

# Evolutionary history of the angiosperm flora of China

Li-Min Lu<sup>1\*</sup>, Ling-Feng Mao<sup>2\*</sup>, Tuo Yang<sup>1\*</sup>, Jian-Fei Ye<sup>1,3,4\*</sup>, Bing Liu<sup>1,5\*</sup>, Hong-Lei Li<sup>6,7\*</sup>, Miao Sun<sup>8,9\*</sup>, Joseph T. Miller<sup>10,11</sup>, Sarah Mathews<sup>10</sup>, Hai-Hua Hu<sup>1,3</sup>, Yan-Ting Niu<sup>1,3</sup>, Dan-Xiao Peng<sup>1,3</sup>, You-Hua Chen<sup>12</sup>, Stephen A. Smith<sup>13</sup>, Min Chen<sup>14</sup>, Kun-Li Xiang<sup>1,3</sup>, Chi-Toan Le<sup>1,3</sup>, Viet-Cuong Dang<sup>1,3</sup>, An-Ming Lu<sup>1</sup>, Pamela S. Soltis<sup>9§</sup>, Douglas E. Soltis<sup>8,9§</sup>, Jian-Hua Li<sup>15§</sup> & Zhi-Duan Chen<sup>1,5§</sup>

**High species diversity may result from recent rapid speciation in a ‘cradle’ and/or the gradual accumulation and preservation of species over time in a ‘museum’<sup>1,2</sup>. China harbours nearly 10% of angiosperm species worldwide and has long been considered as both a museum, owing to the presence of many species with hypothesized ancient origins<sup>3,4</sup>, and a cradle, as many lineages have originated as recent topographic changes and climatic shifts—such as the formation of the Qinghai–Tibetan Plateau and the development of the monsoon—provided new habitats that promoted remarkable radiation<sup>5</sup>. However, no detailed phylogenetic study has addressed when and how the major components of the Chinese angiosperm flora assembled to form the present-day vegetation. Here we investigate the spatio-temporal divergence patterns of the Chinese flora using a dated phylogeny of 92% of the angiosperm genera for the region, a nearly complete species-level tree comprising 26,978 species and detailed spatial distribution data. We found that 66% of the angiosperm genera in China did not originate until early in the Miocene epoch (23 million years ago (Mya)). The flora of eastern China bears a signature of older divergence (mean divergence times of 22.04–25.39 Mya), phylogenetic overdispersion (spatial co-occurrence of distant relatives) and higher phylogenetic diversity. In western China, the flora shows more recent divergence (mean divergence times of 15.29–18.86 Mya), pronounced phylogenetic clustering (co-occurrence of close relatives) and lower phylogenetic diversity. Analyses of species-level phylogenetic diversity using simulated branch lengths yielded results similar to genus-level patterns. Our analyses indicate that eastern China represents a floristic museum, and western China an evolutionary cradle, for herbaceous genera; eastern China has served as both a museum and a cradle for woody genera. These results identify areas of high species richness and phylogenetic diversity, and provide a foundation on which to build conservation efforts in China.**

Species composition within a geographic area is the result of historical processes including speciation, extinction, migration<sup>6</sup> and ongoing ecological interactions. The extent to which each process has contributed to spatial and temporal patterns of biodiversity, as well as community assembly, varies across the landscape. The biodiversity patterns within a region may result from a recent increase in the rate of speciation that has generated a cradle of biodiversity. Alternatively, biodiversity may derive from the presence of numerous surviving

ancient lineages, together forming a museum region. The process of speciation and the maintenance of ancient lineages need not be mutually exclusive, and some regions have features of both cradles and museums.

The evolutionary history of regional floras has typically been addressed using specific taxa as exemplars<sup>7–9</sup> or by examining the entire flora at various taxonomic levels<sup>10–12</sup>. These investigations provide insights into historical factors, including geological history, climatic shifts and evolutionary processes, that might have contributed to modern geospatial patterns of biodiversity<sup>13,14</sup>. Concomitantly, these studies lay the foundation for decision-making in conserving biodiversity. However, few studies have explored the biodiversity patterns of a large region incorporating dated phylogenies and detailed distribution data.

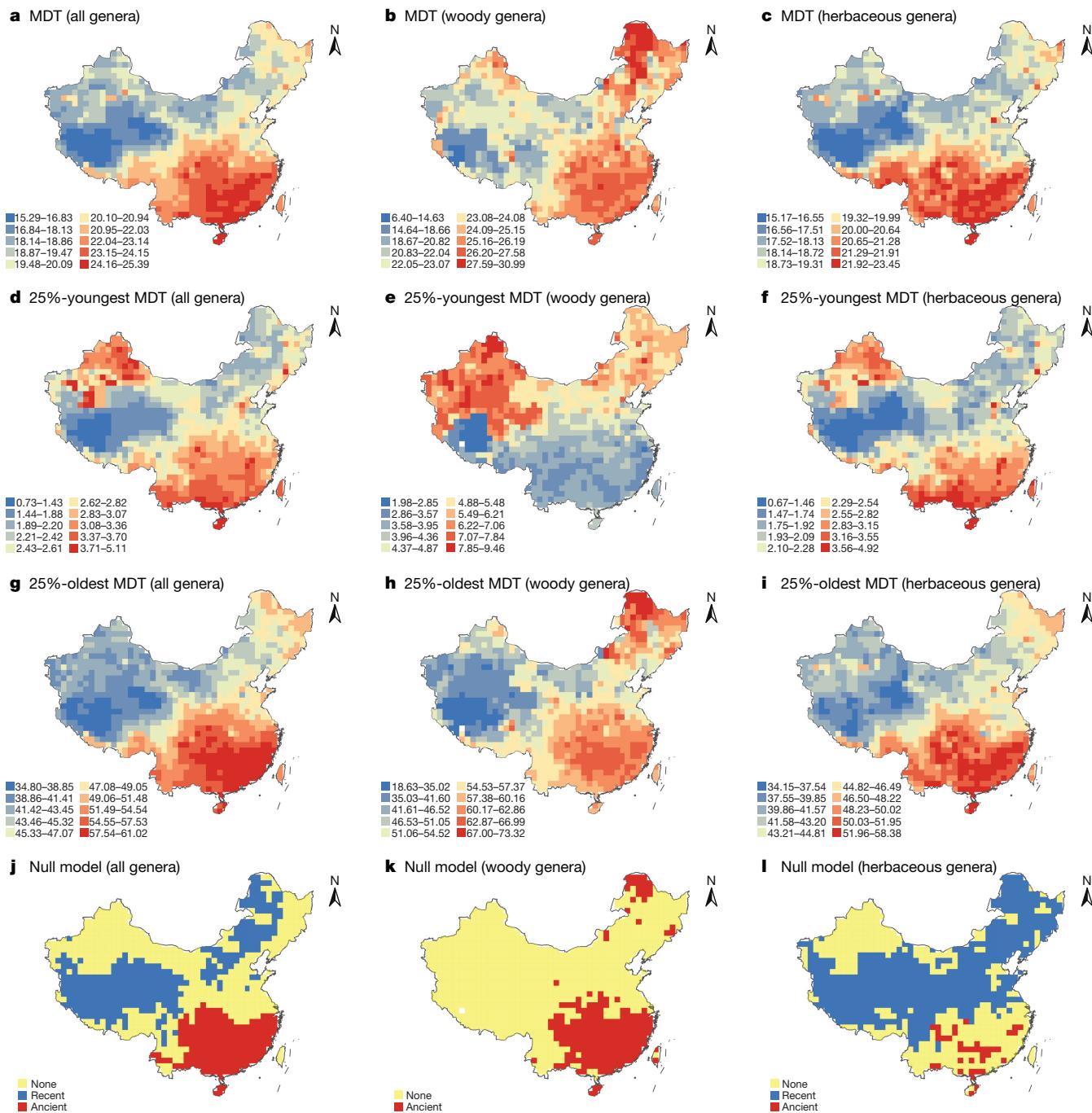
China, which is home to 30,000 of the approximately 350,000–400,000 species of vascular plants<sup>15</sup>, is ideal for investigating patterns of biodiversity because of its large size, range of habitats, considerable biological diversity and heterogeneous physical geography. Whether areas within China serve as cradles or museums remains unclear, as floristic components of putative ancient origin<sup>3,4</sup> and of recent diversification<sup>5</sup> have both been discovered. It has previously been suggested<sup>16</sup> on the basis of comparisons between the taxonomic richness of vascular plants in China and the United States, that the greater species diversity in China reflects the region’s complex topography and long connections with tropical South-East Asia. On the basis of patterns in species richness (using 555 endemic seed plant species), mountainous regions of central and southern China have been identified as the main centres of plant endemism<sup>17</sup>. Previous studies have attributed most of the geographic variation in species richness of woody plants in China to temperature seasonality<sup>18</sup> and the extent of winter cold<sup>19</sup>. Notably, to our knowledge, no previous study has incorporated both phylogenetic and spatial components to address the evolutionary history of the Chinese flora.

We conducted a broad assessment of spatio-temporal divergence patterns and of the assembly of the Chinese angiosperm flora, using a robustly dated phylogeny as well as species distribution data (i) to document the relative proportions of ancient and recent divergences that shaped the extant Chinese angiosperm flora in various geographic regions; (ii) to investigate the differential spatio-temporal divergence patterns of woody and herbaceous genera and their relationships with

<sup>1</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>2</sup>Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China. <sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>4</sup>Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>5</sup>Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China. <sup>6</sup>Chongqing Key Laboratory of Economic Plant Biotechnology/Institute of Special Plants, Chongqing University of Arts and Sciences, Yongchuan 402160, China. <sup>7</sup>Fairylake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, China. <sup>8</sup>Department of Biology, University of Florida, Gainesville, Florida 32611–7800, USA. <sup>9</sup>Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA. <sup>10</sup>CSIRO National Research Collections, Australian National Herbarium, Canberra, Australian Capital Territory 2601, Australia. <sup>11</sup>Office of International Science and Engineering, National Science Foundation, Alexandria, Virginia 22314, USA. <sup>12</sup>Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China. <sup>13</sup>Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>14</sup>Institute of Botany, Jiangsu Province & Chinese Academy of Sciences, Nanjing 210014, China. <sup>15</sup>Biology Department, Hope College, Holland, Michigan 49423, USA.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



**Figure 1 | Patterns of the MDTs for Chinese angiosperm genera.** **a–i**, MDT for all genera, woody genera and herbaceous genera (from left to right), based on all sampled genera (**a–c**), the youngest 25% of genera (**d–f**), and the oldest 25% of genera (**g–i**) in each grid cell. **j–l**, Null-model test to recognize recent (blue grid cells) and ancient (red grid cells) divergence centres. The analyses included 2,592 angiosperm

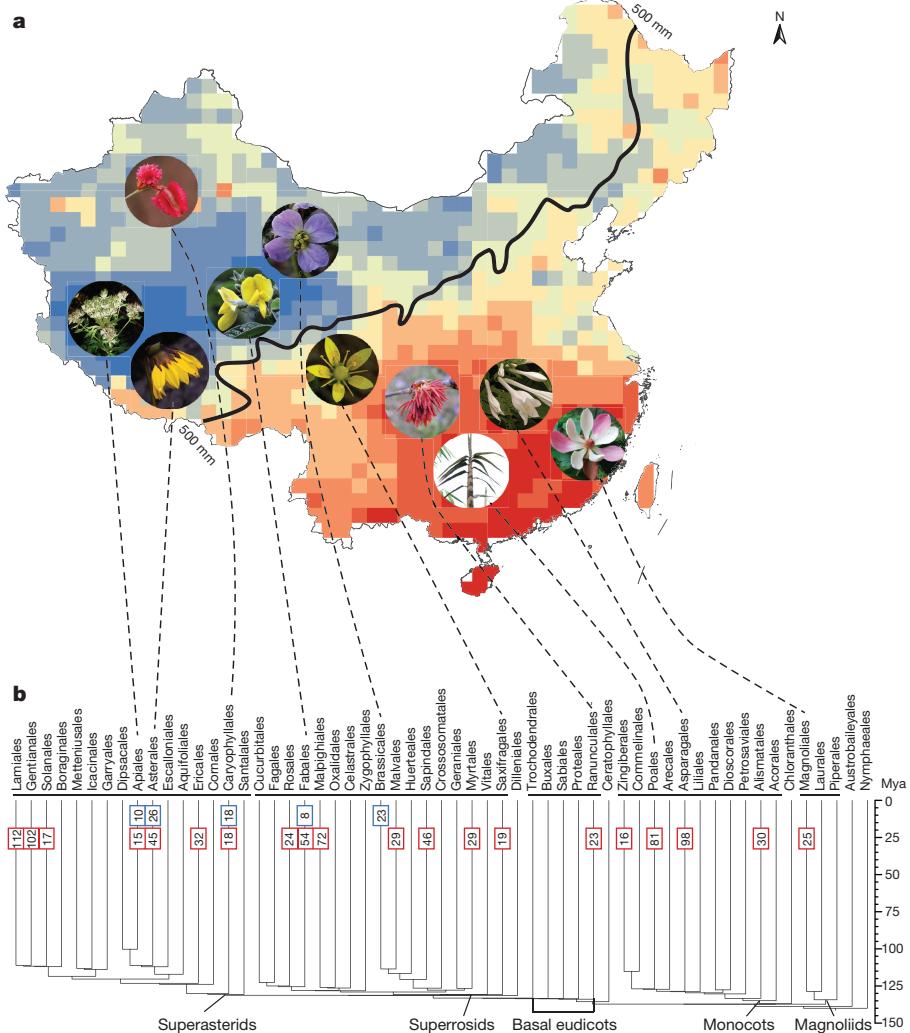
genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ). Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

environmental variables; and (iii) to compare genus- and species-level measures of phylogenetic diversity and explore their conservation implications for the Chinese flora.

