Synchronous effective population size changes and genetic stability of forest trees through glacial cycles

Authors: Pascal Milesi[1,2*†], Chedly Kastally[3,4†], Benjamin Dauphin[5†], Sandra Cervantes[6,7†], Francesca Bagnoli[8], Katharina B Budde[9,10], Stephen Cavers[11], Dario I. Ojeda[12], Bruno Fady[13], Patricia Faivre-Rampant[14], Santiago C González-Martínez[15], Delphine Grivet[16], Felix Gugerli[5], Véronique Jorge[17], Isabelle Lesur-Kupin[15,18], Sanna Olsson[16], Lars Opgenoorth[5,19], Sara Pinosio[8,20], Christophe Plomion[15], Christian Rellstab[5], Odile Rogier[17], Simone Scalabrin[21], Ivan Scotti[13], Giovanni G Vendramin[8], Marjana Westergren[22], GenTree Consortium, Martin Lascoux[1,2*‡], Tanja Pyhäjärvi[3,4*‡]

Affiliations:

[1]Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University; SE-752 36, Uppsala, Sweden.

[2]Science for Life Laboratory (SciLifeLab); Uppsala, Sweden.

[3]Department of Forest Sciences, University of Helsinki; 00014 Helsinki, Finland.

[4]Viikki Plant Science Centre, University of Helsinki; 00014 Helsinki, Finland.

[5]Biodiversity and Conservation Biology, Swiss Federal Research Institute WSL; 8903 Birmensdorf, Switzerland.

[6]Department of Ecology and Genetics, University of Oulu; Oulu, Finland.

[7]Biocenter Oulu, University of Oulu; Oulu, Finland.

[8]Institute of Biosciences and Bioresources, National Reasearch Council of Italy (IBBR-CNR), 50019 Sesto Fiorentino, Florence, Italy.

[9]Department of Forest Genetics and Forest Tree Breeding, Georg-August-University Goettingen; Goettingen, Germany.

[10]Center of Sustainable Land Use (CBL), Georg-August-University Göttingen; 37077 Göttingen, Germany

[11]UKCEH; Bush Estate, Penicuik, Midlothian, EH26 0QB, United Kingdom

[12]Norwegian Institute of Bioeconomy Research.

[13]INRAE, URFM; F-84914, Avignon, France.

[14]Université Paris-Saclay, INRAE, Etude du Polymorphisme des Génomes Végétaux EPGV; 91000 Evry-Courcouronnes, France.

[15]INRAE, Université de Bordeaux, BIOGECO, F-33610, Cestas, France.

[16]Institute of Forest Sciences (ICIFOR-INIA), CSIC; 28040 Madrid, Spain.

[17]INRAE, ONF, BioForA; F-45075, Orléans, France.

[18]Helix Venture; F-33700 Mérignac, France.

[19]Plant Ecology and Geobotany, Philipps-Universität Marburg; 35043 Marburg, Germany.

[20]Institute of Applied Genomics (IGA); 33100 Udine, Italy.

[21]IGA Technology Services S.r.l., I-33100, Udine, Italy

[22]Slovenian Forestry Institute; 1000 Ljubljana, Slovenia.

*Corresponding author. Email: tanja.pyhajarvi@helsinki.fi, pascal.milesi@scilifelab.uu.se, martin.lascoux@ebc.uu.se

†These authors contributed equally to this work

‡These authors contributed equally to this work

Abstract: Past environmental changes have shaped the demographic history and genetic diversity of natural populations, yet the timescale and strength of these effects have not been investigated systematically and simultaneously for multiple phylogenetically distant species. We performed comparative population genomic analyses and demographic inference for seven ecologically contrasting European tree species sampled across their ranges. While patterns of genetic diversity and differentiation were species-specific and best explained jointly by each species' geographic range and dispersal ability, ancient population expansion events were shared and synchronous across species. Effective population sizes increased or remained stable over time, indicating that despite major changes in their geographic ranges, major forest tree species have been remarkably genetically resilient to the environmental challenges of repeated glacial cycles.

One-Sentence Summary: Population genomic analyses reveal demographic events across millions of years for seven forest tree species across Europe.

Main Text: Trees are fundamental to many terrestrial ecosystems, which shelter a major fraction of the Earth's biodiversity (*1*). They provide important ecosystem services and occupy a key position in processes affecting global warming (*2*). The contemporary genetic diversity of a species is vital in determining its future adaptive potential (*3*) but also bears the imprint of the historical factors, such as major demographic changes. Understanding the current distribution of genetic diversity in tree species and the timescale over which demographic events have taken place is therefore of paramount importance to make informed predictions of how rapidly and severely future environmental changes will affect forest ecosystems.

Most extant northern temperate and boreal tree species have existed for millions of years and survived multiple glacial–interglacial cycles. Micro- and macrofossil data indicate that these species have undergone huge range contractions and expansions over this time (*4*). In Europe, these cycles are thought to be the main drivers of the current distribution of genetic diversity (*5*, *6*), with the Last Glacial Maximum (LGM, approximately 27,000 to 19,000 years ago) and ensuing Holocene recolonization events playing a predominant role. Long-lived sessile organisms, such as forest trees, have fewer generations per glacial cycle than annual and herbaceous species, which affects the timescale of genetic footprint of the changes. However, other key biological parameters, such as the ecological niche and dispersal mechanism, also affect the extent and distribution of their genetic diversity. Since evolution is inherently stochastic and has historical contingencies, it is necessary to sample multiple species across broad geographic ranges to identify global drivers of genetic diversity patterns.

We studied seven wind-pollinated tree species (*7*): three conifers (*Picea abies*, *Pinus sylvestris* and *Pinus pinaster*) and four angiosperms (*Betula pendula*, *Fagus sylvatica*, *Populus nigra* and *Quercus petraea*), distributed across Europe from the boreal to the Mediterranean regions. The seeds of these species are dispersed by animals, wind and/or water (Table 1). We conducted targeted nuclear DNA sequencing (~10,000 species-specific probes that covered ~3 Mbp of orthologous sequences) on a total of 3,407 adult trees collected from 19 to 26 locations per species (~25 samples each) across their range (Figs. 1, S1, Supplementary Text, Tables 1, S1–S2) (*8*).

