Journal of Biogeography (J. Biogeogr.) (2014) 41, 1819–1832

SPECIAL PAPER

East Asian pollen database: modern pollen distribution and its quantitative relationship with vegetation and climate

Zhuo Zheng[1,2], Jinhui Wei[1], Kangyou Huang[1,2*], Qinghai Xu[3], Houyuan Lu[4], Pavel Tarasov[5], Chuanxiu Luo[6], Celia Beaudouin[7], Yun Deng[8], Anding Pan[9], Yanwei Zheng[10], Yunli Luo[11], Takeshi Nakagawa[12], Chunhai Li[13], Shixiong Yang[14], Huanhuan Peng[1] and Rachid Cheddadi[15]

[1]School of Earth Science and Geological Engineering, Sun Yat-sen University, Guangzhou 510275, China, [2]Guangdong Provincial Key Laboratory of Mineral Resources & Geological Processes, Guangzhou 510275, China, [3]College of Resources and Environment Science, Hebei Normal University, Shijiazhuang 050016, China, [4]Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China, [5]Palaeontology Department, Institute of Geological Sciences, Free University, Berlin 12249, Germany, [6]Key Laboratory of Marginal Sea Geology, Chinese Academy of Sciences; South China Sea Institute of Oceanology, Guangzhou 510301, China, [7]Laboratoire de Chrono-Environnement, UMR 6249 CNRS, Université de Franche-Comté, UFR des Sciences et Techniques, Besançon 25030, France, [8]Paléoenvironnements et Paléobiosphère, Université Claude Bernard-Lyon 1, Villeurbanne 69622, France, [9]College of Geography Science, Guangzhou University, Guangzhou 510006, China, [10]Guangzhou Institute of Geography, Guangzhou 510070, China, [11]Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing 100093, China, [12]Department of Geography, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK, [13]Nanjing Institute of Limnology and Geography, Chinese Academy of Sciences, Nanjing 210008, China, [14]The key Libertarianism of Marine Hydrocarbon Resources and Environmental Geology, MLR, Institute of Marine Geology, Qingdao 266071, China, [15]Institut des Sciences de l'Evolution, Université Montpellier II, Montpellier 34095, France

*Correspondence: Kangyou Huang, School of Earth Science and Geological Engineering, Sun Yat-sen University, Guangzhou 510275, China.
E-mail: hkangy@mail.sysu.edu.cn

ABSTRACT

Aim Our aims were to provide new pollen data for establishing a sub-continental surface pollen database (East Asian Pollen Database, EAPD) and to study relationships between vegetation and climate.

Location The sample sites covered most regions of East Asia, including China, Mongolia, the Russian Far East, Vietnam, Cambodia and Thailand.

Methods Data quality control procedures were applied, including taxonomic standardization, removal of duplicates, and adjustment of geographical coordinates. Vegetation types and climate parameters were assigned to each sample. Modern pollen distribution maps were drawn using circle scattergrams. The plots of pollen percentages versus climate variables allowed quantitative estimates of climate values. The modern analogue technique (MAT) was used to predict modern biomes and climate parameters.

Results Pollen assemblages extracted from 2858 sites were used to model the geographical distribution of selected taxa and their relationships with climate. For most taxa, the reconstructed range fitted the observed geographical distribution rather well. Arboreal pollen (AP) and *Pinus* dominated the transition zone between forest and steppe. Use of the MAT revealed that the predicted and observed biomes matched in 71% of the cases. The warm temperate evergreen broadleaf forest had the best agreement between predictions and observations. Climate values reconstructed using MAT were highly correlated with observed values in January temperature. The correlation coefficient of the temperature variables ranged from 0.799 to 0.930 and was as high as 0.939 for precipitation.

Main conclusions This paper documents a new modern pollen database for East Asia and makes the data readily available. The reconstructed biomes and climate variables are significantly correlated with the observed values, thus demonstrating the utility of the pollen database for future multiscale palaeoenvironmental studies.

