Resequencing 545 ginkgo genomes across the world reveals the evolutionary history of the living fossil

Yun-Peng Zhao1,13, Guangyi Fan2,13, Ping-Ping Yin1,2,13, Shuai Sun2,13, Ning Li2,3,4,13, Xiaoning Hong2,5,13, Gang Hu2, He Zhang2,2, Fu-Min Zhang3,4, Jing-Dan Han3, Ya-Jun Hao1, Qiwu Xu2,2, Xianwei Yang2, Wenjie Xia2, Wenbin Chen6, Han-Yang Lin1, Rui Zhang2, Jiang Chen2, Xiao-Ming Zheng1, Simon Ming-Yuen Lee7, Joongku Lee8, Koichi Uehara9, Jian Wang6,10, Huanming Yang6,10, Cheng-Xin Fu1, Xin Liu2,11, Xun Xu2,12 & Song Ge3,4

As Charles Darwin anticipated, living fossils provide excellent opportunities to study evolutionary questions related to extinction, competition, and adaptation. Ginkgo (Ginkgo biloba L.) is one of the oldest living plants and a fascinating example of how people have saved a species from extinction and assisted its resurgence. By resequencing 545 genomes of ginkgo trees sampled from 51 populations across the world, we identify three refugia in China and detect multiple cycles of population expansion and reduction along with glacial admixture between relict populations in the southwestern and southern refugia. We demonstrate multiple anthropogenic introductions of ginkgo from eastern China into different continents. Further analyses reveal bioclimatic variables that have affected the geographic distribution of ginkgo and the role of natural selection in ginkgo’s adaptation and resilience. These investigations provide insights into the evolutionary history of ginkgo trees and valuable genomic resources for further addressing various questions involving living fossil species.
Despite numerous efforts on investigations of living fossils, mysteries have remained for centuries regarding why living fossils appear essentially unchanged (i.e., in morphological stasis), whether living fossils are evolutionary dead ends and what roles humans have played in the survival and spread of living fossils. The ginkgo tree is an enigmatic living fossil, characterized by morphological stasis with almost no morphological change for at least 200 million years. A once-diverse and widespread group of gymnosperms, the ginkgo lineage has survived glaciations as a relic in China, has no current living relatives and has recently been redistributed globally via human-aided introductions. As one of the best-known and most distinctive trees worldwide, ginkgo has fascinated humans for centuries by its significance in biology and medicine as well as its power as a source of artistic and religious inspiration.

Although many investigations on the evolutionary history of this mysterious tree have been undertaken, numerous uncertainties and controversies remain, including the identification of relic populations and potential refugia, population dynamics in response to Pleistocene climate change, the roles that humans have played in the dispersal of ginkgo trees in China, and when and how ginkgo trees were introduced to Japan/Korea, Europe, and North America as well as the potential factors contributing to the persistence and resilience of the species. Exploration of these questions has been impeded by difficulties in obtaining sufficient samples that represent the entire gene pool of this living fossil and the limited genetic markers available for such an isolated gymnosperm species. Given that the ginkgo draft genome was completed, genome-wide markers available for such an isolated gymnosperm species. Given that the ginkgo draft genome was completed, genome-wide markers available for such an isolated gymnosperm species.

In this study, we resequence 545 ginkgo genomes and uncover the evolutionary history of this living fossil. Taking advantage of nuclear and plastid genomic data, we investigated unresolved debates on the evolutionary history of ginkgo. Specifically, we identify four ancient genetic components and three glacial refugia of ginkgo populations in China. Our demographic analyses detect multiple cycles of population expansion and reduction along with glacial admixture of relict populations in northern China in response to Pleistocene climate change. We also present a scenario on how ginkgo trees have been dispersed out of refugia and out of China based on multiple lines of evidence. Particularly, we characterize the major bioclimatic variables that have shaped the geographic distribution of ginkgo and addressed the potential roles of natural selection in the survival and resilience of ginkgo populations. These investigations lay an important foundation for further studies of the potential mechanism of ginkgo’s survival and resilience.

**Results**

**A largest genome dataset for a non-model species.** We have made the most extensive collections of the ginkgo samples around the world to date, including old ginkgo trees from most recorded localities (Supplementary Note 1). In the present study, we sampled 545 ginkgo individuals from 51 populations, covering almost all the natural range and growing locations of ginkgo across the world, which represents the most extensive collection of ginkgo samples to date (Fig. 1a, Supplementary Fig. 1, Supplementary Data 1, and Supplementary Table 1). These samples were sequenced using BGISEQ-500 sequencing platform at an average sequencing depth of ~6.3-fold, generating 44.30 Tb data in total (Supplementary Data 2 and Supplementary Fig. 2). Using the chromosome-level ginkgo reference genome, we obtained 214.33 million raw SNPs and 161.04 million high quality SNPs (Dataset 0) using a strict filtering standard (Supplementary Note 2). Distribution of SNP density along chromosomes, minor allele frequency (MAF), gene location, and inter-SNP distances unraveled the genomic landscape of ginkgo whole-population SNPs (Supplementary Fig. 3). To further assess the performance of the BGISEQ-500 sequencing platform for large genomes, we sequenced 14 individuals randomly selected from Chinese populations using the Illumina sequencing platform. More than 98.6% of SNPs were shared between the two sequencing platforms (Supplementary Table 2), suggesting robust stability of the BGISEQ-500 sequencing platform.