Overall, patterns of genetic diversity, differentiation and structure (Fig. 1) reflected neither phylogeny nor environmental preferences, but rather combinations of biological characteristics and demographic history specific to each species. The seven species formed four groups in differentiation–diversity space: highly genetically structured, early maturing, and low diversity *P. pinaster* and *P. nigra*; moderately structured, late maturing, and intermediate diversity *F. sylvatica* and *P. abies*; moderately structured, late maturing, high diversity, hybridizing and long-lived *Q. petraea*; and finally, panmictic boreal, wind dispersed, and moderate diversity *P. sylvestris* and *B. pendula* (Fig. 1I).

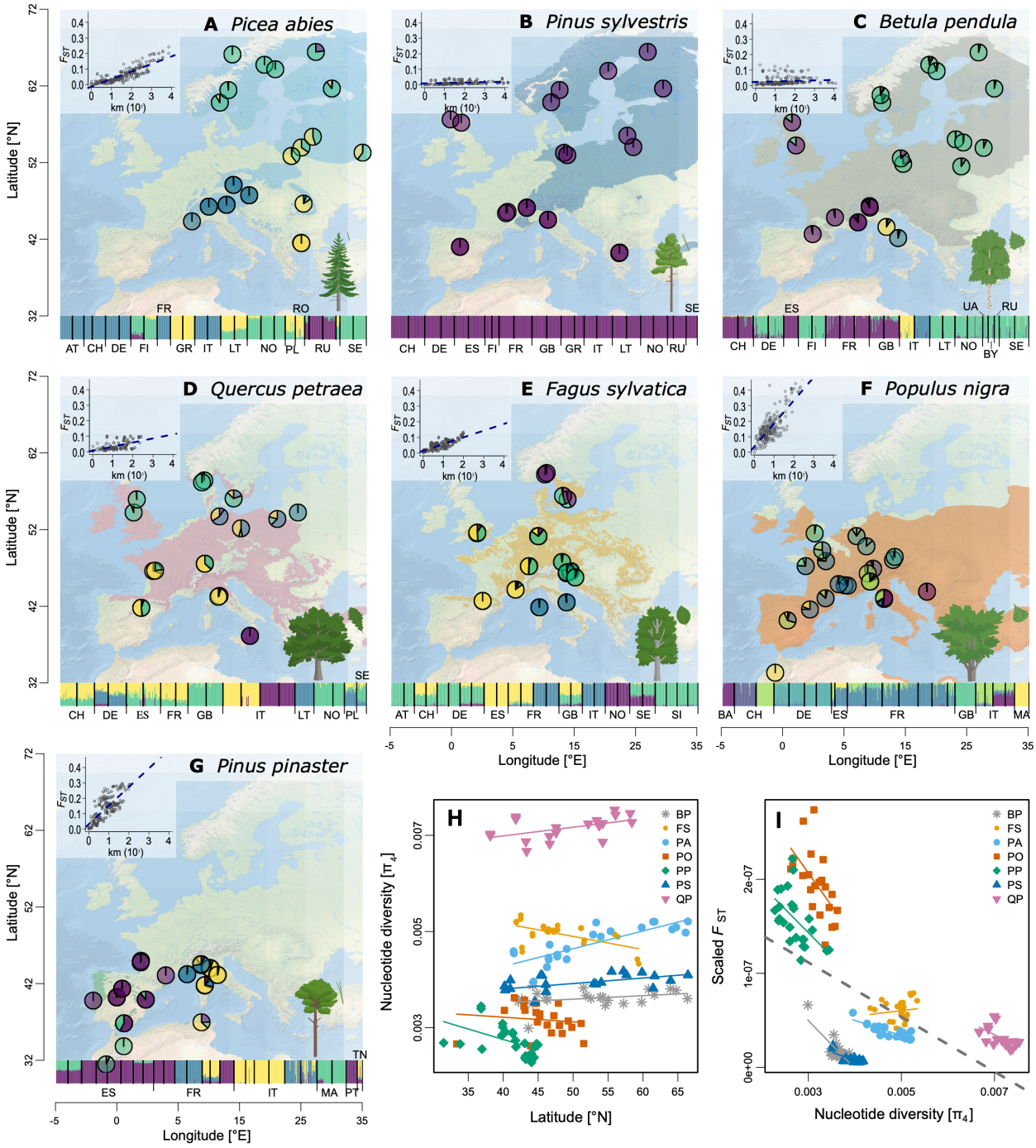

Fig. 1. Admixture patterns and isolation by distance among populations of the seven European tree species. (**A–G**) Pie charts represent population average admixture coefficients. Four genetic clusters are shown to visualize genetic structure, except for *Pinus sylvestris* (*K*=1; B) and *Populus nigra* (*K*=7; F) (see Figs. S3–S9 for other cluster numbers). An admixture plot of all individuals is shown at the bottom of each panel. Background maps represent species' ranges. Inset graphs show patterns of isolation by distance. (**H**) Nucleotide diversity at four-fold sites, π_4, as a function of latitude. (**I**) Population-specific differentiation estimated as the average of the ratios of pairwise F_{ST} over pairwise distance for all population pairs as a function of π_4. Different

symbols and colors represent the populations of the different species with their respective trend lines. Species codes are explained in Table S2.

Nucleotide diversity at four-fold degenerate (non-amino-acid-changing) sites (π_4) ranged from 0.0027 to 0.0072 across the species (Tables 1, S3), as is typical of outcrossing trees (*9*), and was remarkably similar among populations of a given species (Fig. 1H, Table S3). Boreal species exhibited elevated diversity at high latitudes, whereas genetic diversity tended to decrease northwards for the temperate species *P. nigra*, *F. sylvatica* and *P. pinaster* (Fig. 1H, Table S3). Thus, the geographic distribution of genetic diversity did not fit the central–marginal hypothesis, nor did it follow the south–north latitudinal gradient that is often considered as a proxy for postglacial recolonization history (*10*).