Keywords

Biome, climate reconstruction, East Asia, database, modern analogue technique, pollen database, surface pollen, vegetation.

http://wileyonlinelibrary.com/journal/jbi
doi:10.1111/jbi.12361

Journal of Biogeography

INTRODUCTION

Determining the relationship between modern pollen spectra and vegetation and climate is critical for interpreting Quaternary pollen records and for quantitative reconstructions of vegetation and climate (Prentice *et al.*, 1996; Zheng & Guiot, 1999; Whitmore *et al.*, 2005). Inasmuch as climate controls plant distributions, fossil pollen records may be used to make inferences about past climates (Guiot, 1990; Gavin *et al.*, 2003). Many palaeoclimate reconstructions have been conducted in various regions of the world (Bartlein *et al.*, 1984; Guiot, 1990; Tarasov *et al.*, 1998; Nakagawa *et al.*, 2002; Cheddadi *et al.*, 2005). These quantitative climate reconstructions require statistical methods, such as the modern analogue technique (MAT), response surfaces or regression techniques, that are based on the relationship between modern pollen assemblages and abundances and related climate variables (Seppä *et al.*, 2004). The success of the analogue method for reconstructing past climate variables depends on the size and geographical extent of the modern pollen dataset. The method may fail if appropriate modern vegetation is not represented in the modern dataset or modern analogues do not exist (Jackson & Williams, 2004). The MAT has been used in North America, Beringia, Russia, Japan and Europe (Tarasov *et al.*, 1998; Williams *et al.*, 2001; Nakagawa *et al.*, 2002; Williams & Shuman, 2008), where large and geographically extensive modern datasets are available (Gajewski *et al.*, 2000, 2002; Markgraf *et al.*, 2002; Whitmore *et al.*, 2005).

In East Asia, regional studies of modern pollen distribution have been conducted for more than two decades (Zheng *et al.*, 2007). Datasets of surface pollen samples in Asia have been produced by Yu *et al.* (1998), Tarasov *et al.* (1998, 1999), Sun *et al.* (1999) and Gotanda *et al.* (2002). Studies of pollen–plant relationships in China have enabled the first analyses of pollen taxa relative to plant distributions at a sub-continental scale (Yu *et al.*, 2004; Zheng *et al.*, 2008). Maps of the potential modern biomes in China (Members of China Quaternary Pollen Database, 2001; Chen *et al.*, 2010) have been used to reconstruct changes in biomes in China during the late Quaternary (Yu *et al.*, 1998; Members of China Quaternary Pollen Database, 2000). In Japan, modern pollen datasets have been used for both regional biome delineation and climate reconstruction (Gotanda *et al.*, 2002, 2008). The new dataset for East Asia presented here builds upon this prior work and is based on a larger number of samples. Existing pollen datasets for East Asia were compiled from various sources, and new pollen samples, recently collected and analysed by the authors, were added. The objective of this study was to explore spatial, ecological and climatic patterns at a sub-continental scale; we elucidated the geographical distribution of several key taxa and tested climate reconstructions using the MAT.

MATERIALS AND METHODS

The study region and database

The original pollen database covers East Asia, including China, Mongolia, the Russian Far East, Japan, Kazakhstan, Vietnam, Cambodia, Thailand, Malaysia, Indonesia and India (Fig. 1a). Most of the sites in the database lie between latitudes 10° N and 55° N and longitudes 75° E and 150° E. The Himalayas form the south-western edge of East Asia, where many summits exceed 7 km in elevation and the mean elevation is 4 km. Precipitation in East Asia is derived primarily from the Indian Monsoon and the East Asian Monsoon, which originate from the Indian Ocean and the South China Sea (Ding & Chan, 2005), respectively. Beginning in May, the East Asian monsoon moves north-westwards from south-eastern Asia to the Yangtze River region and southern Japan, eventually arriving in North China, Korea and the Russian Far East. Variability in the monsoon transport of moisture coming from the Indian Ocean and the South China Sea is a major component of East Asian climate variability. The arid north-western part of East Asia is beyond the area supplied with moisture by the summer monsoon. The extent of climatic variability across the geographical range of the East Asian Pollen Database is illustrated by the 0–3200 mm range in mean annual precipitation (P_{ann}), and the −8 °C to 28 °C range in mean annual temperature (T_{ann}) (see Fig. S1 in Appendix S1 of the Supporting Information).

