New Phytologist

Research

Check for updates

Evolutionary radiation of the Eurasian *Pinus* species under pervasive gene flow

Wei Zhao[1]*, Jie Gao[2]*, David Hall[3], Bea Angelica Andersson[1], Jade Bruxaux[1], Kyle W. Tomlinson[4], Andreas D. Drouzas[5], Yoshihisa Suyama[6] and Xiao-Ru Wang[1,7]

[1]Department of Ecology and Environmental Science, Umeå Plant Science Center, Umeå University, Umeå, SE-90187, Sweden; [2]CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan, 666303, China; [3]Forestry Research Institute of Sweden (Skogforsk), Sävar, SE-91833, Sweden; [4]Center for Integrative Conservation & Yunnan Key Laboratory for Conservation of Tropical Rainforests and Asian Elephant, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan, 666303, China; [5]Laboratory of Systematic Botany and Phytogeography, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece; [6]Graduate School of Agricultural Science, Tohoku University, Miyagi, 989-6711, Japan; [7]National Engineering Research Center of Tree Breeding and Ecological Restoration, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, 100083, China

Author for correspondence:
Xiao-Ru Wang
Email: xiao-ru.wang@umu.se

Received: *13 September 2023*
Accepted: *4 March 2024*

New Phytologist (2024) **242:** 2353–2368
doi: 10.1111/nph.19694

Key words: demographic history, divergent adaptation, ecological gradients, introgression, phylogeny, *Pinus* evolution, selection.

Summary

- Evolutionary radiation, a pivotal aspect of macroevolution, offers valuable insights into evolutionary processes. The genus *Pinus* is the largest genus in conifers with *c*. 90% of the extant species emerged in the Miocene, which signifies a case of rapid diversification. Despite this remarkable history, our understanding of the mechanisms driving radiation within this expansive genus has remained limited.
- Using exome capture sequencing and a fossil-calibrated phylogeny, we investigated the divergence history, niche diversification, and introgression among 13 closely related Eurasian species spanning climate zones from the tropics to the boreal Arctic.
- We detected complex introgression among lineages in subsection *Pinus* at all stages of the phylogeny. Despite this widespread gene exchange, each species maintained its genetic identity and showed clear niche differentiation. Demographic analysis unveiled distinct population histories among these species, which further influenced the nucleotide diversity and efficacy of purifying and positive selection in each species.
- Our findings suggest that radiation in the Eurasian pines was likely fueled by interspecific recombination and further reinforced by their adaptation to distinct environments. Our study highlights the constraints and opportunities for evolutionary change, and the expectations of future adaptation in response to environmental changes in different lineages.

Introduction

Evolutionary radiation, and adaptive radiation in particular, plays a significant role in macroevolution, and as such, serves as a valuable resource for studying the process of evolution. Adaptive radiations typically exhibit several distinct attributes, including rapid speciation from a common ancestor and a phenotype–environment correlation resulting from divergent natural selection in distinct ecological niches (Schluter, 2000; Gavrilets & Losos, 2009; Losos, 2010). Ecological opportunity and genetic and functional novelties are often considered prerequisites for adaptive radiation (Simpson, 1953; Schluter, 2000; Galis, 2001). An important source of genetic novelty is hybridization and introgression, which is well recognized as an evolutionary stimulus (Anderson & Stebbins, 1954; Arnold, 1997; Rieseberg, 1997). Indeed, genetic investigations into the classical examples of adaptive radiation in many plant groups, as well as Darwin's Finches and African cichlids, have revealed widespread hybridization and recruitment of potentially adaptive variants from ancestral variation (Seehausen, 2006; Lamichhaney *et al*., 2015; Meier *et al*., 2017; Schenk, 2021). However, it is worth noting that the relative significance of genetic and ecological factors, as well as their intricate interplay in driving rapid diversification, is likely to differ across distinct taxonomic groups.

The genus *Pinus*, with *c*. 100 species, is the largest genus of conifers (Mirov, 1967; Price *et al*., 1998; Farjon, 2001). The origin of the genus tracks back to the late Jurassic or early Cretaceous, dating from 177 to 135 million years ago (Ma) (Leslie *et al*., 2018; Jin *et al*., 2021). Remarkably, *c*. 90% of extant species emerged during the Miocene (Saladin *et al*., 2017; Jin *et al*., 2021) which signifies rapid speciation over roughly 200 000 generations. Despite this, *Pinus* species have seldom been considered as a case of adaptive radiation. Explorations into the evolutionary history of the genus and its association with past ecological and geological events have

---

*These authors contributed equally to this study.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

offered a broad perspective on the ecological drivers of *Pinus* evolution (Mirov, 1967; Richardson & Rundel, 1998; Jin *et al.*, 2021). However, an understanding of the interplay between genetic and ecological factors throughout the phylogeny remains incomplete, impeding our ability to discern the unique processes at work within each clade of this expansive genus.

The genus *Pinus* is divided into two subgenera, *Pinus* (the diploxylon pines) and *Strobus* (the haploxylon pines) (Little & Critchfield, 1969; Gernandt *et al.*, 2005). Within the subgenus *Pinus*, most Mediterranean species are classified into subsection *Pinaster*, while other Eurasian species are placed into subsection *Pinus* (Gernandt *et al.*, 2005). The main species of subsection *Pinus* include *P. sylvestris* L., *P. nigra* Arn., *P. densiflora* Sieb. et Zucc., *P. tabuliformis* Carr., *P. henryi* Mast., *P. hwangshanensis* W. Y. Hsia, *P. massoniana* Lamb., *P. yunnanensis* Franch., *P. kesiya* Royle ex. Gordon, and *P. merkusii* Jungh. et de Vriese. With the exception of *P. merkusii*, these species underwent diversification within the past 20 Ma (Jin *et al.*, 2021), adapting to a broad range of environments spanning over 60 degrees of latitude. This group encompasses the most northern (*P. sylvestris*) and southern species (*P. kesiya*, *P. merkusii*) within the genus, forming a series of geographical successions with continuous alternative distributions along latitudes, encompassing climatic zones from tropical to cold temperate (Fig. 1). They grow on sites ranging from low hills to high alpine mountains (Mirov, 1967; Richardson & Rundel, 1998). Among these species, the evolutionary history of the *P. merkusii* has been puzzling. The species has a restricted distribution in the tropical Southeast Asia. Molecular phylogenies based on different datasets inconsistently placed the species in subsections *Pinus* or *Pinaster* (Wang *et al.*, 1999; Gernandt *et al.*, 2005; Parks *et al.*, 2012; Wang & Wang, 2014; Jin *et al.*, 2021), whereas morphology-based classification placed it within subsection *Pinaster* and suggested a Mediterranean origin (Frankis, 1993). Therefore, the subsection *Pinus* represents an interesting group for understanding radiation in conifers. Their recent diversification and broad ecological amplitude raise several intriguing questions regarding the mechanisms of their speciation and adaptive evolution.

Firstly, crossing experiments have shown low genetic incompatibility among these species (Nakai, 1986; Kormutak

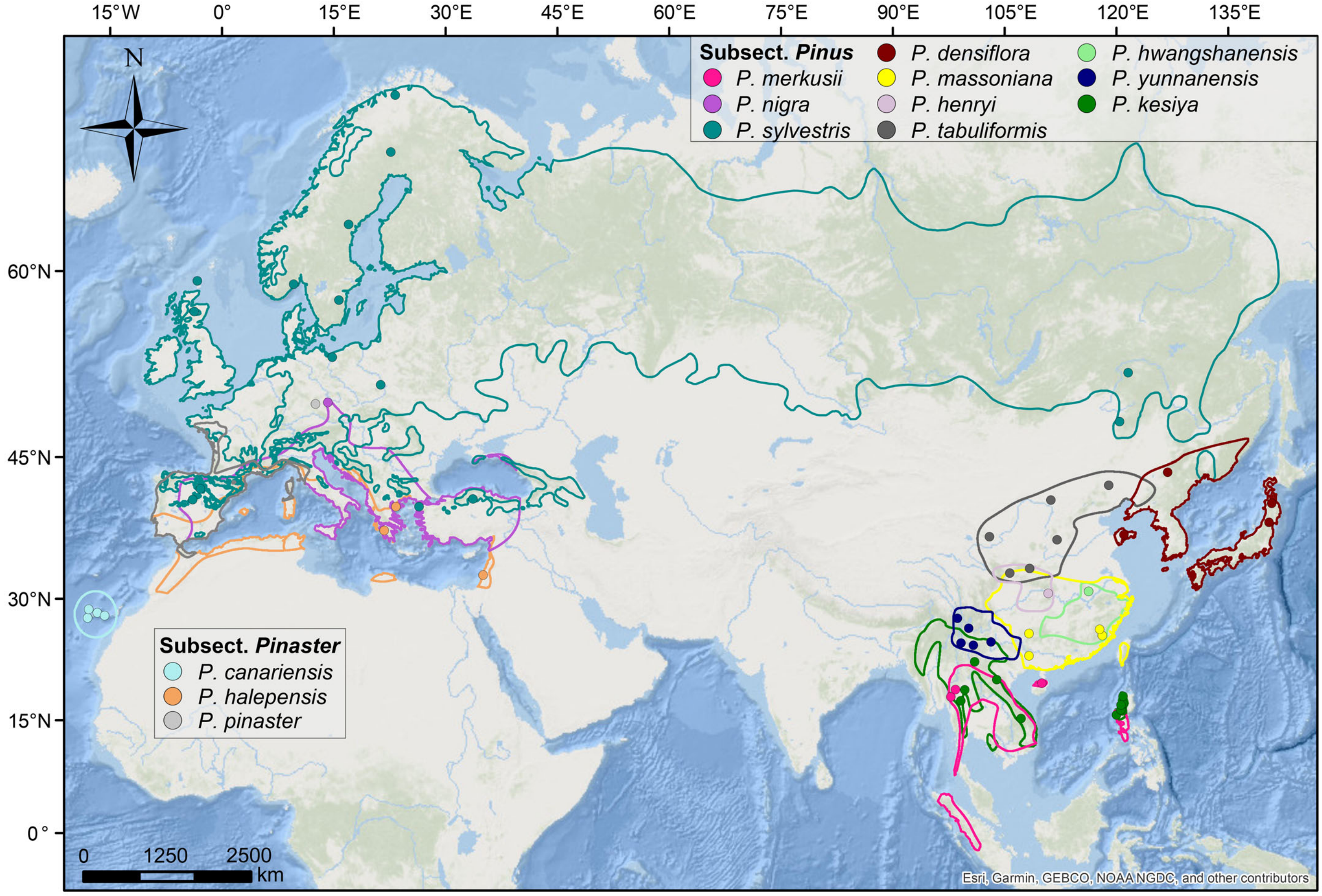

Fig. 1 Geographic distribution of the 13 species with sampling sites. Different color represents different species. The distribution map of *Pinus sylvestris* is from Caudullo *et al*. (2017), reproduced under Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). The distributions of other species are drawn based on presence data in Critchfield & Little (1966), Mirov (1967), Mao & Wang (2011), Liu *et al*. (2012) and Gao *et al*. (2024).