Four ancient genetic components and three refugia of ginkgo. We used multiple approaches to uncover ancient genetic components of ginkgo and infer the genetic structure and relationships of ginkgo populations (Supplementary Note 4). **ADMIXTURE analysis** indicates that the deepest splits occurred between the southwestern populations (the SWEST lineage) and the eastern populations (the EAST lineage) further diverged and that four ancestral genetic components of ginkgo occur in China, with the northern populations (the NORTH lineage) being admixed with three ancient components (Fig. 1a, b and Supplementary Fig. 4). Consistent with ADMIXTURE results, a principal component analysis (PCA) of Chinese samples reveals three major clusters corresponding to the EAST, SWEST and SOUTH lineages, with the NORTH lineage scattered between them (Fig. 1c and Supplementary Fig. 5). A neighbor-joining (NJ) tree further supports such a pattern (Fig. 1d, Supplementary Figs. 6 and 7), i.e., the SWEST, EAST and SOUTH lineages form three major clades while the NORTH lineage consists of multiple subclades divided into three major clades. These results indicate that the Chinese ginkgo populations harbor four ancient genetic components and consist of four major lineages in southwestern, southern, eastern and northern China. In addition to the well-known refugium in southwestern China, we identified a second refugium in eastern China and a third potential refugium in southern China where three ancient genetic components were uncovered (Fig. 1a and Supplementary Fig. 4). It is not unexpected to identify three refugia for ginkgo because these areas are biodiversity hotspots and were the Pleistocene refugia for many seed plants, including other living fossils such as Cathaya argyrophylla, Davidia involucrata, and Metasequoia glyptostroboides.

Phylogenetic analysis based on a plastome dataset, representing the nonrecombinant and maternally inherited genome, showed that 446 Chinese ginkgo trees formed three major clades with significant genetic divergence between them (Supplementary Fig. 8). The first clade consisted exclusively of the individuals from the EAST lineage, and the second comprised mostly the individuals from the SWEST lineage. The third and largest clade included the individuals from all four lineages (i.e., the EAST, SOUTH, NORTH, and SWEST lineages) (Supplementary Fig. 8), which was consistent with the results of previous studies based on plastid markers. Pairwise $F_{ST}$ values calculated based on plastome dataset also revealed deep divergence among the lineages with a large number of identical sequences both within and between lineages (Supplementary Table 3), suggesting seed-mediated and long-distance dispersals of ginkgo trees in China possibly due to human activities. Further network construction...
The x-phylogenetic analysis. (Supplementary Fig. 9), a pattern similar to that revealed by major haplotypes in the largest group shared by all lineages differing by at least a few substitutions, with four major haplotypes in the largest group shared by all lineages (Supplementary Fig. 9), a pattern similar to that revealed by phylogenetic analysis.

A comparably high level of genetic diversity (1.84–2.14 × 10⁻³ for Watterson’s estimator \( \theta_n \)) and 2.19–2.41 × 10⁻³ for the average pairwise diversity within populations (\( \pi \)) was observed for all four lineages of Chinese ginkgo populations (i.e., EAST, SOUTH, NORTH, and SWEST) (Table 1). Diversity estimates based on the whole chloroplast genomes yielded largely congruent results, i.e., the highest diversity for SWEST and the lowest for SOUTH (Supplementary Table 4). These results are in agreement with those of many previous reports indicating high geographical distribution of the sampling locations. The radius of the pies represents population size, and the colors represent ancestral components (according to the substructure at \( K = 4 \)). The map image, derived from ArcGIS Online maps, is the intellectual property of Esri, which is permitted for free use in academic publications (https://doc.arcgis.com/en/arcgis-online/reference/static-maps.htm). Same applies to the other maps in this paper. b Model-based population assignment by ADMIXTURE analysis for \( K = 2-5 \). The x-axis shows populations, and the y-axis quantifies the proportion of inferred ancestral lineages. c Principal component analysis (PCA) of Chinese samples, with the proportion of the variance explained being 8.61% for PC1 and 6.18% for PC2. d A neighbor-joining (NJ) phylogenetic tree of all 545 ginkgo samples around the world constructed using whole-genome SNP data based on pairwise identity-by-state (IBS) genetic distances. The source data of Fig. 1a-c are provided in a Source Data file.