Our phylogeny resolved evolutionary relationships among all major angiosperm lineages in China (Extended Data Fig. 1), yielding topologies that are highly similar to those for angiosperms as a whole<sup>20,21</sup>. Our estimates of divergence times based on penalized likelihood and PATHd8 are congruent with one another, and agree with those obtained in recent studies of angiosperms on a global basis<sup>22,23</sup> (Extended Data Fig. 2). Divergence time estimates show that 66% of Chinese

angiosperm genera originated during the Neogene and Quaternary periods; the remaining genera diverged in the Palaeogene (29%) and Cretaceous (5%) periods. Additionally, the herbaceous genera have diversified much more rapidly than the woody genera during the past 30 million years (Extended Data Fig. 3).

We divided China into 100-km  $\times$  100-km grid cells, evaluated age variance within grid cells (Extended Data Figs 4, 5), and calculated mean divergence times (MDTs) and median divergence times of genera within each grid cell (Fig. 1; Extended Data Figs 6, 7; Supplementary Information). Mapping the MDTs of all genera revealed a transition belt

**Figure 2 | Spatio-temporal divergence patterns of the Chinese angiosperm flora.**

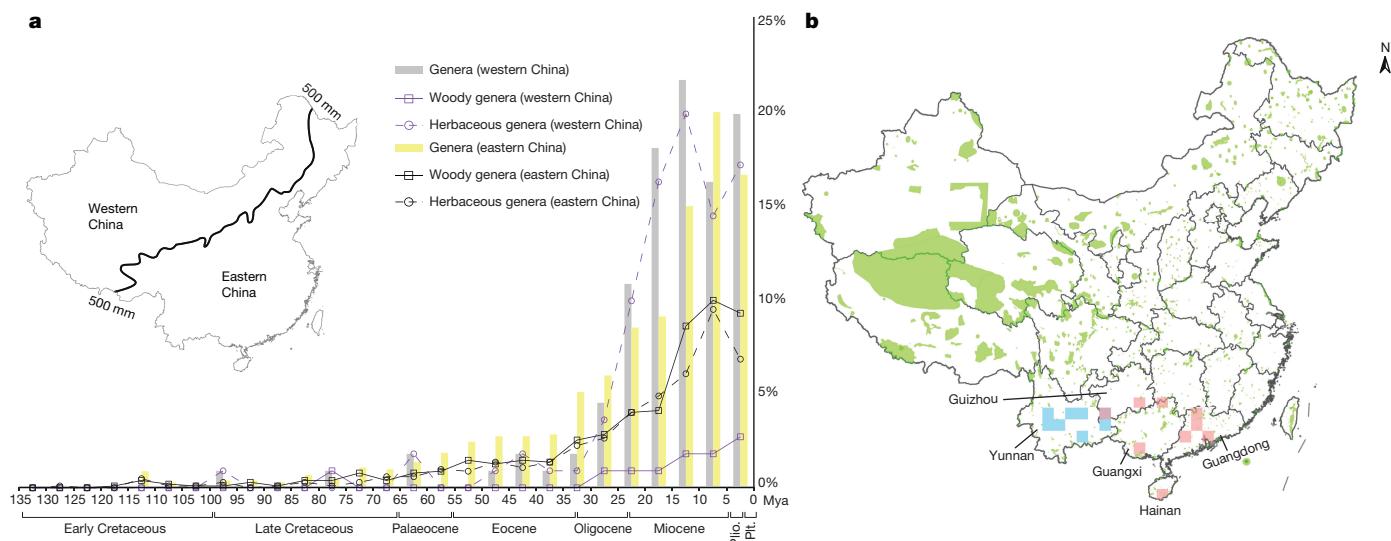
**a**, Patterns of MDTs adapted from Fig. 1a. The dark line represents the 500-mm annual precipitation isoline adapted from ref. 24 (reprinted from ref. 24, with permission from Elsevier). **b**, Ordinal time-tree with the major clades of angiosperms indicated. The top five orders with genera occurring only in western China and top 20 orders with genera occurring only in eastern China are indicated on the tree with blue and red boxes, respectively. The number of genera distributed in western or eastern China from each order is shown within the corresponding box. Map adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

that coincides with the modern 500-mm isoline of annual precipitation, which marks the boundary between humid–semi-humid and arid–semi-arid areas<sup>24</sup> (eastern China versus western China, Fig. 2). Both MDT and null-model analyses indicate that eastern China has older lineages (red grid cells, Fig. 1a, j), particularly in central to southern China. By contrast, western China, and especially the Qinghai–Tibetan Plateau, contains taxa that have diverged more recently (blue grid cells, Fig. 1a, j). Furthermore, our genus-level analyses demonstrate that eastern China is phylogenetically overdispersed with higher phylogenetic diversity, and that western China shows phylogenetic clustering with lower phylogenetic diversity (Extended Data Fig. 8). These findings are also observed in analyses of phylogenetic diversity based on multiple species-level trees, in which taxa that lacked target DNA sequences were provided with meaningful branch lengths using a birth–death clock model (see Methods; Extended Data Fig. 9). The flora of the Cape of South Africa likewise shows phylogenetic structure—the western region is phylogenetically clustered, and the eastern region is overdispersed<sup>10</sup>. However, taxon richness is decoupled from phylogenetic diversity in the Cape of South Africa; in China, taxon richness and phylogenetic diversity are positively correlated.

Western China includes the arid north-western portion of the country and most of the Qinghai–Tibetan Plateau (Fig. 2). A fundamental climate shift may have occurred in western China as recently as the early Miocene, owing to the uplift of the Qinghai–Tibetan Plateau and subsequent development of the Asian monsoon<sup>24,25</sup>. Of the 111 genera that occur only in western China, 76% originated in the past 20 million years and only 24% originated before this time. In western China, a much higher percentage of herbaceous than woody

genera has originated since 30 Mya (Fig. 3a). Moreover, genera that occur only in western China are predominantly members of only a few clades (Apiales, Asterales and Brassicales), most of which have much younger divergence times than the major clades of eastern China (Fig. 2; Extended Data Table 1). MDTs calculated from the youngest 25% of herbaceous genera in each grid cell also indicate that western China—in particular the Qinghai–Tibetan Plateau—has younger lineages (Fig. 1f) than eastern China, which further suggests that western China represents a cradle for herbaceous angiosperms.

Mountainous areas of eastern China have been proposed as refugia for plants that originated in the early Cretaceous or late Jurassic periods<sup>26,27</sup> because their geological environment and climate (including orogenic movements, annual temperature and annual precipitation) may have experienced little change since the Cretaceous<sup>28</sup>. Of the 1,026 genera that occur only in eastern China, 39% originated before 20 Mya and 61% arose more recently than this. Both herbaceous and woody genera diverged at similar rates throughout geological time (Fig. 3a). The 20 major clades with the largest number of genera occurring only in eastern China are distributed throughout the ordinal-level time-tree from early-diverging clades (for example, Alismatales, Asparagales, Magnoliidae and Ranunculales) to later-diverging lineages (for example, Asterales, Gentianales and Lamiaceae) (Fig. 2; Extended Data Table 1). MDTs based on the youngest 25% and oldest 25% of genera in each grid cell reveal that eastern China has old herbaceous lineages (Fig. 1f, i), but has both old and young woody lineages (Fig. 1e, h). Eastern China may have served as a museum for herbaceous genera, but as both a museum and a cradle for woody genera.



**Figure 3 | Angiosperm divergence pattern and conservation priorities in western and eastern China.** **a**, Percentage of genera occurring only in western ( $n = 111$ ) or eastern China ( $n = 1,026$ ) during geological time. Western China has a higher percentage of herbaceous genera (purple dashed line) than woody genera (purple solid line) that have originated since 20 Mya. Western and eastern China are divided by the 500-mm

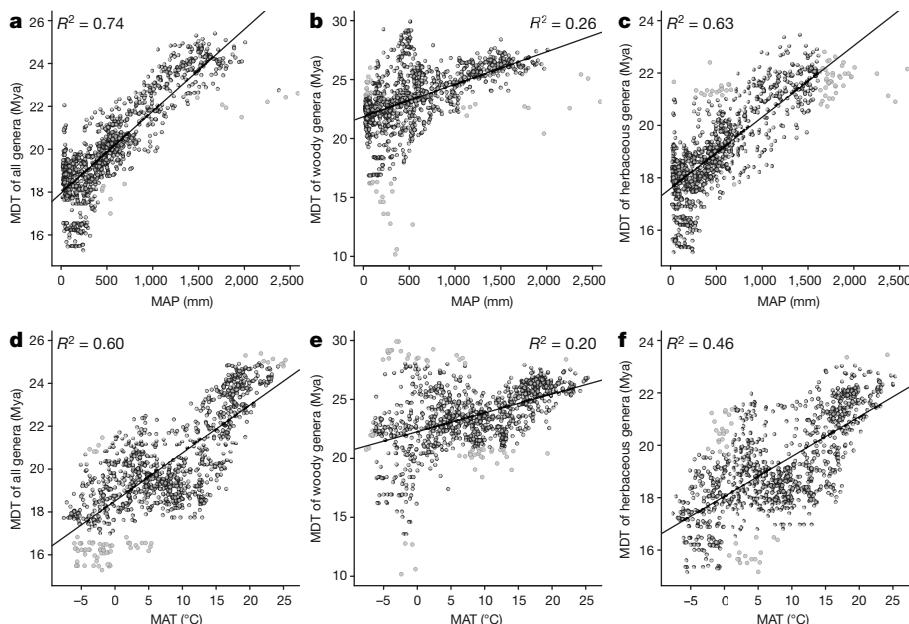
isoline of annual precipitation. Plio., Pliocene epoch; Plt., Pleistocene epoch. **b**, Grid cells with the top 5% highest phylogenetic diversity and SES-PD at genus (pink) and species (blue) levels. Protected areas are highlighted in green. Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

The mean annual precipitation (MAP) and mean annual temperature (MAT) have higher explanatory power for the MDTs of the herbaceous genera (Fig. 4c, f) than of the woody genera (Fig. 4b, e). These patterns may reflect the heterogeneity in rates of evolution between herbaceous and woody lineages. Herbaceous plants are well known to have higher substitution rates owing to their shorter generation times, which perhaps allows them to respond more quickly to environmental change through increased genetic divergence and speciation rates<sup>23,29</sup>.

The spatial divergence and diversity patterns of angiosperms detected here do not precisely reflect the latitudinal gradient in China; MDT and phylogenetic diversity decrease from south-east to north-west (Fig. 2a; Extended Data Fig. 8d, g). Our results show the importance of water and temperature in limiting the dispersal of species from humid and warm regions to drier and colder areas. The effects of topography, with a pronounced altitudinal gradient increasing from east to west, and the monsoon climate in eastern Asia are so extensive that the

decreasing temperature and precipitation gradients from south-eastern to north-western China are not consistent with the latitudinal gradient, as might be expected in flatter regions.

