of $10^{-6} \sim 10^{-2}$ per generation per locus [14, 17] and relatively selective neutrality [18, 19]. Moreover, considering there are thousands of studies that...
have employed microsatellite markers in the last two decades, it provides us opportunity to study population demography from these existing data.

Microsatellites have been widely used in population dynamics since the beginning of this century. Before the likelihood based methods had been developed, most of programs relied on summary statistics calculated from genetic data and tests for departure from their theoretical distribution under a given demographic and mutational model, such as BOTTLENECK [20] and M-RATIO [21], both two software propose tests to detect population declines. However, these methods suffer from a limited statistical power because they do not provide any estimate of the severity and duration of the bottleneck. Likelihood-based methods coupled with Monte Carlo sampling offer a powerful alternative to these moment-based methods ([22–24]). MSVAR [25, 26] and MIGRATE [27] are two of most commonly used methods among them. However, although many models have been developed to describe microsatellite mutation mechanisms [28], MSVAR and MIGRATE only considering strict stepwise mutation models (SMMs) [29]. Unfortunately, it has been widely recognized that deviations from an SMM often cause false detection of population shrinkage [25, 30]. For these reasons, Leblois et al. [31] developed a maximum-likelihood method for the inference of past changes in population size from microsatellite data (MIGRAINE), which allows departure from a strict SMM by using a generalized stepwise mutation model (GSM) [32] with the parameter \( p_{GSM} \) for the geometric distribution of mutation sizes. Additionally, Nikolic and Chevalet [33] developed VarEff, a composite-likelihood approach which can provide multiple mutation models and infer transient changes in population size in the past.

Widespread ecologically dominant forest tree species provide excellent opportunities to explore the relative influences of climate and geography on demographic history [34]. In this study, we selected five temperate tree species that are distributed in East Asia, two oak species, *Quercus mongolica* and *Q. liaotungensis*, and three Asian butternuts species, *Juglans cathayensis*, *J. mandshurica* and *J. ailantifolia*, for evaluation of their recent population dynamics based on microsatellite data. All five species are wind-pollinated, and there is relatively low genetic differentiation among their populations on a large geographical scale [35–37]. For oaks, *Quercus mongolica* is mainly distributed in northeastern China [38], whereas *Q. liaotungensis* is divided into two clades, a northeast group, scattered in the Changbai Mts. and partially distributed in North Korea and the Far East of Russia, and a northwest group, mainly distributed in the Qinling, Liupan and Lüliang Mts. in northern China [39]. No research on the fluctuations in the historical population dynamics of *Q. mongolica* and *Q. liaotungensis* has yet been reported. For butternuts, *J. mandshurica* is distributed in northeastern China, *J. cathayensis* occurs in southern China and Taiwan, and *J. ailantifolia* is only distributed in Japan [40]. Bai et al [41] have estimated the divergence time and ancestral effective population size using DIYABC based on microsatellite data, but the estimation of posterior parameters is greatly influenced by prior value setting [42, 43]. Later, Bai et al [44] inferred their population demography by applying PSMC to genome-wide data, but we know nothing about very recent dynamics because PSMC cannot reveal the continuous changes in species size within the last ten thousand years [9].

The main goal of this study was to reconstruct the recent demographic histories of the above five temperate tree species more accurately with microsatellite data. To that end, we applied two methods implementing a GSM: MIGRAINE, which is based on the maximum likelihood method, and VarEff, which is based on approximate likelihood. Then, we assess the effects of microsatellite mutation models and the multistep mutation proportions of the GSM on demography inference and compared the advantages and disadvantages of the two programs. We hope that our research can aid people’s understanding for recent demographic histories of temperate deciduous trees and shed some light on recent population dynamics analysis for researchers who want to use existing microsatellite data.