The current biome map of East Asia (Fig. 1c) is based on the classification of world terrestrial ecoregions developed by Olson *et al.* (2001). The ecosystems range from tropical to cold temperate forest zones. The distribution of biomes in East Asia is controlled by monsoon precipitation and the latitudinal temperature gradient, ranging from evergreen broadleaf forests in hot humid south-eastern Asia to deserts in the dry north-west. In easternmost Asia, forests stretch continuously from north to south. This belt features a diversity of forest ecosystems from cold needleleaf forests in the north-east to tropical evergreen broadleaf forests (TEBF) in the south-east, separated by temperate deciduous broadleaf forests (TDBF), warm temperate evergreen broadleaf forests (WTEF), and tropical semi-evergreen broadleaf forests (TSBF). A transition belt of shrubs and grasslands stretches from northern Kazakhstan and Inner Mongolia of China to the western flank of north-eastern China. The mountainous Tibetan Plateau is characterized by substantial plant diversity that reflects the changes in elevation. The vegetation of the high plateau includes tropical semi-evergreen forests in the foothills, warm temperate evergreen broadleaf and cool evergreen needleleaf forest in mid-elevation and sub-alpine and alpine shrub and tundra (meadow) in the cold high-elevation regions below the permanent snowline.

Pollen data

The East Asian Pollen Database (EAPD) contains modern pollen spectra, both previously assembled and published and