On the basis of a species-level phylogenetic tree and distribution data with 'county' as the basic unit, we inferred that the species richness and phylogenetic diversity in protected areas cover approximately 88% and 96%, respectively, of the total species richness and phylogenetic diversity in China. For conservation planning, these values may be overestimates that result from the coarse scale of our distributional data, as most nature reserves are smaller in size than Chinese counties. Notably, areas with the top 5% highest phylogenetic diversity and standard effective size of phylogenetic diversity (SES-PD) are mainly located in several provinces of eastern China (Fig. 3b): Guangdong, Guangxi, Guizhou and Hainan for genus-level phylogenetic diversity hotspots, and Yunnan for species-level phylogenetic diversity. These areas are also hotspots for threatened plants in China<sup>30</sup>. However, in contrast



to western China, protected areas in eastern China are fragmented (Fig. 3b), largely as a result of urbanization and administrative division. Our data suggest the need to establish more connections between existing nature reserves and national parks that span provincial borders to conserve plant lineages of ancient and recent origins in eastern China, as well as the other organisms that depend on these floristic elements. These findings should be of broad interest to evolutionary and conservation biologists, and serve to stimulate better-informed conservation planning and research.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 January; accepted 22 December 2017.

Published online 31 January 2018.

1. McKenna, D. D. & Farrell, B. D. Tropical forests are both evolutionary cradles and museums of leaf beetle diversity. *Proc. Natl Acad. Sci. USA* **103**, 10947–10951 (2006).
2. Moreau, C. S. & Bell, C. D. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution* **67**, 2240–2257 (2013).
3. Blackmore, S., Hong, D.-Y., Raven, P. H. & Wortley, A. H. in *Plants of China: A Companion to the Flora of China* (eds Hong, D.-Y. & Blackmore, S.) 1–6 (Science Press, 2013).
4. Sun, G., Dilcher, D. L., Zheng, S.-L. & Zhou, Z.-K. In search of the first flower: a Jurassic angiosperm, *Archaeofructus*, from northeast China. *Science* **282**, 1692–1695 (1998).
5. Wen, J., Zhang, J.-Q., Nie, Z.-L., Zhong, Y. & Sun, H. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front. Genet.* **5**, 4 (2014).
6. Xing, Y.-W. & Ree, R. H. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc. Natl Acad. Sci. USA* **114**, E3444–E3451 (2017).
7. Klak, C., Reeves, G. & Hedderson, T. Unmatched tempo of evolution in Southern African semi-desert ice plants. *Nature* **427**, 63–65 (2004).
8. Mishler, B. D. et al. Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian *Acacia*. *Nat. Commun.* **5**, 4473 (2014).
9. Richardson, J. E., Pennington, R. T., Pennington, T. D. & Hollingsworth, P. M. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**, 2242–2245 (2001).
10. Forest, F. et al. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757–760 (2007).
11. Verboom, G. A. et al. Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both? *Mol. Phylogenet. Evol.* **51**, 44–53 (2009).
12. Thornhill, A. H. et al. Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *J. Biogeogr.* **43**, 2085–2098 (2016).
13. Donoghue, M. J. A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl Acad. Sci. USA* **105**, 11549–11555 (2008).
14. Svenning, J.-C., Eiserhardt, W. L., Normand, S., Ordonez, A. & Sandel, B. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **46**, 551–572 (2015).
15. Wu, Z.-Y., Raven, P. H. & Hong, D.-Y. (eds) *Flora of China*, Vol. 1–25 (Science Press & Missouri Botanical Garden Press, 1994–2013).
16. Qian, H. & Ricklefs, R. E. A comparison of the taxonomic richness of vascular plants in China and the United States. *Am. Nat.* **154**, 160–181 (1999).
17. López-Pujol, J., Zhang, F.-M., Sun, H.-Q., Ying, T.-S. & Ge, S. Centres of plant endemism in China: places for survival or for speciation? *J. Biogeogr.* **38**, 1267–1280 (2011).
18. Qian, H. Environmental determinants of woody plant diversity at a regional scale in China. *PLoS ONE* **8**, e75832 (2013).
19. Wang, Z.-H., Fang, J.-Y., Tang, Z.-Y. & Lin, X. Patterns, determinants and models of woody plant diversity in China. *Proc. R. Soc. Lond. B* **278**, 2122–2132 (2011).
20. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).
21. Soltis, D. E. et al. Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* **98**, 704–730 (2011).
22. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437–453 (2015).
23. Zanne, A. E. et al. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
24. Sun, X.-J. & Wang, P.-X. How old is the Asian monsoon system?—Palaeobotanical records from China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **222**, 181–222 (2005).
25. Favre, A. et al. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol. Rev. Camb. Philos. Soc.* **90**, 236–253 (2015).
26. Axelrod, D. I., Al-Shehbaz, I. A. & Raven, P. H. in *Floristic Characteristics and Diversity of East Asian Plants* (eds Zhang, A.-L. & Wu, S.-G.) 43–55 (China Higher Education, 1996).
27. Benton, M. J. *The Fossil Record 2* (Chapman & Hall, 1993).
28. Wu, Z.-Y., Sun, H., Zhou, Z.-K., Li, D.-Z. & Peng, H. *Floristics of Seed Plants from China* (Science Press, 2010).
29. Smith, S. A. & Beaulieu, J. M. Life history influences rates of climatic niche evolution in flowering plants. *Proc. R. Soc. Lond. B* **276**, 4345–4352 (2009).
30. Zhang, Z.-J., He, J.-S., Li, J.-S. & Tang, Z.-Y. Distribution and conservation of threatened plants in China. *Biol. Conserv.* **192**, 454–460 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J.-Y. Fang, D.-Z. Li, K.-P. Ma, S.-Z. Zhang, H. Sun, J.-Q. Liu, Z.-H. Wang, X.-Q. Wang and H.-Z. Kong for help initiating this study. This research was supported by the National Key Basic Research Program of China (2014CB954100), the National Natural Science Foundation of China (31590822), the Chinese Academy of Sciences International Institution Development Program (SAJC201613), the National Natural Science Foundation of China and US National Science Foundation Dimensions Collaboration Project (31461123001), the US National Science Foundation (Open Tree of Life: DEB-1207915, DEB-1208428; ABI DBI-1458466 and DBI-1458640; iDigBio: EF-1115210 and DBI-1547229; US-China Dimensions of Biodiversity: DEB-1442280) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

**Author Contributions** Z.-D.C., P.S.S., D.E.S and J.-H.L. conceived the paper. L.-M.L., L.-F.M., T.Y., J.-F.Y., B.L., H.-L.L. and M.S. analysed the data. L.-M.L., L.-F.M., T.Y., J.-F.Y., B.L., J.T.M., S.M., P.S.S., D.E.S., J.-H.L. and Z.-D.C. wrote the first draft and finalized the manuscript. H.-H.H., Y.-T.N., D.-X.P., M.C., K.-L.X., C.-T.L. and V.-C.D. contributed data. J.T.M., A.-M.L., Y.-H.C., S.A.S., P.S.S., D.E.S., J.-H.L. and Z.-D.C. contributed substantially to revisions. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to Z.-D.C. ([zhiduan@ibcas.ac.cn](mailto:zhiduan@ibcas.ac.cn)).

**Reviewer Information** *Nature* thanks R. Colwell, V. Savolainen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Phylogeny reconstruction.** Sequences of four plastid genes (*atpB*, *matK*, *ndhF* and *rbcL*) and one mitochondrial gene (*matR*) were used to reconstruct the phylogeny of Chinese vascular plants<sup>31</sup>. Generic circumscriptions were based on ref. 15. We sampled one species for the 1,173 genera with only one species in China. For the 1,736 genera with 2–30 species in China, two species were sampled from each genus. For the 267 genera with more than 30 species in China, approximately 10% of the species of each genus were sampled, reflecting intrageneric diversity. We downloaded all available sequences for the target DNA regions from GenBank; if more than one sequence was available for the same locus for a species, the longest one of good quality was selected. For genera with sequences that were unavailable in the public database (781 genera in total), we generated new sequences from leaf materials, collected from the field for 513 genera and from specimens from the Chinese National Herbarium (PE) for 47 genera. There are 231 genera that remain unavailable because we failed to obtain the materials or amplify the target sequences. Details of DNA extraction, PCR, sequencing, alignment, accession numbers of sequences and phylogeny reconstruction have previously been published<sup>31</sup>.

**Divergence time estimation.** We used the penalized likelihood method as implemented in treePL<sup>32</sup> (<https://github.com/blackrim/treePL>) to date divergence times of Chinese angiosperms based on the optimal maximum likelihood phylogram obtained with RAxML 8.0.22<sup>33</sup> in the CIPRES Science Gateway<sup>34</sup>, after excluding the outgroups (for example, lycophytes, ferns and gymnosperms). Our dated phylogeny included 5,864 species native to China, representing 2,665 genera from 273 families or approximately 92% of the angiosperm genera of China. We validated the available fossils and selected 138 calibrations for dating analyses (Supplementary Table 1 in Supplementary Information). The ‘prime’ option was applied to identify the best optimization parameters, and a ‘thorough’ analysis was then carried out with the optimal parameters determined above (opt = 1, optad = 1 and optcvad = 4). To identify the best smoothing parameter that affects the penalty for rate variation over the phylogram, a ‘random subsample and replicate cross-validation’ was conducted with treePL. Confidence intervals for each node were calculated following previously published methods<sup>22</sup>. To accommodate for variation in branch length estimates, we calculated 100 bootstrap replicates with topology fixed to the above maximum likelihood phylogram but with varying branch lengths. We then conducted treePL on these 100 replicates. Age statistics for all nodes were summarized with TreeAnnotator v.1.8.4<sup>35</sup>.

We also used an alternative dating method, PATHd8<sup>36</sup>, to estimate divergence times of Chinese angiosperms. The calibrations for the PATHd8 analysis were identical to those used for the treePL analysis, except that the crown age of angiosperms was set to 138 Mya instead of a maximum of 140 Mya and a minimum of 136 Mya (as in treePL) because PATHd8 requires one fixed calibration. Both treePL and PATHd8 are rate-smoothing methods, but PATHd8 sequentially takes averages over path lengths from an internode to all its descending terminals, one pair of sister groups at a time<sup>37</sup>, where smoothing is done stepwise for each node separately; by contrast, smoothing in treePL is done simultaneously over the tree. The correlation between ages at all nodes based on the treePL and PATHd8 analyses was assessed with Spearman’s rank correlation analysis in R v.3.2.0<sup>38</sup>.

To evaluate whether dates for this regional time-tree are biased owing to the geographic sampling, we used a correlation analysis to compare our estimated divergence times with recent global-scale angiosperm time-tree reconstructions<sup>22,23</sup>; one of these represents a family-level time-tree with multiple fossil calibration points<sup>22</sup>, and the other is a species-level time-tree with dense taxon sampling (32,223 species) and fewer calibrations<sup>23</sup>. The stem age of each family was extracted for the Spearman’s rank correlation analyses. Only the family ages were compared (circumscription of families, following ref. 20), because different genera and species were included in the three studies. Ages of genera were extracted from our dichotomous time-tree estimated by treePL for the downstream analyses. For monophyletic genera, stem ages were extracted directly by tracing their stem node. For genera that are polyphyletic or paraphyletic (380 out of 2,665), the stem age of each monophyletic lineage was extracted and the oldest one was selected as the age of the genus. The numbers of angiosperm genera that originated during specified geological timespans are provided in Extended Data Fig. 3, with the global temperature changes since 65 Mya<sup>39</sup> indicated.

**Distribution of angiosperm species in China.** The spatial distribution data and information on growth form were assembled from nearly all published national and provincial floras, as well as some local floras, checklists and herbarium records. The spatial distribution data are at the county level (2,377 counties) with an average county-size of approximately 4,000 km<sup>2</sup>. To minimize the sampling bias of unequal sampling areas, we divided the map of China into 100-km × 100-km grid cells, and grid cells on the border that cover less than 50% of the area of a grid cell (that is, 5,000 km<sup>2</sup>) were excluded from the analyses. Maps of China used in this study were

adapted from standard maps released by the National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576). The gridded distribution database contained 1,409,239 occurrence records for 26,978 angiosperm species from 2,845 genera. After matching with the phylogeny, the final dataset included a total of 2,592 angiosperm genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ).

### Spatial distribution of MDTs and null-model test for divergence hotspots.

To explore the spatial divergence patterns of Chinese angiosperm genera, we calculated the weighted MDTs of all genera in each grid cell by integrating spatial distribution data with our dated phylogenetic tree. AGE<sub>i</sub> represented the age of a genus  $i$  ( $i = 1, \dots, n$ ) in a grid cell, and S<sub>i</sub> the species number in genus  $i$  in this grid cell. From this, MDT was calculated as:

$$\text{MDT} = \frac{(\text{AGE}_1 \times S_1) + (\text{AGE}_2 \times S_2) + (\text{AGE}_3 \times S_3) + \dots + (\text{AGE}_n \times S_n)}{S_1 + S_2 + S_3 + \dots + S_n}$$

We further divided the genus dates in each grid cell into quartiles and calculated MDTs on the basis of the youngest and oldest quartiles, separately, in each grid cell. The MDTs based on the youngest quartile allowed us to recognize centres of recent divergence, whereas MDTs based on the oldest quartile detected ancient centres of divergence. To avoid potential bias from grid cells that had either relatively old or young genera, we ranked all genera from youngest to oldest, partitioned them into quartiles based on their ages, computed MDT in each cell for the absolute youngest 25% and the absolute oldest 25% of genera, and then mapped the results across China.