**Results**
Effective population size fluctuation

Both MIGRAINE and VarEff were used to estimate the current and past effective population sizes and the times of the changes in the six "big populations": QM, QL_CBS, and QL_MWC of oaks and JC, JM, and JA of Asian butternuts (see Table 1 and Additional file 1: Table S1). MIGRAINE and VarEff show consistent trends in the population dynamic curves (Figs. 1–2). Focusing on the median $N_e$ of VarEff, all six populations had experienced a rapid population expansion, five populations (except for *Q. mongolica*) then began to shrink, and the decrease continued to the present. However, as shown in the simulation tests of VarEff by Nikolic et al [33], the indicated trends of the recent contractions may not be true, so we will not further discuss them. The population dynamics among the related species were highly similar, while there were obvious differences between the two genera. For butternuts, *J. cathayensis* experienced much more serious bottlenecks in history, resulting in a smaller ancestor effective population size before its expansion (Fig. 2B), while for oaks, *Q. mongolica* only showed a moderate increase in population size and remained stable after the expansion (Fig. 1D).
| Populations (sample size) | Methods | $p_{\text{GSM}}$ | $\theta$ | $\theta_{\text{anc}}$ | $T = G \times \mu$ | $N_{\text{ratio}} = \theta / \theta_{\text{anc}}$ | $N(x \times 10^4)$ | $N_{\text{anc}}(x \times 10^4)$ | $T_{\text{years}}$ (kya) |
|--------------------------|---------|----------------|----------|----------------|-------------------|---------------------|-----------------|----------------|-----------------|
| QL_CBS (139)             | MIGRAINE| 0.53           | 18.58    | 5.37           | 0.67              | 3.46                | 0.46            | 0.13            | 33.65           |
|                          |         | [0.45–0.60]    | [13.55–27.93] | [2.07–9.11]    | [0.25–1.80]       | [1.74–8.16]         | [0.34–0.70]     | [0.05–0.23]     | [12.55–89.75]   |
|                          | VarEff  | 12.50          | 4.79     | [0.20–0.40]    | 2.61              | 0.31                | 0.12            | [10.00–20.00]   |
| QL_NWC (240)             | MIGRAINE| 0.58           | 21.47    | 2.48           | 1.32              | 8.67                | 0.54            | 0.06            | 66.15           |
|                          |         | [0.53–0.63]    | [17.64–27.35] | [0.001–5.34]   | [0.73–8.17]       | [3.98–14696.00]    | [0.44–0.68]     | [0.0003–0.13]   | [36.30–40.845]  |
|                          | VarEff  | 17.40          | 2.46     | [0.30–0.40]    | 7.08              | 0.43                | 0.06            | [15.00–20.00]   |
| QM (502)                 | MIGRAINE| 0.62           | 20.88    | 3.52           | 1.19              | 5.93                | 0.52            | 0.09            | 59.45           |
|                          |         | [0.56–0.66]    | [17.76–25.85] | [0.01–6.77]    | [0.68–6.02]       | [2.98–1314.00]     | [0.44–0.65]     | [0.0003–0.17]   | [34.10–301.00]  |
|                          | VarEff  | 12.23          | 3.24     | [0.30–0.50]    | 3.78              | 0.31                | 0.08            | [15.00–25.00]   |
| JC (596)                 | MIGRAINE| 0.64           | 14.76    | 0.64           | 2.95              | 23.21               | 0.37            | 0.02            | 88.41           |
|                          |         | [0.60–0.68]    | [12.41–17.60] | [0.02–4.08]    | [1.06–9.72]       | [3.58–569.60]      | [0.31–0.44]     | [0.0006–0.10]   | [31.86–291.63]  |
|                          | VarEff  | 17.44          | 1.11     | [0.40–0.80]    | 15.72             | 0.44                | 0.03            | [12.00–24.00]   |
| JM (399)                 | MIGRAINE| 0.62           | 15.65    | 4.35           | 0.99              | 3.60                | 0.39            | 0.11            | 29.64           |
|                          |         | [0.56–0.66]    | [12.69–20.46] | [0.82–8.50]    | [0.29–3.36]       | [1.81–17.41]       | [0.32–0.51]     | [0.02–0.21]     | [8.61–168.20]   |
|                          | VarEff  | 19.30          | 4.01     | [0.30–0.60]    | 4.81              | 0.48                | 0.10            | [9.00–18.00]    |
| JA (107)                 | MIGRAINE| 0.62           | 18.12    | 1.29           | 1.80              | 14.08               | 0.45            | 0.03            | 54.03           |
|                          |         | [0.56–0.66]    | [14.02–24.39] | [0.13–5.01]    | [0.68–5.08]       | [3.67–141.50]      | [0.35–0.61]     | [0.003–0.13]    | [20.28–152.25]  |
|                          | VarEff  | 19.75          | 3.41     | [0.30–0.60]    | 5.79              | 0.49                | 0.09            | [9.00–18.00]    |

NOTE. The estimates of past and current population sizes and times in years ($T_{\text{years}}$) obtained after a conversion of MIGRAINE and VarEff results using a fixed mutation rate of $10^{-3}$ mutation per locus per generation and a generation time of 50 years for oaks, 30 years for Asian butternuts. QL_CBS, Northeast Q. liaotungensis group; QL_NWC, Northwest Q. liaotungensis group; QM, Q. mongolica; JC, J. cathayensis; JM, J. mandshurica; JA, J. alicantifolia. The population size calculated by VarEff is median. $\theta$, scaled current effective population size; $\theta_{\text{anc}}$, scaled ancestor effective population size; $N_{\text{ratio}} = \theta / \theta_{\text{anc}}$; $N$, current effective population size (individual number); $N_{\text{anc}}$, ancestor effective population size (individual number); $G$, time measured by generations; $\mu$, mutation rate per locus per generation.
Although the time of the expansions \((T = G \times \mu)\) were estimated in both programs (Figs. 1–2), the result for MIGRAINE would be slightly less precise because of its simpler model and wide confidence interval, as shown in Table 1, we prefer to consider the VarEff estimation in the follow-up discussion. Based on the posterior distribution of median \(N_e\) in VarEff, we speculated that the two oaks began to expand in the time interval of \(T \sim (0.2–0.5)\) and three Asian butternuts in \(T \sim (0.3–0.8)\). For both oaks and butternuts, the estimate of \(\theta_{\text{anc}}\) was extremely consistent across the two methods. The estimate of \(\theta\) produced by VarEff was much smaller than that from MIGRAINE for QL_CBS, QL_NWC and QM, whereas for Asian butternuts, the \(\theta\) estimate was consistent for both programs.