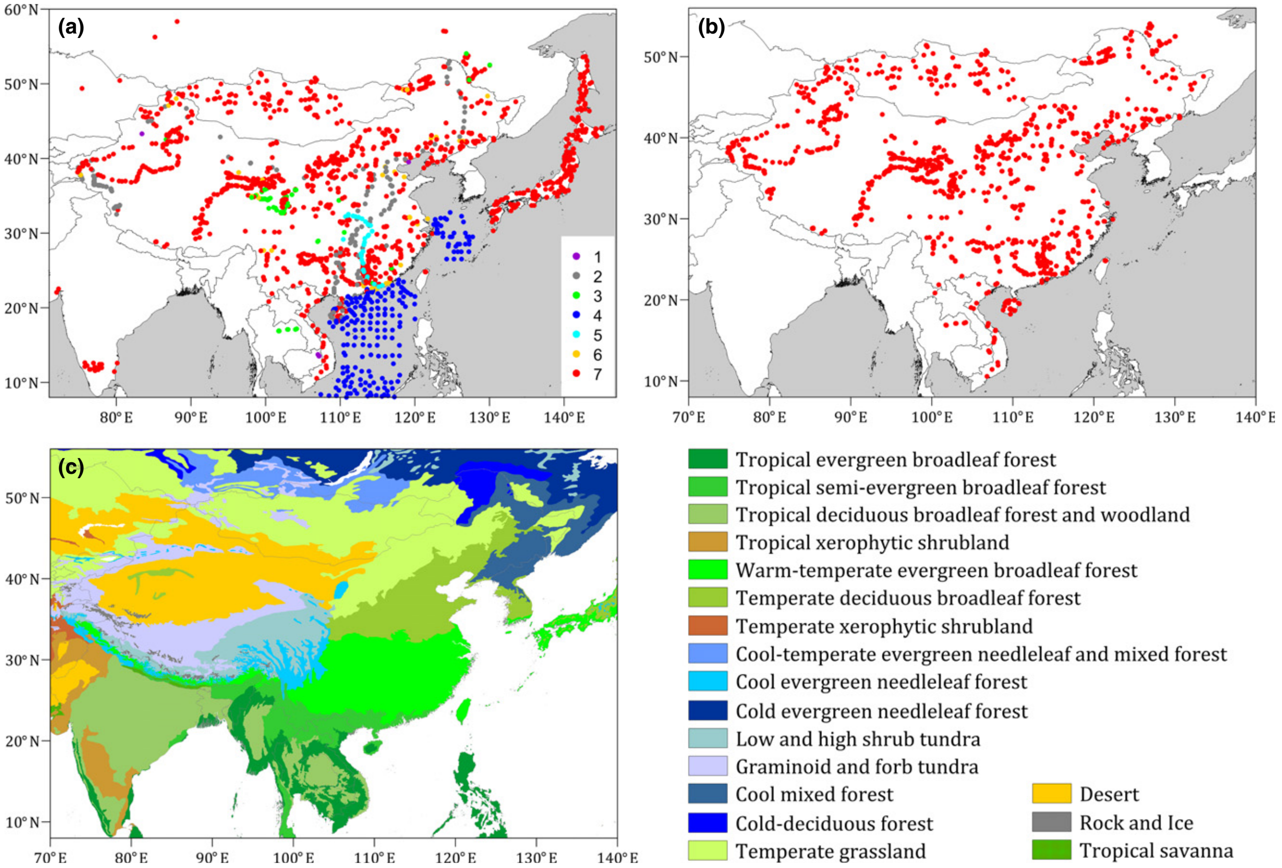

Figure 1 Surface pollen sites and the biome map of East Asia. (a) Locations of the 2858 sites in the East Asian Pollen Database (EAPD). Dots in (1) purple: moss bog samples; (2) grey: dust samples (Zheng *et al.*, 2007); (3) green: seabed surface sediments; (4) blue: core top and profile top; (5) cyan: soil samples collected around cultivated land; (6) gold: terrestrial lake and swamp surface sediments including clay, peat, ice cap samples, and samples collected from the local pond and swamp; (7) red: surface soil and sand. (b) Locations of the 1748 sites chosen for the current study. (c) The biome map of East Asia (modified from the terrestrial ecoregions published by Olson *et al.*, 2001).

newly acquired, from throughout East Asia. This database is intended to facilitate international scientific cooperation. The database comprises 2858 pollen spectra, many of which are derived from newly collected and analysed samples. Other datasets are also incorporated, including 465 spectra from the China Quaternary Pollen Database (Sun *et al.*, 1999), 102 from the Mongolia dataset compiled by Tarasov *et al.* (1998), 343 from Japan (Gotanda *et al.*, 2002), 6 from the European Pollen Database, and 86 from the published literature (Flenley, 1979; Anupama *et al.*, 2000; Barboni & Bonnefille, 2001; Maxwell, 2001; Penny, 2001) (Fig. 1a). The samples were collected from surface soil, moss and surface sediment from lakes, sea bottoms, swamps and littoral mangroves, with a few from core tops. To establish a reliable relationship between the pollen sampling site and the natural vegetation, we obtained new data from protected areas, avoiding potential anthropogenic influences to the extent possible.

The database includes all known information about the sampling sites, including latitude and longitude, elevation, local geomorphology, surrounding vegetation, quantitative measures of vegetation composition, depositional environment (sample types), sampling technique, date and time of fieldwork, analyst names, and related references. However, accurate geographical metadata are only available for recently collected samples and are frequently lacking for earlier datasets. For sites lacking accurate GPS information, geographical calibration was carefully performed to avoid inconsistencies between each site's elevation and the elevation values extracted from several digital elevation models. The original pollen and spore taxa in the database were retained to the extent possible. The database includes 740 pollen and spore types, including angiosperms, gymnosperms, Pteridophyta, algae and indeterminable palynomorphs. Pollen types that contribute more than 0.001% of the total pollen in the EAPD are shown in Appendix S2. The original data are archived in the database formats of ASCII and Microsoft Access and contain climate, taxon names, and raw pollen counts, with the exception of a few earlier sources that provide percentages rather than counts.

This paper analyses the relationship between modern pollen data and observed vegetation and related climate variables. Therefore, 1748 modern surface pollen samples were chosen from the EAPD (http://eapd.sysu.edu.cn/) (Fig. 1b);

we excluded sample sites obviously affected by human disturbance and those that included marine sediments. The data from Japan (Gotanda *et al.*, 2002), which include only arboreal pollen, were also excluded from this study.

Data quality control

The EAPD integrates modern surface pollen samples from various datasets that were collected and analysed by a number of palynologists for different research purposes. Several pollen samples were collected two or three decades ago, whereas others were collected in recent years; therefore, the data quality is uneven. Quality control procedures included verification of taxonomic synonyms, removal of duplicates and low pollen-count samples (< 100 grains), and re-adjustment of coordinates.