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al., 1992; Zhao *et al.*, 2014), suggesting incomplete reproductive isolation. Consequently, instances of gene exchange between parapatric and sympatric species have been documented (Szmidt & Wang, 1993; Wang & Szmidt, 1994; Wang *et al.*, 2011; Ren *et al.*, 2012; Zhou *et al.*, 2017; Gao *et al.*, 2024). This poses the possibility that admixture may have played a role in the evolution of this subsection. However, the extent to which ancient hybridization events were involved in the diversification remains uncertain. Secondly, due to their geographical proximity, ongoing gene flow could potentially homogenize genetic differentiation among these species. This, in turn, prompts inquiries into the mechanisms that uphold species identity within this group. Lastly, considering their distribution across Eurasia, it is reasonable to assume that these species experienced varying degrees of impact from Quaternary glaciations, with northern boreal and temperate species likely facing more substantial effects than tropical species. However, our comprehension of the influences of the last ice age on genetic diversity and population dynamics among these species remains limited due to the scarcity of comparative demography studies.

To address these questions, we conducted phylogenomic analyses using exome capture sequencing to investigate the prevalence of admixture in the evolution of subsection *Pinus*. Additionally, we examined the extent of niche differentiation among these species. Finally, we reconstructed the recent demographic histories of each species to gain insights into their population dynamics. This allowed us to explore their consequential impacts on selection efficacy and genetic diversity. Our study provides new perspectives on the role of introgression in the evolution of Eurasian pines, shedding light on the genetic and ecological factors that likely played a role in their rapid diversification and their resilience in adapting to climate and demographic events. This knowledge is particularly important as we attempt to manage and prepare forests for climate change.

Materials and Methods

Sampling and exome capture sequencing

We sampled 51 sites of 10 major species in subgenus *Pinus* subsection *Pinus*, and eight sites of three species in subsection *Pinaster*, including *P. pinaster* Ait., *P. halepensis* mill. and *P. canariensis* Chr. Sm. ex DC. (Fig. 1). Specifically, 22 sites were sampled for six species in Europe, and 37 sites for eight species in Asia. The three species of subsection *Pinaster* were used mainly as an outgroup to the subsection *Pinus*. Samples for subsection *Pinus* were selected to cover the distribution of each species, but for three species (*P. hwangshanensis*, *P. henryi* and *P. nigra*) with confined distributions, only one population was included for each (Fig. 1). The name, location and sample size of each population is provided in Supporting Information Table S1. Seeds were sampled from either individual trees (i.e. one seed per tree) or stand bulk collections, with 2–17 seeds per population (Table S1). Haploid genomic DNA was extracted from the megagametophyte tissue of each seed using a Plant Genomic DNA kit (Tiangen, Beijing, China).

We performed probe capture sequencing using a set of 40 000 exome probes. The probes, each 120 nt long, were designed based on the *Pinus taeda* UniGenes (Neves *et al.*, 2013). The majority of the probes aligned with *c.* 29 000 genes, while 9800 probes aligned with intergenic regions. The analytical procedure consisted of three major steps: (1) genomic DNA fragmentation; (2) hybridization to the capture probes; and (3) paired-end sequencing (2 × 101 bp) of the captured DNA fragments on HiSeq 2000 (Illumina) (Neves *et al.*, 2013). The library preparation, probe hybridization, and sequencing were conducted by RAPiD Genomics (Gainesville, FL, USA). In total, 467 samples were genotyped.

Bioinformatics

The sequence read quality was checked using FastQC v.0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences and low-quality bases (Phred quality < 20) were removed using Trimmomatic v.0.39 (Bolger *et al.*, 2014). Reads shorter than 36 bases after trimming were discarded. The resulting clean reads were then mapped to the *P. taeda* L. v.1.01 genome (Neale *et al.*, 2014) using the Bwa-Mem v.0.7.17 with default parameters (Li, 2013), generating BAM files.

To improve computational efficiency in the genome analysis toolkit (GATK) pipeline (Van der Auwera *et al.*, 2013), we adopted a strategy of preparing reduced BAM files and corresponding reduced reference for each sample, as previously described (Zhao *et al.*, 2020; Guo *et al.*, 2023b). In summary, variants calling was performed using the BCFtools with default parameters (Li, 2011) on all 467 samples. Scaffolds that contained at least one SNP in $> 50\%$ of the samples were selected for inclusion in the reduced reference. Subsequently, the reduced BAM files for each sample were generated using the 'samtools view' command based on the new reduced reference. The PCR duplicates were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard/). Additionally, reads around putative insertions and deletions were locally realigned using RealignerTargetCreator and IndelRealigner in the Gatk v.3.8-1. Finally, variant calling was performed individually using HaplotypeCaller, and GenotypeGVCFs was utilized for multisample joint aggregation and genotype likelihood correction with the parameter '-includeNonVariantSites'.

To minimize SNP calling errors, we implemented several filtering steps. Genotypes with genotype quality < 20, a read depth < 3, or the depth of the most covered allele $< 90\%$ were masked as missing. Additionally, SNPs that met any of the following criteria were removed: missing rate $> 20\%$, singleton, mapping quality < 40, allele number > 5 or located in repeat regions. Individuals with missing rate > 40 were also removed. We further identified and removed samples with a relatedness greater than or equal to first cousin following the procedure outlined by Hall *et al.* (2020). The remaining SNPs and samples were used in subsequent principal component analysis, fineRADstructure, TreeMix and Dsuite analyses (see the following sections for details). For site frequency spectrum (SFS)-based, nucleotide diversity, and phylogenomic analyses, all sites, including invariant sites, were included.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Genetic relationships and diversity

We used principal component analysis (PCA) implemented in EIGENSOFT v.6.1.4 (Price *et al*., 2006) to examine the distribution of genetic variance in the samples. Furthermore, we quantified the ancestry sources in each population using a Markov chain Monte Carlo method implemented in fineRADstructure (Malinsky *et al*., 2018b). This method captures patterns of genomic similarity and infers population structure based on shared haplotype relationships. Because fineRADstructure is sensitive to missing data, we selected 6 samples per species with the least missing and representing different sites (i.e. if multiple sites were collected for that species), and retained only the SNPs without missing data across all individuals. We used the 'RADpainter' module to generate the coancestry matrix, which summarizes nearest neighbor haplotype relationships across the dataset. This coancestry matrix served as input for the 'finestructure' module, which was run using the default parameters. The results were visualized using the fineRADstructurePlot.R and FinestructureLibrary.R scripts available at https://github.com/millanek/fineRADstructure.

To estimate genetic diversity in each species, we first identified zerofold and fourfold degenerate coding sites using the *P. taeda* v.1.01 genome annotation and the 'get_degeneracy.py' script (https://github.com/zhangrengang/degeneracy). We then calculated pairwise nucleotide diversity at all sites (π) as well as at zerofold (π_0) and fourfold (π_4) degenerate coding sites using ANGSD v.0.925 (Korneliussen *et al*., 2014). Initially, we utilized the -doSaf approach in ANGSD to compute the site allele frequency likelihood, using the GATK genotype likelihood model, for all sites. Subsequently, we applied the -realSFS method to obtain a maximum-likelihood estimate of the folded SFS. We then used the -doThetas function to estimate thetas from posterior probability of allele frequency (global folded SFS), based on a maximum-likelihood approach. To calculate the average pairwise theta across the entire genome, we summed the estimated per-site thetas and divided the sum by the total number of sites.

Estimating the fitness effects of amino acid-changing mutations

To assess the levels of purifying and positive selection in each species, we estimated the distribution of fitness effects (DFE) of nonsynonymous mutations using the maximum-likelihood procedure implemented in DFE-α (Keightley & Eyre-Walker, 2007; Eyre-Walker & Keightley, 2009). This approach assumes that synonymous sites evolve neutrally while mutations at nonsynonymous sites may have a fitness effect. To estimate the DFE, we constructed the folded one-dimensional SFS for each species by applying the down projection method using the Python script 'easySFS' (https://github.com/isaacovercast/easySFS). We categorized sites into two classes: putatively neutral reference sites (i.e. fourfold synonymous degenerate sites) and selected sites (i.e. zerofold nonsynonymous degenerate sites). We modeled the effects of recent demographic change on neutral SFS by assuming a one-step population size change, and inferred the fitness of new mutations at the selected sites from a gamma distribution while fitting the estimated parameters for the demographic model. The strength of purifying selection is defined as the product of the effective population size N_e and the selection coefficient s ($-N_e s$).

Conditional to the estimated DFE, we determined the proportion of adaptive amino acid substitutions fixed by positive selection (α) and the relative rate (ω) of adaptive substitution at zerofold nonsynonymous sites. We used the *P. taeda* reference as outgroup to calculate nucleotide divergence between species at fourfold and zerofold sites. The Jukes–Cantor multiple hits correction was applied to the divergence estimates (Jukes & Cantor, 1969).

We performed 999 bootstrap resampling of SNPs in each site class to generate the 95% confidence intervals (CI) of $-N_e s$, α and ω under one-step population change model. We further validated the estimated DFEs using a two-step demographic model, but due to computational constraints, we limited the bootstrapping to 99 iterations.

Phylogenetic inferences and molecular dating

To infer the species tree, we employed both concatenation and multispecies coalescent methods. Initially, we converted the VCF file, containing informative and invariant sites, into multiple-FASTA alignments using the vcf-to-tab module in VCFTOOLS package and vcf_tab_to_fasta_alignment.pl script (https://github.com/JinfengChen/vcf-tab-to-fasta/blob/master/vcf_tab_to_fasta_alignment.pl). Loci that were > 1000 bp apart were considered separate fragments. In order to root the phylogenetic tree, we included the sequence of the reference genome (*P. taeda*) in the alignments. We removed sites with over 20% Ns and alignments shorter than 100 bp. We then concatenated the remaining 14 605 multiple sequence alignments for the construction of maximum-likelihood (ML) phylogenies using RAxML v.8.2.11 (Stamatakis, 2014), with parameters '-m GTRGAMMA -f a -x $RANDOM -N autoMRE -p $RANDOM'. For the coalescent species tree, individual gene trees were constructed from the multiple sequence alignments using IQ-TREE v.1.6.12 with default parameters and 1000 ultrafast bootstrap replicates (Nguyen *et al*., 2015; Hoang *et al*., 2018). ASTRAL-III (Zhang *et al*., 2018) was then used to reconstruct species trees from the bootstrapped trees, employing a 'coalescence' framework with default parameters and 100 multi-locus bootstrap replicates. To reduce the computational load, we included only 6 individuals per species, consistent with the individuals used in the fineRADstructure analysis, for the aforementioned phylogenetic analyses.