We designed a null model to identify ancient and recent divergence hotspots for the angiosperm flora of China. The mean ages of the youngest and oldest quartiles in each grid cell were selected as the observed values for the null models, and then we shifted the genera randomly using all genera investigated in China as a sampling pool to obtain the null distributions of ages for the youngest and oldest quartiles for each grid cell. The standardized effect size of the mean divergence time (SES-MDT) of genera for each grid cell was calculated as:

$$\text{SES-MDT} = \frac{\text{MDT}_{\text{observed}} - \text{MDT}_{\text{random}}}{\text{s.d.}(\text{MDT}_{\text{random}})}$$

where MDT<sub>observed</sub> is the observed MDT; MDT<sub>random</sub> is the expected MDT of the randomized assemblages ( $n = 999$ ); and s.d.(MDT<sub>random</sub>) is the s.d. of the MDT for the randomized assemblages. Grid cells with values of SES-MDT for the youngest quartile below −1.96 were identified as notable hotspots of recent divergence, whereas grid cells with SES-MDT for the oldest quartile above 1.96 were identified as notable hotspots of ancient divergence. Considering that the evolutionary history of herbaceous and woody plants may differ<sup>40</sup>, the above analyses were also conducted separately for herbaceous and woody genera. Analyses of MDT were implemented in R and ArcGIS 10.1 (<http://www.esri.com/>).

Previous studies have demonstrated that the overall species richness patterns of birds are largely determined by the geographically wide-ranging species<sup>41–43</sup>, indicating that patterns may be driven by a subset of taxa and may not be representative of an entire biota. To explore whether MDT patterns for China are influenced largely by values for widespread species, we ranked genera from the most restricted to most widespread in China, partitioned the genera into quartiles on the basis of their range size and mapped MDT for each quartile following a previously published description<sup>41</sup>.

**Spatial distribution of median divergence times.** Age variation within grid cells was evaluated by plotting divergence times in each grid cell (Extended Data Fig. 4) and calculating the skewness and kurtosis of divergence times (Extended Data Fig. 5). To verify the results of MDT, we also investigated the distribution patterns of the Chinese angiosperm genera by mapping the median divergence times (medianDT) based on all genera, and the youngest and oldest quartiles in each grid cell. The null model for the median divergence time applied a modified effective-size statistic<sup>44–46</sup> and was calculated as:

$$\text{SES-medianDT} = \frac{\text{medianDT}_{\text{observed}} - \text{medianDT}_{\text{random}}}{1.4826 \times \text{MAD}_{\text{random}}}, \text{ if } \text{MAD}_{\text{random}} > 0$$

$$\text{SES-medianDT} = \frac{\text{medianDT}_{\text{observed}} - \text{medianDT}_{\text{random}}}{1.2553 \times \text{meanAD}_{\text{random}}}, \text{ if } \text{MAD}_{\text{random}} = 0$$

where medianDT<sub>observed</sub> is the observed median divergence time; medianDT<sub>random</sub> is the expected median divergence time of the randomized assemblages ( $n = 999$ ); MAD<sub>random</sub> is the median absolute deviation of the divergence times for the randomized assemblages; and meanAD<sub>random</sub> is the mean absolute deviation of the divergence times for the randomized assemblages.

**Richness and phylogenetic diversity.** We calculated the generic richness, Faith's phylogenetic diversity<sup>47</sup> and SES-PD of the Chinese angiosperm genera on the basis of our ultrametric chronogram using the 'picante' package in R. Faith's phylogenetic diversity is the sum of all phylogenetic branch lengths that connect species in a community. We calculated phylogenetic diversity as the length of the subtree that joins the genera in each grid cell to the root of the chronogram. SES-PD was calculated because phylogenetic diversity is usually positively correlated with species richness<sup>48</sup>. We first obtained a null distribution of the expected phylogenetic diversity values by shuffling taxa labels across the tips of the tree 999 times for each grid cell. SES-PD was then calculated by dividing the difference between the observed ( $\text{PD}_{\text{observed}}$ ) and expected phylogenetic diversity ( $\text{PD}_{\text{random}}$ ) by the s.d. of the null distribution ( $\text{s.d.}(\text{PD}_{\text{random}})$ ):

$$\text{SES-PD} = \frac{\text{PD}_{\text{observed}} - \text{PD}_{\text{random}}}{\text{s.d.}(\text{PD}_{\text{random}})}$$

**Phylogenetic structure.** The net relatedness index (NRI) and the nearest taxon index (NTI) were calculated to investigate the phylogenetic structure (clustering or overdispersion) of angiosperm genera across China<sup>49</sup>. NRI is based on the mean phylogenetic distance (MPD), an estimate of the average phylogenetic relatedness between all possible pairs of taxa within a grid cell, and primarily reflects structure at deeper parts of the phylogeny. NTI is based on mean nearest taxon distance (MNTD), an estimate of the mean phylogenetic relatedness between each pair of taxa in a grid cell and its nearest relative in the phylogeny, and reflects shallower parts of the phylogeny. NRI and NTI were calculated as follows:

$$\text{NRI} = -1 \times \frac{\text{MPD}_{\text{observed}} - \text{MPD}_{\text{random}}}{\text{s.d.}(\text{MPD}_{\text{random}})}$$

$$\text{NTI} = -1 \times \frac{\text{MNTD}_{\text{observed}} - \text{MNTD}_{\text{random}}}{\text{s.d.}(\text{MNTD}_{\text{random}})}$$

where  $\text{MPD}_{\text{observed}}$  and  $\text{MNTD}_{\text{observed}}$  are the observed MPD and MNTD;  $\text{MPD}_{\text{random}}$  and  $\text{MNTD}_{\text{random}}$  are the averages of the expected MPD and MNTD of the randomized assemblages ( $n=999$ ); and  $\text{s.d.}(\text{MPD}_{\text{random}})$  and  $\text{s.d.}(\text{MNTD}_{\text{random}})$  are the standard deviation of  $\text{MPD}_{\text{random}}$  and  $\text{MNTD}_{\text{random}}$  for the randomized assemblages. The null distributions of MPD and MNTD were created by randomly selecting the observed number of genera in each grid cell 999 times, with all genera in the phylogeny as a sampling pool. Positive values of NRI and NTI indicate phylogenetic clustering, whereas negative values indicate phylogenetic overdispersion in a grid cell. NRI and NTI for woody and herbaceous genera were calculated separately to compare their phylogenetic structures across China.

**Regression analyses between MDT and two climatic variables.** To explore the underlying mechanisms of spatial divergence patterns of the Chinese angiosperms, MDT in each grid cell was regressed against the respective mean values of MAP and MAT in each grid cell using the linear regression model in R. The adjusted  $R^2$  was used to indicate the explanatory power of each variable, although it is clear that these associations do not necessarily indicate causation of the climatic variables in determining MDT. Climatic data were downloaded from the WorldClim database Version 1.4 (<http://www.worldclim.org/>) with a spatial resolution of 10 min<sup>50</sup>.

**Species tree reconstruction and conservation implications.** With our dated genus-level chronogram as the backbone, a species-level tree including 26,978 Chinese angiosperm species was generated by inserting species that were not sampled in our generic tree within the genera to which they belong using the R package 'SPhyloMaker'<sup>51</sup>. Our species-level tree included approximately 96% of all known angiosperm species native to China; 1,098 aquatic species were not sampled. To mitigate the effect of polytomies on the calculation of phylogenetic diversity, we resolved polytomies in the reconstructed tree with a birth–death clock model<sup>52</sup>. We constructed constraints based on the tree constructed with molecular data, and unresolved taxa were then placed within the relevant constraints. Node heights for each constraint were fixed on the basis of divergence time estimates. We then conducted a Bayesian analysis using MrBayes v.3.2<sup>53</sup> with the topological and node height constraints and with the birth–death (speciation and extinction) priors as uniform (0.0, 10.0). Two analyses were run for 2,500,000 generations, sampling every 500 generations, to ensure convergence and mixing; the first 750,000 generations were discarded as burn-in, and 1,000 of the post-burn-in trees were retained for further analyses. The species-level phylogenetic diversity and SES-PD were calculated on the basis of 10 trees randomly selected from the 1,000 trees. The Spearman's rank correlation was used to assess the consistency of phylogenetic diversity or SES-PD patterns based on different trees. Grid cells with the top 5% highest values of both phylogenetic diversity and SES-PD were identified as hotspots of phylogenetic diversity (Fig. 3b). MDT analyses were not conducted on the species tree as the missing data rendered the variation between

replicates uninformative. Once additional molecular information is collected for these species, further analyses can be performed.

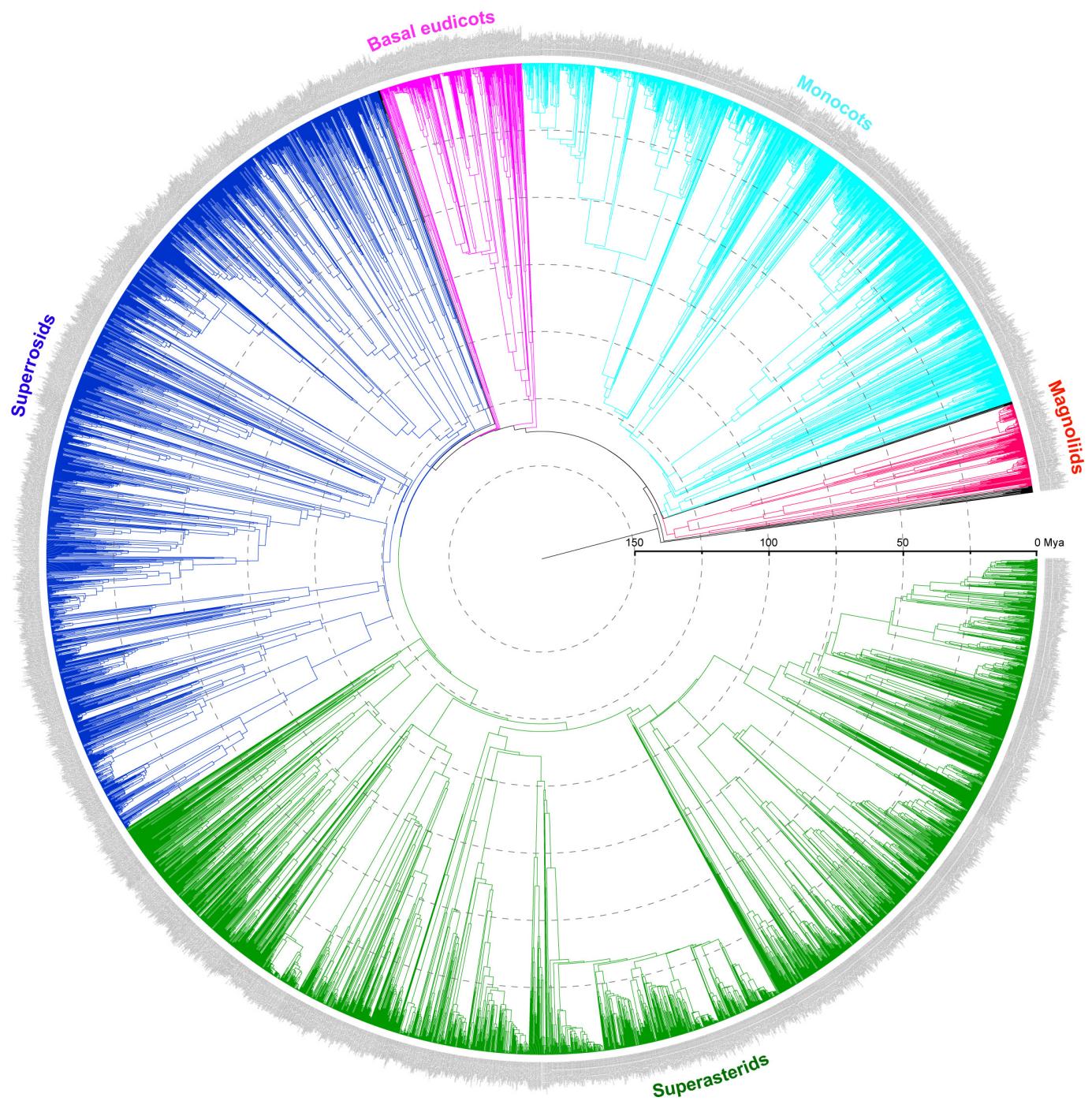
Spatial data of protected areas in China were compiled from two sources: (i) a previous publication<sup>30</sup> that digitized nature reserves in mainland China, which included 334 national, 857 provincial and 1,431 prefectural or county-level nature reserves (provided by Z.-Y. Tang); and (ii) 92 protected areas in Taiwan, downloaded from the Database of Protected Areas (<https://www.protectedplanet.net/>; accessed August 2017). Considering that most of the nature reserves were designed according to administrative units, we calculated richness and phylogenetic diversity in the protected areas with 'county' as the basic unit rather than dividing China into grid cells. Each conservation area was intersected with the map of China to produce the protected areas in ArcGIS. Species occurring in these counties are supposed to be protected, but counties with protected areas that covered less than 10% of the area of a county were excluded to reduce sampling bias.