Taxonomic standardization

Taxonomy was standardized to facilitate continental-scale vegetation and climate reconstruction. The pollen databases include order, family, genus and species as taxonomic identifiers, but not all pollen types follow this rank-order classification. For example, a morphological hierarchy, such as 'cereal-type' (pollen grains > 50 μm) and 'other type' (smaller size), was used to distinguish different Poaceae pollen. For taxonomic (homotypic, objective) synonyms, such as Compositae and Asteraceae, modern names following the APG II family classification (Angiosperm Phylogeny Group, 2003) were used. The book *Flora of China* (Flora of China Editorial Committee, 1989–2013) was also used for specific taxonomy of the local flora. Several taxa with morphologically similar pollen types were merged; for example, *Castanopsis* and *Lithocarpus* were combined into *Castanopsis*-type. *Quercus* is commonly identified as two types, deciduous and evergreen, according to pollen morphology. In this study, however, we used *Quercus* only for the deciduous type; the evergreen type is listed as *Cyclobalanopsis*-type owing to similarities in the morphological features of the pollen. Although *Pinus* pollen can be distinguished as subgenus *Pinus* (Diploxylon) and subgenus *Strobus* (Haploxylon) (Fægri *et al.*, 1989), many contributors to our database did not differentiate them; we therefore combined these two subgenera. Amaranthaceae was used for Amaranthaceae–Chenopodiaceae because of the similarity in pollen morphology; additionally, recent phylogenetic work suggests that the Chenopodiaceae are paraphyletic with respect to the Amaranthaceae (Angiosperm Phylogeny Group, 2003; Haston *et al.*, 2009). Two types of Poaceae were designated: cereal-type and other-type. The family of Asteraceae, except for *Artemisia*, was divided into two subfamilies: Cichorioideae (syn. Compositae–Liguliflorae/Lactuceae) and Asteroideae (syn. Compositae–Tubuliflorae). Because of taxonomic inconsistencies, including duplications, synonyms, and differences in taxonomic circumscription, this study combined 387 taxa from the original 740 pollen types for the climate reconstruction. The original taxon names have been retained in the archival database.

Duplicate pollen spectra removal

The assembly of pollen data from multiple sources can lead to duplicate records. If sample sites shared the same geographical coordinates, pollen assemblages were compared to check for duplicates. One sample was retained when duplicates were found; all samples were otherwise retained. To identify potential duplicate samples that were archived with different coordinates, all samples were compared with all others, and samples with exactly the same pollen assemblages were further verified before being deleted.

Calibration of geographical coordinates

Samples that were collected decades ago usually lack precise geographical coordinates because accurate positioning equipment, such as handheld GPS devices, was not available. Before GPS devices were available, the field positioning of pollen sites has inaccurate latitude–longitude coordinates that were typically approximated from printed maps, but relatively reliable elevation measurements using pressure altimeter. The printed maps were not accurate enough to precisely locate sampling sites, which is critical in mountainous regions when interpolating climate variables. To locate these early sites more accurately, we imported all sites into a GIS (http://www.qgis.org/en/site/) and extracted their elevation values from a 90-m resolution digital elevation model (DEM, SRTM v4.1; Jarvis *et al.*, 2008). The results show that in the western parts of China with high topographic relief (e.g. the Tibetan plateau, Kunlun and the Tianshan Mountains), the difference between the DEM elevation and the one provided at the sample site can be up to 2000–3000 m. A few sites with obviously incorrect coordinates and elevations were excluded from the data.

A coordinate calibration technique was created by generating a topographical contour layer around the original pollen sample sites, based on the 90-m resolution DEM. This layer was imported into Google Earth as a layer over the satellite images. The site location was adjusted according to these criteria: (1) must be as close to the original coordinates as possible; (2) must fall exactly on the contour line of original elevation; (3) must have an elevation concordant with that of Google Earth; and (4) must be a realistic sampling site (avoiding areas with water, unreachable locations, towns or artificial buildings as much as possible). Additional information, such as site location, sampling route and geomorphology, was also considered when correcting the site position.

