New Phytologist

Research

Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus *Juglans* (Juglandaceae) revealed by whole-genome sequences

Wei-Ning Bai[1]*, Peng-Cheng Yan[1,2]*, Bo-Wen Zhang[1]*, Keith E.Woeste[3], Kui Lin[1] and Da-Yong Zhang[1]

[1]State Key Laboratory of Earth Surface Processes 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]Beijing Key Laboratory of Cloud Computing Key Technology and Application, Beijing Computing Center, Beijing 100094, China; [3]USDA Forest Service Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA

Authors for correspondence:
Wei-Ning Bai
Tel: +86 10 58805852
Email: baiwn@bnu.edu.cn

Da-Yong Zhang
Tel: +86 10 58804534
Email: zhangdy@bnu.edu.cn

Received: *4 September 2017*
Accepted: *24 October 2017*

New Phytologist (2017)
doi: 10.1111/nph.14917

Key words: diversification, effective population size, *Juglans*, pairwise sequentially Markovian coalescent (PSMC), temperate forests.

Summary

- Whether species demography and diversification are driven primarily by extrinsic environmental changes such as climatic oscillations in the Quaternary or by intrinsic biological interactions like coevolution between antagonists is a matter of active debate. In fact, their relative importance can be assessed by tracking past population fluctuations over considerable time periods.
- We applied the pairwise sequentially Markovian coalescent approach on the genomes of 11 temperate *Juglans* species to estimate trajectories of changes in effective population size (N_e) and used a Bayesian-coalescent based approach that simultaneously considers multiple genomes (G-PhoCS) to estimate divergence times between lineages.
- N_e curves of all study species converged 1.0 million yr ago, probably reflecting the time when the walnut genus last shared a common ancestor. This estimate was confirmed by the G-PhoCS estimates of divergence times. But all species did not react similarly to the dramatic climatic oscillations following early Pleistocene cooling, so the timing and amplitude of changes in N_e differed among species and even among conspecific lineages.
- The population histories of temperate walnut species were not driven by extrinsic environmental changes alone, and a key role was probably played by species-specific factors such as coevolutionary interactions with specialized pathogens.

Introduction

Phylogeography involves the analysis of the spatial distribution of genealogical lineages, especially within and among conspecific populations and closely related species (Avise *et al.*, 1987, 2016). Early studies in this field have emphasized the role of climatic fluctuations during the Quaternary period (2.6 million yr ago (Ma) to the present) in shaping the spatial and temporal patterns of genetic variation (Avise *et al.*, 1987, 2016; Hewitt, 2000, 2004, 2011). Although this emphasis on abiotic factors has offered invaluable insights into drivers of population divergence, speciation, and the formation of communities, the contribution of taxon-specific factors such as life-history traits or biotic interactions with other organisms is largely overlooked (Papadopoulou & Knowles, 2016). The paradigm that abiotic factors drove species demography has gained wide acceptance among phylogeographers, as exemplified by the recent upsurge in the popularity of ecological niche models (ENMs) that predict the present day and past geographic range of the species of interest using observed correlation between species occurrence and abiotic variables (Alvaradoserrano & Knowles, 2014).

Closely related species tend to have similar physiologies and life histories (Wiens & Graham, 2005), especially in higher plants (e.g. Prinzing *et al.*, 2001; Hawkins *et al.*, 2014). Furthermore, if the influences of biotic interactions on species distribution and abundance are negligible relative to the influences of the environment, as commonly assumed following the lead of Gleason (1926), then congeneric species that share evolutionary adaptations to climate would be expected to show concurrent cycles of population expansions and contractions, largely coinciding with climate cycles (Hewitt, 2000, 2004). Closely related species sharing similar ecological or life-history traits rarely present congruent demographic histories, however, even in equivalent environments (Burbrink *et al.*, 2008; Ruane *et al.*, 2015; Prates *et al.*, 2016), rendering climate-based explanations of demography unlikely. Faced with this mismatch between theory and observation, we note with interest that Ricklefs (2015a), although in a very different context, has suggested the alternative hypothesis that long-term population dynamics is primarily driven by the changing balance of coevolved interactions between hosts and

*These authors contributed equally to this work.

their specialized pathogens. If so, disparate demographic histories should predominate, even among closely related taxa with similar ecological or life-history traits. By understanding the relative contribution of extrinsic environmental changes and intrinsic biological interactions in driving species' temporal dynamics, we can also gain an insight into the historical (e.g. vicariance, dispersion) and demographic (e.g. patterns of gene flow, population bottlenecks) processes responsible for the observed distribution of genetic variation and species diversity at a range of spatial scales.

The contrasting hypotheses predict different demographic responses to climate change, which facilitate empirical comparison. If population size is primarily influenced by environmental change accompanied by glacial cycles, then it should fluctuate with historical climate periods. Above all, we expect to observe that the demographic trends of closely related species living in similar environments will be temporally congruent unless the species' physiology or life history evolves under strong selection (Hansen, 1997; but see Prinzing *et al.*, 2001; Blomberg *et al.*, 2003; Hawkins *et al.*, 2014). By contrast, if population size fluctuates mainly in response to episodic, species-specific pathogenic or antagonistic interactions, as postulated by Ricklefs (2015a), then closely related species would exhibit idiosyncratic responses to the glacial cycles of the Pleistocene. Thus, information on historical fluctuations in the effective population size (N_e) of closely related species is invaluable for inferring the relative importance of extrinsic environmental change vs intrinsic biological interactions in controlling population dynamics.

Temperate tree species in the northern hemisphere were critically influenced by global climatic oscillations during the Quaternary period (Qian & Ricklefs, 2000; Harrison *et al.*, 2001; Hewitt, 2004). Nonetheless, if populations in this system were found to fluctuate independently of climate, the hypothesis that intrinsic biological interactions have an important, and possibly even dominant, role in demography would be strongly supported. In this sense, temperate tree species such as the walnuts and butternuts constitute an excellent study system for determining the importance of species interactions relative to climate change. Members of the genus *Juglans* L. (walnuts and butternuts) grow mostly in the northern hemisphere. *Juglans* consists of *c.* 21 extant species (Manning, 1978; Stanford *et al.*, 2000; Aradhya *et al.*, 2007), and it is usually subdivided into three sections: *Rhysocaryon*, *Cardiocaryon*, and *Dioscaryon* (syn. *Juglans*). *Rhysocaryon* is endemic to temperate and tropical America, with five North American temperate species. *Cardiocaryon* contains three species native to East Asia, and a single species (*Juglans cinerea*) native to eastern North America, which is sometimes distinguished as a separate section, *Trachycaryon* (Manning, 1978; Hoban *et al.*, 2010). *Dioscaryon* includes the cultivated Persian walnut, *Juglans regia*, and the semidomesticated iron walnut, *Juglans sigillata*.

Our goal was to better understand the relative importance of extrinsic vs intrinsic interactions in controlling population dynamics. To that end, we generated the whole-genome sequences of 31 individuals from the 11 temperate species of the genus *Juglans*. We then applied the pairwise sequentially Markovian coalescent (PSMC) method (Li & Durbin, 2011) to infer their N_e dynamics and history of lineage separation. We also sought to determine whether speciation in *Juglans* – and thus perhaps in temperate tree species generally – occurred during glacial cycles, a subject of much renewed interest (Levsen *et al.*, 2012; Wang *et al.*, 2014). Range contraction as a result of cooling climate can lead to the isolation of parts of previously expanded populations, which in turn establishes conditions for species formation through divergence in allopatry (Hewitt, 2011). If Pleistocene climatic oscillations did trigger speciation events, we expect that most diversification would date to the major glacial cycles that characterized the Middle and Late Pleistocene (~0.90–0.012 Ma).