Divergence times among species were estimated using a tip-dating strategy under the fossilized birth–death (FBD) process (Stadler *et al*., 2018) with the uncorrelated lognormal relaxed model in BEAST v2.6.3 (Bouckaert *et al*., 2019). This process integrates the distribution of fossil taxa through time as a means of calibrating the rate of the molecular clock. We concatenated fragments ≥ 1000 bp, resulting in a total length of 416 870 bp. Only one individual per species was included in the FBD analysis to reduce computation time. To calibrate the divergence time, we used 6 fossil occurrence data in the tip-dating analyses: (1)

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

P. baileyi: 45 Ma (Erwin & Schorn, 2006), constrained as monophyletic to all the 13 species; (2) *P. canariensis*: 16.3–12.8 Ma (Klaus, 1989), constrained as monophyletic to the extant *P. canariensis*; (3) *P. halepensis*: 16.3–12.8 Ma (Klaus, 1989), constrained as monophyletic to the extant *P. halepensis*; (4) *P. prekesiya*: 11.6–5.3 Ma (Xing *et al.*, 2010), constrained as monophyletic to *P. yunnanensis* + *P. kesiya* clade; (5) *P. premassoniana*: 11.6–5.3 Ma (Ding *et al.*, 2013), constrained as monophyletic to *P. massoniana* + *P. hwangshanensis* + *P. henryi* + *P. tabuliformis* + *P. yunnanensis* + *P. prekesiya* + *P. kesiya* clade; and (6) *P. densiflora*: 2.1 Ma (Yamada *et al.*, 2014), constrained as monophyletic to *P. densiflora*. For the FBD analysis, we followed the FBD-tutorial (https://taming-the-beast.org/tutorials/FBD-tutorial/) with the exception of using bModelTest v.1.2.1 (Bouckaert & Drummond, 2017) to measure the substitution model. We executed three independent runs with different random seeds and assessed convergence by ensuring an effective sample size (ESS) of > 200 in Tracer v.1.7.1 (Rambaut *et al.*, 2018). We combined parameter log files and tree files from different runs using logcombiner and summarized the posterior sample of trees to produce a maximum clade credibility tree with median node heights using TreeAnnotator (Bouckaert *et al.*, 2019). Finally, we visualized the results using the geoscale_plot.R script (https://github.com/alexeid/fossilDating/blob/master/geoscale_plot.R).

Estimation of niche divergence

To characterize the ecological preference of each species, we collected the distribution of presence-only data of the *Pinus* species from literature (Mao & Wang, 2011; Jin *et al.*, 2021; Bruxaux *et al.*, 2024; Gao *et al.*, 2024). We carefully evaluated the dataset and removed any erroneous records, duplicates, and records within 1 km of each other. The final dataset consisted a total of 1796 distribution records that evenly distributed over the range of the 13 species (Table S2).

We extracted 45 environmental variables, including 19 bioclimatic variables and one topographical layer (elevation) from WorldClim (https://www.worldclim.org/data/worldclim21.html), 11 soil variables within 30 cm of soil horizon (https://files.isric.org/soilgrids/latest/data/), 2 landcover classes (https://www.earthenv.org/landcover), 6 global UV radiation variables (http://www.ufz.de/gluv/), annual potential evapotranspiration and aridity index (https://cgiarcsi.community), growing degree days (http://nelson.wisc.edu/sage/data-and-models/atlas/maps.php), as well as ground-frost frequency, vapor pressure and wet-day frequency (http://www.ipcc-data.org/observ/clim/cru_climatologies.html) (Table S3). All environmental layers were standardized to a resolution of 30 arc-s (*c.* 1 km^2). Using the extracted environmental variables corresponding to the 1796 occurrence sites, we assessed the pairwise correlations among these variables using Spearman's rank correlation coefficient. Among the 45 variables, we retained 15 with correlation coefficients $(\rho) \leq |0.75|$ (Table S3), including annual mean air temperature (bio1), mean diurnal range (bio2), isothermality (bio3), air temperature seasonality (bio4), mean temperature of the wettest quarter (bio8), annual precipitation (bio12), precipitation of the driest month (bio14), precipitation seasonality (bio15), precipitation of the coldest quarter (bio19), evergreen/deciduous needleleaf trees (consensus1), mixed/other trees (consensus4), elevation values (elev), UV-B seasonality (uvb2), soil organic carbon content in the fine earth fraction (soc), and soil pH (phh2o). We then conducted PCA using these 15 retained variables to evaluate niche divergence among the pine species.

History of admixture

To evaluate reticulate forms of evolution that cannot be explained by a classical species tree, we estimated the history of population splits and admixtures using TreeMix (Pickrell & Pritchard, 2012). Maximum-likelihood trees were constructed with blocks of 50 SNPs to account for linkage disequilibrium. We tested the addition of 0 to 10 migration events by generating 100 replicate trees for each scenario. Potential migration events were inferred based on the stabilization of the proportion of explained covariance among groups toward their maximum asymptotic values. All the analyses were performed and summarized using the BITE pipeline (Milanesi *et al.*, 2017).

We also used the Dsuite v.0.5r52 (Malinsky *et al.*, 2021) to perform *D* statistic (ABBA–BABA tests) for detecting introgression. This method assigns evidence of gene flow to specific, potentially internal, branches on a phylogeny while taking into account incomplete lineage sorting. The test involves four populations or taxa in the form (((P1, P2), P3), outgroup) and assesses potential gene flow between P3 and P1 or P2 based on the relative site patterns of ABBA and BABA. Initially, the f_4-ratio was calculated on the basis of the ML phylogenetic tree for all species. This calculation was performed using the 'Dtrios' program, with all samples in subsection *Pinaster* serving as the outgroup. The ML tree was derived from the TreeMix analysis assuming zero migration events. A total of 120 trios of (((P1, P2), P3), outgroup) were tested (Table S4). The significance of each test was assessed using 100 jackknife resampling runs. Subsequently, the *f*-branch statistic values for each phylogenetic branch were estimated using the 'Fbranch' program with the parameter '-p 0.01'. Finally, the obtained *f*-branch statistics were visualized using the 'dtools.py' script.

Demographic history

We used Stairway Plot v.2 (Liu & Fu, 2020) to estimate the demographic history of each species. This approach employs a flexible multi-epoch coalescent approach to infer population size changes over time without assuming any specific demographic model. We generated the folded one-dimensional SFS for each species through the projecting down method using the Python script 'easySFS' (https://github.com/isaacovercast/easySFS). For this analysis, we included only species with > 12 individuals, which excluded *P. pinaster*, *P. nigra*, *P. hwangshanensis* and *P. henryi*. We ran Stairway Plot for each of the remaining 9 species using default parameters and tested four different numbers of random breakpoints using 200 replicates. The mutation rate was set to 7×10^{-10} per site per year (Willyard *et al.*, 2007), and the

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

generation time was assumed to be 50 yr. Two hundred subsamples of 67% of all sites were generated to estimate the median and 95% CI of the effective population size (N_e) over time.

Results

Genetic diversity in the Eurasian pines

Exome capture sequencing of the 467 samples from 13 pine species resulted in 2.88 billion paired-end reads, with an average of 6.17 million reads per sample. (Table S5). The alignment of these reads to the reference genome yielded an average of 27.73 Mbp of genomic sequence covered by at least three reads per individual (Table S5). After removing highly missing and highly related individuals and performing SNP filtering, 5589 715 sites were retained for the remaining 416 samples. Further filtering for nonpolymorphic sites resulted in 598 134 SNPs, which were then reduced to 419 783 SNPs after filtering out singletons.

Nucleotide diversity (π) was relatively consistent among the three species in subsection *Pinaster*, ranging from 0.0035 to 0.0041. However, this value varied with a threefold magnitude among species of subsection *Pinus*, ranging from 0.0023 to 0.0065 (Table 1). Species distributed at low latitudes, for example *P. merkusii*, *P. kesiya*, *P. yunnanensis* and *P. massoniana*, exhibited lower π (0.0023–0.0038) compared to the species distributed at higher latitudes, for example *P. tabuliformis*, *P. hwangshanensis*, *P. henryi*, *P. densiflora*, *P. sylvestris* and *P. nigra* (π: 0.0.0042–0.0065) (Table 1). Nucleotide diversity at zerofold (π_0) and fourfold (π_4) degenerate sites followed a similar pattern.

The haplotype-based coancestry analysis by fineRADstructure revealed distinct separation of the two subsections (Fig. 2a). The three species of subsection *Pinaster* are each discrete with much less shared ancestry between species than within each species, suggestive of their long evolution history in isolation. By contrast, the species in subsection *Pinus* are less differentiated. This is particularly visible among *P. tabuliformis*, *P. henryi* and *P. hwangshanensis*, which together formed one weakly differentiated block, suggesting a very recent divergence and/or high level of gene flow. Similarly, less differentiation was observed between *P. yunnanensis* and *P. kesiya*. The remaining four species, *P. densiflora*, *P. sylvestris*, *P. nigra*, and *P. massoniana* each had a relatively higher degree of identity. One individual of *P. densiflora* showed high affinity with *P. sylvestris*, indicating localized introgression. *Pinus merkusii* was distinct from all species, but exhibited a slightly higher degree of coancestry with the Mediterranean species than with the Asian species (Fig. 2a).

Table 1 Mean nucleotide diversity across all sites (π), zerofold degenerate (π_0), fourfold degenerate sites (π_4) and the ratio of mean π_0 to mean π_4 ($\pi_0 : \pi_4$) of the 13 *Pinus* species included in this study.

Species	π	π_0	π_4	$\pi_0 : \pi_4$
Subsection *Pinaster*				
Pinus canariensis	0.0041	0.0021	0.0050	0.4133
Pinus pinaster	0.0035	0.0017	0.0042	0.4020
Pinus halepensis	0.0037	0.0018	0.0045	0.3971
Subsection *Pinus*				
Pinus merkusii	0.0023	0.0014	0.0028	0.4835
Pinus nigra	0.0064	0.0029	0.0079	0.3712
Pinus sylvestris	0.0042	0.0017	0.0051	0.3362
Pinus densiflora	0.0052	0.0025	0.0067	0.3781
Pinus massoniana	0.0035	0.0017	0.0042	0.3933
Pinus hwangshanensis	0.0063	0.0036	0.0089	0.4076
Pinus henryi	0.0065	0.0034	0.0088	0.3869
Pinus tabuliformis	0.0053	0.0023	0.0066	0.3503
Pinus yunnanensis	0.0038	0.0019	0.0048	0.3915
Pinus kesiya	0.0028	0.0013	0.0034	0.3711

The PCA revealed that the 1st PC axis primarily highlighted the separation of the two subsections, with *P. merkusii* occupying an intermediate position (Figs 2b, S1). The 2nd, 3rd and 4th axes reflected the differentiation of species within subsection *Pinaster* and their relationship with *P. merkusii*. The separation of species within subsection *Pinus* was evident only along PC5 and PC6 (Figs 2c, S1c). The first six eigenvalues significantly (Tracy–Widom test, $P < 0.001$) explained 40.46% of the total genetic variance (12.32, 7.87, 7.03, 6.83, 4.42 and 1.99% for PC1 to 6, respectively) (Figs 2b,c, S1). This analysis corroborates the fineRADstructure result, suggesting more recent diversification among species in subsection *Pinus*, with the exception of *P. merkusii*, which is distinctly more isolated from the other members.

Phylogeny and molecular dating

Unpartitioned ML analyses of the concatenated matrix yielded a highly resolved topology (bootstrap support ≥ 95% for most nodes, Fig. S2a). Notably, all currently defined species were found to be monophyletic. Specifically, the 13 species were split into two distinct clades, consistent with their classifications into subsections *Pinaster* and *Pinus*. *Pinus merkusii* was the first to diverge in the subsection *Pinus* clade, being sister to the rest of the members (Fig. S2a). *Pinus hwangshanensis*, *P. tabuliformis* and *P. henryi* formed a strongly supported group while *P. yunnanensis* and *P. kesiya* formed a separate group; *P. massoniana* appeared as a sister to these two groups. The Eurasian *P. sylvestris* was closely related to the Asian *P. densiflora*, with the European *P. nigra* being sister to them. The coalescent-based phylogeny was mostly consistent with the ML tree; however, *P. tabuliformis*, *P. henryi*, and *P. hwangshanensis* did not form a monophyletic group (Fig. S2b). Additionally, *P. nigra* appeared as sister to the rest of the members of subsection *Pinus* excluding *P. merkusii* (Fig. S2b). These discrepancies may arise from introgression resulting in unstable placements (See results on introgression).