**Effects of microsatellite mutation models on demography inference**

First, we used MIGRAINE to estimate the multistep mutation proportions of the GSM of six populations. The point estimation for \(p_{\text{GSM}}\) were quite high, at 0.53 for QL_CBS, 0.58 for QL_NWC, 0.63 for QM, 0.64 for JC, 0.62 for JA, and 0.62 for JM (Table 1). Then, we implemented VarEff to assess the effects of microsatellite mutation models and the multistep mutation proportions of the GSM on demography inference (Fig. 3). When the mutation model was set as an SMM (Fig. 3A) or the probability of multistep mutations for the GSM was set as 0.22 in VarEff (Fig. 3B), the six populations showed different degrees of population contraction dynamics.

**Discussion**

**Historical dynamics of populations driven by climate fluctuations**

The repeated Quaternary glacial cycles had a profound impact on the geographic history of temperate plant groups, especially in LGM (20–18 kya) [2–4], during which temperate trees were either forced to retreat southward [45] or just restricted in their northern refugia [46–49] with population decreasing dramatically, and then, these trees are expected to have reconolized their modern distribution areas and increased their population size during interglacial periods. Our result of population dynamics based on microsatellite markers using two complementary software showed a strong recent expansion after the LGM in five temperate deciduous trees, which confirms previous hypothesis that the distribution range of these trees increased with the warming of the climate after the ice age, as revealed by ecological niche modeling [41, 50]. After conversing VarEff results using a fixed mutation rate of \(1 \times 10^{-3}\) (per locus per generation) [14, 17] and generation times of 50 years for the Quercus genus [51, 52] and 30 years for Asian butternuts [44], we estimated the expansion times for all populations were approximately 25–10 kya (Table 1). Since the absolute times of the \(N_e\) fluctuations of the six populations were highly consistent with each other, we believe that the dynamic changes in the above populations are most likely attributed to climate fluctuations after the LGM.

Among all the populations, *J. cathayensis* has experienced the most serious bottleneck in its history, leading to the smallest effective population size of ancestor before the latest expansion event, which indicated that *J. cathayensis* may be more susceptible to environmental changes than the other two species of butternuts. According to VarEff, the ancestral population sizes of *J. mandshurica* and *J. ailantifolia* before expansion were similar to each other, but much larger than that of *J. cathayensis*. This finding is basically consistent with the results of PSMC based on whole-genome data [44], in which *J. cathayensis* decrease faster to an extremely small size than *J. mandshurica* and *J. ailantifolia* before LGM. Population dynamics of *J. mandshurica* and *J. ailantifolia* were highly coincident with each other, reflecting the close phylogenetic relationship. Among three populations of oak, trends of the dynamics of the two populations of *Q. liaotungensis* were very similar, but different from *Q. mongolica* which maintained a constant size after the expansion according to the dynamic curve. This may because *Q. mongolica* is more acclimate to cold but not to the dry warming climate[53].
Effects of the microsatellite mutation model on inferring population dynamics

A mutation model of microsatellite evolution is needed for the estimation of population parameters such as the number of migrants, population structure and $N_e$ [14]. Although an increasing amount of data show that most microsatellite loci do not conform to an SMM [54, 55], many programs for analyzing microsatellite assumed SMM as the only model, such as software package MSVAR, which has been one of the most commonly used to detect population demography in the last two decades [56–58]. It has been acknowledged that violations of SMM assumptions might generate severe bias in the inference of demographic history because mutations of more than one step in the GSM can produce gaps in the allele length distribution, which is typically observed after a population decline under SMM model [31, 59]. Moreover, when estimating relatively old evolution events, the error rate of population parameters for the microsatellite is much higher than that for SNPs if we assume an SMM and an infinite allele model (IAM) [15]. Therefore, when we are not sure about the microsatellite mutation process in a species, we should be very cautious to choose SMM as the mutation model [16]. In our study, the proportions of multistep mutations in the GSM calculated by MIGRAINE were from 0.50–0.65 in the oaks and 0.60–0.65 in the Asian butternuts, indicating GSM model was more suitable for microsatellites, at least for some of temperate tree species. In fact, Bai et al. [41] estimated a similar value by using DIYABC. We also found that different settings for the proportions of multistep mutations may lead to completely opposite trends of population size change, which suggested it is a key parameter in the GSM (Fig. 3).