Materials and Methods

Genome sequencing and assembly for three walnut species

Sequencing libraries were generated using Illumina's TruSeq Sample Preparation kit following the manufacturer's instructions (Illumina, San Diego, CA, USA; www.illumina.com) and fragment library of 450 bp was performed with a paired-end 250 bp strategy on the Illumina HiSeq 2500 sequencing platform at a depth of 90× for *Juglans mandshurica*, 57× for *J. regia* and 53× for *Juglans nigra*. Jumping libraries with insert sizes of 3K bp, 5K bp and 10K bp were sequenced with a paired-end 150 bp strategy on Illumina HiSeq X ten. The total depth of jump libraries was *c.* 50× for *J. mandshurica*, *J. regia* and *J. nigra*. All libraries were assembled and scaffolded using ALLPATHSLG (v.474 117) with default parameters after filtering reads containing sequence adaptors or reads not paired properly. GAPCLOSER (v.1.12) was applied to fill gaps in initial assembly results (see Supporting Information Methods S1). We also predicted and annotated repeat regions of each genome (see Methods S2).

Resequencing and mapping

We resequenced 31 individuals from different populations of 11 *Juglans* species' (*Rhysocaryon*: *J. californica*, *J. hindsii*, *J. microcarpa*, *J. major*, *J. nigra*; *Cardiocaryon*: *J. cinerea*, *J. mandshurica*, *J. ailantifolia*, *J. cathayensis*; *Dioscaryon*: *J. regia* and *J. sigillata*) (Tables S1, S2; Fig. S1) genomes to an average of *c.* 30–40 depth and > 80% coverage using Illumina HiSeq 4000 paired-end sequencing libraries with insert sizes of 350 bp. To get the heterozygosity state for each base, the trimmed clean reads of each species were mapped back onto its newly created genome by BWA-MEM using default parameters (*Rhysocaryon* species were mapped onto the *J. nigra* genome, *Cardiocaryon* species onto the *J. mandshurica* genome, and *Dioscaryon* species onto the *J. regia* genome).

We were constrained by the availability of only three reference genomes for the 11 *Juglans* species we studied. It is likely this constraint biased variant calling when we mapped reads interspecifically to a reference genome, but its impact was somewhat mitigated by the availability of a reference genome in each of the three sections of *Juglans* and by our following filtering strategy of using only uniquely mapped reads for calling variants.

Reconstructing the past population dynamics

The PSMC model, originally developed to study the demographic history of humans (Li & Durbin, 2011), can be used to estimate population size changes well into the early Pleistocene by utilizing information from the whole genome of a single diploid individual. Hence, PSMC could represent an appropriate tool with which to test the aforementioned hypotheses. The PSMC method has been successfully applied to animals covering a broad phylogenetic range from vertebrates to arthropods (e.g. Li & Durbin, 2011; Hung *et al.*, 2014; Wallberg *et al.*, 2014; Nadachowska-Brzyska *et al.*, 2015), but seldom to plants (e.g. AmborellaGenomeProjects, 2013).

Filtering is likely to be critical to PSMC analysis because PSMC is prone to biases that are the result of sequencing/genotyping errors and missing data. Nadachowska-Brzyska *et al.* (2016) recommended sequencing data with a mean genome coverage of $\geq 18\times$, a per-site filter of ≥ 10 reads and no more than 25% of missing data when using PSMC. It is well known that repetitive sequences (or repeats) are particularly vulnerable to mapping errors, resulting in false-positive and incorrect single nucleotide polymorphism (SNP) calls (Treangen & Salzberg, 2011). To avoid this problem, Nadachowska-Brzyska *et al.* (2016) simply masked repeats out from the reference genome. We could not use this approach, however, because of the high proportions of repeats in the assembled genomes (*J. nigra*, 47.3%; *J. mandshurica*, 50.1%; *J. regia*, 46.7%; *Populus trichocarpa*, 41.95%; and *Fraxinus excelsior*, 41.98%; Table S3); simple repeat masking would have resulted in *c.* 50% missing data, a proportion much higher than the recommended cutoff value of 25% (Nadachowska-Brzyska *et al.*, 2016). Thus, in addition to the basic filter strategies in PSMC, which include setting quality adjuster -C to 50, setting the minimal mapping quality to 10, minimum depth to one-third of the mean genome coverage, and maximum depth to two times the mean genome coverage, we minimized the influence of repeats by retaining only uniquely mapped, properly paired, and nonduplicated reads (see Methods S3).

The consensus autosomal sequence for each species was used as the input for the PSMC modeling. The settings were chosen manually according to suggestions given by Li & Durbin (2011; https://github.com/lh3/psmc) (see Methods S4). A mutation rate and a generation time were needed to convert the demographic time to years, and the output of PSMC is biased if these parameters are under- or overestimated. Unfortunately, the mutation rates and generation times of perennial trees are notoriously difficult to estimate.

We generally calculated the mutation rate using the number of mismatches per nucleotide and the divergence time inferred from fossil data. The accuracy of this method depends strongly on reliable estimates of fossil age, however, and the identification and interpretation of fossils are often fraught with uncertainty. Fortunately, because some plant species experienced a whole-genome duplication (WGD) event near the Cretaceous–Tertiary (K–T) boundary (Tuskan *et al.*, 2006; Huang *et al.*, 2009; Velasco *et al.*, 2010; Young *et al.*, 2011), the timing of a species' WGD can be used to calibrate its mutation rate. Luo *et al.* (2015) uncovered a WGD in the genome of *J. regia*, so we used K_s of paralogous genes at the whole-genome level (Martinez-Garcia *et al.*, 2016) and the timing of WGD (66–80 Ma) of *Juglans* to estimate the mutation rate (see Methods S5). We considered 80 Ma most likely for the *Juglans* WGD event (see Methods S5), so a mutation rate of 2.06×10^{-9} nucleotide yr^{-1} was adopted throughout this study. For comparison, we also presented the PSMC results in Fig. S2–S4 using 66 Ma as the time for the WGD, which translated into an estimated mutation rate of 2.50×10^{-9} nucleotide yr^{-1}.

Generation time can vary among species within a genus and even among different geographic populations of the same species. Thus, in addition to the assumed generation time of 30 yr in our analysis (Fig. 1) we also generated the PSMC plots based on two other generation times (20 yr, Fig. S5; 40 yr, Fig. S6; and all generation times overlaid, Fig. S7). Moreover, we used a random generation time for each species within the range of 20–40 yr to generate the plots (Fig. S8; see Methods S6). Comparing Fig. 1 with Figs S2–S8, it is readily apparent that different estimates of mutation rate (2.06×10^{-9} vs 2.50×10^{-9}) and generation time (20, 30, and 40 yr) lead to similar qualitative conclusions about N_e dynamics because, as is well known, they do not change the shape of the N_e curve, but only move the curve along the axes (Nadachowska-Brzyska *et al.*, 2016).