**Table 1 Sample size and genetic diversity of five lineages of ginkgo trees**

| Lineage  | Region                  | No. populations/No. trees | No. SNPs | \( \pi (SD) \times 10^{-3} \) | \( \theta (SD) \times 10^{-3} \) |
|----------|-------------------------|---------------------------|----------|-------------------------------|-------------------------------|
| EAST     | Eastern China           | 11/96                     | 126,719,777 | 1.84 (0.84)                   | 2.41 (0.93)                   |
| SOUTH    | Southern China          | 10/92                     | 124,187,675 | 2.07 (0.98)                   | 2.38 (0.92)                   |
| NORTH    | Northern China          | 15/152                    | 145,809,400 | 2.11 (0.95)                   | 2.57 (0.93)                   |
| SWEST    | Southwestern China      | 9/118                     | 119,088,666 | 2.14 (1.00)                   | 2.19 (0.88)                   |
| OVERSEAS | Out of China            | 6/87                      | 116,979,767 | 2.05 (0.96)                   | 2.26 (0.90)                   |
| Total    | Global                  | 51/545                    | 160,887,036 | 2.11 (0.95)                   | 2.36 (0.82)                   |

* a average number of pairwise nucleotide differences per site, \( \theta_n \) Watterson’s estimator of \( \theta \) per base pair.

*All samples were grouped based on geographical distributions.*

showed that 13 haplotypes could be divided into three major and distinct groups differing by at least five substitutions, with four major haplotypes in the largest group shared by all lineages (Supplementary Fig. 9), a pattern similar to that revealed by phylogenetic analysis.
level of genetic diversity in ginkgo at both the species and local scales.\textsuperscript{9,10,13} This result provides additional evidence that no correlation exists between genetic diversity and morphological variation in living fossils.\textsuperscript{8} Notably, high genetic diversity was observed in the NORTH lineage ($\pi_e = 2.11 \times 10^{-3}$ and $\pi = 2.57 \times 10^{-3}$), which reflects the characteristics of admixed origin of ginkgo populations in northern and central China and supports the argument of glacial admixture in relict populations in many species in eastern Asia.\textsuperscript{25}

Multiple cycles of expansion and reduction of ginkgo populations. An important question about living fossils is whether their extinction is underway.\textsuperscript{8} To reconstruct the demographic history of ginkgo, we applied the pairwise sequentially Markovian coalescent (PSMC)\textsuperscript{26} to assess population size changes and obtained a well-defined demographic history from 20 million to 60,000 years ago (Fig. 2a and Supplementary Fig. 10). We found multiple substantial demographic fluctuations, i.e., three peaks in population size at ~15 million years ago (mya), ~1.05 mya, and ~0.5 mya, as well as three significant population bottlenecks at ~4 mya, 0.1 mya, and 0.07 mya. The result indicated a few cycles of population expansions and reductions in ginkgo populations during the Pleistocene glaciations. Notably, the demographic fluctuations since 2 mya showed opposing changes in the amount of atmospheric dust, as inferred by the mass accumulation rate (MAR) of Chinese loess\textsuperscript{27}, which indicates that population size reductions were correlated with a cold climate (high MAR).

Because of the insufficient power of PSMC in inferring more recent demographic history due to the limited number of recombination events in a single genome,\textsuperscript{28} we simulated the demographic fluctuations since 0.4 mya (Supplementary Tables 5 and 6) based on SNPs of the ginkgo populations using \textit{fastsimcoal2}\textsuperscript{29}. The model with the best score supports the earliest divergence between the SWEST and EAST + SOUTH lineages being 515,780 years ago followed by the split-off of the SOUTH lineage 318,120 years ago (Fig. 2b and Supplementary Table 5), which are in well agreement with an estimate based on whole-plastid-genome sequence data\textsuperscript{12}. It is evident that the NORTH lineage originated 139,260 years ago as a result of admixture between the SWEST (28.45%) and SOUTH (71.55%) lineages (Fig. 2b), probably resulting from repeated range expansions and contraction during the warming and cooling phases since 0.5 mya\textsuperscript{12}. A moderate reduction in effective population sizes ($N_e$) was observed for the four descendant lineages ($N_e = 24,819 \sim 32,093$) relative to their most recent common ancestor ($N_e = 50,514$) (Fig. 2b), consistent with the PSMC inference (Fig. 2a). No significant recent gene flow was detected among the three ancient lineages (i.e., the EAST, SOUTH, and SWEST lineages).

The plastome dataset also indicated deep divergence among the three ancient lineages with possible seed-mediated dispersal, given that a large number of identical chloroplast sequences were found among lineages (Supplementary Figs. 8 and 9, Supplementary Table 3). The fact that individuals from the NORTH lineage clustered with all three other lineages supported the admixed nature of the NORTH lineage (Supplementary Fig. 8), as indicated by the nuclear genome dataset. Together, we found that ginkgo populations underwent several cycles of expansion and contraction during the Pleistocene glaciations and revealed a moderate reduction in ginkgo populations in refugia associated with admixture of relict populations in northern and central China.