Climate parameters

The climate variables for each pollen site were obtained from the WorldClim dataset (Hijmans *et al.*, 2005), which provides gridded climate data at a spatial resolution of 30 arc-seconds (*c.* 1-km resolution). This dataset is derived from a variety of sources, including the Global Historical Climate Network Dataset (Peterson & Vose, 1997), World Meteorological Organization (WMO) climatological normals (WMO, 1996), and the FAOCLIM 2.0 global climate database (FAO, 2001). The WorldClim dataset includes monthly values for temperatures and precipitation (minimum, maximum, mean) and 19 bioclimatic variables widely used for predictive models of plant distributions. To identify the climate thresholds of each pollen taxon, we used the T_{ann}, mean January temperature (T_{Jan}), mean July temperature (T_{Jul}), P_{ann}, mean January precipitation (P_{Jan}), and mean July precipitation (P_{Jul}). Representative pollen taxa were selected to explore their relationship with these climate variables.

Vegetation data

The database includes vegetation metadata, such as tree cover, community structure, and dominant plant lists, for 752 samples collected by the authors. Plant cover data were collected from 100 m × 100 m plots in forested zones and from 5 m × 5 m plots in arid and semi-arid regions in north-western China. Species names, numbers of trees, tree height (m), diameter at breast height (cm), and canopy width (m) were recorded. Only species names and cover (%) were recorded for the herbaceous layer. To better understand the relationship between modern pollen assemblages and vegetation, we assigned a specific biome to each sample.

The vegetation map of East Asia was compiled from the terrestrial ecoregions of the world (Olson *et al.*, 2001). Other published sources were collected and compiled to improve the accuracy of the map. The biome classification for the Chinese region was adjusted according to the *Vegetation of China* (Wu, 1980), the *Vegetation Atlas of China* (1:1,000,000) (Editorial Board of Vegetation Map of China, Chinese Academy of Sciences, 2001), and biomes based on plant affinity groups (Huang *et al.*, 2009). For comparability with the global biome classification, we followed the BIOME 4 model (Kaplan, 2001) for most biomes. The final dataset includes 14 biomes of East Asia (Table 1, Fig. 1c).

Modern analogue technique

The modern analogue technique (MAT) is among the most widely applied statistical approaches for reconstructing past climates and vegetation. The method is based on matching individual fossil pollen spectra with a large array of modern pollen data (Guiot, 1990; Whitmore *et al.*, 2005). The squared chord distance (SCD) is considered one of the best measures of dissimilarity between fossil and modern pollen assemblages (Overpeck *et al.*, 1985; Gavin *et al.*, 2003) and has been used in many terrestrial climate reconstructions (Jackson & Williams, 2004; Sawada *et al.*, 2004).

We used the MAT to assess the reliability of our dataset for the biome and climate reconstruction. The SCD is used for computing the dissimilarity between pairs of pollen spectra. For the biome prediction, only one 'best analogue' with the lowest SCD is selected, regardless of 'good', 'possible' or 'non-analogue' (Anderson *et al.*, 1989; Gajewski *et al.*, 2002; Whitmore *et al.*, 2005). Pteridophyte and bryophyte spores and algae cysts were excluded from the MAT analysis. Four climate parameters, T_{ann}, T_{Jul}, T_{Jan} and P_{ann}, were reconstructed in this study. The accuracy of MAT-based climate reconstructions from modern pollen data can be evaluated by comparing currently observed climate values to those predicted from the best analogues, i.e. those with the smallest SCDs (Whitmore *et al.*, 2005). The evaluation was set in the 'leave-one-out' mode, i.e. a modern pollen sample was not allowed to find itself as the best analogue. Therefore, the reconstructions were distinct from and could be evaluated against the observed values (Tarasov *et al.*, 2011). Calculations were performed in R 2.15.2 (R Core Team, 2013) and the library ANALOGUE, version 0.8-2 (Simpson, 2007).

Table 1 The major biomes in East Asia and the number of pollen sample sites for each.