Divergence analysis using PSMC and G-PhoCS

Pairwise sequentially Markovian coalescent can be used not only to infer changes in N_e over time, but also to estimate divergence times between lineages by means of overlaying the inferred PSMC plots for each lineage (Li & Durbin, 2011). When reading the PSMC plots from the present into the past, between-lineage divergence is inferred to have occurred at the point in time when the trajectories of two overlaid plots become identical. It should be noted, however, that this approach does not account for the possibility that demographic processes other than lineage divergence, such as population structure (probably weak for wind-pollinated trees), or that the two lineages might have had the same effective population size as a result of chance, can produce a similar-appearing outcome.

In contrast to the PSMC method, which analyzes individual genomes in isolation, the G-PhoCS program (Gronau *et al.*, 2011) simultaneously considers multiple genomes for a direct estimation of divergence times between the genomes/lineages under study. As a result, G-PhoCS and PSMC complement each other methodologically for the inference of lineage divergence times. Because we do not know the phylogenetic relationship among the three sections of *Juglans*, we performed three separate analyses for each pair of sections. The dataset consisted of 2370 (*Rhysocaryon-Cardiocaryon)*, 2334 (*Rhysocaryon-Dioscaryon*) and 2425 (*Cardiocaryon-Dioscaryon*) 1000 bp nuclear loci, and each section was represented by one species (see Methods S7).

PSMC analysis of *Populus trichocarpa* and *Fraxinus excelsior*

The genome sequences of *P. trichocarpa* (Tuskan *et al.*, 2006) and *F. excelsior* (Sollars *et al.*, 2017) were derived from previous

New Phytologist

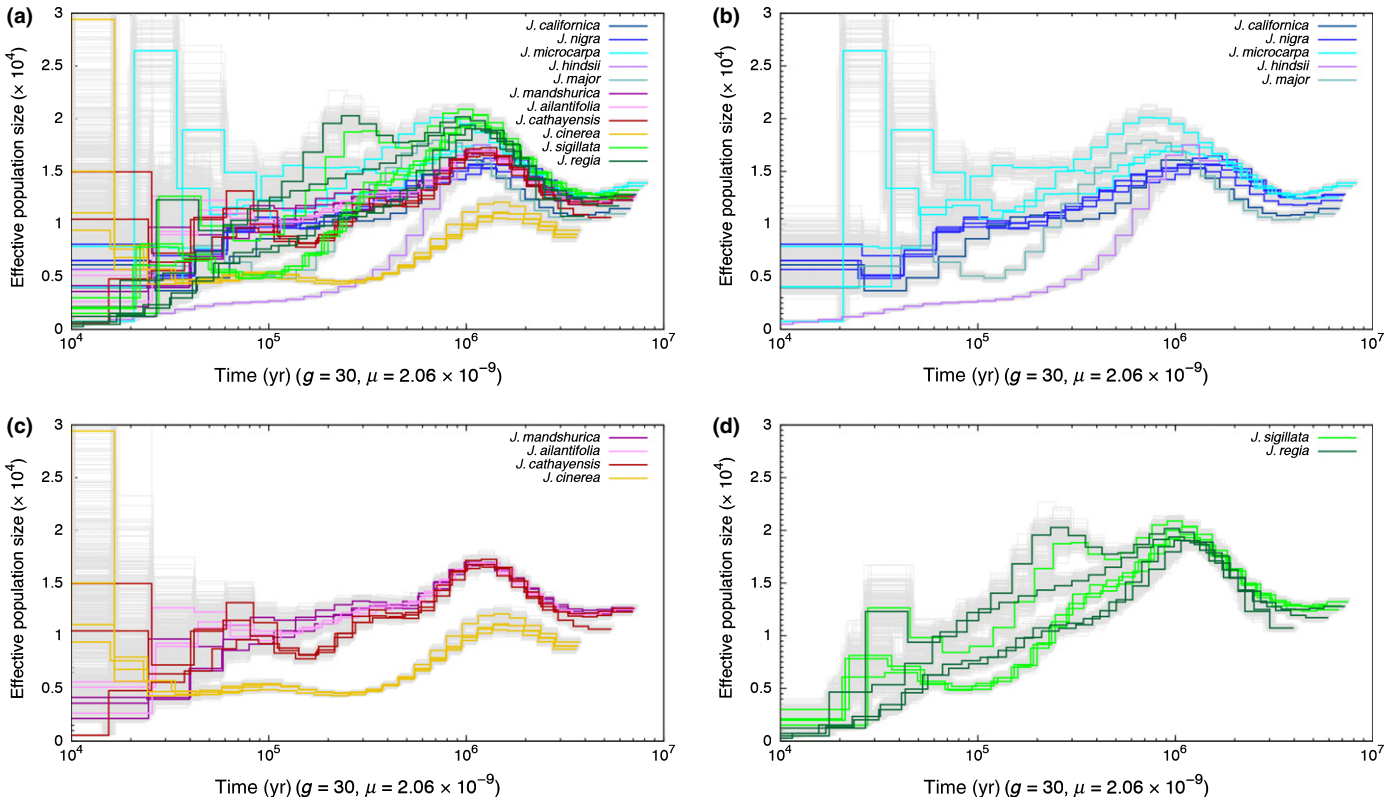

Fig. 1 Pairwise sequentially Markovian coalescent estimates of the changes in effective population size over time for 11 temperate species of *Juglans*. (a) All the individuals; (b) section *Rhysocaryon*; (c) section *Cardiocaryon*; (d) section *Dioscaryon*. Each line represents one individual. Gray lines indicate 95% confidence intervals. *g*, generation time (yr); μ, mutation rate.

publications and retrieved from the NCBI (accession no. GCA_000002775.2 and GCA_900149125.1). Similar to *Juglans* species, *P. trichocarpa* and *F. excelsior* are widespread, cold-temperate deciduous trees, so we also performed PSMC analysis on these species using the method described earlier (see Methods S8). A mutation rate of 2.50×10^{-9} nucleotide yr^{-1} and a generation time of 15 yr (Tuskan *et al.*, 2006) were used to convert the demographic time to years for both species. Previously, Sollars *et al.* (2017) used a much smaller substitution rate of 0.5×10^{-9} nucleotide yr^{-1}, which probably explains why the N_e they reported is generally an order of magnitude higher than ours (Fig. 2).

Data accessibility

All DNA sequences have been deposited in the National Center for Biotechnology Information (NCBI) under the BioProject accession no. PRJNA356989.

Results

We first sequenced and *de novo* assembled the genome of *J. mandshurica*, *J. regia* and *J. nigra* to serve as the reference. The final genome assembly of *J. mandshurica* was 558 Mb long (genomic coverage = 145×) and contained 13 810 scaffolds (N50 = 496.9 Kb). The final genome assembly of *J. nigra* was 682 Mb long (genomic coverage = 109×) and contained 18583 scaffolds (N50 = 232.8 Kb). The final genome assembly of *J. regia* was 635 Mb long (genomic coverage = 112×) and contained 25 670 scaffolds (N50 = 310.8 Kb) (Table S4). The draft genomes of *J. mandshurica*, *J. regia* and *J. nigra* included a total of 279.6 Mb (50.10%), 296.6 Mb (46.70%), and 323.1 Mb (47.33%) of repetitive sequences, respectively. Whole-genome sequences from 31 individuals in 11 temperate *Juglans* species generated sequencing depths from 23.3 to 38.8× (Table S2). Heterozygosity ranged from 0.0009 to 0.0024 across species and populations; *J. cinerea* was the most variable and *J. regia* was the least variable (0.0009) (Table S2).