Human-aided dispersal of ginkgo out of China. Unlike many other living fossils that are confined to an isolated area, ginkgo has been thought to be extinct but has actually flourished after being introduced by humans to different areas around the world.\textsuperscript{5,6,8} It remains speculative that ginkgo trees underwent human-aided spread out of the refugia in China and were further introduced to various areas in East Asia, Europe and the Americas\textsuperscript{1–6}. Our ADMIXTURE result indicates that non-Chinese populations (the OVERSEAS samples) comprise three ancient genetic components that are only found in the EAST and SOUTH lineages (Fig. 3a, Supplementary Fig. 4), excluding the possibility that the SWEST lineage was the donor of ginkgo trees outside of China. PCA and NJ analyses produced congruent results showing that the OVERSEAS samples formed a cluster mainly with EAST that consisted of two subgroups (Fig. 3a and Supplementary Fig. 11). It is noted from Fig. 3a, that the Japanese and Korean samples form a dense cluster with the main group of the EAST samples, while those from Europe are mainly clustered with other EAST samples that have a rare genetic component (Fig. 1b). The American samples form three minor clusters that overlap with those from Japan/Korea, EAST and Europe. These observations

---

**Fig. 2** Demographic history of ginkgo populations. a Demographic history of four ginkgo lineages including the populations in eastern China (EAST), southern China (SOUTH), northern China (NORTH), and southwestern China (SWEST) inferred by PSMC. The mass accumulation rate (MAR) of Chinese loess is shown in black line. b Schematic of demographic scenarios modeled using \textit{fastsimcoal2}, with the ancestral population shown in brown. Column width represents the relative effective population size, with the NORTH lineage inferred as a mixture of 71.55% of the SOUTH lineage and 28.45% of the SWEST lineage. The numbers on the vertical axis indicate the estimated time of population divergence.
support the speculation that old ginkgo trees in the USA might have originated multiple times from eastern China\textsuperscript{4,5} but are inconsistent with the argument that ginkgo in Europe was introduced from Japan\textsuperscript{6}. To better understand the dispersal history of ginkgo trees, we calculated the pairwise identity by state (IBS) genetic distance between individuals (Fig. 3b). We found that the genetic distance between individuals did not follow a strict isolation-by-distance model, in which the genetic distance between individuals correlates with their geographic distance. One extreme case comprising individuals with the smallest IBS values (IBS < 0.03) (Fig. 3b) consists of groups of nearly identical individuals, reflective of recent dispersal or introduction. As shown in Fig. 3c, the pairs with an IBS < 0.03 (red lines) occur mainly among EAST, Europe and the USA, supporting the recent introduction of ginkgo from eastern China to different continents\textsuperscript{4,5}. Additional pairs of individuals with an IBS < 0.03 were found in China, Japan, and Europe (Fig. 3c). The pairs with slightly larger IBS values (0.03 ≤ IBS < 0.07, pairs without connected lines) were found mainly between China and Japan/Korea (Fig. 3c), suggesting earlier introductions of Chinese ginkgo into Japan/Korea than into Europe and the USA.

Interestingly, the pairs with the largest IBS values (IBS > 0.09, green lines) only occurred in China, mainly between the EAST and SWEST lineages (Fig. 3c), a reflection of the highest genetic differentiation between the EAST and SWEST lineages Table 1). However, several pairs with almost identical samples (IBS < 0.03, red lines) are also observed in China, linking samples that are geographically distant (Fig. 3c). This result is consistent with the results of analyses based on the plastome dataset, suggesting seed-mediated dispersal (Supplementary Table 3) and supports previous arguments that recent human activity helped in the effective dispersal of ginkgo trees in China\textsuperscript{14,15}.

Environmental adaptation and natural selection. To reveal the potential impact of climate fluctuations on the distribution of the species, we simulated distribution patterns of ginkgo at present (1970–2000), during the last glacial maximum (LGM: c. 21 thousand years before present, kyr BP), and the last interglacial (LIG: c. 115–130 kyr BP) with all 19 bioclimatic variables using species

---

**Fig. 3** Genetic and geographic distances between all samples. a Principal component analysis of global samples. Triangles indicated by black arrows are the samples containing more than 50% of a rare genetic component (blue in Fig. 1b). This component is widespread in the European and American samples. b Distribution of pairwise identity-by-state (IBS) genetic distances, with three cutoffs indicted by vertical dashed lines. The map image was derived from ArcGIS. c Geographic locations of 545 global samples, showing the pairs with the shortest genetic distance (IBS < 0.03, connected by red lines) and pairs with the longest genetic distance (IBS > 0.09, connected by green lines). All the pairs without a line connection have the IBS values between 0.03 and 0.09. Source data are provided in a Source Data file.
distribution modeling (SDM)\textsuperscript{30} (Supplementary Note 6). It is evident that the overall distribution pattern of the species predicted for the present is largely consistent with its actual distribution and that suitable habitats for ginkgo contracted during the LGM and then expanded northward after the LGM, with the suitable habitats being much larger at present than during the LGM (Fig. 4a). We then estimated the relative importance of the climatic variables on the species distribution and found that seven bioclimatic variables, i.e., mean temperature of the coldest quarter (BIO11), temperature seasonality (BIO4), mean temperature of the wettest quarter (BIO8), precipitation of the driest month (BIO14), precipitation seasonality (BIO15), precipitation of the warmest quarter (BIO18) and mean diurnal range (BIO2), strongly affected the distribution of ginkgo trees (Supplementary Table 7). Notably, the variables determining the ginkgo distribution differ between eastern and southwestern China (Supplementary Table 7), implying the differentiation of habitat preferences between ginkgo populations in the two refugia.