Genetic differentiation between populations (F_{ST}) was low for all species (Table 1) and isolation-by-distance was significant for most species, likely reflecting the domination of long-range, wind-mediated pollen dispersal over effects of genetic drift (Fig. 1A–G). However, the level of divergence was not uniform across the species distributions. The most genetically divergent populations were found at southern latitudes (Figs. 1A–G, S2, Table S3). Additionally, divergence systematically rose with increasing elevation for those species with a stronger population divergence towards southern latitudes. This trend means that populations at higher elevations were more differentiated from the rest of the range (Fig. S2, Table S4), as previously found in a study of 22 angiosperm trees and shrubs using chloroplast DNA (*6*). Genetic structure often coincided with the main discontinuities in the species' distributions, but with considerable variation across species (Figs. 1A–G, S3–S9).

To understand the origin and timing of the subtle but significant divergence observed among populations (Table 1), especially with latitude, we used demographic models implemented in fastsimcoal2 (*11*). We analyzed two non-admixed populations representing the main southern and northern clusters (see Table S5 for a list of selected populations) to estimate the timeframe of population separation (Fig. S10). In all species, divergence models with migration had better support than models without migration (Table S5), and the estimated divergence times between major clusters (0.6–17 Mya) largely predated the LGM (Fig. S11, Tables S5–S7). Hence, for all species the formation of the main genetic groups was likely the outcome of events through multiple glacial cycles. Consequently, the overall pattern of differentiation reflects topography and other factors creating persistent barriers to movement of populations, rather than just separation during the LGM and post-LGM re-colonization processes. For example, the mountains of southern Europe drove the recurrent formation of 'sky islands', isolated high-elevation regions to which cold-adapted species repeatedly shifted during inter-glacial periods (*12*).

To infer the timescale of changes and the historical effective population size (N_e) and to compare them across species, we used Stairway Plot 2, a method based on the site frequency spectrum (SFS) (*13*). Because Stairway Plot 2 is model-flexible and can be biased by intrinsic noise in the data, such as sampling effects, we then assessed the robustness of the results with the complementary but more constrained fastsimcoal2 2-epoch model. To account for the effect of sampling (*14*), we conducted analyses at the species, population (in the sense of sampling location), and one-sample-per-population levels. With the last level, the analysis was focused on the 'collecting' phase of the genealogy, which dominates the history of the metapopulation sample (*15*, *16*).

All species, except for *F. sylvatica*, showed an excess of rare variants in the SFS relative to levels expected under the standard coalescent model (Fig. 2). This is in line with findings from earlier studies on *P. abies*, *P. sylvestris, P. pinaster,* and *B. pendula* (*17–20*). Consequently, all species had a signal of ancient population growth, which was consistent across the sampling schemes, as well as between the Stairway Plot 2 and fastsimcoal2 2-epoch model inferences (Figs. 3A–B, S12). Strikingly, very few populations exhibited a signal of decreasing N_e through time, and those populations were often disconnected from the rest of the range and thus likely to have experienced stronger genetic drift (Fig. S13, Table S8). The magnitude of increase in N_e varied across species and was largest for *P. sylvestris* (from ~5,000 individuals to 500,000) and weakest for *F. sylvatica*, for which the proportional increase was only two-fold (from 100,000 to 200,000). Crucially, these patterns suggest that the overall genetic diversity of each species has been maintained even during the range contractions caused by glacial cycles. In other words, forest trees – with their large ranges, large census sizes, and efficient gene flow – have been able to maintain genetic diversity over long periods, despite the loss of single populations or even regional eradication during ecologically unfavorable periods. All estimates of past and present N_e were in the range of tens or hundreds of thousands; these values are much smaller than any reasonable estimate of current species census size, which is in the scale of billions of individuals for most species (*21*).

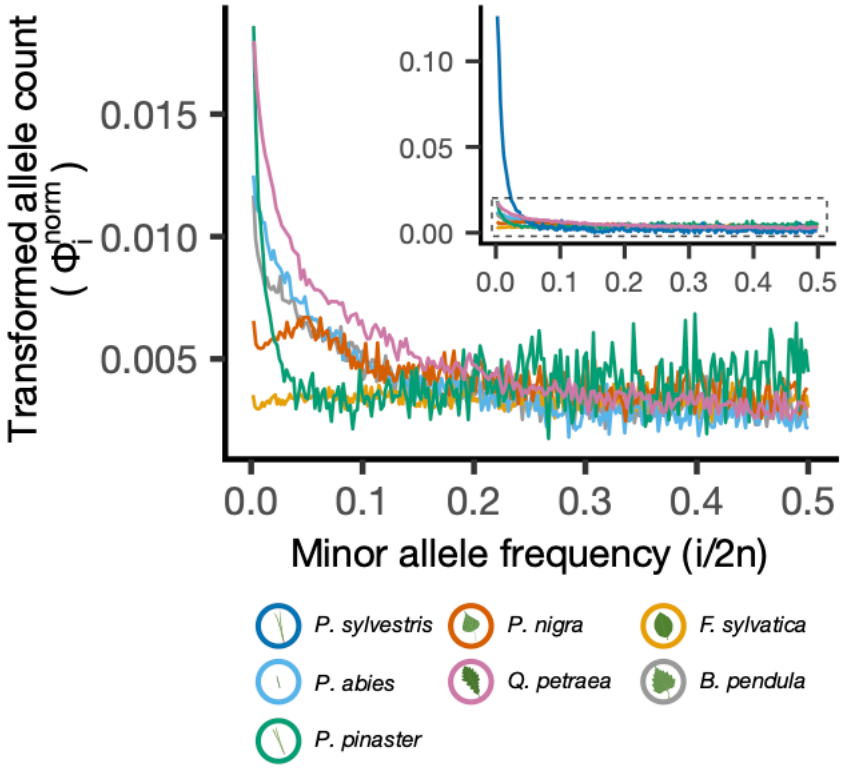

Fig. 2 Folded site frequency spectra (SFS) of all seven studied species (inset) and of the remaining six species after excluding *Pinus sylvestris*. The SFS were transformed and normalized following (*28*) (see Material & Methods). The transformed and normalized SFS are flat under the standard neutral model, facilitating the visualization of the SFS that depart from expectations of the standard neutral model (*29*, *30*).