**Statistics and reproducibility.** No statistical methods were used to predetermine sample size. Spearman's rank correlation and linear regression analyses were conducted in R. Precise  $P$  values are provided to show statistical significance. Null-model tests (999 random replicates) were used to assess the significance of spatial diversity and divergence distributions with  $-1.96$  and  $1.96$  as significant boundaries.

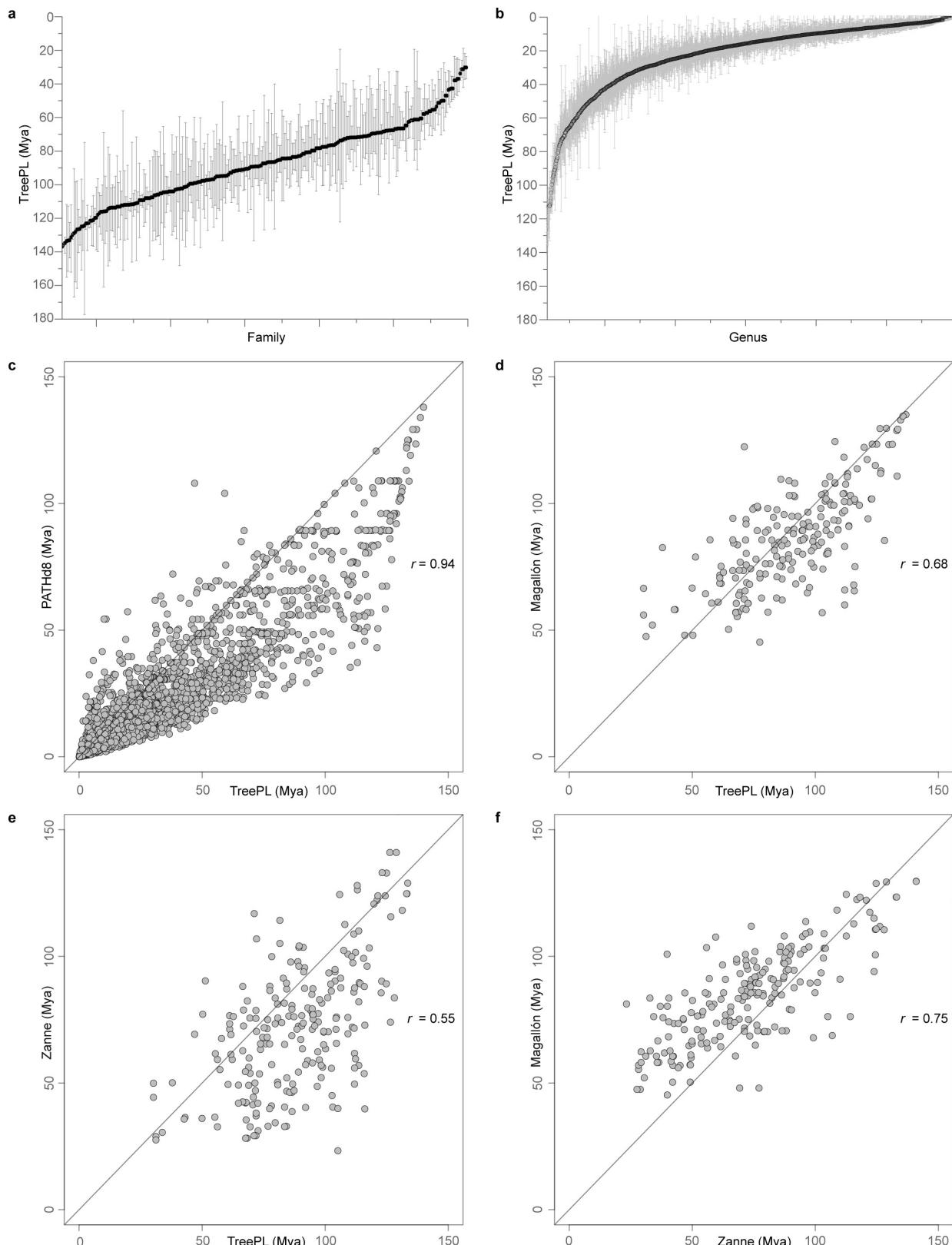
**Code availability.** Example code used to conduct null-model test (written in R) can be found at Dryad: <http://datadryad.org/resource/doi:10.5061/dryad.p89m3>.

**Data availability.** Sequences for phylogenetic analyses have previously been published<sup>31</sup> and deposited in GenBank. The dated phylogeny is archived in Dryad: <http://datadryad.org/resource/doi:10.5061/dryad.p89m3>. The spatial distribution data are available from: [http://www.darwintree.cn/resource/spatial\\_data](http://www.darwintree.cn/resource/spatial_data). All other additional data are available from the corresponding author upon reasonable request.

31. Chen, Z.-D. et al. Tree of life for the genera of Chinese vascular plants. *J. Syst. Evol.* **54**, 277–306 (2016).
32. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
33. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
34. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop, IEEE*. <http://ieeexplore.ieee.org/document/5676129/> (2010).
35. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
36. Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S. & Bremer, K. Estimating divergence times in large phylogenetic trees. *Syst. Biol.* **56**, 741–752 (2007).
37. Anderson, C. L. *Dating Divergence Times in Phylogenies*. PhD thesis, Uppsala Univ. (2007).
38. R Core Team. R: a language and environment for statistical computing. <http://R-project.org/> (2014).
39. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
40. Smith, S. A. & Donoghue, M. J. Rates of molecular evolution are linked to life history in flowering plants. *Science* **322**, 86–89 (2008).
41. Jetz, W. & Rahbek, C. Geographic range size and determinants of avian species richness. *Science* **297**, 1548–1551 (2002).
42. Rahbek, C. et al. Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proc. R. Soc. Lond. B* **274**, 165–174 (2007).
43. Lennon, J. J., Koleff, P., Greenwood, J. J. D. & Gaston, K. J. Contribution of rarity and commonness to patterns of species richness. *Ecol. Lett.* **7**, 81–87 (2004).
44. Iglewicz, B. & Hoaglin, D. *How to Detect and Handle Outliers* (ASQC Quality, 1993).
45. Rousseeuw, P. J. & Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **88**, 1273–1283 (1993).
46. Cleophas, T. J. Clinical trials: robust tests are wonderful for imperfect data. *Am. J. Ther.* **22**, e1–e5 (2015).
47. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
48. Rodrigues, A. S. L., Brooks, T. M. & Gaston, K. J. in *Phylogeny and Conservation* (eds Purvis, A., Gittleman, J. L., & Brooks, T.) 101–119 (Cambridge Univ. Press, 2005).
49. Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
50. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
51. Qian, H. & Jin, Y. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J. Plant Ecol.* **9**, 233–239 (2016).
52. Kuhn, T. S., Mooers, A. Ø. & Thomas, G. H. A simple polytomy resolver for dated phylogenies. *Methods Ecol. Evol.* **2**, 427–436 (2011).
53. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

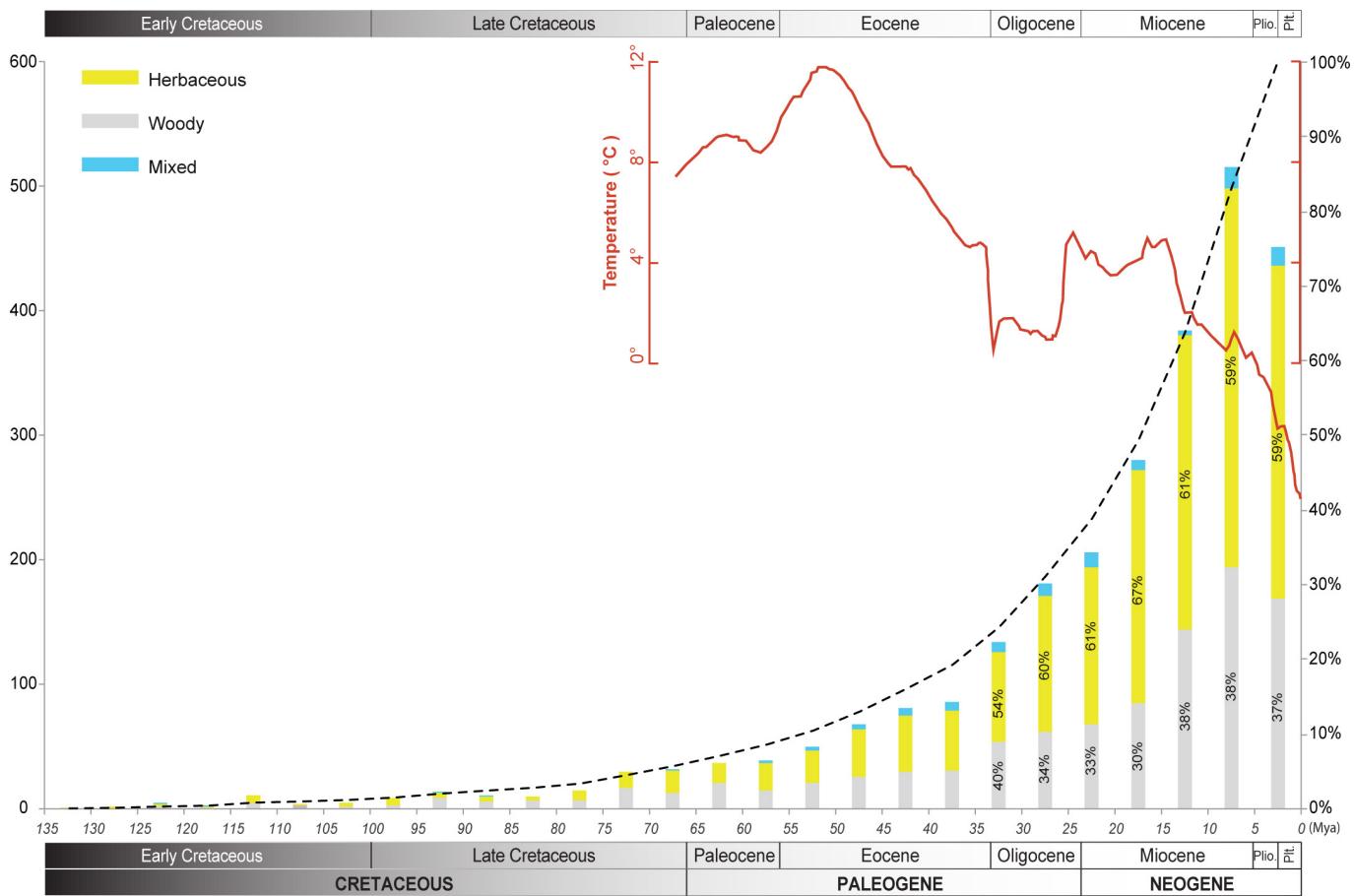


**Extended Data Figure 1 | Dated megaphylogeny of the Chinese angiosperms.** Major clades, including magnoliids, monocots, superrosids and superasterids, as well as the basal eudicot grade, are indicated with different colours. Divergence times were estimated using treePL.