To examine the potential mechanism of adaptation to environmental changes during ginkgo’s evolution, we investigated genomic signals of adaptation at the local scale. We selected one group of individuals (58 trees) each from the EAST and SWEST lineages, herein named EAST-g and SWEST-g, respectively (Fig. 1a, Supplementary Table 8) to represent the ginkgo lineages in two refugia and searched the genome for potential regions with signatures of selection using parameter-based statistics and a likelihood-based program SweeD\textsuperscript{31} (Supplementary Note 7). In total, we identified 7 windows in EAST-g and 46 windows in SWEST-g, which showed signature of selection (Fig. 4b and Supplementary Data 3) and most of these regions could also be identified under different sliding widows (Supplementary Data 4 and 5). To avoid the potential impact of low genetic variation at the linked loci on the detection of selective signals, we implemented composite likelihood ratio (CLR) statistics to identify regions that are significant deviations from neutral the site frequency spectrum (SFS) (Supplementary Note 7). The results showed that a total of 910 and 949 putative regions that showed signatures of selection and contained 643 and 504 candidate genes, in EAST-g and SWEST-g, respectively (Fig. 4b, Supplementary Data 6 and 7). We further performed gene ontology (GO) enrichment analysis of the genes in these putative regions and obtained 14 and 17 significantly enriched terms (Fisher’s exact test, $P < 0.05$) in EAST-g and SWEST-g, respectively, including many terms involving responses to various abiotic and biotic stress (Supplementary Table 9). Sequence homology analysis of 25 genes identified by both approaches revealed some important genes involved in insect and fungal defenses and responses to abiotic stress such as dehydration, low temperature and high salt (Supplementary Note 7 and Supplementary Table 10), consistent with previous findings that ginkgo possessed unusually high resistance or tolerance to both abiotic and biotic stress, particularly herbivores and pathogens\textsuperscript{1,6,16}. Further studies on the function of these genes and their roles in ginkgo adaptation are highly encouraged.

**Discussion**

This study uncovered the evolutionary history of a species with a very large genome (10.6 Gb\textsuperscript{16}) based on resequencing of
hundreds of samples covering all known localities of the species. Here we used an updated genome assembly using HiC technology to conduct substantial investigations on population genomics and evolutionary history of ginkgo. Such a chromosome-level reference provides the distribution information of SNPs and thus represents a valuable resource for addressing numerous questions involving biologic diversity, population genetics, and evolutionary history. To evaluate the applicability of SNPs called based on draft genome where no position information is available, we used the SNPs called based on the draft genome to perform various analyses. As demonstrated by ADMIXTURE analyses (Supplementary Fig. 4c), PCA and NJ trees of both Chinese (Supplementary Figs. 5c, d and 6b) and global (Supplementary Figs. 11c, d and 7b) samples as well as the distribution of the IBS genetic distances (Supplementary Fig. 12), the two datasets (i.e., the SNPs called based on the draft and updated genomes) generated almost identical results, implying that the SNPs called based on a draft genome (or virtual pseudo-genomes) generated almost identical results, implying that the morphological stasis of living fossils is an effective adaptation strategy in response to environmental change, although the underlying mechanisms are unclear. The reasons for and potential factors contributing to ginkgo’s resilience have been speculated about in many previous studies, which showed that ginkgo trees maintained outstanding resistance or tolerance to both herbivores and pathogens, partly accounting for the longevity and high vegetative propagation ability of individual trees. Consistently, our analyses of adaptation revealed numerous pathways and specific genes that are involved in responses to abiotic and biotic stresses (Supplementary Note 7), which deserves further in-depth investigations.

Whether living fossils and relict species are undergoing extinction is an intriguing question. As an example, the giant panda has long been regarded as an evolutionary dead-end but proved to be a successful species highly adapted to environmental changes. Wei et al. argued that extensive speculation about the giant panda arose from an unsystematic and unsophisticated understanding of its biology. The present study provides multiple lines of evidence explicitly refuting the notion that ginkgo is an evolutionary dead end in terms of adaptation to abiotic and biotic stress (Supplementary Note 7) as well as the high level of genetic diversity (Table 1), repeated resurgence of population size in response to glaciations (Fig. 2), and a wide distribution of potential habitats (Fig. 4a). The dawn redwood tree is another famous living fossil that was believed to have been extinct for several million years and re-discovered in the 1940s in several isolated areas in central and southern China. Similar to ginkgo, dawn redwood has successfully expanded its distribution to approximately 50 countries via decades of human efforts, even over a much wider range than the fossils indicated. Living gymnosperms that were considered as ancient and living fossils are convincing examples of successful plant lineages, in which many extant species occupy diverse habitats and survive dramatic environmental changes by adaptive shifts. For example, in a study on the cycad lineage, various evolutionary adaptations were observed. Nagalingum et al. found that cycads underwent global rediversification beginning in the late Miocene, indicating that the species diversity of today’s cycads arose from a relatively recent radiation, less than 12 mya. Ginkgo’s resilience, along with other similar cases such as that in dawn redwood and cycads, as well as some other gymnosperms supports the argument that intuitively fragile living fossils may possess the ability to survive successfully and even flourish if they are reintroduced to suitable habitats, particularly when aided by humans. Together, the results of our study provide not only a comprehensive evolutionary framework for ginkgo research and conservation but also
insights into the evolutionary history and conservation of other living fossil species.