According to randomization tests, changes in N_e were more synchronous than expected by chance among the seven species (Fig. 3B–D, Table S9), likely reflecting a shared response to environmental drivers. However, the resolution and exact timing of expansion varied among the species. In particular, the recovered demographic history of *Q. petraea* and *F. sylvatica* goes back much further in time (i.e. ~32 and ~17 Mya, respectively) than for the other species. This partly reflects their longer generation times, but the depth of the *Q. petraea* genealogy and its relatively large N_e are also likely a consequence of continuous hybridization with its sister species, *Q. robur* and *Q. pubescens* (*22*). Notably, the two pine species, *P. pinaster* and *P. sylvestris*, showed patterns very distinct from each other, despite similar sampling, mutation rates and generation times, reflecting their different ecological requirements, levels of population structure, and geographic ranges.

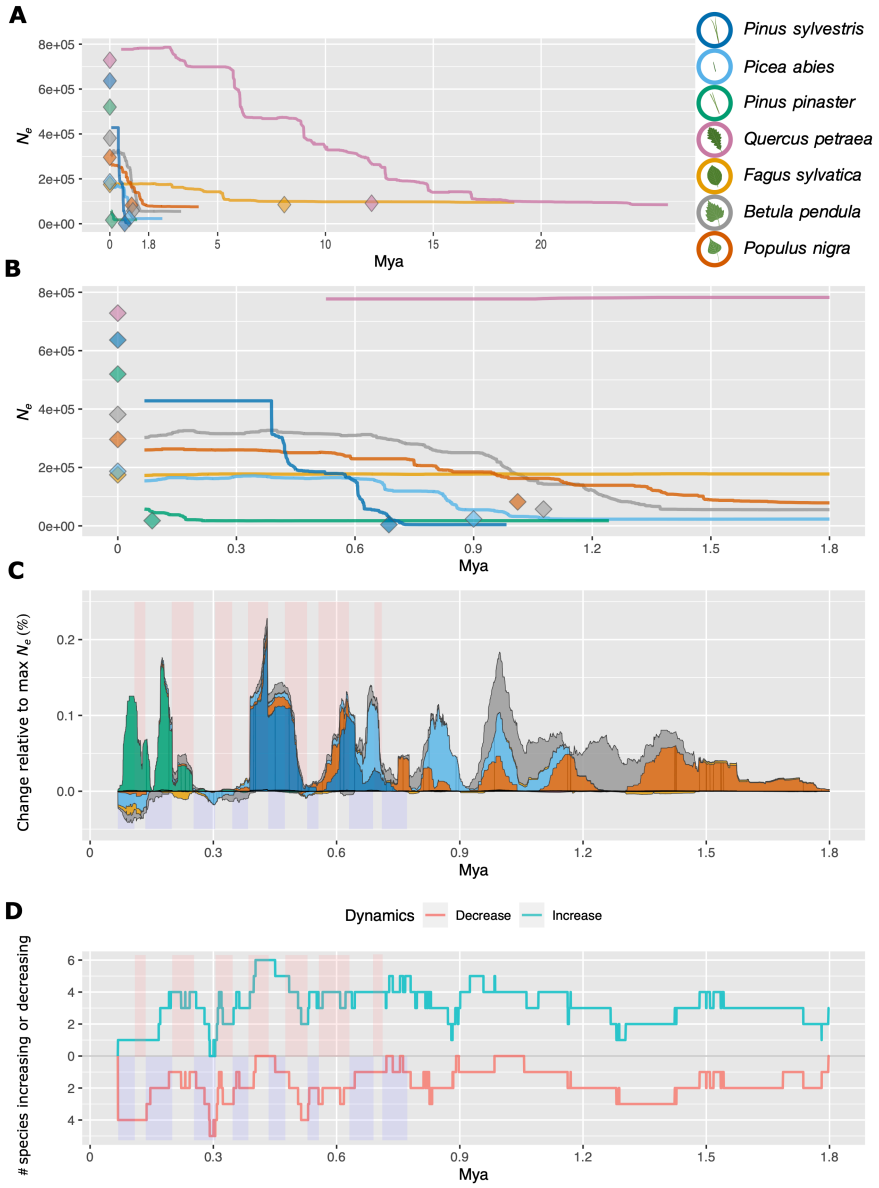

Figure 3. Synchronicity of demographic change across the seven species. (A) Change of effective population size (*Ne*) through time (million years ago, Mya), inferred with Stairway Plot 2 (lines, one-sample-per-population dataset, see Table S8) or with fastsimcoal2 (diamonds, 2-epoch model and one-sample-per-population dataset, see Tables S6–S7). The median change in *Ne* is reported for both methods. **(B)** Same as A but focused on the 0–1.8 Mya period. **(C)** Stacked area plot representing species-specific changes in *Ne* relative to species-specific maximum *Ne* (*Ne* changes were averaged across 250 time points using sliding windows) over the 0–1.8 Mya period (see Fig. S14 for 0–32 Mya). **(D)** Synchronicity of *Ne* changes through time, with solid lines representing the number of species experiencing an increase (blue) or decrease (red) in *Ne*. Species that experienced neither were excluded. The direction of change is given by the average change in *Ne* across 250 time points using sliding windows. In (C) and (D), blue and red shaded rectangles delineate glacial and interglacial periods, respectively.

Since we focused on past demographic events influencing current patterns of genetic diversity, we disregarded individuals with a high degree of admixture with closely related species. However, in many natural populations of forest trees, as in most plant species, hybridization contributes significantly to genetic diversity (*23*). For demographic inference, hybridization introduces signals of even older evolutionary events and leads to elevated N_e. For predicting population responses to climate change, more information on groups of closely related species will be essential, especially as introgression can be important in environmental adaptation (*24*).

In line with recent studies showing the ability of tree species to rapidly respond to environmental challenges (*25*) and to swiftly colonize new areas as they become suitable (*26*), our study demonstrates that tree species were also able to retain their evolutionary potential through multiple glacial cycles. This is likely a reflection of their unique biological features. First, very large and genetically connected populations have allowed tree species to retain genetic diversity through time, even after range contractions. Second, this diversity has evolved over very long periods, involving multiple glacial cycles and not just one main event. While the use, in previous studies, of late Pleistocene and Holocene pollen and macrofossil records, and/or low-resolution uniparentally inherited organellar DNA markers may have revealed post-LGM dispersal patterns, these approaches have not given a full picture of the dynamics of genetic diversity (*6*). Importantly, our results additionally suggest that tree species that have survived glacial cycles and currently have large ranges might be more capable of responding to climate change than one might have feared. However, genetic stability across millions of years does not exclude drastic changes in the short term, e.g., in species distributions and local abundances, which can still have major impacts on ecosystem services and forest functions.