The phylogenetic tree inferred by the FBD analysis in a Bayesian framework recovered a topology consistent with the ML tree. The divergence time estimates, utilizing six fossil records, were generally in agreement with the study by Jin *et al.* (2021) (Fig. 2d). According to the estimates, the *Pinus* and *Pinaster* subsections diverged during the Eocene period, *c.* 45.4 Ma (95% CI: 57.1–34.2 Ma). The three species in subsection *Pinaster* diverged at 41.3–39.0 Ma, while the species in subsection *Pinus* diverged in the Miocene ranging from 22.1 to 9.4 Ma, except for *P. merkusii*, which had an origin in the Eocene at 36.4 Ma. Our results, along with those of Jin *et al.* (2021), are based on

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

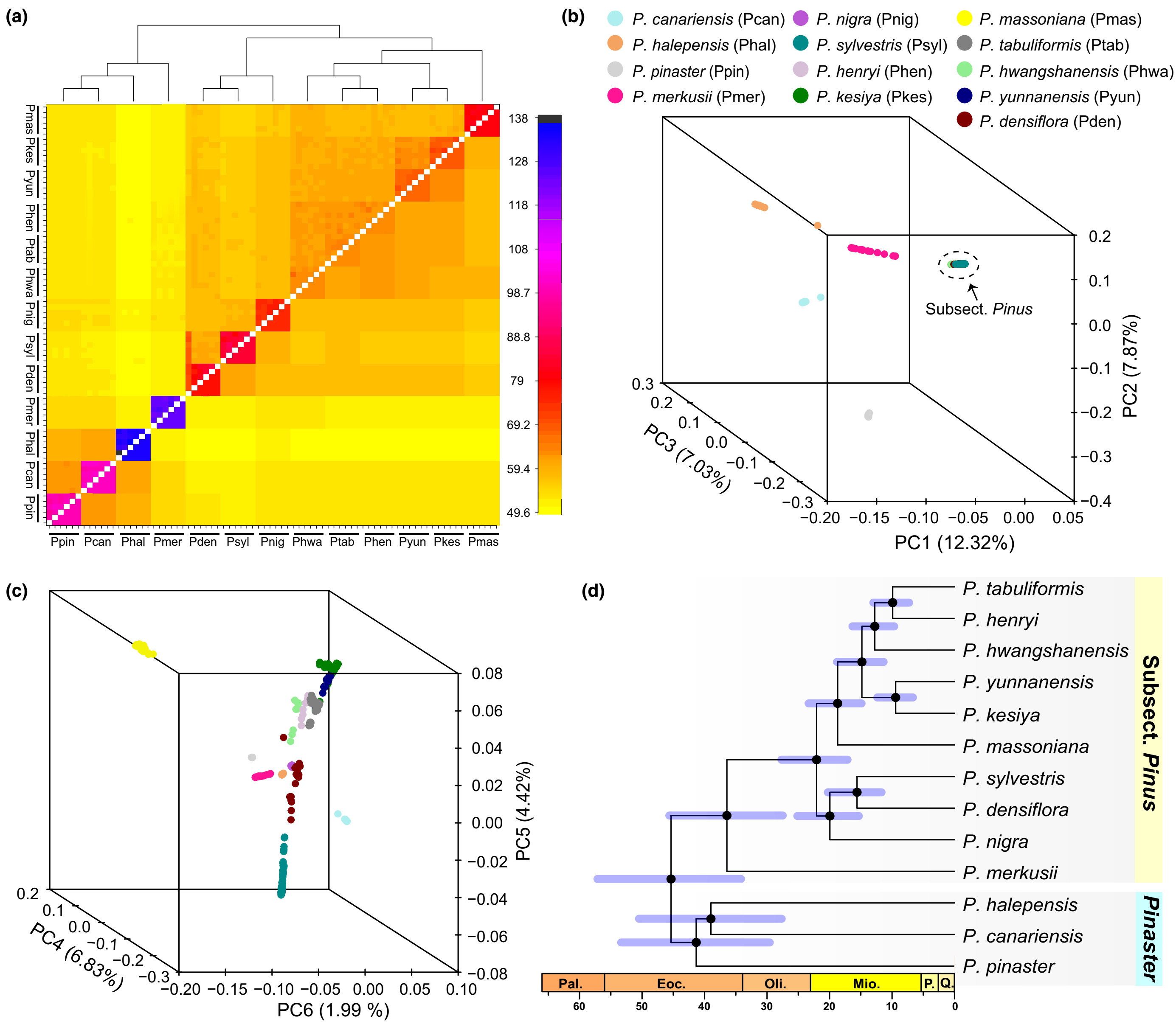

Fig. 2 Relationships and phylogeny of the sampled species. (a) fineRADstructure results showing coancestry coefficients between individuals (above diagonal) and average (below diagonal). The level of recent coancestry is color coded from low (yellow) to high (black). The species abbreviations are described in (b). (b, c) Principal component analysis of genetic variations in the 13 pine species, with different color denotes different species. (d) Phylogeny based on 0.4 Mbp alignments using the fossilized birth–death model in Beast 2. Horizontal bars at nodes correspond to the 95% highest posterior density of divergence time. Pal., Paleocene; Eoc., Eocene; Oli., Oligocene; Mio., Miocene; P., Pliocene; Q., Quaternary.

different datasets (exome capture sequencing and RNA-Seq, respectively) and dating methods. Despite these differences, they have consistently yielded similar results, providing robust evidence for the reliability of the divergence estimations.

DFE and adaptive evolution

To gain insight into the efficacy of selection, we examined the fitness effects of nonsynonymous mutations in nine species, each with a sample size of ≥12 individuals. The estimated distribution of fitness effects suggests that 25.7–35.5% of new amino acid mutations were weakly deleterious that behave as nearly neutral ($-N_es < 1$), 6.3–16.6% were moderately deleterious ($1 < -N_es < 10$), and 53.9–64.0% were strongly deleterious ($-N_es > 10$), which likely subjected to high levels of purifying selection across all species (Fig. 3a). The estimations of DFE under one-step and two-step population size changes were very similar, confirming the accuracy of the estimation (Figs 3a, S3). The proportion (α) and the rate (ω) of adaptive substitutions varied largely among species, with *P. tabuliformis* (α: 0.415; ω: 0.176) and *P. sylvestris* (α: 0.432; ω: 0.178) showed the highest

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

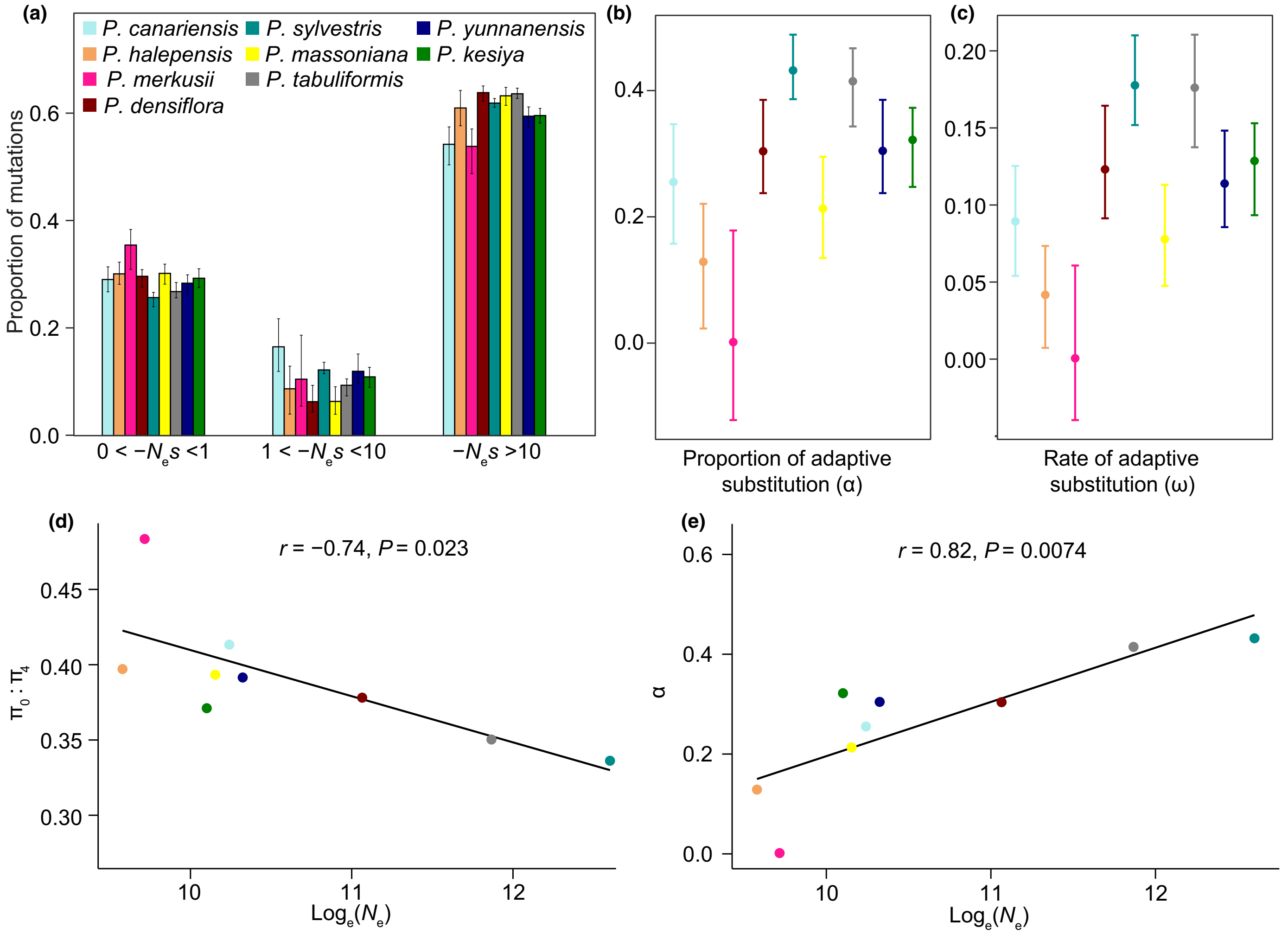

Fig. 3 Efficacy of purifying and positive selection. (a) Estimates of purifying selection at zerofold degenerate sites in nine *Pinus* species. Error bars represent 95% bootstrap confidence intervals. (b, c) the proportion of adaptive substitution (α) and the rate of adaptive nonsynonymous-to-synonymous substitutions (ω) in each species. (d, e) Correlation of the ratio $\pi_0 : \pi_4$ and α with effective population size (N_e). The color scheme in (b–e) aligns with that in (a).

values, and *P. merkusii* the lowest (α: 0.002; ω: 0.005) among all species (Fig. 3b,c).

The ratio of $\pi_0 : \pi_4$ reflects the efficacy of purifying selection, with lower values indicating stronger selection. We plotted the ratio of $\pi_0 : \pi_4$ and α against the estimated effective population size (N_e) of each species (Fig. 3d,e). The relationship demonstrates a tight correlation of both purifying and positive selection with N_e ($r = -0.74$ for π_0/π_4, $r = 0.82$ for α), where species with larger distributions (e.g. *P. tabuliformis* and *P. sylvestris*) experienced stronger selection compared to species with restricted distributions (e.g. *P. merkusii*). These findings are in line with theoretical expectations that selection operates more effectively in large populations.