Methods

Sample collection and resequencing. A total of 545 individual ginkgo trees were carefully selected to represent most of the known localities of large ginkgo trees around the world, including 51 populations from nine countries (Supplementary Fig. 1 and Supplementary Note 1). Genomic DNA extraction, library construction, and amplification followed standard protocols (Supplementary Note 1). All samples were sequenced using BGISEQ-500 with a pair-end read length of 50 bp or 100 bp. We filtered raw data using SOAPnuke (ver. 1.5.4)41 and obtained clean sequencing reads with an average depth of up to 6.1-fold, ranging from 4- to 10-fold for each sample (Supplementary Data 2, Supplementary Figs. 2, 13).

SNP calling, quality control, and validation. We used the DRAGEN (https://www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html) toolkit to conduct read mapping and SNP calling, with the clean reads of FASTQ files used for alignment, sorting, duplicate removal, and variant calling. About 99% of the samples were aligned to the reference genome with a mapping ratio of more than 85% (Supplementary Data 2 and Supplementary Fig. 14). The hidden Markov model and Smith-Waterman alignment in GATK 4.0 Haplotype Variant Caller43 were used for the SNP calling. The raw SNP dataset was filtered using variant quality score recalibration (VQSR), with the majority of the SNVs showing high quality scores and low missing rates (Supplementary Fig. 15). Five datasets were compiled to satisfy the requirements of various analyses (Supplementary Note 2). To assess potential biases resulting from the use of different sequencing platforms, we sequenced 14 randomly selected samples using the Illumina platform. By comparing the two SNP datasets generated by the Illumina HiSeq2000 sequencing platform and BGISEQ-500, we found that more than 98.6% of SNPs were shared by the two platforms with different filtering standards, suggesting high consistency between the two sequencing platforms (Supplementary Table 2).

Chloroplast genome assembly and SNP calling. Being maternally inherited in ginkgo44, the chloroplast genome provides additional information on population genetics and phylogenetic relationships. The obtained SNP data from chloroplast genomes, representing the maternally inherited markers, to perform analyses of phylogenetic relationships and population genetic structure of ginkgo trees. We filtered raw resequencing data to obtain reads mapping to the chloroplast sequence AB684440.1 of ginkgo45 using BWA and SAMtools. A total of 481 chloroplast genomes were completely assembled using the GetOrganelle pipeline46, of which 446 plastomes of Chinese ginkgo trees were used in various analyses. SNP calling and quality control for the chloroplast genomes were performed following processes similar to those mentioned above (Supplementary Note 3).

Analysis of population genetics and phylogeny. We performed routine analyses of population genetic structure and phylogenetic relationships of ginkgo with both the resequencing and plastome datasets (Supplementary Notes 4 and 5). The population structure and admixture of our global samples were inferred using ADMIXTURE (ver. 1.3.0)18 with five replicates and 10-fold cross-validation (CV) from K = 2 to 10. PCA was conducted to study the relatedness and clustering of populations or samples. The top 10 PCs of the variance-standardized relationship matrix were extracted using PLINK (ver. 1.90)37 with the –pca parameter to quantify the relatedness between individuals, the pairwise identity-by-state (IBS) genetic distance matrix of all 545 samples was calculated using PLINK (ver. 1.90)37 with the parameter–distance 1-ibs. We constructed a neighbor-joining (NJ) phylogenetic tree using MEGA (ver. 4.0)38 based on the distance matrix. A haplotype network was constructed for the plastome dataset using PopART39 based on codon-based alignments from chloroplast genomes.

Demographic history inference. PSMC was used to infer historical dynamics of effective population size and the divergence timing of ginkgo lineages51 based on 8 samples from four lineages with a high sequencing depth (~30-fold) (Supplementary Table 11). The whole-genome diploid consensus sequences for each sample were generated using SAMtools and BCFtools with the parameter C30. Sites with sequencing depths <10 and >100 (vcfutils.pl vcf2gz -d 10 -D 100) were removed to reduce the probability of false positives. The PSMC parameter (psmc -N25 -r15 -r5 -p "4 + 25*2+4+6") was used to infer the historical effective population size. The estimated generation time and mutation rate were set to 40 and 0.67e-9, respectively (Supplementary Table 16). To infer demographic history using fastsimcoal248, two-dimensional joint SFS (2D-SFS) was constructed from posterior probabilities of sample allele frequencies using easySFSpy (https://github.com/isaacovercast/easySFS). Each model was run 100 times with 100,000 simulations for the calculation of the composite likelihood, and 10–40 expectation-conditional maximization (ECM) cycles (Supplementary Fig. 17). Model comparison was based on the maximum likelihood value across 20 independent runs using the Akaike information criterion and Akaike’s weight of evidence49. The model with the maximum Akaike weight value was chosen as the optimal model. Finally, we calculated confidence intervals of parameter estimates from 100 para-metric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters on the pairwise identity by state (IBS) genetic distance calculated by PLINK (ver. 1.9), we used a histogram to display the IBS distribution. Pairs with short genetic distances (IBS < 0.03 or IBS > 0.07) may reflect recent dispersal or introgression.