References and Notes:

1. J. E. M. Watson, T. Evans, O. Venter, B. Williams, A. Tulloch, C. Stewart, I. Thompson, J. C. Ray, K. Murray, A. Salazar, C. McAlpine, P. Potapov, J. Walston, J. G. Robinson, M. Painter, D. Wilkie, C. Filardi, W. F. Laurance, R. A. Houghton, S. Maxwell, H. Grantham, C. Samper, S. Wang, L. Laestadius, R. K. Runting, G. A. Silva-Chávez, J. Ervin, D. Lindenmayer, The exceptional value of intact forest ecosystems. *Nature Ecology and Evolution* **2**, 599-610 (2018).
2. G. Popkin, How much can forests fight climate change. *Nature* **565**, 280-282 (2019).
3. M. Kardos, E. E. Armstrong, S. W. Fitzpatrick, S. Hauser, P. W. Hedrick, J. M. Miller, D. A. Tallmon, W. C. Funk, The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences* **118**, e2104642118 (2021).
4. H. J. B. Birks, T. W, ”Past forests of Europe” in *European Atlas of Forest Tree Species* (Publication Office of the European Union, 2016), pp. e010c45+.
5. P. Taberlet, L. Fumagalli, A. Wust-Saucy, J. Cosson, Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464 (1998).
6. R. J. Petit, I. Aguinagalde, J.-l. D. Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos, S. Fineschi, D. Grivet, A. Mohanty, G. Müller-Starck, B. Demesure-Musch, J. P. Martín, S. Rendell, G. G. Vendramin, Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**, 1563-1565 (2003).
7. E. Pennisi, Massive effort to document the genetics of European forests bears fruit. *Science* (2020); doi:10.1126/science.abb0632

8. L. Opgenoorth, B. Dauphin, R. Benavides, K. Heer, P. Alizoti, E. Martínez-Sancho, R. Alía, O. Ambrosio, A. Audrey, F. Auñón, The GenTree Platform: growth traits and tree-level environmental data in 12 European forest tree species. *GigaScience* **10**, giab010 (2021).
9. J. Chen, S. Glémin, M. Lascoux, Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution* **34**, 1417-1428 (2017).
10. C. G. Eckert, K. E. Samis, S. C. Lougheed, Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* **17**, 1170-1188 (2008).
11. L. Excoffier, M. Foll, Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332-1334 (2011).
12. A. S. Hirao, M. Watanabe, S. Tsuyuzaki, A. Shimono, X. Li, T. Masuzawa, N. Wada, Genetic diversity within populations of an arctic-alpine species declines with decreasing latitude across the Northern Hemisphere. *Journal of Biogeography* **44**, 2740-2751 (2017).
13. X. Liu, Y.-X. Fu, Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* **21**, (2020).
14. L. Chikhi, V. C. Sousa, P. Luisi, B. Goossens, M. A. Beaumont, The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* **186**, 983-995 (2010).
15. J. Wakeley, N. Aliacar, Gene genealogies in a metapopulation. *Genetics* **159**, 893-905 (2001).
16. J. Wakeley, Nonequilibrium migration in human history. *Genetics* **153**, 1863-1871 (1999).
17. J. Chen, L. Li, P. Milesi, G. Jansson, M. Berlin, B. Karlsson, J. Aleksic, G. G. Vendramin, M. Lascoux, Genomic data provide new insights on the demographic history and the extent of recent material transfers in Norway spruce. *Evolutionary Applications* **12**, 1539-1551 (2019).
18. T. Pyhäjärvi, M. R. García-Gil, T. Knürr, M. Mikkonen, W. Wachowiak, O. Savolainen, Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* **177**, 1713-1724 (2007).
19. J. Salojärvi, O.-P. Smolander, K. Nieminen, S. Rajaraman, O. Safronov, P. Safdari, A. Lamminmäki, J. Immanen, T. Lan, J. Tanskanen, P. Rastas, A. Amiryousefi, B. Jayaprakash, J. I. Kammonen, R. Hagqvist, G. Eswaran, V. H. Ahonen, J. A. Serra, F. O. Asiegbu, J. de Dios Barajas-Lopez, D. Blande, O. Blokhina, T. Blomster, S. Broholm, M. Brosché, F. Cui, C. Dardick, S. E. Ehonen, P. Elomaa, S. Escamez, K. V. Fagerstedt, H. Fujii, A. Gauthier, P. J. Gollan, P. Halimaa, P. I. Heino, K. Himanen, C. Hollender, S. Kangasjärvi, L. Kauppinen, C. T. Kelleher, S. Kontunen-Soppela, J. P. Koskinen, A. Kovalchuk, S. O. Kärenlampi, A. K. Kärkönen, K.-J. Lim, J. Leppälä, L. Macpherson, J. Mikola, K. Mouhu, A. P. Mähönen, Ü. Niinemets, E. Oksanen, K. Overmyer, E. T. Palva, L. Pazouki, V. Pennanen, T. Puhakainen, P. Poczai, B. J. H. M. Possen, M. Punkkinen, M. M. Rahikainen, M. Rousi, R. Ruonala, C. van der Schoot, A. Shapiguzov, M. Sierla, T. P. Sipilä, S. Sutela, T. H. Teeri, A. I. Tervahauta, A. Vaattovaara, J. Vahala, L. Vetchinnikova, A. Welling, M. Wrzaczek, E. Xu, L. G. Paulin, A. H. Schulman, M. Lascoux, V. A. Albert, P. Auvinen, Y. Helariutta, J. Kangasjärvi, Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nature Genetics* **49**, 904-912 (2017).
20. J. P. Jaramillo-Correa, F. Bagnoli, D. Grivet, B. Fady, F. A. Aravanopoulos, G. G. Vendramin, S. C. González-Martínez, Evolutionary rate and genetic load in an emblematic