Niche divergence among species

Principal component analysis on 15 environmental factors identified four components with eigenvalues > 1. These four components collectively explained 72.40% of the observed variation in the 1796 species occurrence records of all 13 species, with the individual components accounting for 26.95, 21.58, 13.81 and 10.06% of the total variation respectively (Fig. S4). The PCA distance biplot illustrated the relative contribution of each environmental variable to PC1 and PC2 (Fig. S4). We then examined the ecological space of each species within each subsection. For subsection *Pinus*, most species showed clear niche differentiation (Fig. 4a), for example *P. sylvestris*, a dominant Eurasian boreal forest species is clearly differentiated from the tropical *P. merkusii* and *P. kesiya* on annual mean temperature (bio1), temperature variability (bio3 and bio4), and annual precipitation (bio12). For *P. tabuliformis*, *P. hwangshanensis* and *P. henryi*, which are concentrated in central China, clear differentiation was detected for mean diurnal range (bio2), precipitation of driest month (bio14), precipitation of coldest quarter (bio19) and soil pH (phh2o). However, the three longitudinally distributed species, *P. nigra*, *P. densiflora* and *P. sylvestris*, had more

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

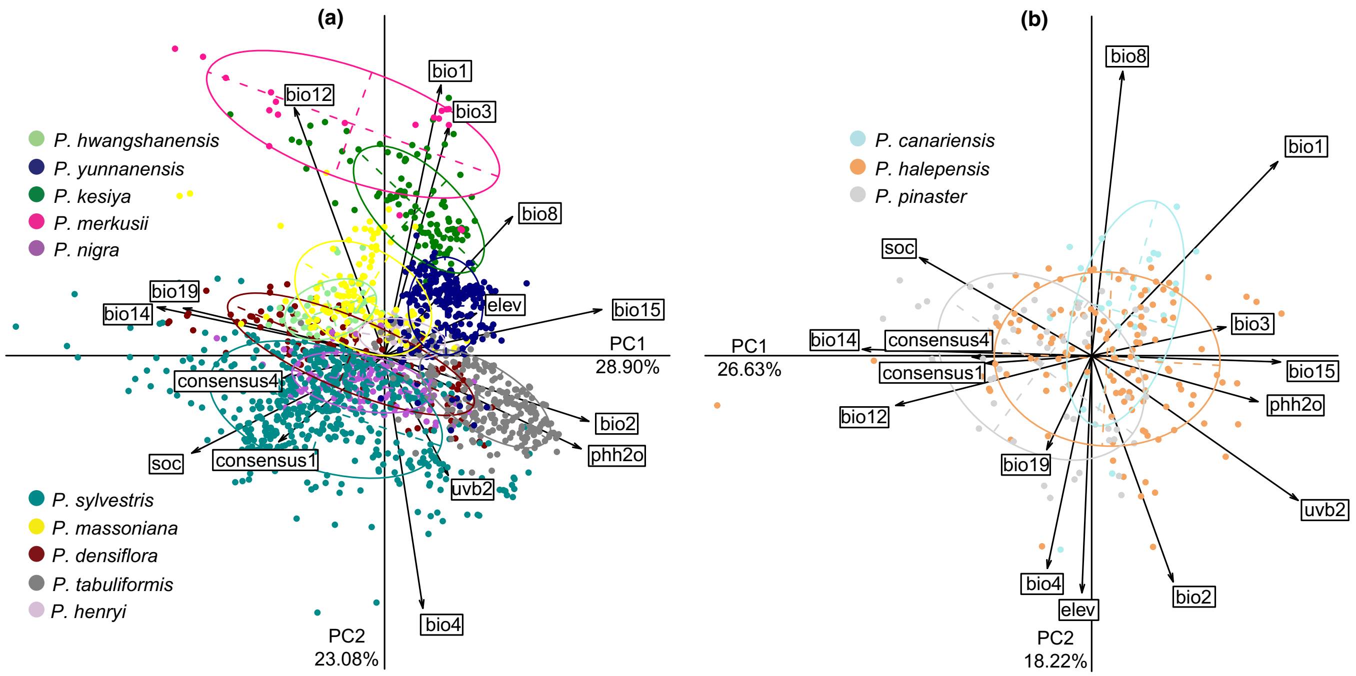

Fig. 4 Niche divergence in subsections *Pinus* and *Pinaster*. Principal component analysis biplot of the 15 selected environmental variables in subsection *Pinus* (a) and *Pinaster* (b).

overlapping niche space (Fig. 4a). The three species of subsection *Pinaster* showed a large degree of shared niche space (Fig. 4b). These findings illustrate higher niche divergence among the species of subsection *Pinus* (Figs 4a, S4).

Introgression among species

We applied two methods to infer introgression in the evolutionary history of subsection *Pinus*. Detection of introgression using TreeMix suggested that migration events (m) of 6 and 7 among species explained 99.95 and 99.98% of the variation in the data, respectively (Fig. S5a). These migration edges appeared on various points of the phylogeny (Figs 5a, S5b). Under 6 migration events, we detected a gene exchange between the Mediterranean *cannariensis*–*halepensis* branch and *P. merkusii* (Fig. S5b). This signal disappeared when more migration event was added. At $m = 7$, four strong signals of gene flow were shown between *P. densiflora* and *P. sylvestris* (40.45%; i.e. the estimated fraction of ancestry in the receiving population derived from the donor), between *P. henryi* and *P. tabuliformis* (41.22%), between *P. massoniana* and the ancestor of *hwangshanensis* – *henryi* (31.77%), and between *P. kesiya* and *P. massoniana* (20.77%) (Fig. 5a). The other three events were between *P. massoniana* and *P. hwangshanensis*, between *P. massoniana* and *P. densiflora*, and between the ancestors of *P. massoniana* and (*yunnanensis* – *kesiya* – *henryi* – *hwangshanensis*). The discrepancy in the position of *P. densiflora* under $m = 0$ and $m = 7$ suggests the probability of introgression involving *P. densiflora* with the clade consisting of *yunnanensis* – *kesiya* – *henryi* – *hwangshanensis* – *tabuliformis* (Figs 5a, S5c). All weights were significantly different than zero ($P << 0.001$).

ABBA–BABA tests identified 80 of the 120 trios with significant D values ($P < 0.01$, Table S4), indicating massive introgression within subsection *Pinus*. By summarizing the f_4 ratio values (an estimation of the ancestry proportions in an admixed population) through the calculation f-branch, we narrowed down the number of potential acceptor and donor lineages involved in a gene flow event (Fig. 5b). Strong signals of introgression were detected between the ancestor of the clade *hwangshanensis* – *henryi* – *tabuliformis* – *kesiya* – *yunnanensis* with each of *P. densiflora*, *P. sylvestris* and *P. nigra*. Other strong signals were between *P. henryi* and *P. hwangshanensis*, between *P. yunnanensis* and *P. tabuliformis*, and between *P. densiflora* and all other East Asian species except *P. merkusii* (Fig. 5b). Slightly weaker signals were scattered on both internal and terminal branches of the phylogeny. Overall, these results complement the TreeMix analysis and support pervasive admixture events occurred both historically and recently within subsection *Pinus* (Fig. 5b).

Demographic history

The demographic history of each species is presented as N_e changes over time (Figs 6, S6). Notably, there is no clear reduction in N_e in any species around the last glacial maximum (LGM) and during the last glacial period (LGP). On the contrary, the N_e of *P. sylvestris*, *P. tabuliformis*, *P. yunnanensis*, *P. halepensis* and *P. cannariensis* increased during the LGP. The oceanic *P. canariensis* and the Mediterranean *P. halepensis* exhibited distinct demographic histories. *Pinus canariensis* showed a gradual expansion starting *c.* 0.23 Ma, while *P. halepensis* declined sharply from 0.53 to 0.15 Ma followed by a recovery

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

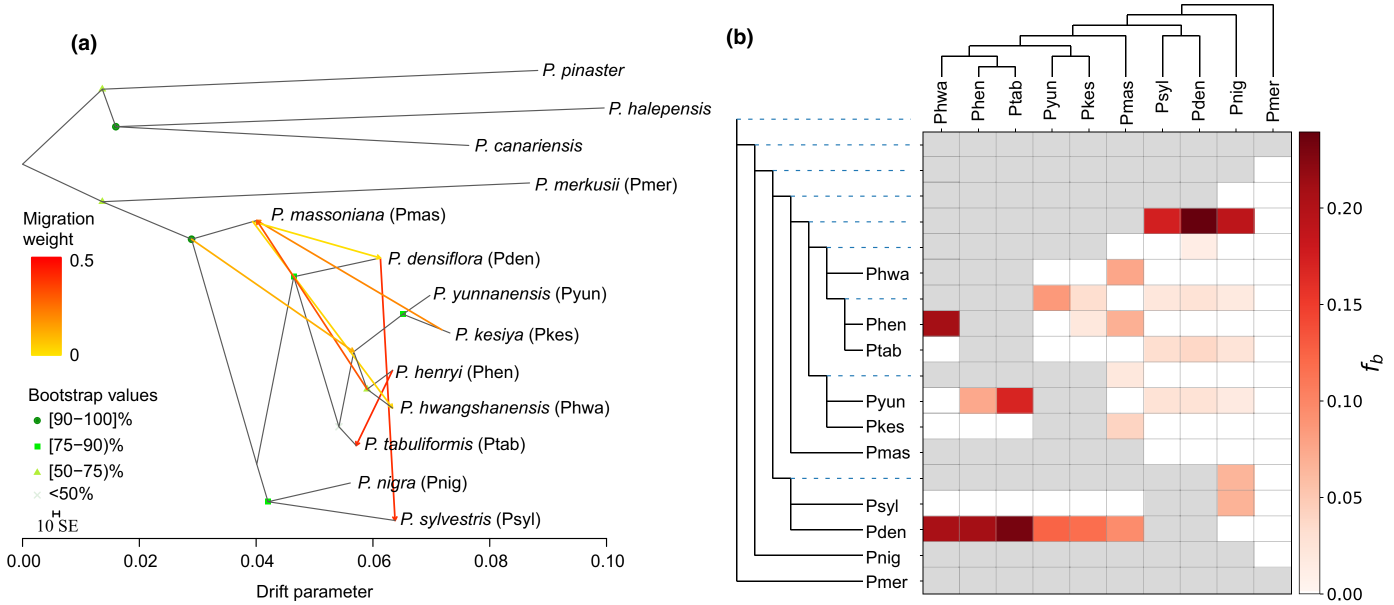

Fig. 5 Signatures of introgression in subsection *Pinus*. (a) Gene flow inferred by TreeMix at migration events ($m_{\text{best}} = 7$). Arrows indicate the direction of gene flow and are colored according to the percentage (weight) of alleles from that source. Scale bars represent a 10-fold average SE for the entries in the sample covariance matrix. (b) ABBA–BABA tests of introgression based on *D*-statistic with *D*suite. The *f*-branch (f_b) statistic identifies excess sharing of derived alleles between the branch of the tree on the *y*-axis and the species on the *x*-axis. Gray cells are empty where comparisons cannot be made. The abbreviations are described in (a).