Analysis of the species distribution and local adaptation. We performed species distribution modeling (SDM) using the software MAXENT (v3.3.1)50,56 to explore the impacts of climate on the species distribution and to detect major climatic variables that contribute to the species distribution (Supplementary Note 6). We detected genomic regions under selection in two groups of individuals (EAST-g and SWEST-g), representing the EAST and SWEST lineages, respectively to explore the genomic basis of local adaption using two approaches (Supplementary Note 7, Supplementary Figs. 18 and 19), including one based on population genetics summary statistics (H2 and Fst)37 and the other in the likelihood-based program SweeD (version 3.1)59. We performed gene ontology (GO) enrichment analysis for the candidate genes in EAST-g and SWEST-g using a perl script60.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this study are available within the paper and its supplementary information files. The resequencing reads of ginkgo in this study have been deposited in NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA78810. The datasets reported in this study are also available in the CNGB Nucleotide Sequence Archive under accession number CNP0000136. The source data underlying Figs. 1a–c, 3a–c, 4b and Supplementary Figs. 2, 3b, 5a, b, 6a, 11a, b, 14, 15, and 18 are provided as a Source Data file.

Received: 8 November 2018 Accepted: 23 August 2019
Published online: 13 September 2019

References

1. Major, R. T. The ginkgo, the most ancient living tree: the resistance of Ginkgo biloba L. to pests accounts in part for the longevity of this species. Science 157, 1270–1273 (1967).
2. Zhou, Z. & Zheng, S. Palaeobiology: the missing link in Ginkgo evolution. Nature 423, 821 (2003).
3. Hori, T. et al. in Ginkgo Biloba A Global Treasure - From Biology to Medicine (Springer, 1997).
4. Zhao, Y., Pu, J., Fu, C. X. & Koch, M. A. Out of China: Distribution history of Ginkgo biloba L. Taxon 59, 495–504 (2010).
5. Crane, P. Ginkgo: The Tree That Time Forgot (Yale University Press, 2013).
6. Del Tridici, P. Ginkgos and People—a thousand years of interaction. Arnoldia 51, 1–25 (1991).
7. Zhou, Z. Y. An overview of fossil Ginkgoales. Palaeoworld 18, 1–22 (2009).
8. Werth, A. J. & Shear, W. A. The evolutionary truth about living fossils. Am. Sci. 102, 434 (2014).
9. Gong, W., Chen, C., Dobeš, C., Fu, C. X. & Koch, M. A. Phylogeography of a living fossil: Pleistocene glaciations forced Ginkgo biloba L. (Ginkgoaceae) into two refuge areas in China with limited subsequent postglacial expansion. Mol. Phylogenetics Evol. 48, 1094–1105 (2008).
10. Shen, L. et al. Genetic variation of Ginkgo biloba L. (Ginkgoaceae) based on cpDNA PCR-RFLPs: inference of glacial refugia. Heredity 94, 396–401 (2005).
11. Yang, C. Q. et al. Evidence for the persistence of wild Ginkgo biloba (Ginkgoaceae) populations in the Dalou Mountains, southwestern China. Am. J. Bot. 99, 1408–1414 (2012).
12. Hohmann, N. et al. Ginkgo biloba’s footprint of dynamic Pleistocene history dates back only 390,000 years ago. BMC Genom. 19, 299 (2018).
13. Zhao, Y. P. et al. Incongruent range dynamics between co-occurring Asian temperate tree species facilitated by life history traits. Ecol. Evol. 6, 2346–2358 (2016).
14. Gong, W. et al. Glacial refugia of Ginkgo biloba and human impact on its genetic diversity: evidence from chloroplast DNA. J. Integr. Plant Biol. 50, 368–374 (2008).
15. He, S. A., Yin, G. & Pang, Z. J. Resources and Prospects of Ginkgo biloba in China. (Springer Japan, 2014).
16. Guan, R. et al. Draft genome of the living fossil Ginkgo biloba. Gigascience 5, 49 (2016).
17. Guan, R. et al. Updated draft genome assembly of Ginkgo biloba. Gigascience Database. https://doi.org/10.5524/100613 (2019).
Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 (2009).

Lepage-Pajot, J., Arist-Sun, F.-M., Sun, H. Q., Xiao, T. S. & Ge, S. Centres of plant endemism in China: places for survival or for speciation? J. Biogeogr. 38, 1267–1280 (2011).

Lu, L. M. et al. Evolutionary history of the angiosperm flora of China. Nature 554, 234–238 (2018).

Qu, X. F., Gu, C. X., Coves, H. F. Plant molecular phylogeny in China and adjacent regions: tracing the genetic imprint of Quaternary climate and environmental change in the world’s most diverse temperate flora. Mol. Phylogenetics Evol. 59, 225–244 (2011).

Wang, H. W. & Ge, S. Phylogeny of the endangered Cathaya argyrophylla (Pinaceae) inferred from sequence variation of mitochondrial and nuclear DNA. Mol. Ecol. 15, 4109–4122 (2006).

Ma, Q. et al. Phylogeny of Davidaia involuta (Davidaeae) inferred from cpDNA haplotypes and nSSR data. Syst. Bot. 40, 796–810 (2015).