There are thousands of studies that have employed microsatellite markers in the last two decades, it is worth continuing to mine the information from these existing data. As we noted earlier, microsatellites are well suited to detect recent population changes due to their high mutation rate and neutrality [15, 16], and they can provide a good complement to the program with the inability to estimate recent population dynamics [9]. The development of bioinformatic methods in recent years will allow us to understand more about the distribution and mutation mechanisms of microsatellites [14]. For all of these reasons, microsatellites would remain an effective and cost-efficient marker to study population demography in the genomic age [60].

Comparison of MIGRAINE vs. VarEff for calculating population dynamics based on microsatellite markers

MIGRAINE is a software that based on the importance sampling of gene genealogies and the coalescent theory under the maximum likelihood framework to estimate current and past $N_e$, the time of occurrence of past changes in population size. It is more flexible and robust than other programs because it provides several demographic models, i.e. distance isolation model (IBD), the single population dynamic variation model (OnePopVarSize), the founder model (OnePopFounderFlush) and multiple microsatellite mutations, SMMs, GSMs and infinite sites mutation models (ISMs). In addition, MIGRAINE can estimate the multistep mutation proportion ($p_{GSM}$) of the GSM, which is a key parameter and needed to be set for other programs (e.g., VarEff). However, MIGRAINE is based on the maximum likelihood method, and when running models of OnePopVarSize or OnePopFounderFlush, more runs (more than 2000 is recommended) per sampling point are needed, which often makes the calculation time long. Furthermore, MIGRAINE assumes that a single isolated population has only undergone a single past size change, which definitely oversimplifies actual complex population dynamics.

Compared to MIGRAINE, the VarEff program relies on an approximation of the likelihood of data from which a fast algorithm allows size variations to be efficiently detected without any prior hypothesis about demographic history, such as monotonous growth or decline. Unlike traditional Bayesian or approximate Bayesian methods, VarEff did not
depend much on the priors and could explore their dependence on the assumed mutation model [33]. VarEff also has included several microsatellite mutation models, SMMs, two-phase models (TPMs) and GSMs, but it needs to set the multistep mutation proportion parameter of GSM as a prior parameter. Disadvantages of VarEff are that it cannot detect ancient population changes (e.g., \( G \times \mu > 20 \)) and a false contraction would be detected when SMM model is wrongly chosen or a strong gene flow is ignored. Considering the advantages and disadvantages of the two programs, we believe that applying them in combination can effectively reduce unnecessary errors caused by inappropriate model selection and prior parameter setting when implementing demography inference basing on microsatellite data.

**Conclusions**

As we mentioned, even though whole-genome sequencing is prevalent today, the low mutation rate of SNPs and the limited sampling size are two main factors preventing us from inferring recent population dynamics accurately. Under these circumstances, microsatellites may be a good choice for reflecting recent historical events due to their high mutation rates and extensively accumulated data in last three decades. Our study showed five cool-temperate tree species in East Asia have indeed experienced expansions after the Last Glacial Maximum (LGM) using microsatellites data. This confirms previous hypothesis that the distribution range of these trees increased with the warming of the climate after the ice age, as revealed by ecological niche modeling [41, 50]. The study also indicated that when inferring demographic history using microsatellite data, both the microsatellite mutation model and the parameter settings can have a significant impact, and it is better to assume a GSM rather than a SMM and apply multiple programs to achieve a more accurate microsatellite analysis.

**Methods**

**Data source**

All data used were from previous studies; the data on oaks came from Zeng et al [39] and those on Asian butternuts from Bai et al [41].

Based on previous studies, we considered six groups of these five tree species (two groups of \( Q. liaotungensis \), northeast group and northwest group), and excluded any possible hybrids with related species from our analysis. The investigated areas covering the whole range of \( Q. mongolica \) and most distribution areas of \( Q. liaotungensis \) in China, included 502 individuals from 17 \( Q. mongolica \) populations (QM), 139 individuals from five populations of northeast \( Q. liaotungensis \) (QL_CBS) and 240 individuals from eight populations of northwest \( Q. liaotungensis \) (QL_NWC). The sample distribution range of Asian butternuts also covered the entire distribution area of the three species, including 596 individuals from 25 \( J. cathayensis \) (JC) populations, 399 individuals from 14 \( J. mandshurica \) populations (JM) and 107 individuals from five \( J. ailantifolia \) populations (JA) (Fig. 4).