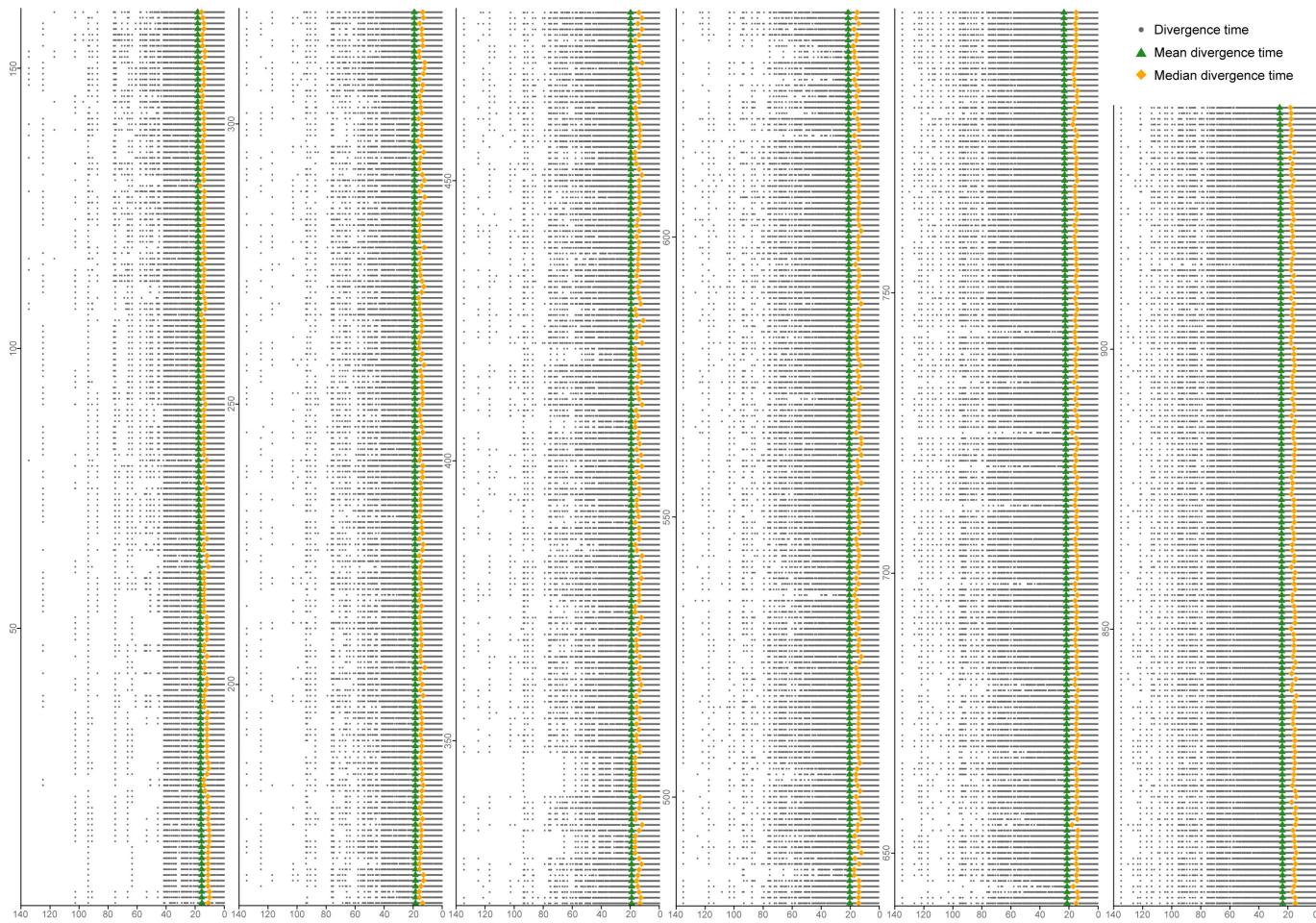
**Extended Data Figure 2 | The 95% confidence intervals of divergence times and the Spearman's rank correlation between our dating and those of recent publications.** **a, b**, Plots of divergence times and 95% confidence intervals (grey bars) for each family (**a**,  $n = 273$ ) and genus (**b**,  $n = 2,909$ ). The centre values are ages calculated based on the optimal maximum likelihood tree. **c**, Correlation of nodal ages between treePL and

PATHd8 in this study ( $n = 5,863$ ;  $r = 0.94$ ,  $P = 0$ ). **d**, Correlation of family ages between treePL and ref. 22 ( $n = 236$ ,  $r = 0.68$ ,  $P = 1.17 \times 10^{-33}$ ). **e**, Correlation of family ages between treePL and ref. 23 ( $n = 257$ ;  $r = 0.55$ ,  $P = 4.54 \times 10^{-22}$ ). **f**, Correlation of family ages between ref. 22 and ref. 23 ( $n = 235$ ;  $r = 0.75$ ,  $P = 2.11 \times 10^{-43}$ ). The solid line is  $y = x$ .

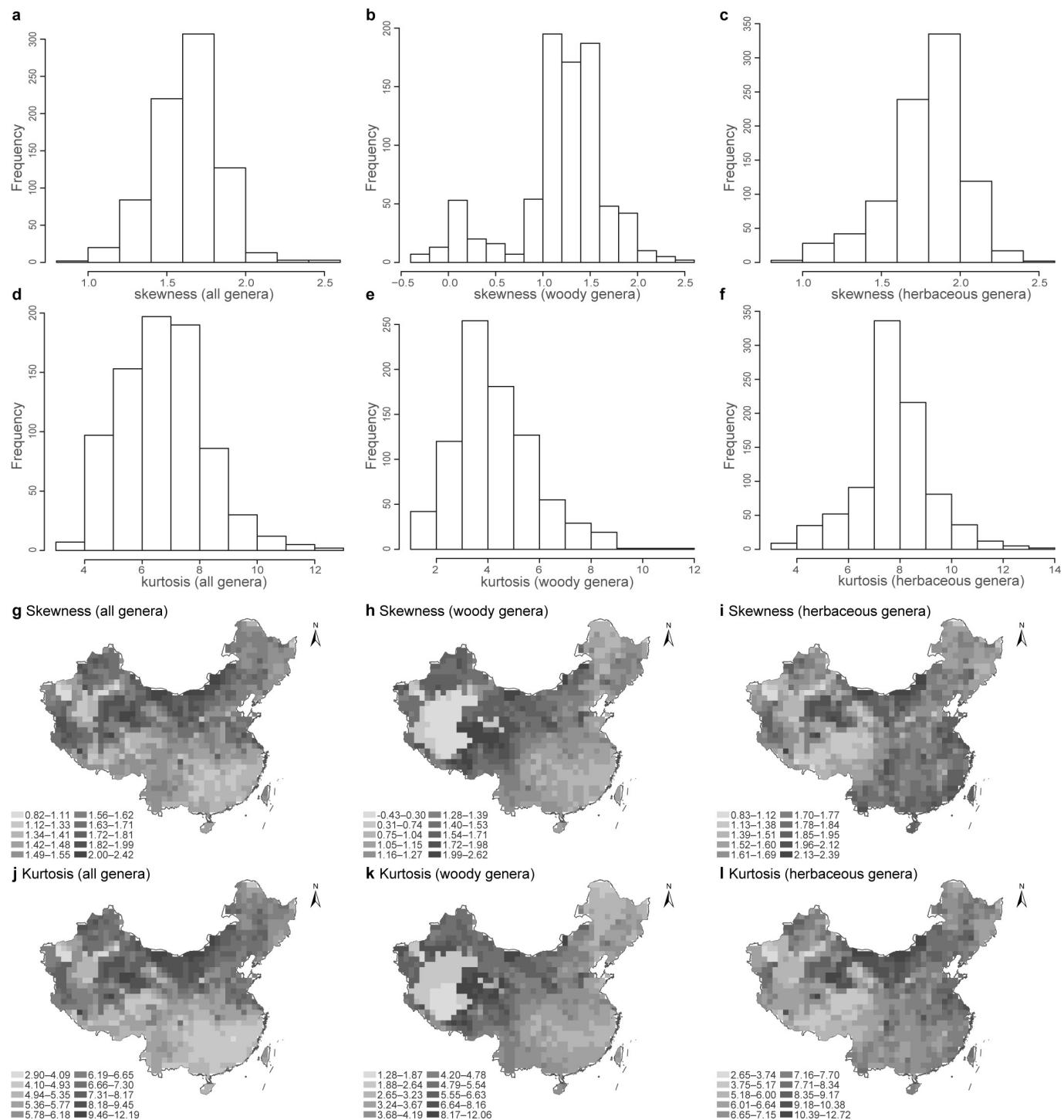


**Extended Data Figure 3 | Number of angiosperm genera that originated during specified geological timespans.** Column with three colours shows the number of woody (grey), herbaceous (yellow) and mixed genera (light blue) that originated within a specific geological timespan. Number of woody genera,  $n = 995$ ; number of herbaceous genera,  $n = 1,569$ ; mixed genera (genera with both woody and herbaceous species),  $n = 101$ . The dashed line indicates the accumulated percentage of genera that have

originated since the Early Cretaceous. Global temperature changes that have occurred since the Palaeogene are shown by the red curve (from ref. 39; reprinted with permission from AAAS). The x axis indicates the geological period and time in millions of years. The left y axis shows the total number of genera that have originated by any given time period; the right y-axis represents the accumulated percentage of genera that originated within a geological time period.

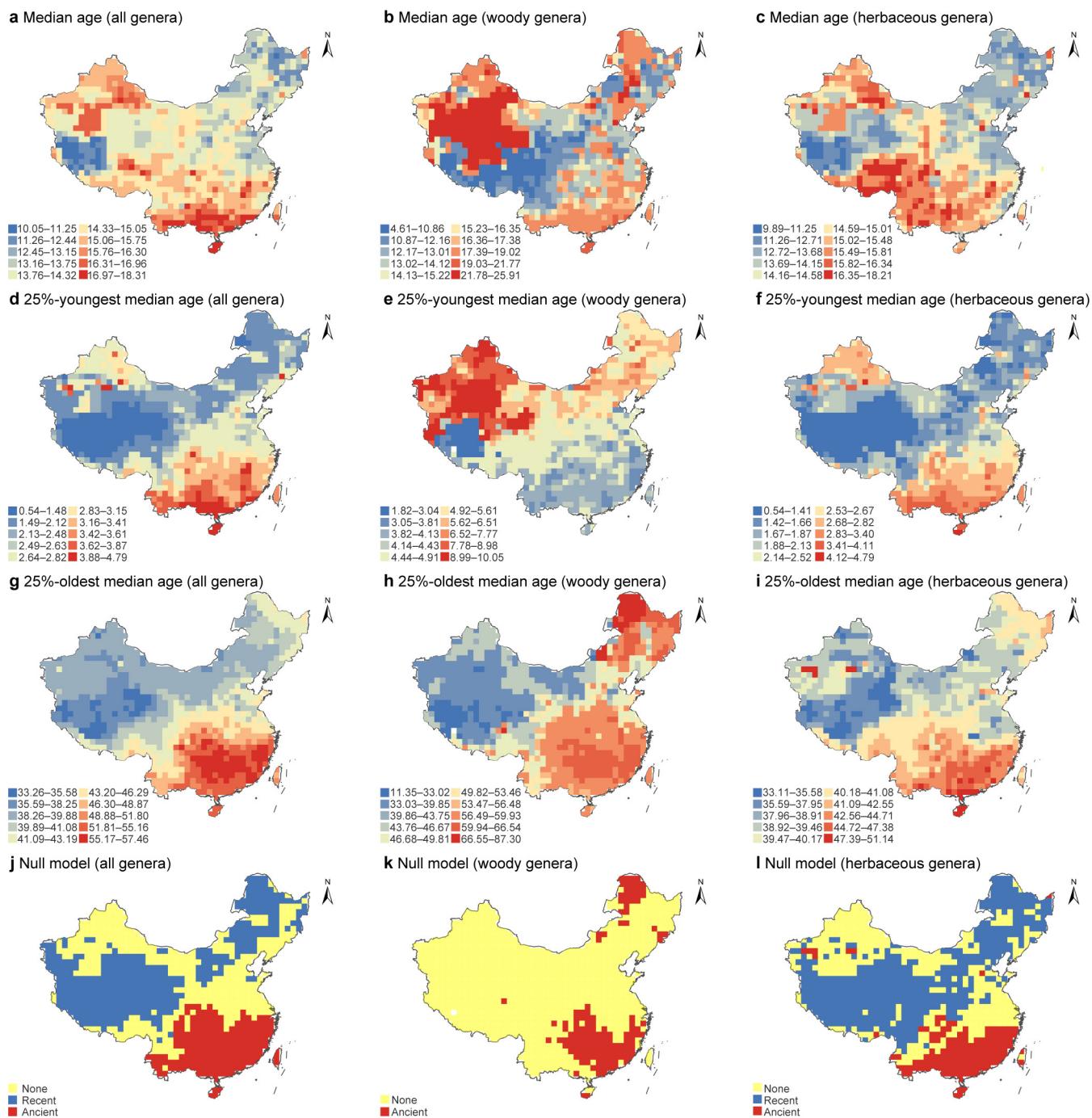


**Extended Data Figure 4 | Plot of divergence times of the Chinese angiosperm genera in each grid cell.** Mean and median values of the divergence times are indicated.