Variation in patterns of N_e dynamics among species and populations

During the period from 3.0 to 1.0 Ma (early Pleistocene), the similarity of the N_e dynamics of all 11 temperate species of *Juglans* was pronounced. Most species' N_e started to increase from $\approx 1.0 \times 10^4$ at *c.* 3.0 Ma, and most species reached their largest population size ($N_e \approx 1.5$–2.0×10^4) at *c.* 1.0–1.5 Ma.

Within section *Rhysocaryon*, *J. microcarpa* had dramatically variable N_e curves among its three lineages: one lineage experienced one cycle of population expansion and decline but the other two lineages experienced a second cycle of dramatic population expansion and decline between 100 000 and 20 000 yr ago (100–20 kya). *Juglans hindsii*, *Juglans major*, and *Juglans californica* declined rapidly, while the N_e of *J. nigra* declined slowly (Fig. 1b).

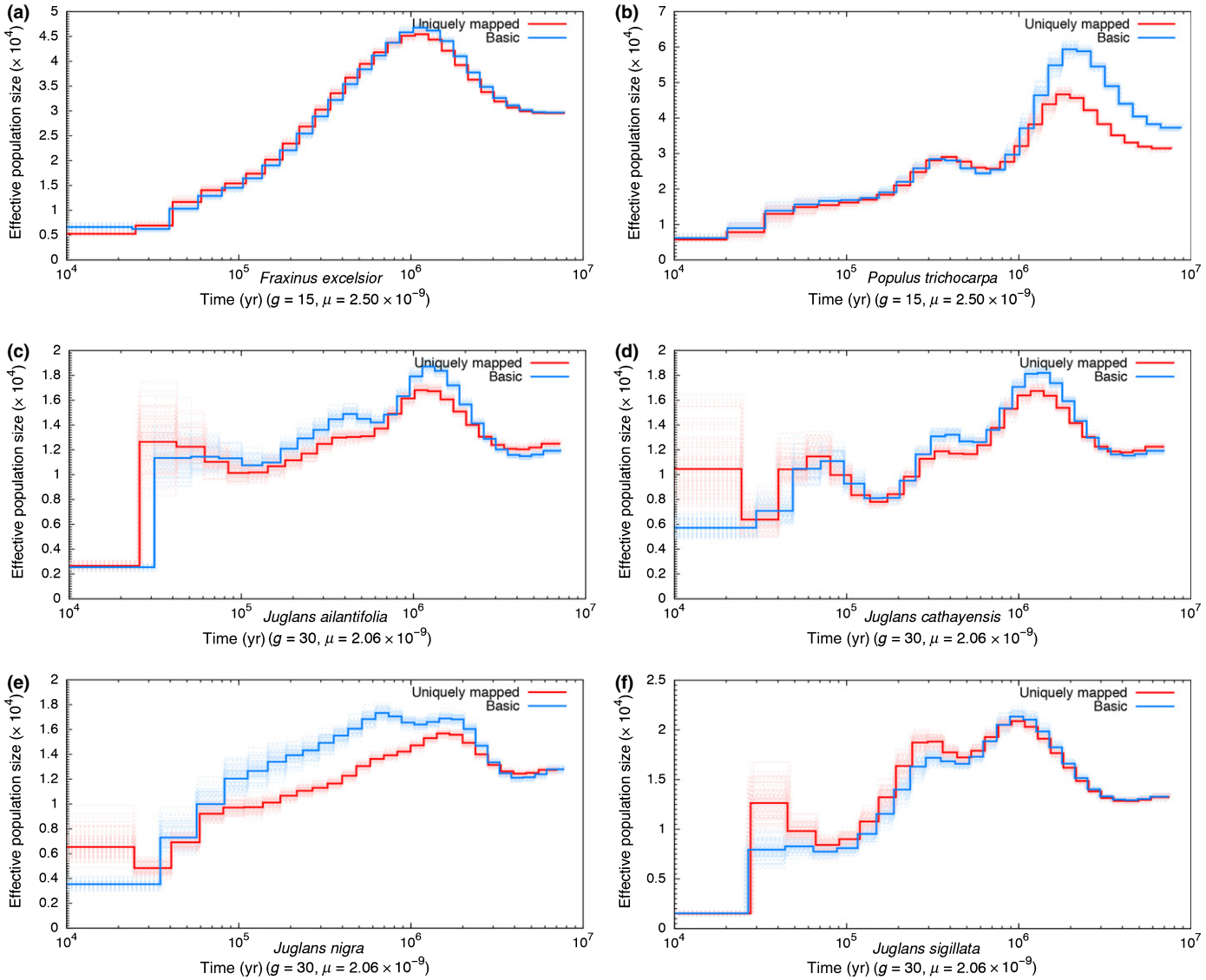

Fig. 2 Comparison of pairwise sequentially Markovian coalescent estimates of the changes in effective population size over time for *Fraxinus excelsior* (a), *Populus trichocarpa* (b), *Juglans ailantifolia* (c), *Juglans cathayensis* (d), *Juglans nigra* (e), and *Juglans sigillata* (f) when using a basic strategy of data filtering (blue) vs using only uniquely mapped repeats (red). *g*, generation time (yr); μ, mutation rate.

Within section *Cardiocaryon*, the N_e of *J. cinerea* declined rapidly and its trajectory was distinctly different than all other species after 2.0 Ma. Four lineages of *J. cinerea* showed a pattern of population increase in the period 20–10 kya. *Juglans cathayensis* declined slowly for most of the last million yr and then experienced a second cycle of population expansion and decline between *c.* 200 and 40 kya, and two lineages increased again during 30–20 kya (Fig. 1c). The inferred history of N_e of *J. cathayensis* was unlike that of *J. mandshurica* and *J. ailantifolia* at *c.* 300 kya. The N_e of *J. mandshurica* and *J. ailantifolia* remained stable between 300 and 50 kya and then declined to $N_e \approx 0.5 \times 10^4$ during 20–10 kya (Fig. 1c).

Within section *Dioscaryon*, the N_e curves began to diverge *c.* 1.0 Ma–700 kya, not only between *J. regia* and *J. sigillata* but also within each species (Fig. 1d). Starting at 1.0 Ma, two lineages of *J. regia* declined rapidly, one lineage declined slowly, and the fourth lineage experienced a second and third expansion at *c.* 300 and 40 kya, respectively. Three lineages of *J. sigillata* rapidly declined between 900 and 100 kya and increased slowly between 100 and 20 kya; the fourth lineage of *J. sigillata* experienced a second cycle of population expansion since 400 kya and a third cycle of expansion during the period 50–30 kya (Fig. 1d). All lineages in *Dioscaryon* experienced population declines during the period 20–10 kya when the N_e of *J. regia* became smaller than *J. sigillata*.

The impact of unique-mapping filter on the estimates of N_e

In addition to basic filters recommended in the original PSMC program (Li & Durbin, 2011), we retained only those reads that were uniquely mapped to the reference genome (unique-mapping filter) to mitigate the concern caused by high percentages of repeats in plant genomes. Upon comparison, we noticed that

with or without the unique-mapping filter, very similar N_e curves resulted in most cases, but definitely not in all cases (Figs 2, S9–S11), which illustrated the importance of performing a unique-mapping filter for PSMC analysis. In general, curves generated using uniquely mapped filtered data showed lower N_e in the distant past and higher N_e in the recent past as compared with the corresponding curves derived without the filter (Fig. 2).