**Data analysis**

**Analysis of historical population dynamics based on MIGRAINE**

We explored the demographic history of the above five species using MIGRAINE 0.5.4 (http://kimura.univ-montp2.fr/~rousset/Migraine.htm), which is based on importance sampling of gene genealogies under a maximum likelihood framework and is extended for GSM which notably allows departure from the strict SMM with the parameter \( p_{GSM} \) for the geometric distribution of mutation sizes. The present and ancestral scaled population sizes \((\theta = 4N\mu, \theta_{anc} = 4N_{anc}\mu)\), the scaled time of occurrence of the past change in mutation rate \((T = G \times \mu)\) was calculated. Here, \( N_{anc} \) is the
ancestral effective population size; and $N$ is the current effective population size; $\mu$ is the mutation rate per locus per generation; $G$ is the time measured by generations. The detection of significant past change in the population is based on the population size ratio ($\theta_{\text{ratio}} = \theta / \theta_{\text{anc}}$). If $\theta_{\text{ratio}} < 1$, we can infer that the population has experienced contraction; otherwise, if $\theta_{\text{ratio}} > 1$, the population has experienced expansion. Also, we estimated the multistep mutation proportion of the GSM ($p_{\text{GSM}}$) using MIGRAINEs.

The OnePopVarSize model was used, which considers a single isolated population with a unique past size change. In order to ensure there are enough points with high likelihoods for the smoothing procedure, we consider relatively high values of parameter $NRunsPerPoint$ to get reliable results. For the data of QL_CBS, QL_NWC, QM and JM, we considered 2,000 trees, with 200 points in each iteration and a total of 16 iterations. For the data of JC and JA, MIGRAINE was run using 20,000 trees, with 200 points in each iteration and a total of 16 iterations, because when run 2000 trees, it is hard to get a convergent result for them.

Analysis of historical population dynamics based on VarEff

Since MIGRAINE infers dynamics under a model of a single panmictic population with one exponential change in population size, the highly simplified demographic model may be different from the actual situation. Here, we used VarEff [33], another model for estimating the past changes in $N_e$ from microsatellite data by using approximate likelihoods under an MCMC approach, which has been found to be especially useful for providing evidence of transient changes in population size in the past. The VarEff method was implemented in the R package VarEff (https://ggsp.jouy.inra.fr). By using the functions in the VarEff package, we could extract several global statistics of $N_e$ (arithmetic and harmonic means, mode, median, and quantiles). Since the median of the posterior distribution was found to be the most robust estimator [33], we visualized the posterior distribution of median at different times in the past.

Using the GSM as the microsatellite mutation mechanism, we set the multistep mutation proportion, $C$ (which is same as parameter $p_{\text{GSM}}$ in MIGRAINE) as the value calculated from MIGRAINE’s estimation. In order to compare with the results of MIGRAINE, we set the parameter $TMAX$ (Length of the period for which the distributions of $N_e$ in the past) in VarEff with reference to the estimation results of parameter $T$ in MIGRAINE. For the parameter settings related to MCMC, NumberBatch, LengthBatch, and SpaceBatch were set to 10000, 1, and 100, respectively, and Burnin is set to 10,000.

Abbreviations

SMM: stepwise mutation model; GSM: generalized stepwise mutation model; TPMs: two-phase models; IAM: infinite allele model; LGM: Last Glacial Maximum; $p_{\text{GSM}}$: the multistep mutation proportion in the generalized stepwise mutation model; SNP: single nucleotide polymorphism.

Declarations

Acknowledgements

We thank Chen J, Zhang WP and Ding YM for their helpful suggestions and valuable comments on the manuscript. We thank Wu Y and Liang Y for their help in applying R software. The authors have no conflicts of interest.

Funding
This work was financially supported by the National Natural Science Foundation of China (31421063) and the “111” Program of Introducing Talents of Discipline to Universities (B13008). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets analysed during the current study (SSR genotypes and geographical locations of collections) are available in the Dryad Digital Repository,

https://datadryad.org/stash/share/b05Zbkr00AEkDx0yBDKWo5QJ5xBSznPE72gOLrh8zTQ.

Authors’ contributions

Da-Yong Zhang and Wei-Ning Bai conceived the ideas, designed methodology, Wei-Ning Bai and Yan-Fei Zeng provided data, Yu Cao analyzed data; Yu Cao drafted the article, Da-Yong Zhang, Wei-Ning Bai and Yan-Fei Zeng revised it critically for important intellectual content. All authors contributed critically to the drafts and gave final approval for publication.

Ethics approval and consent to participate

All the plant materials were sampled from natural populations in North and Northeast China and no specific permission was needed to collect such samples. This study was conducted in accordance with local legislation and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1 State Key Laboratory of Earth Surface Process and Resource Ecology and Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China. 2 State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China.

References

1. Nie GZ, Liu JQ, Guo ZT. The major stratigraphic boundaries and climatic events in Weinan loess section since 0.15Ma B. P.: based on chronological evidences. Quaternary Sciences. 1996; 16:221–231 http://www.dsjyj.com.cn/CN/Y1996/V16/I3/221.

2. Hewitt, Godfrey. The genetic legacy of the Quaternary ice ages. Nature. 2000; 405:907–913 https://doi.org/10.1038/35016000.

3. Hu FS, Hampe A, Petit RJ. Paleoecology meets genetics: deciphering past vegetational dynamics. Frontiers in Ecology and the Environment. 2009; 7:371–379 https://doi.org/10.1890/070160.

4. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci. 2004; 359:183-195; discussion 195 https://www.ncbi.nlm.nih.gov/pubmed/15101575.
5. Melo WA, Freitas CG, Bacon CD, Collevatti RG. The road to evolutionary success: insights from the demographic history of an Amazonian palm. Heredity (Edinb). 2018; 121:183–195 https://doi.org/10.1038/s41437-018-0074-1.

6. Evans LM, Allan GJ, DiFazio SP, Slavov GT, Wilder JA, Floate KD, Rood SB, Whitham TG. Geographical barriers and climate influence demographic history in narrowleaf cottonwoods. Heredity (Edinb). 2015; 114:387–396 https://doi.org/10.1038/hdy.2014.115.

7. Zhou L, Bawa R, Holliday JA. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (Populus trichocarpa). Molecular Ecology. 2014; 23:2486–2499 https://doi.org/10.1111/mec.12752.

8. Lin RC, Yeung CK, Li SH. Drastic post-LGM expansion and lack of historical genetic structure of a subtropical fig-pollinating wasp (Ceratosolen sp. 1) of Ficus septica in Taiwan. Molecular Ecology. 2008; 17:5008-5022 https://www.ncbi.nlm.nih.gov/pubmed/19120988.

9. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011; 475:493–496 https://doi.org/10.1038/nature10231.

10. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nature Genetics. 2014; 46:919-925 https://www.ncbi.nlm.nih.gov/pubmed/24952747.

11. Liu X, Fu YX. Exploring population size changes using SNP frequency spectra. Nature Genetics. 2015; 47:555–559 https://www.ncbi.nlm.nih.gov/pubmed/25848749.

12. Salmena J, Heller R, Lascoux M, Shafer A. Inferring Demographic History Using Genomic Data. In: Population Genomics. Edited by Rajora OP: Springer International Publishing; 2017: 511–537.

13. Patton AH, Margres MJ, Stahlke AR, Hendricks S, Lewallen K, Hamede RK, Ruiz-Aravena M, Ryder O, McCallum Hl, Jones ME et al. Contemporary Demographic Reconstruction Methods Are Robust to Genome Assembly Quality: A Case Study in Tasmanian Devils. Mol Biol Evol. 2019; 36:2906-2921 https://www.ncbi.nlm.nih.gov/pubmed/31424552.

14. Hoshino AA, Bravo JP, Nobile PM, Morelli KA. Microsatellites as Tools for Genetic Diversity Analysis. In: Genetic Diversity in Microorganisms. Edited by Caliskan M. Rijeka: InTech; 2012: 149–170.https://doi.org/10.5772/35363.

15. Freeland JR, Anderson S. Molecular Ecology. In: eLS. John Wiley & Sons, Ltd (Ed.); 2019https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0003268.pub3.

16. Bai WN, Zhang DY. Current status and future directions in plant phylogeography. Chinese Bulletin of Life Sciences. 2014; 26:125–137 https://doi.org/10.13376/j.cblls/2014020.

17. Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters. 2006; 9:615–629 https://doi.org/10.1111/j.1461-0248.2006.00889.x.

18. Pokhriyal B, Thorat K, Limaye D, Joshi Y, Kadam V, R D. Microsatellite markers - a novel tool in molecular genetics. International Journal of Research in Pharmacy and Chemistry. 2012; 2:397–412.

19. Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: a review. Molecular Ecology. 2002; 11:1–16 https://doi.org/10.1046/j.0962-1083.2001.01418.x.

20. Cornuet JM, Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 1996; 144:2001–2014.

21. Garza JC, Williamson EG. Detection of reduction in population size using data from microsatellite loci. Molecular Ecology. 2001; 10:305–318 https://doi.org/10.1046/j.1365-294x.2001.01190.x.

22. Felsenstein J. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genetics Research. 1992; 59:139–147.
23. Griffiths RC, Tavare S. Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond B Biol Sci. 1994; 344:403-410 https://www.ncbi.nlm.nih.gov/pubmed/7800710.

24. Emerson BC, Paradis E, Thébaud C. Revealing the demographic histories of species using DNA sequences. TRENDS in Ecology & Evolution. 2001; 16:707–716.

25. Girod C, Vitalis R, Leblois R, Freville H. Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method. Genetics. 2011; 188:165–179 https://doi.org/10.1534/genetics.110.121764.

26. Beaumont MA. Detecting population expansion and decline using microsatellites. Genetics. 1999; 153:2013–2029.

27. Beerli P. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics. 2006; 22:341–345 https://doi.org/10.1093/bioinformatics/bti803.

28. Bhargava A, Fuentes FF. Mutational dynamics of microsatellites. Mol Biotechnol. 2010; 44:250–266 https://doi.org/10.1007/s12033-009-9230-4.

29. Ohta T, Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population*. Genetics Research. 2007; 89:367–370 https://doi.org/10.1017/S0016672308009531.