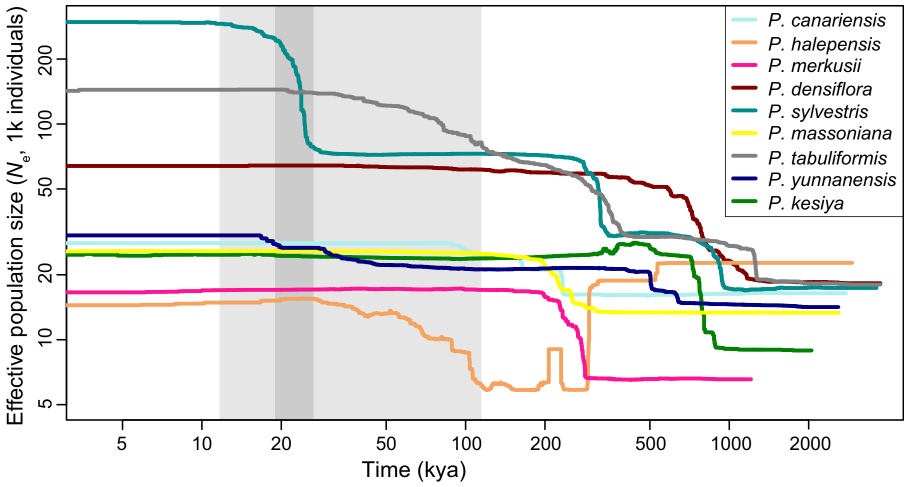

Fig. 6 Demographic histories of nine *Pinus* species inferred by Stairway Plot 2 including singletons. The last glacial maximum (LGM, *c.* 26.5–19 kya) and last glacial period (LGP, *c.* 115–11.7 kya) were colored in gray and light gray.

during the LGP. This strong bottleneck in *P. halepensis* is also discovered by Olsson *et al.* (2021).

All species in subsection *Pinus* experienced population expansions before LGP, although the timing, ranging from 1.30–0.28 Ma, and initial N_e varied. Generally, *P. massoniana*, *P. yunnanensis*, *P. kesiya* and *P. merkusii*, which are distributed at low latitudes, had smaller initial N_e compared to those in high latitudes, that is *P. tabuliformis*, *P. sylvestris* and *P. densiflora*. The initial N_e of *P. sylvestris*, *P. densiflora*, and *P. tabuliformis* were similar, and their first expansion started at 1.30–0.95 Ma, earlier than the four southern species. Both *P. sylvestris* and *P. tabuliformis* experienced multiple population increases, reaching current N_e values of 2.97×10^5 and 1.42×10^5, respectively. *Pinus merkusii* stands out due to its very small initial N_e of 0.65×10^4, but it underwent rapid increase at *c.* 0.28 Ma, reaching a N_e of 1.66×10^4 at *c.* 0.2 Ma before LGP and has remained stable ever since.

Removing singletons from the datasets revealed similar timing of the demographic events, but a decline of *P. sylvestris*, *P. tabuliformis* and *P. halepensis*, and expansion of *P. kesiya* during LGP (Fig. S7). The estimated N_e values were generally lower compared to those estimated from the SFS that included singletons.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Discussion

Pines are an important ecological lineage, and their deep history of diversification across wide ecological settings make them an excellent model for studying biogeography, introgression, and the genetics of adaptation. Despite the extensive research on the evolutionary history of *Pinus*, there is still a gap in our understanding of how introgression has systematically impacted their diversity and speciation. Species of subsection *Pinus* are major components of conifer forests across the Eurasian continent. Their recent diversification and broad ecological range raise several compelling questions about the drivers of this radiation. Here we discuss a few factors that may have played a role in this process.

Widespread introgression and reticulate evolution in subsection *Pinus*

Rapid diversification often implies insufficient time for establishing strong genetic incompatibility between species. Evidence for this can be seen in crossing experiments between *P. tabuliformis* and *P. yunnanensis*, which yielded a seed rate as high as 24% (Zhao *et al.*, 2014). Given the close geographic proximity of these pine species, hybridization is likely to have been a prominent factor in their evolution. In this study, we employed two approaches to infer introgression among species. The detected patterns of gene flow were not fully consistent across the two methods, rather, they complemented each other. This could be attributed to the complicated patterns of recurrent hybridization across various time frames and the sensitivity of each method to different intensities of gene flow at different time scales (Malinsky *et al.*, 2018a). A joint inference from the two methods supports a complex network in subsection *Pinus*, with evidence of admixture among almost all branches (Fig. 5). This suggests that hybridization was common during all radiation stages of this group, leaving a detectable signal within the genomes of many extant species.

Introgression among recently diversified sympatric species is expected, as observed in the *tabuliformis – henryi – hwangshanensis – massoniana* complex. The recognition of *P. hwangshanensis* and *P. henryi* as separate species is argued by their morphological and anatomical characteristics that differentiate them from *P. tabuliformis* and *P. massoniana* (Li & Xu, 1989; Liu *et al.*, 2014). Our results suggest that the speciation of *P. hwangshanensis* and *P. henryi* was strongly influenced by hybridization events with *P. tabuliformis* and *P. massoniana*. Additionally, we also recovered gene flow among allopatric species, for example between *P. densiflora* and other eastern Asian species, as well as hybridization on several internal ancestor branches. These results indicate more ancient gene flow, and are important for understanding ancient species distributions and their impact on shaping genetic diversity and the formation of new species.

Supporting evidence for widespread hybridization and introgression among the Asian pines comes from case-specific studies between *P. tabuliformis* and *P. yunnanensis* (Wang & Szmidt, 1994; Wang *et al.*, 2011; Gao *et al.*, 2012), *P. sylvestris* var. *mongolica* and *P. densiflora* (Szmidt & Wang, 1993; Ren *et al.*, 2012), *P. massoniana* and *P. hwangshanensis* (Zhou *et al.*, 2014, 2017), and *P. yunnanensis* and *P. kesiya* (Gao *et al.*, 2024). Hybridization between *P. tabuliformis* and *P. yunnanensis* is involved in the origin of another Asian pine, *P. densata*, in the late Miocene, a species distributed on the eastern Tibetan Plateau. This case has been studied in-depth, and represents a prominent example of speciation promoted by interspecific hybridization and niche divergence in conifers (Wang & Szmidt, 1994; Mao & Wang, 2011; Wang *et al.*, 2011; Gao *et al.*, 2012; Zhao *et al.*, 2014, 2020; Guo *et al.*, 2023b).

Overall, our findings support a scenario of divergence with gene flow in subsection *Pinus*, raising the intriguing possibility that radiation in this group was fueled by genetic variants generated through interspecific recombination. Genetic resources that fuel adaptive diversification include new mutations, ancestral variation, and adaptive introgression. Although recent radiations allow limited time to generate novel genetic variation, ancestral polymorphisms can facilitate the rapid assembly of advantageous allele combinations, and introgression becomes particularly important in cases where standing genetic variation and mutation within a species offer limited potential for adaptation (Barrett & Schluter, 2008; Suarez-Gonzalez *et al.*, 2018; Marques *et al.*, 2019). These genetic resources could enhance the potential of selection by increasing both phenotypic and genetic variation, thus accelerating evolution (Barrett & Schluter, 2008; Pease *et al.*, 2016; Marques *et al.*, 2019; Slovák *et al.*, 2023).

Ecological and geological factors

Weak genetic barriers and frequent introgressions between species could potentially lead to genetic assimilation over time, unless there are geographical and ecological barriers that facilitate pre- and postmating reproductive isolation (Levin *et al.*, 1996). Eight Asian species studied here, *P. sylvestris*, *P. densiflora*, *P tabuliformis*, *P. henryi*, *P. hwangshanensis*, *P. massoniana*, *P. yunnanensis*, and *P. kesiya*, form a continuous geographic distribution, with one replacing the other from north to south. Despite extensive historical and recent gene flow between these species, clear species identity is observed. For instance, all the samples from the same species form a monophyletic group with high support. Additionally, common garden experiments often reveal significant species-specific adaptation to climate conditions (Savolainen *et al.*, 2007; Zhao *et al.*, 2014; Sun *et al.*, 2019; Ramírez-Valiente *et al.*, 2022). Thus, geographic isolation and ecological adaptation are both likely important in maintaining the species boundaries.

This hypothesis is supported by the niche characterization of these species, which provided a quantitative evaluation of their ecological divergence. Additional support is the increased rates of niche evolution in several branches of subsection *Pinus* revealed on the whole genus phylogeny (figs S19, S20 in Jin *et al.*, 2021). In contrast to the Mediterranean pines that share much ecological space, the Asian species show clear differentiation along several climate gradients. Even for the *tabuliformis – hwangshanensis –*

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

henryi–*massoniana* complex, which shared a high degree of ancestry and are sympatric in central China, clear differentiation on temperature, precipitation, soil properties, altitudes and vegetation types are detected (Fig. 4a,b) (Fu *et al.*, 1999; Zhou *et al.*, 2014, 2017). The region where these species reside is marked by complex mountain systems (Qingling and Daba Mountains) that define the geography and climate at both macro- and micro-scales. Thus, regional geographic barriers and ecological separation could act as isolation forces operating during the incipient stages of their diversification.

The current distribution of *P. merkusii* partly overlaps with *P. kesiya* in tropical Southeast Asia. However, natural hybridization has not been documented between them, possibly due to their long divergence leading to greater genetic incompatibility than that among the other Asian species. This could also explain the low crossability of the currently sympatric Mediterranean *P. pinaster* and *P. halepensis* (Duffield, 1952; Kormutak *et al.*, 1992). As suggested by Klaus (1989), the Mediterranean pines represent an extremely heterogeneous assembly consisting mainly of relic pines from the Cretaceous–Tertiary period. Paleogeography of Europe in Eocene (50 Ma; Deep Time Maps™, https://deeptimemaps.com/map-lists-thumbnails/europe/) indicates that the Mediterranean region was very fragmented and separated from the Eurasian plate. Thus, the pine species in this region likely evolved allopatrically and developed genetic barriers along their long isolation history, and remained reproductively isolated in secondary contact as we see today. Overall, the contrasting patterns in niche diversification in the two subsections suggest that introgression and environment-specific sorting of ancestral/introgressed variation are important forces in the diversification of sympatric or parapatric sister species, whose speciation would be driven by divergent adaptation to different local environments. Under such speciation scenarios, gene flow among incipient species is expected to be common, and postmating selection determines the evolutionary outcome of the populations (Guo *et al.*, 2023b).

Diversity and purifying and positive selection

The diverse evolutionary histories and interactions with environments are expected to leave genetic footprints in each species. We found a threefold difference in nucleotide diversity among species of the subsection *Pinus* (Table 1). Outcrossing wind-pollinated conifers with large distributions generally exhibit higher genetic diversity than those with limited ranges (Petit & Hampe, 2006; Ellegren & Galtier, 2016). Our findings are in general agreement with this prediction, for example large species such as *P. tabuliformis* and *P. sylvestris* had greater π values than those with restricted distributions, such as *P. merkusii* and *P. kesiya*. However, several species with very confined ranges showed high diversity, such as *P. hwangshanensis*, and *P. henryi*. This could be due to the high degree of introgression with *P. tabuliformis* and *P. massoniana*. The sample sizes for *P. nigra*, *P. hwangshanensis* and *P. henryi* were limited, so our results need further validation in future studies.