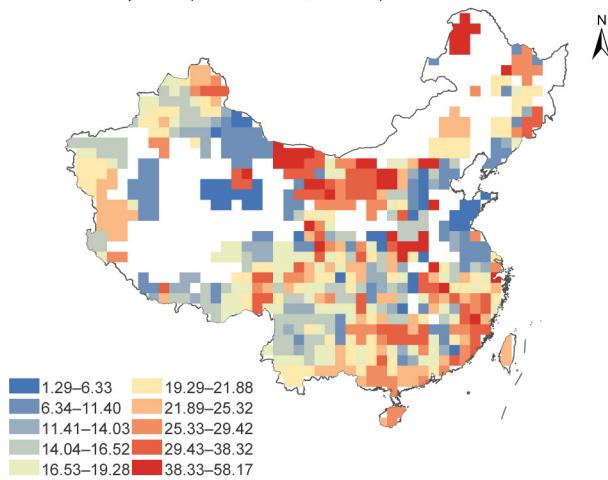
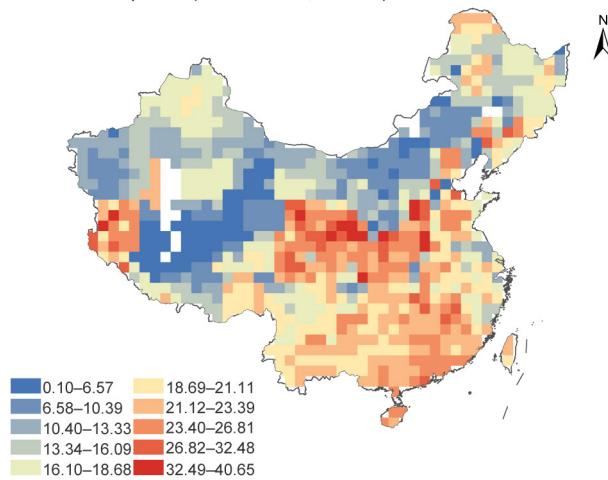
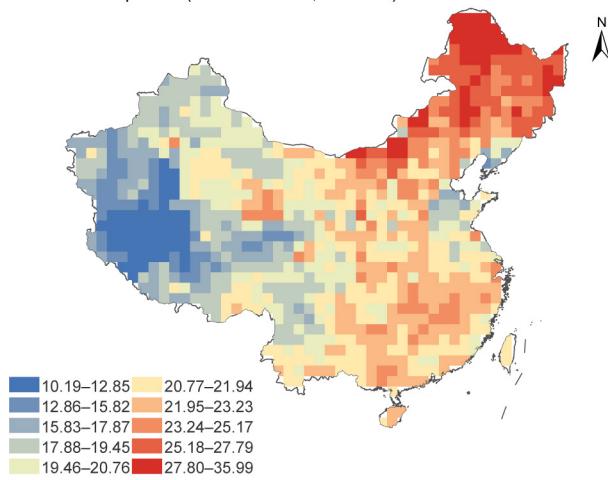
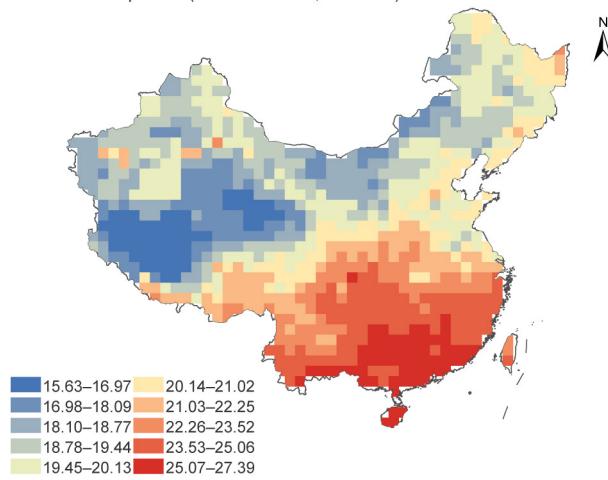
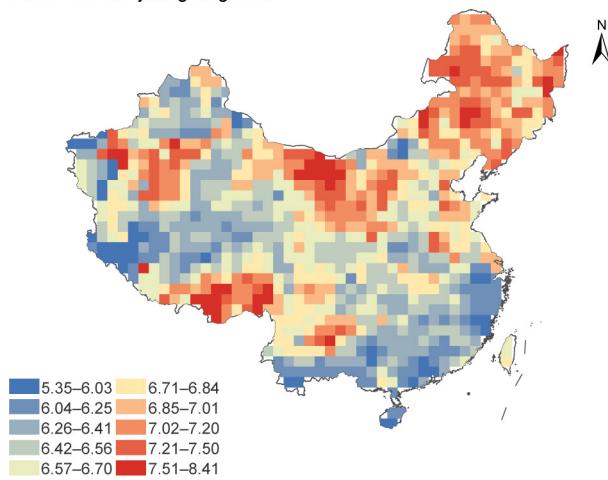
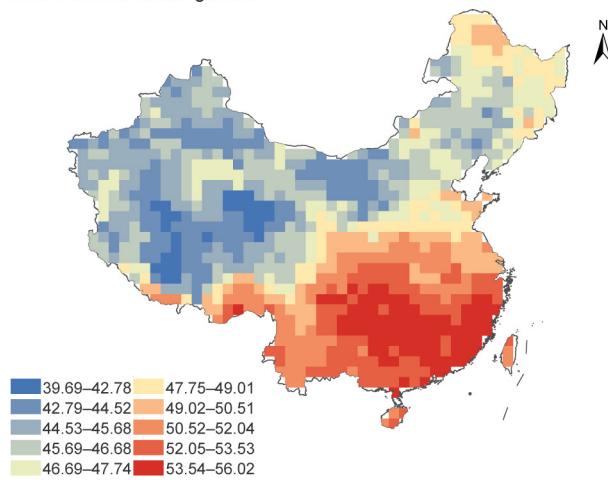
**Extended Data Figure 5 | Histograms and distribution of skewness and kurtosis for divergence times in each grid cell.** **a–c**, Range of skewness for all genera (**a**), woody genera (**b**) and herbaceous genera (**c**). **d–f**, Range of kurtosis (computed as the fourth standardized moment) for all genera (**d**), woody genera (**e**) and herbaceous genera (**f**). **g–i**, Spatial distribution of skewness for all genera (**g**), woody genera (**h**) and herbaceous genera (**i**). **j–l**, Spatial distribution of kurtosis for all genera (**j**), woody genera (**k**) and herbaceous genera (**l**). Skewness values in most grid cells are positive and

around 1–2, which implies that divergence times of genera are slightly right-skewed (there are more young ages in each grid cell). Kurtosis values in most grid cells are within a range of 4–8, larger than the value (3) for a normal distribution, which implies that the distribution of divergence times has more extreme outliers than the normal distribution. For eastern China, kurtosis values of approximately 4 for all genera are consistent with grid cells having a range of divergence times—including very young and very old ages—as expected for an area that is both a cradle and a museum.



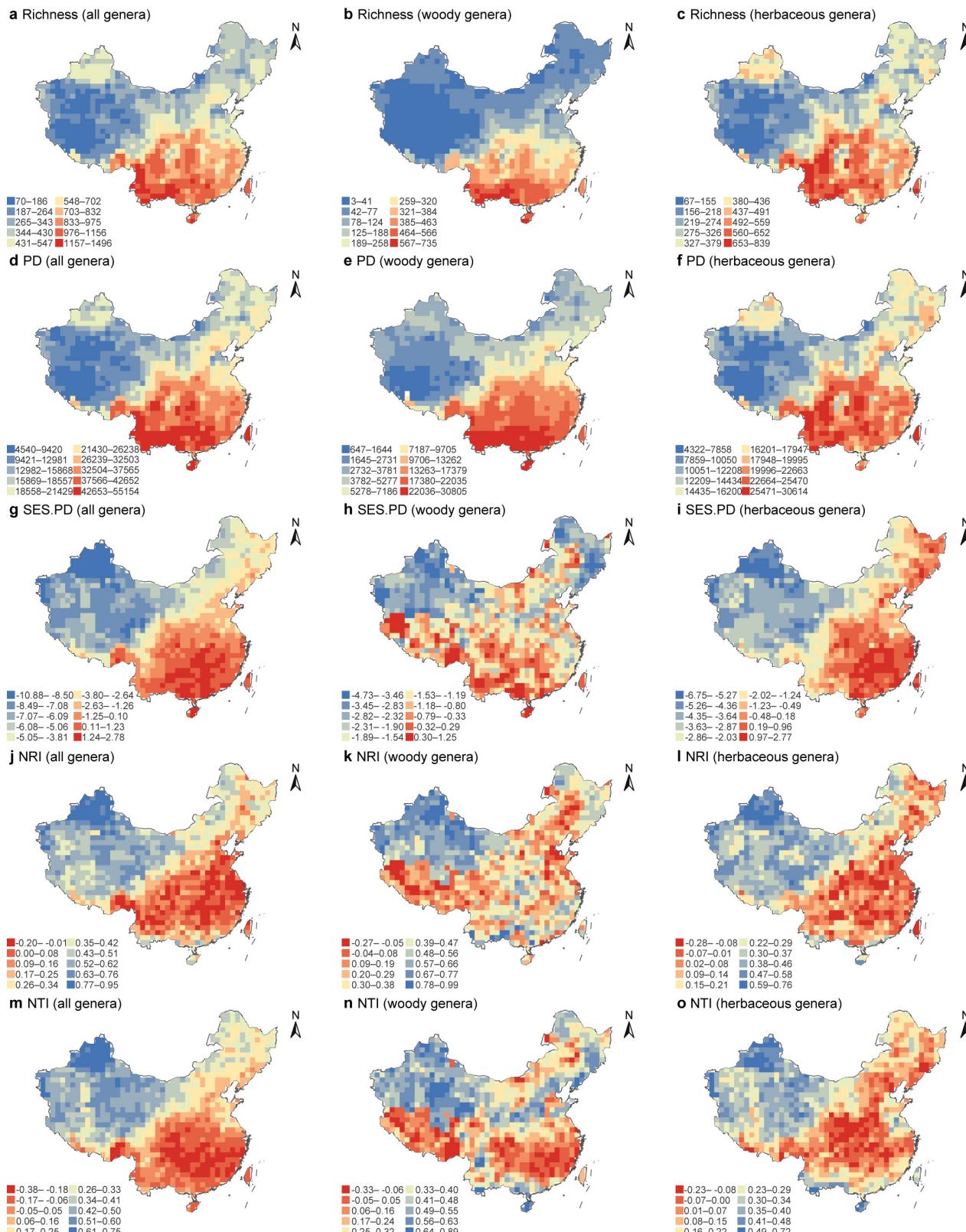
**Extended Data Figure 6 | Geographic patterns of median ages for the Chinese angiosperm genera.** **a–i**, Median ages for all genera, woody genera and herbaceous genera (from left to right), based on all sampled genera (**a–c**), the youngest 25% of genera (**d–f**), and the oldest 25% of genera (**g–i**) in each grid cell. **j–l**, Null-model test to identify recent (blue grid cells) and ancient (red grid cells) divergence centres for all genera (**j**),

woody genera (**k**) and herbaceous genera (**l**). The analyses include 2,592 angiosperm genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ). Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

**a** MDT of 1<sup>st</sup> quartile (0.8% records,  $r = 0.12$ )**b** MDT of 2<sup>nd</sup> quartile (3.9% records,  $r = 0.59$ )**c** MDT of 3<sup>rd</sup> quartile (14.3% records,  $r = 0.43$ )**d** MDT of 4<sup>th</sup> quartile (81.1% records,  $r = 0.99$ )**e** MDT of 25%-youngest genera**f** MDT of 25%-oldest genera