30. Faury S, Pertoldi C. The consequences of the unlikely but critical assumption of stepwise mutation in the population genetic software, MSVAR. Evol Ecol Res. 2012; 14:859–879 https://www.researchgate.net/publication/257924528.

31. Leblois R, Pudlo P, Neron J, Bertaux F, Reddy Beeravolu C, Vitalis R, Rousset F. Maximum-likelihood inference of population size contractions from microsatellite data. Molecular Biology and Evolution. 2014; 31:2805–2823 https://doi.org/10.1093/molbev/msu212.

32. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. Molecular Biology and Evolution. 1999; 16:1791–1798 https://doi.org/10.1093/oxfordjournals.molbev.a026091.

33. Nikolic N, Chevalet C. Detecting past changes of effective population size. Evolutionary Applications. 2014; 7:663–681 https://doi.org/10.1111/eva.12170.

34. Savolainen O, Pyhäjärvi T, Knürr T. Gene Flow and Local Adaptation in Trees. Annual Review of Ecology, Evolution, and Systematics. 2007; 38:595-619 https://doi.org/10.1146/annurev.ecolsys.38.091206.095646.

35. Bai WN, Liao WJ, Zhang DY. Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytologist. 2010; 188:892-901 https://doi.org/10.1111/j.1469-8137.2010.03407.x.

36. Zeng YF, Liao WJ, Petit RJ, Zhang DY. Exploring Species Limits in Two Closely Related Chinese Oaks. Plos One. 2010; 5:1–15 https://doi.org/10.1371/.

37. Bai WN, Wang WT, Zhang DY. Contrasts between the phylogeographic patterns of chloroplast and nuclear DNA highlight a role for pollen-mediated gene flow in preventing population divergence in an East Asian temperate tree. Molecular Phylogenetics and Evolution. 2014; 81:37–48 https://doi.org/10.1016/j.ympev.2014.08.024.

38. Wang LM, Ren XW, Liu YQ. Geographic distribution of deciduous oaks in China. J Beijing For Coll. 1985; 2:57–69.

39. Zeng YF, Liao WJ, Petit RJ, Zhang DY. Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. Molecular Ecology. 2011; 20:4995–5011 https://doi.org/10.1111/j.1365-294X.2011.05354.x.

40. Aradhya MK, Potter D, Simon CJ. Origin, evolution, and biogeography of Juglans: a phylogenetic perspective Acta Horticulturae. 2005; 705:85-94 https://doi.org/10.17660/ActaHortic.2005.705.8.
41. Bai WN, Wang WT, Zhang DY. Phylogeographic breaks within Asian butternuts indicate the existence of a phytogeographic divide in East Asia. New Phytologist. 2016; 209:1757–1772 https://doi.org/10.1111/nph.13711.

42. Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, Estoup A. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics. 2014; 30:1187–1189 https://doi.org/10.1093/bioinformatics/btt763.

43. Navascues M, Leblois R, Burgarella C. Demographic inference through approximate-Bayesian-computation skyline plots. PeerJ. 2017; 5:13 https://doi.org/10.7717/peerj.3530.

44. Bai WN, Yan PC, Zhang BW, Woeste KE, Lin K, Zhang DY. Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus Juglans (Juglandaceae) revealed by whole-genome sequences. New Phytologist. 2018; 217:1726–1736 https://doi.org/10.1111/nph.14917.

45. Bennett KD, Tzedakis PC, Willis KJ. Quaternary Refugia of North European Trees. Journal of Biogeography. 1991; 18:103–115 https://doi.org/10.2307/2845248.

46. Qian H, Ricklefs RE. Large-scale processes and the Asian bias in species diversity of temperate plants. Nature. 2000; 407:180–182 https://doi.org/10.1038/35025052.

47. Harrison SP, Yu G, Takahara H, Prentice IC. Palaeovegetation - Diversity of temperate plants in east Asia. Nature. 2001; 413:129–130 https://doi.org/10.1038/35093166.

48. Qiu YX, Fu CX, Comes HP. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Molecular Phylogenetics and Evolution. 2011; 59:225–244 https://doi.org/10.1016/j.ympev.2011.01.012.

49. Ye J, Yuan Y, Cai L, Wang X. Research progress of phylogeographic studies of plant species in temperate coniferous and broadleaf mixed forests in Northeastern China. Biodiversity Science. 2017; 25:1339–1349 https://doi.org/10.17520/biods.2017265.

50. Zeng YF, Wang WT, Liao WJ, Wang HF, Zhang DY. Multiple glacial refugia for cool-temperate deciduous trees in northern East Asia: the Mongolian oak as a case study. Molecular Ecology. 2015; 24:5676–5691 https://doi.org/10.1111/mec.13408.

51. Gregorius HR, Degen B, König A. Problems in the Analysis of Genetic Differentiation Among Populations – a Case Study in Quercus robur. Silvae Genetica. 2007; 56:190–199.