No.	Biomes	Abbr.	Number of sites
1	Cold evergreen needleleaf forest	CDNF	0
2	Cold-deciduous forest	CDDF	48
3	Cool-temperate evergreen needleleaf and mixed forest	CNMF	6
4	Cool evergreen needleleaf forest	CENF	74
5	Cool mixed forest	CMF	117
6	Temperate deciduous broadleaf forest	TDBF	199
7	Warm-temperate evergreen broadleaf forest	WTEF	357
8	Tropical deciduous broadleaf forest and woodland	TDFW	4
9	Tropical semi-evergreen broadleaf forest	TSBF	51
10	Tropical evergreen broadleaf forest	TEBF	85
11	Temperate xerophytic shrubland	TEXS	0
12	Tropical xerophytic shrubland	TRXS	0
13	Low and high shrub tundra	LHST	135
14	Graminoid and forb tundra	GFT	179
15	Temperate grassland	TG	313
16	Tropical savanna	TS	0
17	Desert	DS	188

RESULTS AND DISCUSSION

Pollen geographical distribution and related climate

The database includes 267 taxa, representing 1748 samples, that each contribute at least 0.001% of the total pollen. The distribution maps for the 120 most abundant taxa, as represented by pollen, are shown in Appendix S3, and more useful resources of pollen database can be found on the EAPD website (http://eapd.sysu.edu.cn/). Three characteristic taxa were selected for mapping, owing to limited space. These were *Abies*, Amaranthaceae and *Betula* (Fig. 2).