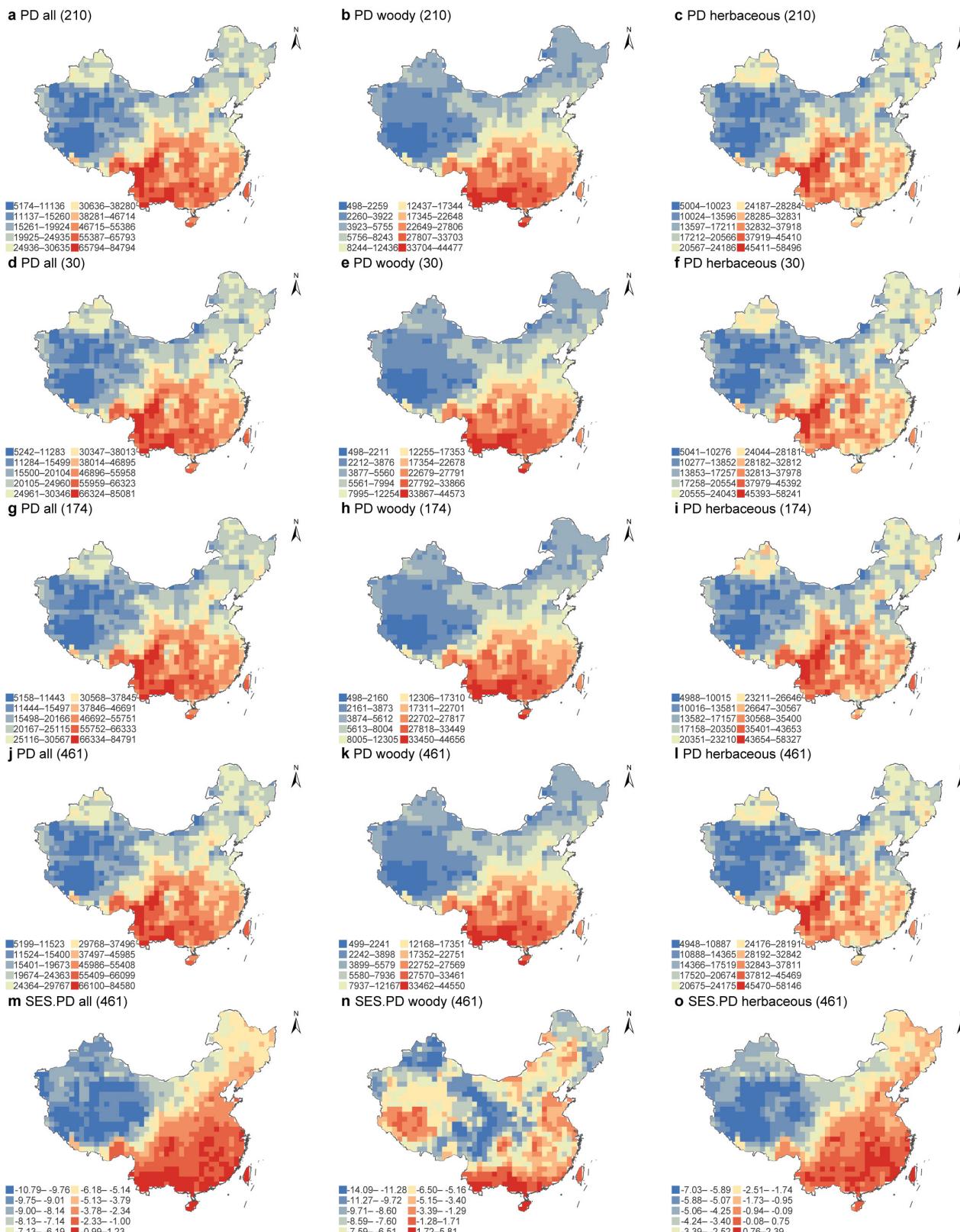
**Extended Data Figure 7 | Spatial distribution of MDTs based on geographic range-size quartiles and the youngest 25% and oldest 25% of genera in China.** **a-d**, MDT patterns of the first (a), second (b), third (c) and fourth quartiles (d) of the sampled Chinese angiosperm genera. The first, second, third and fourth quartiles range from the narrowest to the widest geographic distribution, and represent 0.6%, 3.5%, 13.7% and 82.1% of 1,409,239 records, respectively. The Spearman's rank correlation coefficients between the overall MDT (including all genera)

and MDT of the first, second, third and fourth geographic quartile are 0.12 ( $P=1.46 \times 10^{-3}$ ), 0.59 ( $P=1.21 \times 10^{-87}$ ), 0.43 ( $P=2.51 \times 10^{-43}$ ) and 0.99 ( $P=0$ ), respectively. **e**, MDT pattern of the youngest 25% of genera in China, showing that there are young genera in both western and eastern China. **f**, MDT pattern of the oldest 25% of genera in China, confirming that older genera mainly occur in eastern China. Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).



**Extended Data Figure 8 | Patterns of generic richness, phylogenetic diversity and phylogenetic structure for the Chinese angiosperm genera.** **a–c**, Richness for all genera (**a**), woody genera (**b**) and herbaceous genera (**c**). **d–f**, Phylogenetic diversity for all genera (**d**), woody genera (**e**) and herbaceous genera (**f**). **g–i**, SES.PD for all genera (**g**), woody genera (**h**) and herbaceous genera (**i**). **j–l**, NRI for all genera (**j**), woody genera (**k**) and

herbaceous genera (**l**). **m–o**, NTI for all genera (**m**), woody genera (**n**) and herbaceous genera (**o**). The analyses include 2,592 angiosperm genera (woody genera,  $n = 925$ ; herbaceous genera,  $n = 1,501$ ; genera with both woody and herbaceous species,  $n = 166$ ). Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).



**Extended Data Figure 9 | Patterns of species-level phylogenetic diversity for all Chinese angiosperms.** **a–l**, Observed phylogenetic diversity for all species (**a, d, g, j**), woody species (**b, e, h, k**) and herbaceous species (**c, f, i, l**) based on species trees 210, 30, 174 and 461 (species trees were randomly selected from 1,000 post-burn-in trees). **m–o**, SES-PD for all species (**m**), woody species (**n**) and herbaceous species (**o**) based on

species tree 461. The analyses include 26,978 angiosperm species (*woody*,  $n = 10,169$ ; *herbaceous*,  $n = 16,809$ ). Phylogenetic diversity and SES-PD based on 10 species trees produce similar patterns; Spearman's rank correlation coefficients,  $r > 0.99$ ,  $P < 2.20 \times 10^{-16}$ . Maps adapted from National Administration of Surveying, Mapping and Geoinformation of China (<http://www.sbsm.gov.cn>; review drawing number: GS(2016)1576).

**Extended Data Table 1 | Number of genera that occur only in western or eastern China, with the number of woody, herbaceous and mixed genera in each order indicated**

Rank	Order	No. of genera	Woody	Herbaceous	Mixed
Western China	Asterales	26	0	26	0
	Brassicales	23	0	23	0
	Caryophyllales	18	3	14	1
	Apiales	10	0	10	0
	Fabales	8	4	3	1
	Poales	6	2	4	0
	Lamiales	5	0	5	0
	Malvales	3	0	3	0
	Asparagales	3	0	3	0
	Boraginales	2	0	2	0
	Alismatales	1	0	1	0
	Ranunculales	1	0	1	0
	Malpighiales	1	0	1	0
	Myrtales	1	0	1	0
	Rosales	1	1	0	0
	Saxifragales	1	0	1	0
	Zygophyllales	1	1	0	0
Eastern China	Lamiales	112	25	72	15
	Gentianales	102	82	18	2
	Asparagales	98	1	96	1
	Poales	81	16	65	0
	Malpighiales	72	60	9	3
	Fabales	54	37	16	1
	Sapindales	46	43	3	0
	Asterales	45	7	35	3
	Ericales	32	27	5	0
	Alismatales	30	0	30	0
	Malvales	29	24	3	2
	Myrtales	29	24	3	2
	Magnoliales	25	25	0	0
	Rosales	24	20	4	0
	Ranunculales	23	14	9	0
	Saxifragales	19	14	5	0
	Caryophyllales	18	3	14	1
	Solanales	17	8	6	3
	Zingiberales	16	1	15	0
	Apiales	15	8	6	1
	Cucurbitales	13	2	11	0
	Laurales	13	12	1	0
	Arecales	13	13	0	0
	Santalales	12	12	0	0
	Cornales	10	7	3	0
	Brassicales	9	5	4	0
	Celastrales	9	9	0	0
	Icacinales	7	7	0	0
	Oxalidales	6	5	1	0
	Fagales	6	6	0	0
	Commelinaceae	6	0	6	0
	Boraginales	5	1	4	0
	Liliiales	4	0	4	0
	Aquifoliales	3	2	1	0
	Pandanales	3	1	2	0
	Dipsacales	3	3	0	0
	Piperales	3	0	3	0
	Dioscoreales	2	0	2	0
	Dilleniales	2	2	0	0
	Chloranthales	2	0	1	1
	Huerteales	2	2	0	0
	Nymphaeales	1	0	1	0
	Proteales	1	1	0	0
	Petrosaviales	1	0	1	0
	Metteniusales	1	1	0	0
	Escalloniales	1	1	0	0
	Vitales	1	1	0	0

Mixed, genera with both woody and herbaceous species.

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

## ► Experimental design

### 1. Sample size

Describe how sample size was determined.

1. Our dated phylogeny included 5,864 species representing 2,665 genera native to China (2,665/2,884, ca. 92% of angiosperm genera in China) from 273 families.
2. Our gridded distribution data included 1,409,239 occurrence records for 26,978 angiosperm species representing 2,845 genera. After matching with the phylogeny, the final data set included a total of 2,592 angiosperm genera.
3. Our species-level time tree included 26,978 Chinese angiosperm species, covering ca. 96% of all angiosperm species native to China.

### 2. Data exclusions

Describe any data exclusions.

1. We divided the map of China into 100-km × 100-km grid cells, and grid cells on the border covering less than 50% of the area of a grid cell were excluded from the analyses to minimize the sampling bias of unequal sampling areas. The criteria were pre-established.
2. Our species tree excluded 1,098 aquatic species because phylogenetic diversity and the environmental driving factors of terrestrial plants usually differ from those of aquatic plants. The criteria were pre-established.
3. We inferred richness and PD in protected areas based on a species-level phylogenetic tree and distribution data with “county” as the basic unit. Considering that most nature reserves are smaller in size than Chinese counties, we did not count counties with protected areas covering less than 10% of the area of a county.

### 3. Replication

Describe whether the experimental findings were reliably reproduced.

1. We conducted the dating analysis of Chinese angiosperms based on the penalized likelihood and PATHd8 methods. Divergence times estimated with treePL and PATHd8 are highly congruent with each other (Extended Data Fig. 2c). Furthermore, a range of age estimates was computed using treePL based on 100 maximum likelihood bootstrap trees to better account for errors around the ages (Extended Data Fig. 2a, b). We then compared these age estimates to those inferred in recent papers by other authors.
2. To explore the spatial divergence patterns of Chinese angiosperms, we mapped the mean and median divergence times of all genera in each grid cell and designed null-model analyses to identify the ancient and recent divergence hotspots (Fig. 1; Extended Data Fig. 6). We also mapped the mean age for the relative and absolute 25% youngest and oldest genera in each grid cell over the entire map (Extended Data Fig. 7e–f vs. Fig. 1d, g), and both analyses show similar patterns and support our conclusion.
3. We generated 1,000 post-burnin trees using the birth-death model to account for alternative resolutions of polytomies in our species tree (see Methods for details). The species-level PD and SES.PD were calculated based on 10 trees randomly selected from the 1,000 trees, which indicate that the patterns of phylogenetic diversity are robust to variation of trees generated by the birth-death model (Extended Data Fig. 9).

### 4. Randomization

Describe how samples/organisms/participants were

1. To identify the best smoothing parameter that affects the penalty for rate

allocated into experimental groups.

variation over the phylogram, a random subsample and replicate cross-validation were conducted with treePL.

2. A randomized assemblage ( $n = 999$ ) was used when we calculated SES.MDT, SES.MedianDT, SES.PD, NRI and NTI (see Methods for details).

## 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g.  $P$  values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

## 7. Software

Describe the software used to analyze the data in this study.

RAxML 8.0.22, treePL, TreeAnnotator 1.8.4, PATHd8, R 3.2.0, ArcGIS 10.1, MrBayes 3.2

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). [Nature Methods guidance for providing algorithms and software for publication](#) provides further information on this topic.

## ► Materials and reagents

Policy information about [availability of materials](#)

## 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No restrictions. Most of the data that support the findings of this study are available within the paper (and its Extended Data and Supplementary Information). Sequences for phylogenetic analyses and divergence time estimation have been deposited in GenBank. Detailed phylogenetic relationships within each major clade are provided in Chen et al. (reference 31). Information on 138 calibrations used for divergence time estimates of Chinese angiosperms is provided in Supplementary Table 1. We also include a statement on data availability in our manuscript, as required by Nature policy.

## 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Not applicable

**10. Eukaryotic cell lines**

- a. State the source of each eukaryotic cell line used.
- b. Describe the method of cell line authentication used.
- c. Report whether the cell lines were tested for mycoplasma contamination.
- d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

Not applicable

Not applicable

Not applicable

Not applicable

**► Animals and human research participants**

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

**11. Description of research animals**

- Provide details on animals and/or animal-derived materials used in the study.

Not applicable

Policy information about [studies involving human research participants](#)

**12. Description of human research participants**

- Describe the covariate-relevant population characteristics of the human research participants.

Not applicable