Divergence time between sections

When the PSMC plots in Fig. 1 were overlaid, they revealed that divergence among lineages of *Juglans* initiated *c.* 1.0 Ma, a surprisingly recent estimate. Analysis using G-PHOCS indicated that the divergence times for *Rhysocaryon-Cardiocaryon*, *Rhysocaryon-Dioscaryon* and *Cardiocaryon-Dioscaryon* were 1.64, 2.77 and 2.71 Ma, respectively (Tables 1, S5), and that N_e estimates were $2.77–3.27 \times 10^4$ for *J. nigra* (*Rhysocaryon*), $7.52–7.88 \times 10^4$ for *J. mandshurica* (*Cardiocaryon*), and $4.12–6.79 \times 10^4$ for *J. sigillata* (*Dioscaryon*) (Table S5).

N_e dynamics of *Populus trichocarpa* and *Fraxinus excelsior*

Like *Juglans* species, *P. trichocarpa* and *F. excelsior* are deciduous tree species with broad, cold-temperate distributions. But unlike most *Juglans* species, there are high-quality reference genomes for both *P. trichocarpa* and *F. excelsior*, so the results of PSMC analysis for these species served as a type of control for our analysis of *Juglans*. About 3.0 Ma, the N_e of *F. excelsior* started to increase from $\approx 3.0 \times 10^4$ until the species reached its largest effective population size ($N_e \approx 4.5–5.0 \times 10^4$), *c.* 1.0–1.5 Ma. Beginning 1.0 Ma and continuing until 20–10 kya, *F. excelsior* N_e declined steadily and reached a low ebb ($N_e \approx 0.5 \times 10^4$) (Fig. 2a). The N_e of *P. trichocarpa* started to increase from $\approx 3.0 \times 10^4$ at *c.* 4–5 Ma, and the species reached its largest N_e ($\approx 6.0 \times 10^4$) at *c.* 2.0 Ma. From 2.0 Ma to 600 kya, the N_e of *P. trichocarpa* declined rapidly to $\approx 2.5 \times 10^4$. A second cycle of population expansion and decline occurred between 600 and 200 kya, leading to a final decline to $N_e \approx 0.5 \times 10^4$ at 20–10 kya (Fig. 2b).

Discussion

Conclusions based on comparative demography can now be used to evaluate the relative importance of extrinsic and intrinsic factors in controlling population dynamics. We found that the timing and amplitude of changes in N_e differed among temperate walnut species, and even among conspecific lineages within the same species for *J. microcarpa*, *J. regia* and *J. sigillata*. This finding supports the conclusion that population histories of temperate walnut species were not driven by extrinsic environmental changes alone, and a key role was probably played by species-specific factors such as coevolutionary interactions with specialized pathogens. Our study contributes new insights into the historical (e.g. vicariance, dispersion) and demographic (e.g. patterns of gene flow, population bottlenecks) processes responsible for the observed distribution of genetic variation and species diversity at a range of spatial scales.

Table 1 Lineage divergence times estimated using G-PHOCS

Section pair	τ ($\times 10^4$)	Divergence time (yr $\times 10^6$)
Cardiocaryon-Rhysocaryon	33.74 (31.31–36.36)	1.64 (1.52–1.77)
Disocaryon-Rhysocaryon	57.14 (54.15–60.04)	2.77 (2.63–2.91)
Disocaryon-Cardiocaryon	55.89 (52.68–59.19)	2.71 (2.56–2.87)

The divergence time parameter τ is scaled by the mutation rate of 2.06×10^{-9} yr^{-1}. The posterior 95% credible interval for each estimate is in parentheses.

Idiosyncratic responses of temperate walnut lineages to climate change

Using the PSMC method, we reconstructed the fluctuations in N_e for 11 walnut species starting 10 kya and extending as far back as 4.0–5.0 Ma (Fig. 1). The PSMC results clearly indicated that the walnut species shared strongly similar trajectories of N_e before 1.0 Ma, but the shapes of the N_e curves diverged *c.* 1.0 Ma, not only within each section (Fig. 1b–d) but also within several species. The common pattern of population expansion during the period of ~1.0–3.0 Ma might have been the result of a congruent response by each species to early Pleistocene cooling. But a more parsimonious explanation is that these walnut species shared a common ancestry until ~1.0 Ma when the trajectories of population size began to vary considerably between sections (Fig. 1a), within each section (Fig. 1b–d), and even among different populations of species, as shown by *J. microcarpa*, *J. regia*, and *J. sigillata*. Within species, *J. microcarpa*, *J. regia* and *J. sigillata* had distinctly different N_e curves among different lineages during most of the middle and late Pleistocene. This could mean that different lineages within *J. microcarpa*, *J. regia* and *J. sigillata* might actually constitute distinct (cryptic) species, but this conjecture needs to be confirmed in future studies.

Even during the Last Glacial Maximum (LGM; 10–20 kya), over half the *Juglans* species we studied did not show the commonly expected, glacially induced population decline. Moreover, lineages within many species had strongly different patterns of population size changes during LGM (Fig. 1a). Thus, closely related walnut species in broadly similar environments responded idiosyncratically to Pleistocene climate changes; some species' N_e increased whereas others decreased or remained stable. This result lends strong support to Ricklef's hypothesis that intrinsic factors such as coevolution between antagonists can be the major drivers of long-term population dynamics. Reassuringly, it is well known that *J. cinerea* became rare throughout its native North American range over the past 100 yr because of a fungal pathogen (Broders *et al.*, 2015), *Ophiognomonia clavigignenti-juglandacearum* (*Oc-j*). *Oc-j* kills *J. cinerea* at all stages of growth and has caused extensive mortality of butternut (Parks *et al.*, 2013), but it is fatal to *J. cinerea* only; the closely related Japanese walnut (*J. ailantifolia*) appears resistant to it (Ostry & Woeste, 2004). This system probably constitutes a living example of pathogen–host coevolution and a host jump that caused rapid tree population decline, as

envisaged by Ricklefs (2015a). Similar declines in other genera of forest trees occurred with the arrival in eastern North America of Dutch elm disease, chestnut blight, and the emerald ash borer (Ricklefs, 2015b).

Recent studies have used PSMC to reveal variable population dynamics of related taxa, independent of their evolutionary relationship. Kozma *et al.* (2016) found asynchronous population expansions and contractions among three grouse species. In particular, the most cold-adapted rock ptarmigan (*Lagopus muta*) showed clear evidence of consistent population decline throughout the last ice age, whereas the most temperate black grouse (*Lyrurus tetrix*) experienced a cycle of population decline and growth (with substantial population expansion before the last glacial maximum). Similarly, Nadachowska-Brzyska *et al.* (2016) found that the timing and amplitude of changes in N_e differed among four closely related *Ficedula* flycatcher species, and even among conspecific populations within species. Several related studies (e.g. Burbrink *et al.*, 2016; Fan *et al.*, 2016) have examined how species responded to climate change by focusing on codistributed taxa less closely related than those we studied. Comparing phylogeographic patterns across species within the same region, one may likewise discern whether changes in distribution and population size represent the influence of extrinsic factors affecting whole communities, or whether they can be ascribed to intrinsic factors particular to each species. For example, using a hierarchical approximate Bayesian computational approach, Burbrink *et al.* (2016) found asynchronous demographic responses to Pleistocene climate change across lineages of snakes, lizards, turtles, mammals, birds, salamanders and frogs of eastern North America. Similarly, Fan *et al.* (2016) revealed incongruent demographic histories among three dominant species of subtropical evergreen broadleaved forest trees in China. Such studies provide complementary evidence for our inference that intrinsic biological interactions play an important role in determining effective population size of temperate species.