Mediterranean tree following an ancient and prolonged population collapse. *Molecular Ecology* **29**, 4797-4811 (2020).
21. T. Pyhäjärvi, S. T. Kujala, O. Savolainen, 275 years of forestry meets genomics in *Pinus sylvestris*. *Evolutionary Applications* **13**, 11-30 (2020).
22. T. Leroy, C. Roux, L. Villate, C. Bodénès, J. Romiguier, J. A. Paiva, C. Dossat, J. M. Aury, C. Plomion, A. Kremer, Extensive recent secondary contacts between four European white oak species. *New Phytologist* **214**, 865-878 (2017).
23. T. G. Whitham, P. A. Morrow, B. M. Potts, Conservation of hybrid plants. *Science* **254**, 5033 (1991).
24. A. Suarez-Gonzalez, C. Lexer, Q. C. B. Cronk, Adaptive introgression: A plant perspective. *Biology Letters* **14**, (2018).
25. D. Saleh, J. Chen, J. C. Leplé, T. Leroy, L. Truffaut, B. Dencausse, C. Lalanne, K. Labadie, I. Lesur, D. Bert, F. Lagane, F. Morneau, J. M. Aury, C. Plomion, M. Lascoux, A. Kremer, Genome-wide evolutionary response of European oaks during the Anthropocene. *Evolution Letters* **6**, 4-20 (2022).
26. R. J. Dial, C. T. Maher, R. E. Hewitt, P. F. Sullivan, Sufficient conditions for rapid range expansion of a boreal conifer. *Nature* **608**, 546-551 (2022).
27. D. De Rigo, G. Caudullo, T. Houston Durrant, J. San-Miguel-Ayanz, ”The European Atlas of Forest Tree Species: modelling, data and information on forest tree species” in *European Atlas of Forest Tree Species* (Publication Office of the European Union, 2016), pp. e01aa69.
28. M. Lapierre, A. Lambert, G. Achaz, Accuracy of demographic inferences from the site frequency spectrum: The case of the Yoruba population. *Genetics* **206**, 439-449 (2017).
29. N. Nawa, F. Tajima, Simple method for analyzing the pattern of DNA polymorphism and its application to SNP data of human. *Genes and Genetic Systems* **83**, 353-360 (2008).
30. G. Achaz, Frequency spectrum neutrality tests: one for all and all for one. *Genetics* **183**, 249-258 (2009).
31. Z. Ning, A. J. Cox, J. C. Mullikin, SSAHA: a fast search method for large DNA databases. *Genome Research* **11**, 1725-1729 (2001).
32. C. Del Fabbro, S. Scalabrin, M. Morgante, F. M. Giorgi, An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE* **8**, e85024 (2013).
33. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
34. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
35. G. J. McKinney, R. K. Waples, L. W. Seeb, J. E. Seeb, Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources* **17**, 656-669 (2017).
36. A. R. Carson, E. N. Smith, H. Matsui, S. K. Brækkan, K. Jepsen, J.-B. Hansen, K. A. Frazer, Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics* **15**, 1-15 (2014).
37. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164-e164 (2010).
38. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664 (2009).

39. C. Rellstab, A. Bühler, R. Graf, C. Folly, F. Gugerli, Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (Quercus spp.). *Annals of Forest Science* **73**, 669-679 (2016).
40. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, M. Foll, Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9**, e1003905 (2013).
41. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 1-6 (2011).
42. R. J. Hijmans, J. van Etten, raster: Geographic data analysis and modeling, R package version. (2022); https://cran.r-project.org/web/packages/raster/index.html
43. B. S. Weir, C. C. Cockerham, Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370 (1984).
44. L. W. Pembleton, N. O. I. Cogan, J. W. Forster, St AMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* **13**, 946-952 (2013).
45. Z. N. Kamvar, J. F. Tabima, N. J. Grünwald, Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281 (2014).
46. F. Rousset, Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228 (1997).
47. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
48. Hijmans R. J., Spherical Trigonometry, R package geosphere version 1.5-10. (2019); https://cran.r-project.org/web/packages/geosphere/index.html
49. M. Nei, W.-H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* **76**, 5269-5273 (1979).
50. A. Keinan, A. G. Clark, Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**, 740-743 (2012).
51. Z. Xie, L. Wang, L. Wang, Z. Wang, Z. Lu, D. Tian, S. Yang, L. D. Hurst, Mutation rate analysis via parent–progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20161016 (2016).
52. S. Ossowski, K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, M. Lynch, The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92-94 (2010).
53. M. Krasovec, M. Chester, K. Ridout, D. A. Filatov, The mutation rate and the age of the sex chromosomes in *Silene latifolia*. *Current Biology* **28**, 1832-1838. e4 (2018).
54. V. C. T. Hanlon, S. P. Otto, S. N. Aitken, Somatic mutations substantially increase the per-generation mutation rate in the conifer *Picea sitchensis*. *Evolution Letters* **3**, 348-358 (2019).
55. A. Willyard, J. Syring, D. S. Gernandt, A. Liston, R. Cronn, Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for Pinus. *Molecular Biology and Evolution* **24**, 90-101 (2007).
56. T. R. Gaunt, S. Rodríguez, I. N. M. Day, Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool'CubeX. *BMC Bioinformatics* **8**, 1-9 (2007).