52. Leroy T, Rougemont Q, Dupouey JL, Bodenes C, Lalanne C, Belser C, Labadie K, Le Provost G, Aury JM, Kremer A et al. Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. New Phytol. 2020; 226:1183-1197 https://www.ncbi.nlm.nih.gov/pubmed/31264219.

53. Li T, He XY, Chen ZJ. Tree-ring growth responses of Mongolian oak (Quercus mongolica) to climate change in southern Northeast: A case study in Qianshan Mountains. Chinese Journal of Applied Ecology. 2014; 25:1841-1848.

54. Ellegren H. Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics. 2004; 5:435-445 https://www.ncbi.nlm.nih.gov/pubmed/15153996.

55. Seyfert AL, Cristescu ME, Frisse L, Schaack S, Thomas WK, Lynch M. The rate and spectrum of microsatellite mutation in Caenorhabditis elegans and Daphnia pulex. Genetics. 2008; 178:2113-2121 https://www.ncbi.nlm.nih.gov/pubmed/18430937.

56. Goossens B, Chikhi L, Ancrenaz M, Lackman-Ancrenaz I, Andau P, Bruford MW. Genetic signature of anthropogenic population collapse in orang-utans. PLoS Biol. 2006; 4:e25 https://www.ncbi.nlm.nih.gov/pubmed/16417405.
57. Sharma R, Arora N, Goossens B, Nater A, Morf N, Salmona J, Bruford MW, Van Schaik CP, Krutzen M, Chikhi L. Effective population size dynamics and the demographic collapse of Bornean orang-utans. Plos One. 2012; 7:e49429 https://www.ncbi.nlm.nih.gov/pubmed/23166666.

58. Frantz AC, McDevitt AD, Pope LC, Kochan J, Davison J, Clements CF, Elmeros M, Molina-Vacas G, Ruiz-Gonzalez A, Balestrieri A et al. Revisiting the phylogeography and demography of European badgers (Meles meles) based on broad sampling, multiple markers and simulations. Heredity (Edinb). 2014; 113:443-453 https://www.ncbi.nlm.nih.gov/pubmed/24781805.

59. Jentzsch IMV, Bagshaw ATM, Buschiazzo E, Merkel A, Gemmell NJ. Evolution of Microsatellite DNA. In: eLS. Edited by Wiley J, Sons. Chichester: John Wiley & Sons, Ltd; 2013.

60. Hodel RG, Segovia-Salcedo MC, Landis JB, Crowl AA, Sun M, Liu X, Gitzendanner MA, Douglas NA, Germain-Aubrey CC, Chen S et al. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. Applications in Plant Sciences. 2016; 4:1–13 https://doi.org/10.3732/apps.1600025.

Supplementary Information

Additional file 1: Table S1. Detailed population parameters estimated by VarEff. QL_CBS, Northeast Q. liaotungensis group; QL_ NWC, Northwest Q. liaotungensis group; QM, Q. mongolica; JC, J. cathayensis; JM, J. mandshurica; JA, J. ailantifolia. θ, scaled current effective population size; θ_anc, scaled ancestor effective population size; θ/θ_anc (H), the ratio of harmonic means of effective population size; θ/θ_anc (M), the ratio of median of effective population size.

Figures
Figure 1

Fluctuation of historical effective population size of three oak populations. (A) All three populations; (B) Northeast Q. liaotungensis group; (C) Northwest Q. liaotungensis group; (D) Q. mongolica. The solid line represents the calculation results of VarEff, and the dotted line represents the calculation results of MIGRAINE. G, time scaled by generations; μ, mutation rate per locus per generation. T = G × μ, scaled time in the past in mutation rate. θ(T), scaled current effective population size at past time T.
Figure 2

Fluctuation of historical effective population size of three Asian butternut populations. (A) All three populations; (B) J. cathayensis; (C) J. mandshurica; (D) J. ailantifolia. The solid line represents the calculation results of VarEff, and the dotted line represents the calculation results of MIGRAINE. $G$, time measured by generations; $\mu$, mutation rate per locus per generation. $T = G \times \mu$, scaled time in the past in mutation rate. $\theta(T)$, scaled current effective population size at past time $T$. 
Figure 3

Effects of microsatellite mutation models on demography inference. (A) SMM model; (B) The proportion of multi-step mutation of GSM is set at 0.22. Each line is associated to the posterior distribution of one population, QL_CBS in red line, QL_NWC in orange line, QM in green line, JC in pink line, JM in blue line and JA in cyan line. G, time measured by generations; $\mu$, mutation rate per locus per generation. $T = G \times \mu$, scaled time in the past in mutation rate. $\theta(T)$, scaled current effective population size at past time $T$. 
Figure 4

Geographic distribution of the six sampled populations. According to Zeng et al [39], Bai et al [41].

Supplementary Files

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

- 3.Additionalfile1.docx