Regional environmental effects could potentially drive patterns of N_e across species and confound predicted associations between global climate and demography over evolutionary time. Thus, we cannot exclude some region-specific effects from confounding our interpretations of the idiosyncratic responses. Two exceptions are *J. regia*–*J. sigillata* (species that hybridize) and *J. cinerea*–*J. nigra* (which do not hybridize). It is interesting, nonetheless, that although *J. nigra* and *J. cinerea* codistribute in eastern North America, their N_e curves were distinct (Fig. 1b,c), and the codistributed lineages of *J. regia* and *J. sigillata* from southwestern China also had distinctive N_e curves beginning *c.* 500 kya (Fig. 1d). Therefore, it was unlikely that specific, regional effects strongly influenced our analysis of the relative importance of extrinsic environmental change vs intrinsic biological interactions in controlling long-term population dynamics.

Explosive ice age divergence of walnut lineages

When the plots of individual species (Fig. 1) were overlaid, it revealed that divergence among lineages of *Juglans* initiated at *c.* 1.0 Ma, a much more recent estimate than the fossil data appeared to indicate (Manchester, 1987), i.e., an initial split into *Rhysocaryon* and *Cardiocaryon* at *c.* 45 Ma in North America, which clearly resolved ~38 Ma. Because of the great disparities between the interpretations of the fossil data and the results of PSMC, we also used an independent Bayesian demographic inference method, G-PhoCS (Gronau *et al.*, 2011), to estimate the divergence time between sections. Interestingly, the G-PhoCS results were in striking conformity with our finding based on PSMC plots (Tables 1, S3), which greatly enhance our confidence in the molecular estimates. The apparent inconsistency between molecular data and fossil records in the timing of divergence between sections of *Juglans* is not insurmountable, given the general uncertainty in the identification and interpretation of fossils. The fossil record of *Juglans* is no exception.

Fossil nuts of *Juglans* were usually identified via detailed comparisons with extant walnuts, but fossil morphology is often too variable to allow unambiguous identification of fossil remains to species or section level (cf. Manchester, 1987). Although the earliest known fossil walnut fruit was thought to belong to *Rhysocaryon* (Manchester, 1987), it had a simple seed and locule morphology, more similar to that of modern *Cardiocaryon*, and variable development of the secondary septum, which ranges from absent, as in modern *Cardiocaryon*, to present and well developed, as in modern *Rhysocaryon* and *Dioscaryon*. In fact, van der Ham (2015) suggests that 15 fossil *Cardiocaryon* species described so far should be merged into a single, widespread fossil species, *Juglans bergomensis*, with sharply ridged nuts not unlike the extant *J. cinerea*. *J. bergomensis* would have been continuously distributed across North America, East Asia and Europe in the Neogene and early Pleistocene.

There is paleoecological evidence supporting a recent divergence of *Juglans* as well. Giterman & Golubeva (1967) showed that several tree species typically considered temperate, including walnut, persisted in northern Siberia through the early Pleistocene. Millennial-scale pollen studies of the sediments in Lake El'gygytgyn ($67^0$30′N, $172^0$50′E), *c.* 100 km to the north of the Arctic Circle, clearly demonstrate that the pollen of many temperate forest species, including *Juglans*, was present at that location from 2.7 to 2.5 Ma (Andreev *et al.*, 2016). Thus it is highly plausible that early Pleistocene glacial cycles did not prevent walnut trees from surviving in high latitudes. This finding is important for the timing of divergence of walnut lineages because Beringia may have been a corridor connecting the fauna and flora (including trees) (Hoffecker & Elias, 2003) of the two continents (Hundertmark *et al.*, 2002; Goebel *et al.*, 2008). Because walnut trees are wind-pollinated and extensive gene flow via pollen seems very likely, it is plausible that ancestral *Juglans* could persist as a single cohesive unit throughout much of its evolutionary history until 1.0–1.5 Ma. In this connection, it is interesting to observe that the very large census of *J. nigra* covering much of the eastern US effectively constitutes a single homogeneous population, which offers a clear example of how *Juglans* can resist local genetic differentiation (Victory *et al.*, 2006).

Despite mounting evidence for the influence of Pleistocene glacial cycles on patterns of intraspecific genetic diversity, it still remains unclear whether these climatic oscillations also acted as

drivers of speciation, especially among lineages with long generation times, such as temperate trees. Our analysis using both PSMC (Fig. 1) and G-PHOCS (Table 1) indicated that all major walnut diversification events dated to the major glacial advances that characterized the latter half of the Pleistocene (Augustin *et al.*, 2004). These results, together with some recent investigations in the genus *Populus* (Levsen *et al.*, 2012; Wang *et al.*, 2014), provide strong evidence for the importance of the middle and late Pleistocene glaciations in driving tree speciation. As is well known, the Earth's climate underwent a fundamental change around ~0.9 Ma when the dominant periodicity of climate cycles changed from 41–100 kya (Augustin *et al.*, 2004), leading to prolonged isolation of populations in glacial refugia, which in turn increased the likelihood of speciation. Needless to say, the hypothesis of very recent divergence of walnut lineages (*c.* 1.0 Ma) awaits further testing, especially by means of phylogenomic analysis.

Unexpected small effective population size for temperate deciduous trees

The maximum N_e for most lineages of *Juglans* estimated over the time span we analyzed was $\sim 2.0 \times 10^4$, and the minimum N_e was less than 1.0×10^3 (Fig. 1). N_e estimated by G-PHOCS was *c.* $2.8–7.9 \times 10^4$ (Table S5). We performed a PSMC analysis for two widespread, cold-temperate deciduous trees, *P. trichocarpa* and *F. excelsior*, which have whole-genome sequences available. Their estimated N_e were *c.* 5.0×10^3–5×10^4 (Fig. 2), surprisingly similar to our estimate for *Juglans*. This finding is in contrast to the common perception that N_e for temperate forest trees is generally quite large, normally on the order of 10^5 or even higher (Petit & Hampe, 2006).

The expectation of large N_e for temperate trees is reasonable because trees can have huge global census sizes (N_c). For instance, the census size of mature beech (*Fagus sylvatica*) in European forests is estimated at 1.5–2 billion (Petit & Hampe, 2006). The rule of thumb is that N_e is on average one-tenth of N_c (Frankham, 1995), but there is considerable variability in N_e /N_c. Hung *et al.* (2014) found that for passenger pigeon, N_e was several orders of magnitude lower than the census population size, once estimated at 3–5 billion. They explained that a low N_e /N_c is typical for outbreak species, which are characterized by dramatic and recurrent changes in population size. If temperate trees fit the outbreak model, only occasionally numbering in the billions during interglacial 'outbreak' phases and then experiencing severe bottlenecks during glaciation, a low N_e would be expected over the last million yr, a time period characterized by large climatic oscillations (especially warm interglacials and harsh, cold glacials). During interglacial periods of Pleistocene, there would be dramatic population expansions resulting in large N_c, but during unfavorable glacial periods, temperate trees in general would experience severe range contraction and fragmentation, leading to small population sizes, which in turn would facilitate the evolution of numerous new species through allopatric speciation, as theoretically demonstrated by Nei *et al.* (1983). Mayr (1970) had examined the relationships between speciation and population size in many different groups of organisms and concluded that the evolution of reproductive isolation occurs faster in small populations than in large populations. Nevertheless, the generality of small effective population size for temperate trees remains uncertain, given that whole-genome sequencing information is only available for a limited number of tree species.