We detected strong purifying and positive selection in the pine species. Consistent with the findings in many other plant species (Gossmann *et al.*, 2010; Wang *et al.*, 2016; Chen *et al.*, 2017), a majority of new amino acid-altering mutations were strongly deleterious, subjecting to strong purifying selection. We found a strong correlation between the efficacy of purifying selection and the effective population size N_e of each species, implying purifying selection operates more efficiently in widely distributed species, for example *P. tabuliformis* and *P. sylvestris*, than in small N_e species, for example *P. merkusii*, leading to accumulation of more weakly deleterious mutations in the latter species. This result aligns with the nearly neutral theory of molecular evolution (Ohta, 1992), which predicts that species with a small N_e is less effective in purging weakly deleterious mutations, resulting in their accumulation as segregating polymorphisms and fixation due to drift.

Interestingly, we further detected high proportions of adaptive substitution in all species except *P. merkusii*, along with a strong positive correlation between the rates of adaptive evolution in the species and their N_e. The estimates of the proportion of amino acid substitutions driven to fixation by positive selection were highest in *P. sylvestris* (43.2%) and *P. tabuliformis* (41.5%) and lowest in *P. merkusii* (0.2%). Similar high estimates have been reported in *Populus* (Wang *et al.*, 2016), *Capsella* (Slotte *et al.*, 2010) and *Helianthus* species (Strasburg *et al.*, 2011). By contrast, little evidence of widespread adaptive evolution is found in many other plants (Gossmann *et al.*, 2010). The positive correlation between effective population size and the rate of adaptive evolution is expected because larger populations are more probable to generate advantageous mutations, and selection becomes more effective on those mutations that are weakly selected. We speculate that the high rates of adaptive evolution detected in subsection *Pinus* are driven by adaptation to diverse ecological conditions, which is consistent with the trend that positive selection is more frequent in rapid adaptive evolution (Nevado *et al.*, 2019). However, establishing causal links between genetic variants and ecological traits is beyond the scope of the current investigation, partly due to the scattered short sequence fragments of the dataset and the poor annotation of the reference genome.

Demography of Eurasian *Pinus* species

We recovered distinct recent demographic histories of the sampled Eurasian pine species. *Pinus merkusii*, for instance, initially had a very small effective population size followed by a sharp increase at 0.28 Ma. The very small N_e, supports a theory of founder event during migration out of the Mediterranean region. The mtDNA- and cpDNA-based phylogenetic positions of *P. merkusii* are unstable relative to its placement in either the Asian subsections *Pinus* or Mediterranean subsection *Pinaster* (Wang *et al.*, 1999; Parks *et al.*, 2012; Wang & Wang, 2014). To this discordance, Wang & Wang (2014) suggests that *P. merkusii* captured a mtDNA variant from the ancestor of Asian pines during its eastward migration. Morphological, chemical, and population studies have revealed distinct differentiation of *P. merkusii* from other Asian hard pines (Cooling, 1968; Weißmann & Lange, 1987; Szmidt *et al.*, 1996). All these lines of evidence support an early

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

separation leading to heightened reproductive isolation of this species with other Asian members.

Other Asian pine species experienced growth starting from *c.* 1.30 Ma, with *P. sylvestris* and *P. tabuliformis* showing multiple stages of increase leading to high contemporary N_e. The increase in N_e detected by STAIRWAY PLOT could result from either population growth and/or meeting of genetically differentiated populations due to range shift during repeated glaciations. While the N_e varied among these species, there was no observed population contraction during the LGM. This holds true for both southern warm zone species and for northern cold zone species. Similar findings have been reported for seven European tree species (Milesi *et al.*, 2023; Bruxaux *et al.*, 2024). The emerging picture from the recent genomic-based demographic inferences for Eurasian pine species suggests resilience and survival in scattered microrefugia during the Quaternary glaciations, and rapid population rebound in inter-glacial times (Gao *et al.*, 2012; Xia *et al.*, 2018; Guo *et al.*, 2023a,b; Bruxaux *et al.*, 2024).

Intriguingly, contrary to the expectation of population growth after the LGP, most species either remained stable in N_e as during the LGP or even declined slightly (Figs 6, S6). Tree populations decrease in size for several reasons, including the level of standing genetic variation and the pace of environmental change. When N_e is linked to accumulation of deleterious mutations and genetic load, prolonged isolation can constrain evolutionary potential and increase risk of extinction due to genetic predispositions or ecological factors, as could be the case for *P. merkusii*. Although it is too short evolutionary time to observe any impact of decreased connectivity due to habitat fragmentation and land conversion on N_e and adaptive evolution, the effect is unlikely to be positive, especially when gene exchange has been an important source of genetic variation in past adaptation. Overall, our study of past adaptive radiation in pines highlights the constraints and opportunities for evolutionary change, and the expectations of future adaptation in response to environmental changes in different lineages.

Conclusion

Despite the relatively brief diversification history of the species in subsection *Pinus*, these species have remarkably adapted to a wide array of environments, stretching from equatorial regions to the Arctic. Our study uncovered evidence of hybridization at all stages throughout the radiation of this group, yet each species formed a monophyletic group, accompanied by significant niche differentiation and adaptive molecular evolution among them. The widespread gene exchange and ecological differentiation likely played a pivotal role in driving the radiation, with reproductive barriers gradually evolving in concert with the diversification history. Building upon these results, future studies should include systematic fine sampling spanning species boundaries to refine our understanding of the past and ongoing genetic and ecological dynamics underpinning species identity and adaptive evolution. With the availability of chromosome-level genome assemblies for pine species and more comprehensive genome annotations, functional inferences of introgressed loci and targets of selection, and their link to ecological traits, would offer more powerful insights into conifer tree evolution.

Acknowledgements

We thank the Biodiversity Management Bureau-Department of Environment and Natural Resources (BMB-DENR) for the permit (Gratuitous Permit no. 2016-253) for collecting pine samples in the Philippines. Genomic data processing and analyses were performed using resources provided by the Swedish National Infrastructure for Computing (SNIC), through the High Performance Computing Centre North (HPC2N) at Umeå University. This study was supported by grants from the Swedish Research Council (VR 2017-04686), Formas (2021-02155) and T4F program, Sweden.

Competing interests

None declared.

Author contributions

XRW designed the study. WZ, JG and DH performed genotyping. JG, ADD, YS, KWT and XRW provided samples. WZ and JG analyzed data with input from DH, BAA and JB. WZ and XRW wrote the manuscript draft. All authors contributed to the revision of the manuscript. WZ and JG contributed equally to this work.

Data accessibility

All sequencing data are archived in the NCBI SRA database under BioProject accession no. PRJNA1084782.

ORCID

Bea Angelica Andersson https://orcid.org/0000-0002-7657-6933
Jade Bruxaux https://orcid.org/0000-0001-9326-4299
Jie Gao https://orcid.org/0000-0002-2640-5116
David Hall https://orcid.org/0000-0001-8152-2713
Yoshihisa Suyama https://orcid.org/0000-0002-3136-5489
Kyle W. Tomlinson https://orcid.org/0000-0003-3039-6766
Xiao-Ru Wang https://orcid.org/0000-0002-6150-7046
Wei Zhao https://orcid.org/0000-0001-9437-3198

References

Anderson E, Stebbins GL. 1954. Hybridization as an evolutionary stimulus. *Evolution* **8**: 378–388.

Arnold ML. 1997. *Natural hybridization and evolution*. New York, NY, USA: Oxford University Press.

Barrett RD, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**: 38–44.

Bolger AM, Lohse M, Usadel B. 2014. TRIMMOMATIC: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N *et al.* 2019. BEAST

2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15: e1006650.

Bouckaert RR, Drummond AJ. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology* 17: 42.

Bruxaux J, Zhao W, Hall D, Curtu AL, Androsiuk P, Drouzas AD, Gailing O, Konrad H, Sullivan AR, Semerikov V *et al.* 2024. Scots pine – panmixia and the elusive signal of genetic adaptation. *New Phytologist.* doi: 10.1111/nph.19563.

Caudullo G, Welk E, San-Miguel-Ayanz J. 2017. Chorological maps for the main European woody species. *Data in Brief* 12: 662–666.

Chen J, Glémin S, Lascoux M. 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution* 34: 1417–1428.

Cooling ENG. 1968. *Pinus merkusii.* Oxford, UK: Commonwealth Forestry Institute, University of Oxford.

Critchfield WB, Little EL. 1966. *Geographic distribution of the pines of the world.* Washington, DC, USA: U.S. Department of Agriculture, Forest Service.

Ding ST, Wu JY, Chen JL, Yang Y, Yan DF, Sun BN. 2013. Needles and seed cones of *Pinus premassoniana* sp. nov., and associated pollen cone from the upper Miocene in East China. *Review of Palaeobotany and Palynology* 197: 78–89.

Duffield JW. 1952. Relationships and species hybridization in the genus *Pinus. Zeitschrift Fur Forstgenetik Und Forstpflanzenzuchtung* 1: 93–97.

Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nature Reviews Genetics* 17: 422–433.

Erwin DM, Schorn HE. 2006. *Pinus baileyi* (section *Pinus*, Pinaceae) from the Paleogene of Idaho, USA. *American Journal of Botany* 93: 197–205.

Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution* 26: 2097–2108.

Farjon A. 2001. *World checklist and bibliography of conifers.* Richmond, UK: Royal Botanical Gardens Kew.

Frankis MP. 1993. Morphology and affinities of *Pinus brutia.* In: Tashkin M, ed. *International Symposium Pinus brutia.* Ankara, Turkey: Ministry of Forestry, 11–18.

Fu L, Li N, Elias TS. 1999. Pinaceae. In: Wu Z, Raven PH, eds. *Flora of China.* Beijing, China: Science Press, 15–90.

Galis F. 2001. Key innovations and radiations. In: Wagner GP, ed. *The character concept in evolutionary biology.* San Diego, CA, USA: Academic Press.

Gao J, Tomlinson KW, Zhao W, Wang B, Lapuz RS, Liu JX, Pasion BO, Hai BT, Chanthayod S, Chen J *et al.* 2024. Phylogeography and introgression between *Pinus kesiya* and *Pinus yunnanensis* in Southeast Asia. *Journal of Systematics and Evolution* 62: 120–134.

Gao J, Wang B, Mao JF, Ingvarsson P, Zeng QY, Wang XR. 2012. Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. *Molecular Ecology* 21: 4811–4827.

Gavrilets S, Losos JB. 2009. Adaptive radiation: contrasting theory with data. *Science* 323: 732–737.

Gernandt DS, López GG, García SO, Liston A. 2005. Phylogeny and classification of *Pinus. Taxon* 54: 29–42.

Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution* 27: 1822–1832.

Guo JF, Wang B, Liu ZL, Mao JF, Wang XR, Zhao W. 2023a. Low genetic diversity and population connectivity fuel vulnerability to climate change for the Tertiary relict pine *Pinus bungeana. Journal of Systematics and Evolution* 61: 143–156.

Guo JF, Zhao W, Andersson B, Mao JF, Wang XR. 2023b. Genomic clines across the species boundary between a hybrid pine and its progenitor in the eastern Tibetan Plateau. *Plant Communications* 4: 100574.

Hall D, Zhao W, Wennström U, Andersson Gull B, Wang XR. 2020. Parentage and relatedness reconstruction in *Pinus sylvestris* using genotyping-by-sequencing. *Heredity* 124: 633–646.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.

Jin WT, Gernandt DS, Wehenkel C, Xia XM, Wei XX, Wang XQ. 2021. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *Proceedings of the National Academy of Sciences, USA* 118: e2022302118.