57. C. C. Chang, C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015 (2015).
58. I. Lesur, A. Bechade, C. Lalanne, C. Klopp, C. Noirot, J. Leplé, A. Kremer, C. Plomion, G. Le Provost, A unigene set for European beech (*Fagus sylvatica* L.) and its use to decipher the molecular mechanisms involved in dormancy regulation. *Molecular Ecology Resources* **15**, 1192-1204 (2015).
59. M. Müller, S. Seifert, T. Lübbe, C. Leuschner, R. Finkeldey, De novo transcriptome assembly and analysis of differential gene expression in response to drought in European beech. *PLoS ONE* **12**, e0184167 (2017).
60. K. Csilléry, H. Lalagüe, G. G. Vendramin, S. C. González-Martínez, B. Fady, S. Oddou-Muratorio, Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in E uropean beech (*Fagus sylvatica*) populations. *Molecular Ecology* **23**, 4696-4708 (2014).
61. A. Chateigner, M.-C. Lesage-Descauses, O. Rogier, V. Jorge, J.-C. Leplé, V. Brunaud, C. Paysant-Le Roux, L. Soubigou-Taconnat, M.-L. Martin-Magniette, L. Sanchez, Gene expression predictions and networks in natural populations supports the omnigenic theory. *BMC Genomics* **21**, 1-16 (2020).
62. P. Milesi, M. Berlin, J. Chen, M. Orsucci, L. Li, G. Jansson, B. Karlsson, M. Lascoux, Assessing the potential for assisted gene flow using past introduction of Norway spruce in southern Sweden: Local adaptation and genetic basis of quantitative traits in trees. *Evolutionary Applications* **12**, 1946-1959 (2019).
63. E. Eveno, C. Collada, M. A. Guevara, V. Léger, A. Soto, L. Díaz, P. Léger, S. C. González-Martínez, M. T. Cervera, C. Plomion, Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* **25**, 417-437 (2008).
64. D. Grivet, F. Sebastiani, R. Alía, T. Bataillon, S. Torre, M. Zabal-Aguirre, G. G. Vendramin, S. C. González-Martínez, Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution* **28**, 101-116 (2011).
65. D. Grivet, K. Avia, A. Vaattovaara, A. J. Eckert, D. B. Neale, O. Savolainen, S. C. González-Martínez, High rate of adaptive evolution in two widespread European pines. *Molecular Ecology* **26**, 6857-6870 (2017).
66. J.-P. Jaramillo-Correa, I. Rodríguez-Quilón, D. Grivet, C. Lepoittevin, F. Sebastiani, M. Heuertz, P. H. Garnier-Géré, R. Alía, C. Plomion, G. G. Vendramin, Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics* **199**, 793-807 (2015).
67. J. A. Cabezas, S. C. González-Martínez, C. Collada, M. A. Guevara, C. Boury, N. de María, E. Eveno, I. Aranda, P. H. Garnier-Géré, J. Brach, Nucleotide polymorphisms in a pine ortholog of the Arabidopsis degrading enzyme cellulase KORRIGAN are associated with early growth performance in *Pinus pinaster*. *Tree Physiology* **35**, 1000-1006 (2015).
68. C. Plomion, J. Bartholomé, I. Lesur, C. Boury, I. Rodríguez-Quilón, H. Lagraulet, F. Ehrenmann, L. Bouffier, J.-M. Gion, D. Grivet, High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*). *Molecular Ecology Resources* **16**, 574-587 (2016).
69. R. A. Cañas, I. Feito, C. Ávila, J. Majada, F. M. Cánovas, Transcriptome-wide analysis supports environmental adaptations of two *Pinus pinaster* populations from contrasting habitats. *BMC Genomics* **16**, 1-17 (2015).

70. W. Wachowiak, U. Trivedi, A. Perry, S. Cavers, Comparative transcriptomics of a complex of four European pine species. *BMC Genomics* **16**, 234 (2015).
71. W. Wachowiak, M. J. Salmela, R. A. Ennos, G. Iason, S. Cavers, High genetic diversity at the extreme range edge: Nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity* **106**, 775-787 (2011).
72. C. Plomion, J.-M. Aury, J. Amselem, T. Leroy, F. Murat, S. Duplessis, S. Faye, N. Francillonne, K. Labadie, G. Le Provost, Oak genome reveals facets of long lifespan. *Nature Plants* **4**, 440-452 (2018).
73. C. Rellstab, S. Zoller, L. Walthert, I. Lesur, A. R. Pluess, R. Graf, C. Bodénès, C. Sperisen, A. Kremer, F. Gugerli, Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. *Molecular Ecology* **25**, 5907-5924 (2016).
74. T. Leroy, J. Louvet, C. Lalanne, G. Le Provost, K. Labadie, J. Aury, S. Delzon, C. Plomion, A. Kremer, Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist* **226**, 1171-1182 (2020).
75. B. Mishra, D. K. Gupta, M. Pfenninger, T. Hickler, E. Langer, B. Nam, J. Paule, R. Sharma, B. Ulaszewski, J. Warmbier, J. Burczyk, M. Thines, A reference genome of the European beech (*Fagus sylvatica* L.). *GigaScience* **7**, giy063 (2018).
76. G. A. Tuskan, S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R. R. Bhalerao, R. P. Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G. L. Chen, D. Cooper, P. M. Coutinho, J. Couturier, S. Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroeve, A. Déjardin, C. Depamphilis, J. Detter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjärvi, J. Karlsson, C. Kelleher, R. Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J. C. Leplé, P. Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D. R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryaboy, J. Schmutz, J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C. J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. Van de Peer, D. Rokhsar, The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596-1604 (2006).
77. B. Nystedt, N. R. Street, A. Wetterbom, A. Zuccolo, Y.-C. Lin, D. G. Scofield, F. Vezzi, N. Delhomme, S. Giacomello, A. Alexeyenko, R. Vicedomini, K. Sahlin, E. Sherwood, M. Elfstrand, L. Gramzow, K. Holmberg, J. Hällman, O. Keech, L. Klasson, M. Koriabine, M. Kucukoglu, M. Käller, J. Luthman, F. Lysholm, T. Niittylä, A. Olson, N. Rilakovic, C. Ritland, J. A. Rosselló, J. Sena, T. Svensson, C. Talavera-López, G. Theißen, H. Tuominen, K. Vanneste, Z.-Q. Wu, B. Zhang, P. Zerbe, L. Arvestad, R. Bhalerao, J. Bohlmann, J. Bousquet, R. Garcia Gil, T. R. Hvidsten, P. de Jong, J. MacKay, M. Morgante, K. Ritland, B. Sundberg, S. L. Thompson, Y. Van de Peer, B. Andersson, O. Nilsson, P. K. Ingvarsson, J. Lundeberg, S. Jansson, The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579-584 (2013).
78. D. B. Neale, J. L. Wegrzyn, K. A. Stevens, A. V. Zimin, D. Puiu, M. W. Crepeau, C. Cardeno, M. Koriabine, A. E. Holtz-Morris, J. D. Liechty, P. J. Martínez-García, H. A. Vasquez-Gross, B. Y. Lin, J. J. Zieve, W. M. Dougherty, S. Fuentes-Soriano, L.-S. Wu, D.

Gilbert, G. Marçais, M. Roberts, C. Holt, M. Yandell, J. M. Davis, K. E. Smith, J. F. D. Dean, W. Lorenz, R. W. Whetten, R. Sederoff, N. Wheeler, P. E. McGuire, D. Main, C. A. Loopstra, K. Mockaitis, P. J. deJong, J. A. Yorke, S. L. Salzberg, C. H. Langley, Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* **15**, R59 (2014).