Methodological issues

Like other population genetic methods that use next-generation sequencing data, PSMC is prone to biases that are the result of sequencing/genotyping errors and missing data, so filtering is particularly critical to PSMC analysis. Nadachowska-Brzyska *et al.* (2016) recommended using sequencing data with a mean genome coverage of $\geq 18\times$, a per-site filter of ≥ 10 reads, no more than 25% missing data, and a repeat masked reference genome. These conditions are easily satisfied in humans and animals whose genomes tend to have much lower proportions of repetitive sequences than plants (Tuskan *et al.*, 2006; Myburg *et al.*, 2014; Martinez-Garcia *et al.*, 2016). For many plants, the high prevalence of repeats in assembled genomes (Table S3) is reason for concern about whether data quality permits a reliable PSMC analysis. Because of the high proportion of repeats in our assembled *Juglans* genomes (Table S3), the proportion of missing data after repeat masking was nearly 50%, much higher than the recommended 25%. Thus, we recommend a strategy of using a filter that accepts only uniquely mapped, properly paired, and nonduplicated reads. This approach should minimize the influence of repeats and missing data when applying PSMC to plants. Upon comparison of outcomes based on different filtering strategies, we found that N_e curves were affected by using a filtering strategy that employed only uniquely mapped reads (Figs 2, S9–S11). Most notably, curves generated using the default filtering strategy indicated higher N_e in the distant past and lower N_e in the recent past compared with the corresponding curves based on our filtering criteria. The reason may be that multiply mapped reads would lead to intervals of higher heterozygosity because of false-positive SNPs, especially in repeat-dense regions. Thus, in improperly filtered data, younger states (low heterozygosity) would be mistakenly regarded as older states (showing higher heterozygosity) in PSMC, resulting in overestimated population sizes in ancient times and underestimated population sizes in recent times. We concluded that a filtering strategy of using only uniquely mapped reads is essential for reliable demographic inference, at least when applying PSMC to genomes with high proportions of repeats.

Acknowledgements

This work was supported by the State Key Basic Research and Development Plan (2017YFA0605104), the National Natural Science Foundation of China (31421063 and 41671040), the '111' Program of Introducing Talents of Discipline to Universities (B13008), and a key project of State Key Laboratory of Earth Surface Processes and Resource Ecology. The mention of a trademark, proprietary product, or vendor does not constitute a

guarantee or warranty of the product by the US Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

Author contributions

W-N.B. and D-Y.Z. conceived of the study and wrote the manuscript. W-N.B., B-W.Z. and P-C.Y. assembled the genomes. B-W.Z., W-N.B. and P-C.Y. performed the PSMC and G-PhoCS analyses. K.L. and K.E.W. provided samples, contributed ideas, and assisted in editing the manuscript.

References

Alvaradoserrano DF, Knowles LL. 2014. Ecological niche models in phylogeographic studies: applications, advances and precautions. *Molecular Ecology Resources* **14**: 233–248.

AmborellaGenomeProjects. 2013. The Amborella genome and the evolution of flowering plants. *Science* **342**: 1241089.

Andreev A, Tarasov PE, Wennrich V, Melles M. 2016. Millennial-scale vegetation changes in the north-eastern Russian Arctic during the Pliocene/Pleistocene transition (2.7–2.5 Ma) inferred from the pollen record of Lake El'gygytgyn. *Quaternary Science Reviews* **147**: 245–248.

Aradhya MK, Potter D, Gao F, Simon CJ. 2007. Molecular phylogeny of *Juglans* (Juglandaceae): a biogeographic perspective. *Tree Genetics & Genomes* **3**: 363–378.

Augustin L, Barbante C, Barnes PRF, Barnola JM, Bigler M, Castellano E, Cattani O, Chappellaz J, DahlJensen D, Delmonte B *et al.* 2004. Eight glacial cycles from an Antarctic ice core. *Nature* **429**: 623–628.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA Bridge between population genetics ans systematics. *Annual Review of Ecology and Systematics* **18**: 489–522.

Avise JC, Bowen BW, Ayala FJ. 2016. In the light of evolution X: comparative phylogeography. *Proceedings of the National Academy of Sciences, USA* **113**: 7957–7961.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**: 717–745.

Broders KD, Boraks A, Barbison L, Brown JL, Boland GJ. 2015. Recent insights into the pandemic disease butternut canker caused by the invasive pathogen Ophiognomonia clavigignenti-juglandacearum. *Forest Pathology* **45**: 1–8.

Burbrink FT, Chan YL, Myers EA, Ruane S, Smith BT, Hicherson MJ. 2016. Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates. *Ecology Letters* **19**: 1457–1467.

Burbrink FT, Fontanella FM, Pyron RA, Guiher TJ, Jimenez C. 2008. Phylogeography across a continent: the evolutionary and demographic history of the North American racer (Serpentes: Colubridae: Coluber constrictor). *Molecular Phylogenetics and Evolution* **47**: 274–288.

Fan D, Hu W, Li B, Morris AB, Zheng M, Soltis DE, Soltis PS, Zhang Z. 2016. Idiosyncratic responses of evergreen broad-leaved forest constituents in China to the late Quaternary climate changes. *Scientific Reports* **6**: 31044.

Frankham R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* **89**: 491.

Giterman RE, Golubeva LV. 1967. Vegetation of eastern Siberia during the Anthropogene period. In: Hopkins DM, ed. *The bering land bridge*. Stanford, CA: Stanford University Press, 232–235.

Gleason HA. 1926. The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* **53**: 7–26.

Goebel T, Waters MR, O'Rourke DH. 2008. The Late Pleistocene dispersal of modern humans in the Americas. *Science* **319**: 1497–1502.

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics* **43**: 1031–1034.

van der Ham R. 2015. On the history of the butternuts (*Juglans* section *Cardiocaryon*, Juglandaceae). *Palaeontographica Abteilung B-Palaophytologie* **293**: 125–147.

Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**: 1341–1351.

Harrison SP, Yu G, Takahara H, Prentice IC. 2001. Palaeovegetation – diversity of temperate plants in east Asia. *Nature* **413**: 129–130.

Hawkins BA, Rueda M, Rangel TF, Field R, Diniz-Filho JAF. 2014. Community phylogenetics at the biogeographical scale: cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography* **41**: 23–38.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**: 183–195.

Hewitt GM. 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**: 617–638.

Hoban SM, Borkowski DS, Brosi SL, McCleary T, Thompson LM, McLachlan JS, Pereira MA, Schlarbaum SE, Romero-Severson J. 2010. Range-wide distribution of genetic diversity in the North American tree Juglans cinerea: a product of range shifts, not ecological marginality or recent population decline. *Molecular Ecology* **19**: 4876–4891.

Hoffecker JF, Elias SA. 2003. Environment and archeology in Beringia. *Evolutionary Anthropology* **12**: 34–49.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P. 2009. The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* **41**: 1275–1281.

Hundertmark KJ, Shields GF, Udina IG, Bowyer RT, Danilkin AA, Schwartz CC. 2002. Mitochondrial phylogeography of moose (*Alces alces*): Late Pleistocene divergence and population expansion. *Molecular Phylogenetics and Evolution* **22**: 375–387.

Hung C, Shaner PL, Zink RM, Liu W, Chu T, Huang W, Li S. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proceedings of the National Academy of Sciences, USA* **111**: 10636–10641.