Jukes TH, Cantor CR. 1969. *Evolution of protein molecules.* New York, NY, USA: Academic Press.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177: 2251–2261.

Klaus W. 1989. Mediterranean pines and their history. *Plant Systematics and Evolution* 162: 133–163.

Kormutak A, Vookova B, Gajdošova A, Salaj J. 1992. Hybridological relationships between *Pinus nigra* Arn, *Pinus thunbergii* Parl and *Pinus tabulaeformis* Carriere. *Silvae Genetica* 41: 228–234.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15: 356.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin CJ, Wang C, Zamani N *et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518: 371–375.

Leslie AB, Beaulieu J, Holman G, Campbell CS, Mei W, Raubeson LR, Mathews S. 2018. An overview of extant conifer evolution from the perspective of the fossil record. *American Journal of Botany* 105: 1531–1544.

Levin DA, Francisco-Ortega J, Jansen RK. 1996. Hybridization and the extinction of rare plant species. *Conservation Biology* 10: 10–16.

Li CX, Xu HC. 1989. A numerical analysis on the dividing line of geographic distribution and taxonomy relation between *Pinus henryi* Mast. and *Pinus tabulaeformis* Carr. *Scientia Silvae Sinicae (China)* 25: 14–21.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997.*

Little EL, Critchfield WB. 1969. *Subdivisions of the genus Pinus (pines).* Washington, DC, USA: USDA Forest Service.

Liu X, Fu YX. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* 21: 280.

Liu ZH, Xie Q, Li ZQ. 2014. Comparative palynology and anatomy of *Pinus henryi, Pinus massoniana* and *Pinus tabulaeformis* (Pinaceae) and their taxonomic implications. *Pakistan Journal of Botany* 46: 1593–1600.

Liu ZL, Cheng C, Li J. 2012. High genetic differentiation in natural populations of *Pinus henryi* and *Pinus tabuliformis* as revealed by nuclear microsatellites. *Biochemical Systematics and Ecology* 42: 1–9.

Losos JB. 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism. *The American Naturalist* 175: 623–639.

Malinsky M, Matschiner M, Svardal H. 2021. Dsuite – Fast *D*-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* 21: 584–595.

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018a. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution* 2: 1940–1955.

Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018b. RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution* 35: 1284–1290.

Mao JF, Wang XR. 2011. Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan Plateau. *The American Naturalist* 177: 424–439.

Marques DA, Meier JI, Seehausen O. 2019. A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution* 34: 531–544.

Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* 8: 14363.

Milanesi M, Capomaccio S, Vajana E, Bomba L, Garcia JF, Ajmone-Marsan P, Colli L. 2017. bite: an R package for biodiversity analyses. *BioRxiv.* doi: 10.1101/181610.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Milesi P, Kastally C, Dauphin B, Cervantes S, Bagnoli F, Budde KB, Cavers S, Ojeda DI, Fady B, Faivre-Rampant P *et al.* 2023. Synchronous effective population size changes and genetic stability of forest trees through glacial cycles. *BioRxiv*. doi: 10.1101/2023.01.05.522822.

Mirov NT. 1967. *The genus Pinus*. New York, NY, USA: Ronald Press.

Nakai I. 1986. Crossability in interspecific pollinations among species of hard pine in the genus *Pinus*. *Journal of the Japanese Forest Society* **68**: 406–416.

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD *et al.* 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* **15**: R59.

Nevado B, Wong ELY, Osborne OG, Filatov DA. 2019. Adaptive evolution is common in rapid evolutionary radiations. *Current Biology* **29**: 3081–3086.

Neves LG, Davis JM, Barbazuk WB, Kirst M. 2013. Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* **75**: 146–156.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* **23**: 263–286.

Olsson S, Lorenzo Z, Zabal-Aguirre M, Piotti A, Vendramin GG, González-Martínez SC, Grivet D. 2021. Evolutionary history of the mediterranean *Pinus halepensis-brutia* species complex using gene-resequencing and transcriptomic approaches. *Plant Molecular Biology* **106**: 367–380.

Parks M, Cronn R, Liston A. 2012. Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evolutionary Biology* **12**: 100.

Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology* **14**: e1002379.

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

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**: e1002967.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**: 904–909.

Price R, Liston A, Strauss S. 1998. Phylogeny and systematics of *Pinus*. In: Richardson DM, ed. *Ecology and Biogeography of Pinus*. Cambridge, UK: Cambridge University Press, 49–68.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using TRACER 1.7. *Systematic Biology* **67**: 901–904.

Ramírez-Valiente JA, Santos del Blanco L, Alía R, Robledo-Arnuncio JJ, Climent J. 2022. Adaptation of Mediterranean forest species to climate: lessons from common garden experiments. *Journal of Ecology* **110**: 1022–1042.

Ren GP, Abbott RJ, Zhou YF, Zhang LR, Peng YL, Liu JQ. 2012. Genetic divergence, range expansion and possible homoploid hybrid speciation among pine species in Northeast China. *Heredity* **108**: 552–562.

Richardson DM, Rundel PW. 1998. Ecology and biogeography of *Pinus*: an introduction. In: Richardson DM, ed. *Ecology and biogeography of Pinus*. Cambridge, UK: Cambridge University Press, 3–46.

Rieseberg LH. 1997. Hybrid origins of plant species. *Annual Review of Ecology, Evolution, and Systematics* **28**: 359–389.

Saladin B, Leslie AB, Wüest RO, Litsios G, Conti E, Salamin N, Zimmermann NE. 2017. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology* **17**: 95.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38**: 595–619.

Schenk JJ. 2021. The next generation of adaptive radiation studies in plants. *International Journal of Plant Sciences* **182**: 245–262.

Schluter D. 2000. *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.

Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences* **273**: 1987–1998.

Simpson GG. 1953. *The major features of evolution*. New York, NY, USA: Columbia University Press.

Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Molecular Biology and Evolution* **27**: 1813–1821.

Slovák M, Melichárková A, Štubňová EG, Kučera J, Mandáková T, Smyčka J, Lavergne S, Passalacqua NG, Vďačný P, Paun O. 2023. Pervasive introgression during rapid diversification of the European mountain genus *Soldanella* (L.) (Primulaceae). *Systematic Biology* **72**: 491–504.

Stadler T, Gavryushkina A, Warnock RCM, Drummond AJ, Heath TA. 2018. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of Theoretical Biology* **447**: 41–55.

Stamatakis A. 2014. RAxML v.8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1213–1312.

Strasburg JL, Kane NC, Raduski AR, Bonin A, Michelmore R, Rieseberg LH. 2011. Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Molecular Biology and Evolution* **28**: 1569–1580.

Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018. Adaptive introgression: a plant perspective. *Biology Letters* **14**: 20170688.

Sun YQ, Zhao W, Xu CQ, Xu Y, El-Kassaby YA, De La Torre AR, Mao JF. 2019. Genetic variation related to high elevation adaptation revealed by common garden experiments in *Pinus yunnanensis*. *Frontiers in Genetics* **10**: 1405.

Szmidt AE, Wang XR. 1993. Molecular systematics and genetic differentiation of *Pinus sylvestris* (L.) and *P. densiflora* (Sieb. et Zucc.). *Theoretical and Applied Genetics* **86**: 159–165.

Szmidt AE, Wang XR, Changtragoon S. 1996. Contrasting patterns of genetic diversity in two tropical pines: *Pinus kesiya* (Royle ex Gordon) and *P. merkusii* (Jungh et De Vriese). *Theoretical and Applied Genetics* **92**: 436–441.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J *et al.* 2013. From FASTQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43**: 11.10.11–11.10.33.

Wang B, Mao JF, Gao J, Zhao W, Wang XR. 2011. Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Molecular Ecology* **20**: 3796–3811.

Wang B, Wang XR. 2014. Mitochondrial DNA capture and divergence in *Pinus* provide new insights into the evolution of the genus. *Molecular Phylogenetics and Evolution* **80**: 20–30.

Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related *Populus* species. *Genetics* **202**: 1185–1200.

Wang XR, Szmidt AE. 1994. Hybridization and chloroplast DNA variation in a *Pinus* species complex from Asia. *Evolution* **48**: 1020–1031.

Wang XR, Tsumura Y, Yoshimaru H, Nagasaka K, Szmidt AE. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *MATK*, *RPL20-RPS18* spacer, and *TRNV* intron sequences. *American Journal of Botany* **86**: 1742–1753.

Weißmann G, Lange W. 1987. Zusammensetzung der Neutralteile des Balsamkolophoniums von *Pinus massoniana* Lamb., *Pinus merkusii* Jungh. und *Pinus luchuensis* Mayr. *Holzforschung* **12**: 147–154.

Willyard A, Syring J, Gernandt DS, Liston A, Cronn R. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution* **24**: 90–101.

Xia H, Wang B, Zhao W, Pan J, Mao JF, Wang XR. 2018. Combining mitochondrial and nuclear genome analyses to dissect the effects of colonization, environments and geography on population structure in *Pinus tabuliformis*. *Evolutionary Applications* **11**: 1931–1945.

Xing YW, Liu YS, Su T, Jacques FMB, Zhou ZK. 2010. *Pinus prekesiya* sp. nov. from the upper Miocene of Yunnan, southwestern China and its biogeographical implications. *Review of Palaeobotany and Palynology* **160**: 1–9.

Yamada T, Yamada M, Tsukagoshi M. 2014. Fossil records of subsection *Pinus* (genus *Pinus*, Pinaceae) from the Cenozoic in Japan. *Journal of Plant Research* **127**: 193–208.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.

Zhao W, Meng J, Wang B, Zhang L, Xu Y, Zeng QY, Li Y, Mao JF, Wang XR. 2014. Weak crossability barrier but strong juvenile selection supports ecological speciation of the hybrid pine *Pinus densata* on the Tibetan Plateau. *Evolution* **68**: 3120–3133.

Zhao W, Sun YQ, Pan J, Sullivan AR, Arnold ML, Mao JF, Wang XR. 2020. Effects of landscapes and range expansion on population structure and local adaptation. *New Phytologist* **228**: 330–343.

Zhou Y, Duvaux L, Ren G, Zhang L, Savolainen O, Liu J. 2017. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* **118**: 211–220.

Zhou YF, Zhang LR, Liu JQ, Wu GL, Savolainen O. 2014. Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. *Molecular Ecology* **23**: 3504–3522.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Principal component analysis of genetic variations in the 13 pine species.

Fig. S2 Phylogeny of the 13 studied pine species.

Fig. S3 Estimates of purifying selection at zerofold degenerate sites in nine *Pinus* species, assuming a two-step population size change.

Fig. S4 Principal component analysis biplot on the 15 retained environmental variables of the 13 pine species.

Fig. S5 Models fit in Treemix.

Fig. S6 Demographic histories inferred by Stairway Plot 2 for nine species, including singletons.

Fig. S7 Demographic histories inferred by Stairway Plot 2, without singletons.

Table S1 Geographic locations and sample size (*N*) of the sampled populations of each species.

Table S2 Distribution records for the 13 species.

Table S3 Environmental variables sampled in this study.

Table S4 Introgression events supported by *D* and f_4 statistic calculated by *D*suite.

Table S5 Number of sequenced reads, mean sequencing depth and genome coverage for each sample.

Please note: Wiley is 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.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

14698137, 2024, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19694 by Xishuangbanna Tropical Botanical Garden, Wiley Online Library on [09/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License