Li, Y. Y., Tsang, E. F. K., Cui, M. Y. & Chen, X. Y. Too early to call it success: an evaluation of the natural regeneration of the endangered Metasequoia glyptostroboides. Biol. Conserv. 150, 1–4 (2012).

Qian, H. & Bickel, R. F. Large-scale processes and the Asian bias in species diversity of temperate plants. Nature 407, 180 (2000).

Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. Nat. Genet. 46, 919 (2014).

Maher, B. A. Palaeoecological records of the loess/palaeosol sequences of the Chinese Loess Plateau. Quat. Sci. Rev. 19, 253–264 (2000).

Li, H. & Durbin, R. Inferece of human population history from individual whole-genome sequences. Nature 475, 493 (2011).

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. PLoS Genet. 9, e1003905 (2013).

Weber, S. J., Anderson, R. P. & Schipper, R. E. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259 (2006).

Pavlidis, P., Živkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. 30, 2224–2234 (2013).

Qi, X. S. et al. Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relic Cercidiphyllum (Cercidiphyllaceae). New Phytol. 196, 617–630 (2012).

Tang, C. Q. et al. Population structure of relic Metasequoia glyptostroboides and its habitat fragmentation and degradation in south-central China. Biol. Conserv. 144, 279–289 (2011).

Combosch, D. J., Lemer, S., Ward, P. D., Landman, N. H. & Gribert, G. Genomic signatures of evolution in Nautilia—an endangered living fossil. Mol. Ecol. 26, 5923–5938 (2017).

Royer, D. L., Hickey, L. J. & Wing, S. L. Ecological conservatism in the “living fossil” Ginkgo. Paleobiology 29, 84–104 (2003).

Zhao, S. et al. Whole-genome sequencing of giant pandas provides insights into genomic variation in collared pikas (Ochotona collaris). Mol. Ecol. 24, 6868–7305 (2015).

Jing, W., Street, N. R., Scofield, D. G. & Ingvarsson, P. K. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. Mol. Biol. Evol. 33, 1754–1767 (2016).

Petersen, B. J. & Graves, W. R. Chloroplast phylogeography of Draca palustris L. indicates populations near the glacial boundary at the Last Glacial Maximum in eastern North America. J. Biogeogr. 43, 314–327 (2016).

Call, A. et al. Genetic structure and post-glacial expansion of Cornus florida L. (Cornaceae): Integrative evidence from phylogeography, population demographic history, and species distribution modeling. J. Ecol. 54, 136–151 (2016).

Rubin, C. J. et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587 (2010).

Nielsen, R. et al. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566 (2005).

Pavlidis, P., Živkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. 30, 2224 (2013).

Alexa, A. et al. Smoothing improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22, 1600 (2006).

Acknowledgements
This work was financially supported by National Key Research and Development Program of China (Nos. 2016YFE0122000; 2017YFA0603104), the National Natural Science Foundation of China (Nos. 31870190; 31461123001; J1310002) and the grant from the Chinese Academy of Sciences (No. XDB31000000; XDB31020000; CAS/SAFEA International Partnership Program for Creative Research Teams). We thank Wen-Bin Zhou, Yu-Lu Ru, Nian Tong, Lu-Xian Liu, Min-Min Chen, Yu-Bin Yang, Jing Shang, Dong-Ya Wu from Zhejiang University and You-Chi Watanabe from Chiba University for their invaluable assistance in the sample collections. We also thank Yating Qin, Qian Liu, Longqi Liu for their technical help in Hi-C experiment and data analysis. The authors are grateful to Peter Crane for his insightful comments and suggestions.

Author contributions
S.G., X.X., C.-X.F. and X.L. conceived the project. S.G., G.F., Y.-P.Z., S.M-Y.L., J.W., H.Y. and R.Z. performed analyses of population genetics, phylogeography and nuclear DNA. X.X., C.-X.F. and X.L. conceived the project. S.G., G.F., Y.-P.Z., S.M-Y.L., J.W., H.Y. and R.Z. contributed to sample collection. S.S., P.-Y.Y., N.L., G.H., X.H., Y.-H., F.-M.Z., J.-D.H., H.Z., J.C., W. and R.Z. performed analyses of population genetics, phylogeography and nuclear DNA. Partial funding was provided by the National Natural Science Foundation of China (Nos. 31870190; 31461123001; J1310002) and the grant from the Chinese Academy of Sciences (No. XDB31000000; XDB31020000; CAS/SAFEA International Partnership Program for Creative Research Teams). We thank Wen-Bin Zhou, Yu-Lu Ru, Nian Tong, Lu-Xian Liu, Min-Min Chen, Yu-Bin Yang, Jing Shang, Dong-Ya Wu from Zhejiang University and You-Chi Watanabe from Chiba University for their invaluable assistance in the sample collections. We also thank Yating Qin, Qian Liu, Longqi Liu for their technical help in Hi-C experiment and data analysis. The authors are grateful to Peter Crane for his insightful comments and suggestions.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-12133-5.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/ reprintsandpermissions/

Peer review information Nature Communications thanks Marcus Koch, Alexander Werth and Qingyi Yu for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