79. J. L. Wegrzyn, J. D. Liechty, K. A. Stevens, L. S. Wu, C. A. Loopstra, H. A. Vasquez-Gross, W. M. Dougherty, B. Y. Lin, J. J. Zieve, P. J. Martínez-García, C. Holt, M. Yandell, A. V. Zimin, J. A. Yorke, M. W. Crepeau, D. Puiu, S. L. Salzberg, P. J. de Jong, K. Mockaitis, D. Main, C. H. Langley, D. B. Neale, Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* **196**, 891-909 (2014).

Acknowledgments: The computations were enabled by resources in project SNIC 2017/7-328, SNIC 2020/15-107 and SNIC 2021/5-540 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. CSC – IT Center for Science, Finland, Genotoul Bioinformatics platform Toulouse Occitanie and the Genetic Diversity Centre at ETH Zurich, Switzerland, are acknowledged for generous computational and storage resources, and GBFOR, INRAE, Forest Genetics and Biomass Facility and Natural Resources Institute Finland (LUKE) for sample collection.

Funding:

European Union's Horizon 2020 Research and Innovation Programme grant agreement no. 676876

Academy of Finland grant nr. 287431 (TP)

Swiss Secretariat for Education, Research and Innovation (SERI) contract no. 6.0032 (FG)

Spanish Ministry of Agriculture, Fisheries and Food (MAPAMA) grant no. AEG 17-048 (DG)

European Agricultural Fund for Rural Development (EAFRD) (DG)

Swedish Research Council for Sustainable Development (FORMAS) grant nos. 2016-00780, 2020-01456 (ML)

Spanish Ministry of Economy and Competitiveness (MINECO) contract n. PTA2015-10836-I (SO)

LIFEGENMON (LIFE13 ENV/SI/000148) (MW)

The Slovenian Research Agency, research core funding no. P4-0107 (MW)

Biocenter Oulu

Author contributions:

Conceptualization: StC, SCGM, DG, ML, PM, TP, GGV

Methodology: KB, PFR, SCGM, CK, ML, PM, SP, TP, CR, SS

Software: KB, BD, CK, SP, TP, SS, IS,

Validation: SaC, VJ, PM, TP, SS

Formal Analysis: FB, KB, SaC, BD, PFR, SCGM, DIO, VJ, CK, ILK, PM, LO, SP, CP, TP, CR, OR, IS

Investigation: FB, SCGM, DG, FG, VJ, ML, ILK, PM, SO, LO, CP, TP, CR, IS, MW

Resources: BD, PFR, SCGM, DG, FG, ML, SO, LO, CP, CR, IS,

Data curation: BD, SaC, VJ, CK, PM, LO, TP

Visualization: BD, CK, PM, IS, MW

Funding acquisition: StC, BF, SCGM, DG, FG, ML, SO, LO, TP, GGV, MW

Project administration: StC, BF, SCGM, FG, ML, LO, TP, GGV,

Supervision: FB, SCGM, FG, VJ, ML, LO, CP, TP, CR, GGV, MW

Writing – original draft: KB, SaC, BD, SCGM, VJ, CK, ML, PM, LO, SP, TP, CR, OR, IS

Writing – review & editing: FB, KB, BD, BF, PFR, SCGM, FG, VJ, CK, ML, ILK, PM, LO, CR, SS, IS, GGV, MW

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Short reads are available under NCBI BioProjects: PRJNA602465, PRJNA602466, PRJNA602467, PRJNA602468, PRJNA602470, PRJNA602471, PRJNA602473. Code is available at: https://github.com/GenTree-h2020-eu/GenTree.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1-S14

Table S1-S16

References (*31–79*)

Table 1. Biological characteristics and genetic summary statistics for seven European tree species. Age information was retrieved from the European atlas of tree species (*27*) and the European forest genetic resources program (EUFORGEN).

Species	**Biological characteristics**	**Hybridization with**	**Min. flowering age (years)**	**Max. age known (years)**	F_{ST}	π_4 **(per bp)**	**Ns (individuals)**
Betula pendula (Silver birch)	▪ deciduous ▪ temperate to boreal	*B. platyphylla* *B. pubescens*	10–25	90–150	0.03*	0.0036	235

	▪ wind pollination ▪ wind seed dispersal ▪ large and continuous range						
Fagus sylvatica (European beech)	▪ deciduous ▪ temperate ▪ wind pollination ▪ animal seed dispersal ▪ large and continuous range	*F. orientalis*	40–50	150–300	0.05*	0.0050	65
Populus nigra (Black poplar)	▪ deciduous ▪ Mediterranean to temperate ▪ wind pollination ▪ water and wind seed dispersal ▪ vegetative and sexual reproduction ▪ intermediate and discontinuous range	*P. nigra* 'Italica' *P. deltoides* *P. trichocarpa* *P. maximowiczii* *Populus* sp. hybrid cultivars	4–10	100–400	0.16*	0.0032	8
Quercus petraea (Sessile oak)	▪ deciduous ▪ temperate ▪ wind pollination ▪ animal seed dispersal ▪ large and continuous range	*Q. robur* *Q. pubescens*	40–100	>1000	0.04*	0.0072	115
Picea abies (Norway spruce)	▪ conifer ▪ temperate to boreal ▪ wind pollination ▪ wind seed dispersal ▪ limited clonal reproduction ▪ large and discontinuous range	*P. obovata*	20–40	200–300	0.06*	0.0048	30
Pinus pinaster	▪ conifer		6–20	120–250	0.13*	0.0027	17

(Maritime pine)	▪ Mediterranean ▪ wind pollination ▪ wind seed dispersal ▪ limited and discontinuous range						
Pinus sylvestris (Scots pine)	▪ conifer ▪ temperate to boreal ▪ wind pollination ▪ wind seed dispersal ▪ large and continuous range	*P. mugo* *P. uliginosa*	15–30	400–750	0.01*	0.0039	392

Notes: mean pairwise genetic differentiation (F_{ST}), * denotes significantly differentiated populations based on AMOVA ($P < 0.05$); pairwise nucleotide diversity per site at four-fold sites (π_4); neighborhood size (Ns)