Kozma R, Melsted P, Magnusson KP, Hoglund J. 2016. Looking into the past – the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Molecular Ecology* **25**: 570–580.

Levsen ND, Tiffin P, Olson MS. 2012. Pleistocene speciation in the genus *Populus* (Salicaceae). *Systematic Biology* **61**: 401–412.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**: 493–496.

Luo M, You FM, Li P, Wang J, Zhu T, Dandekar AM, Leslie CA, Aradhya MK, Mcguire PE, Dvorak J. 2015. Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. *BMC Genomics* **16**: 707.

Manchester SR. 1987. The fossil history of Juglandaceae. *Missouri Botanical Garden Monograph* **21**: 1–137.

Manning WE. 1978. The classification within the Juglandaceae. *Annals of the Missouri Botanical Garden* **65**: 1058–1087.

Martinez-Garcia PJ, Crepeau MW, Puiu D, Gonzalezibeas D, Whalen J, Stevens K, Paul R, Butterfield TS, Britton MT, Reagan RL. 2016. The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of non-structural polyphenols. *Plant Journal* **87**: 507–532.

Mayr E. 1970. *Populations, species, and evolution.* Cambridge, MA, USA: Harvard University Press.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D. 2014. The genome of *Eucalyptus grandis*. *Nature* **510**: 356.

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology* **25**: 1058–1072.

Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Current Biology* **25**: 1375–1380.

Nei M, Maruyama T, Wu CI. 1983. Models of evolution of reproductive isolation. *Genetics* **103**: 557–579.

Ostry M, Woeste K. 2004. Spread of butternut canker in North America, host range, evidence of resistance within butternut populations and conservation genetics. In: Michler CH, Pijut PM, Van Sambeek J, Coggeshall MV, Seifert J, Woeste K, Overton R, Ponder F, eds. *Black walnut in a new century, proceedings of the 6th Walnut Council research symposium; 2004 July 25–28*. Lafayette, IN, USA: Gen. Tech. Rep. NC-243, 114–120.

Papadopoulou A, Knowles LL. 2016. Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences, USA* **113**: 8018–8024.

Parks AM, Jenkins MA, Woeste KE, Ostry ME. 2013. Conservation status of a threatened tree species: establishing a baseline for restoration of *Juglans cinerea* L. in the southern Appalachian Mountains, USA. *Natural Areas Journal* **33**: 413–426.

Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics* **37**: 187–214.

Prates I, Xue AT, Brown JL, Alvarado-Serrano DF, Rodrigues MT, Hickerson MJ, Carnaval AC. 2016. Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences, USA* **113**: 7978–7985.

Prinzing A, Durka W, Klotz S, Brandl R. 2001. The niche of higher plants: evidence for phylogenetic conservatism. *Proceedings of the Royal Society B-Biological Sciences* **268**: 2383–2389.

Qian H, Ricklefs RE. 2000. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**: 180–182.

Ricklefs RE. 2015a. Intrinsic dynamics of the regional community. *Ecology Letters* **18**: 497–503.

Ricklefs RE. 2015b. How tree species fill geographic and ecological space in eastern North America. *Annals of Botany* **115**: 949–959.

Ruane S, Torrescarvajal O, Burbrink FT. 2015. Independent demographic responses to climate change among temperate and tropical milksnakes (Colubridae: genus *Lampropeltis*). *PLoS ONE* **10**: e0128543.

Sollars ESA, Harper AL, Kelly LJ, Sambles CM, Ramirez-Gonzalez RH, Swarbreck D, Kaithakottil G, Cooper ED, Uauy C, Havlickova L *et al.* 2017. Genome sequence and genetic diversity of European ash trees. *Nature* **541**: 212–216.

Stanford AM, Harden R, Parks CR. 2000. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *American Journal of Botany* **87**: 872–882.

Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics* **13**: 36–46.

Tuskan GA, Difazio SP, Jansson S, Bohlmann J, Grigoriev IV, Hellsten U, Putnam NH, Ralph SK, Rombauts S, Salamov A. 2006. The genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar S, Troggio M, Pruss D. 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* **42**: 833–839.

Victory ER, Glaubitz JC, Rhodes OE Jr, Woeste KE. 2006. Genetic homogeneity in *Juglans nigra* (*Juglandaceae*) at nuclear microsatellites. *American Journal of Botany* **93**: 118–126.

Wallberg A, Han F, Wellhagen G, Dahle B, Kawata M, Haddad N, Simoes ZLP, Allsopp MH, Kandemir I, La Rua PD. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics* **46**: 1081–1088.

Wang J, Kallman T, Liu J, Guo Q, Wu Y, Lin K, Lascoux M. 2014. Speciation of two desert poplar species triggered by Pleistocene climatic oscillations. *Heredity* **112**: 156–164.

Wiens JJ, Graham CH. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics* **36**: 519–539.

Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Modern distribution of *Juglans* taxa and sampling sites used in this study.

Fig. S2 PSMC estimates of the changes in effective population size over time for 11 species of *Juglans* using a mutation rate of 2.50×10^{-9} nucleotide yr^{-1} and a generation time of 20 yr.

Fig. S3 PSMC estimates of the changes in effective population size over time for 11 species of *Juglans* using a mutation rate of 2.50×10^{-9} nucleotide yr^{-1} and a generation time of 30 yr.

Fig. S4 PSMC estimates of the changes in effective population size over time for 11 species of *Juglans* using a mutation rate of 2.50×10^{-9} nucleotide yr^{-1} and a generation time of 40 yr.

Fig. S5 PSMC estimates of the changes in effective population size over time for 11 species of *Juglans* using a mutation rate of 2.06×10^{-9} nucleotide yr^{-1} and a generation time of 20 yr.

Fig. S6 PSMC estimates of the changes in effective population size over time for 11 species of *Juglans* using a mutation rate of 2.06×10^{-9} nucleotide yr^{-1} and a generation time of 40 yr.

Fig. S7 PSMC estimates of the changes in effective population size over time for 11 species of *Juglans* using a mutation rate of 2.06×10^{-9} nucleotide yr^{-1} and separating generation time.

Fig. S8 PSMC estimates of the changes in effective population size over time for 11 species of *Juglans* using a mutation rate of 2.06×10^{-9} nucleotide yr^{-1} and a random generation time.

Fig. S9 PSMC estimates of the changes in effective population size over time for uniquely mapped and basic filtering strategy of section *Rhysocaryon.*

Fig. S10 PSMC estimates of the changes in effective population size over time for uniquely mapped and basic filtering strategy of section *Cardiocaryon.*

Fig. S11 PSMC estimates of the changes in effective population size over time for uniquely mapped and basic filtering strategy of section *Dioscaryon.*

Table S1 The sources of samples used for analysis

Table S2 Summary statistics for the individual whole-genome sequences

Table S3 Summary of repeat elements found by repeat masking

Table S4 Summary of assembled genomes

Table S5 Lineage divergence times estimated from the dataset using G-PhoCS

Methods S1 Genome sequencing and assembly for three walnut species.

Methods S2 Repeat prediction and annotation.

Methods S3 Resequencing and mapping.

Methods S4 Reconstructing the past population dynamics – the PSMC.

Methods S5 Mutation rate and generation time estimation.

Methods S6 PSMC with disparate generation time between species.

Methods S7 Divergence analysis using G-PhoCS

Methods S8 PSMC of *Populus trichocarpa* and *Fraxinus excelsior*.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.