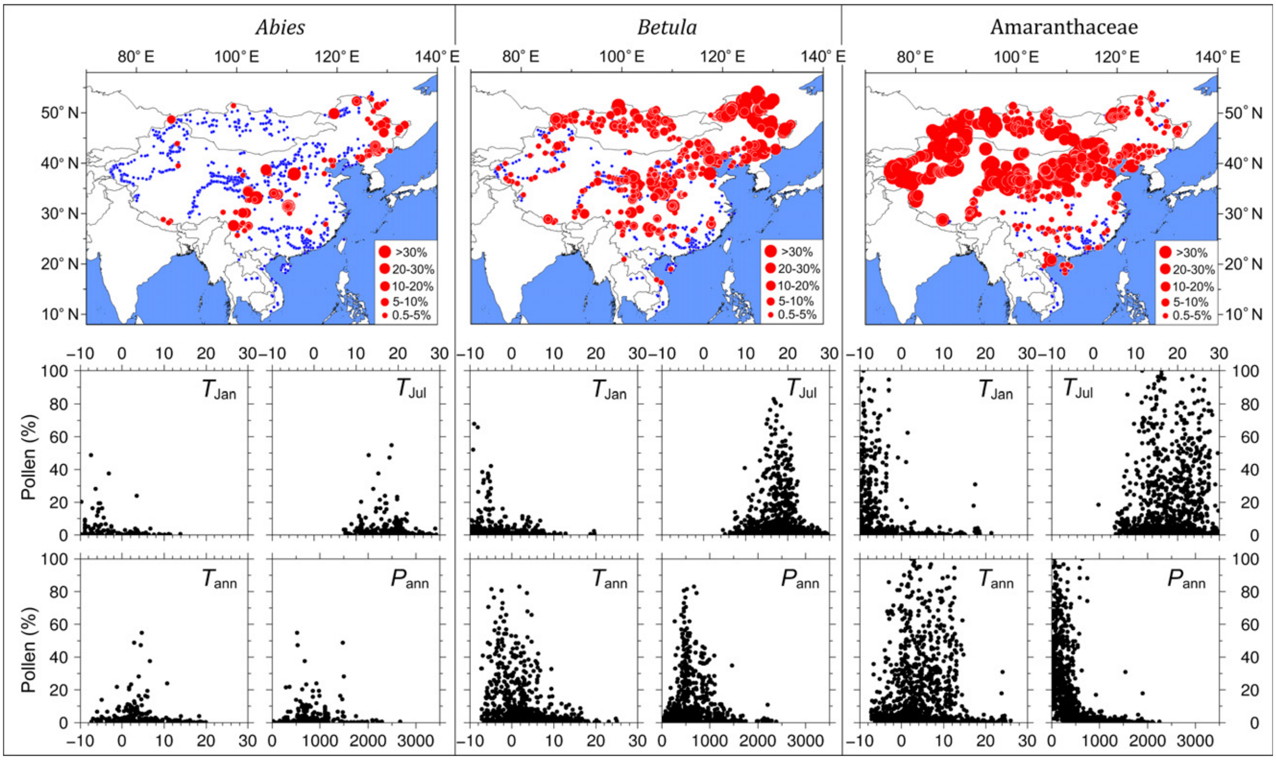

Figure 2 Maps of the distribution of selected pollen of taxa (*Abies*, *Betula* and Amarathaceae) in East Asia, and the relationship between pollen percentages and climate variables (°C for temperature and mm for precipitation) (more taxa can be found in Appendix S3).

The non-arboreal pollen (NAP) types are most abundant in the western part of the temperate zone, where P_{ann} is generally lower than 600 mm. Amaranthaceae and *Artemisia* are most abundant in desert regions and the Mongolian semi-arid steppe (Appendix S3), respectively. The desert–steppe transition is characterized by a gradual change in dominance from Amaranthaceae to *Artemisia* (Zheng *et al.*, 2008). Because of its high pollen productivity, *Artemisia* pollen is abundant even in areas where the plant species is relatively uncommon (Xu *et al.*, 2005; Zheng *et al.*, 2008). Percentages of Poaceae pollen are highest in the tropical to cool temperate vegetation zones. The values are generally less than 20%, which is often less than the values for *Artemisia* and Amaranthaceae. Poaceae is widely distributed in high-latitude grasslands, however, dominating the temperate steppe. Along the eastern forest zone in China, Poaceae pollen percentages decrease from south to north, which may be a percentage-constraint artefact because of the increasing abundance of *Artemisia* along the same gradient (Zheng *et al.*, 2007). Cyperaceae is most abundant in alpine and sub-alpine meadows on the eastern Tibetan Plateau and in wetlands of the Chinese north-eastern plain. The pollen percentages and abundances of the source plants are clearly correlated for shrub taxa in arid regions, particularly for *Ephedra, Tamarix* and *Nitraria* (Zheng *et al.*, 2008).

Pollen from angiosperm trees characterizes the major vegetation communities of East Asia. High pollen percentages from *Betula* occur in the cold temperate zone of north-east Asia and in the high mountains of central China (Appendix S3). Other deciduous broadleaved types, such as *Alnus*, *Carpinus* and *Quercus* (deciduous), span the regions from south-western to north-eastern China. *Alnus* is most abundant in the south-western mountains and in the Daxingan Mountains of north-eastern China. Although *Carpinus* usually constitutes less than 5% of the pollen (Appendix S3), it is widespread in eastern China. *Quercus* occurs in almost all sites of eastern China and is most abundant immediately north of the Yangtze River in the broadleaf mixed forests.

Many taxa, including *Cyclobalanopsis*-type, *Castanopsis*-type, *Liquidambar, Pterocarya,* Hamamelidaceae, Myrtaceae, *Platycarya, Ilex, Symplocos, Syzygium*, *Helicia* and Moraceae, are represented at the edge of the tropical and subtropical areas. High values of *Cyclobalanopsis*-type and *Castanopsis*-type (*c.* 10% and *c.* 30%, respectively) are good indicators of subtropical forests. Many taxa are less abundant but are confined primarily to the tropical zone, such as Dipterocarpaceae, Moraceae, Annonaceae, Lythraceae and *Lagerstroemia*.

The gymnosperm pollen types are among the most abundant taxa in the dataset, and *Pinus* is widely distributed. *Abies*, *Larix* and *Picea* characterize the cold temperate zones and alpine regions in the mid-low latitudes (Appendix S3). Although *Abies* and *Picea* occupy similar ranges, particularly in the cold temperate needleleaf forests, *Abies* is concentrated mainly in the north-eastern mountains of China and the eastern Tibet-Qinghai Plateau; *Picea* is more widespread in the western part of the study area. The 10% isoline of *Abies*

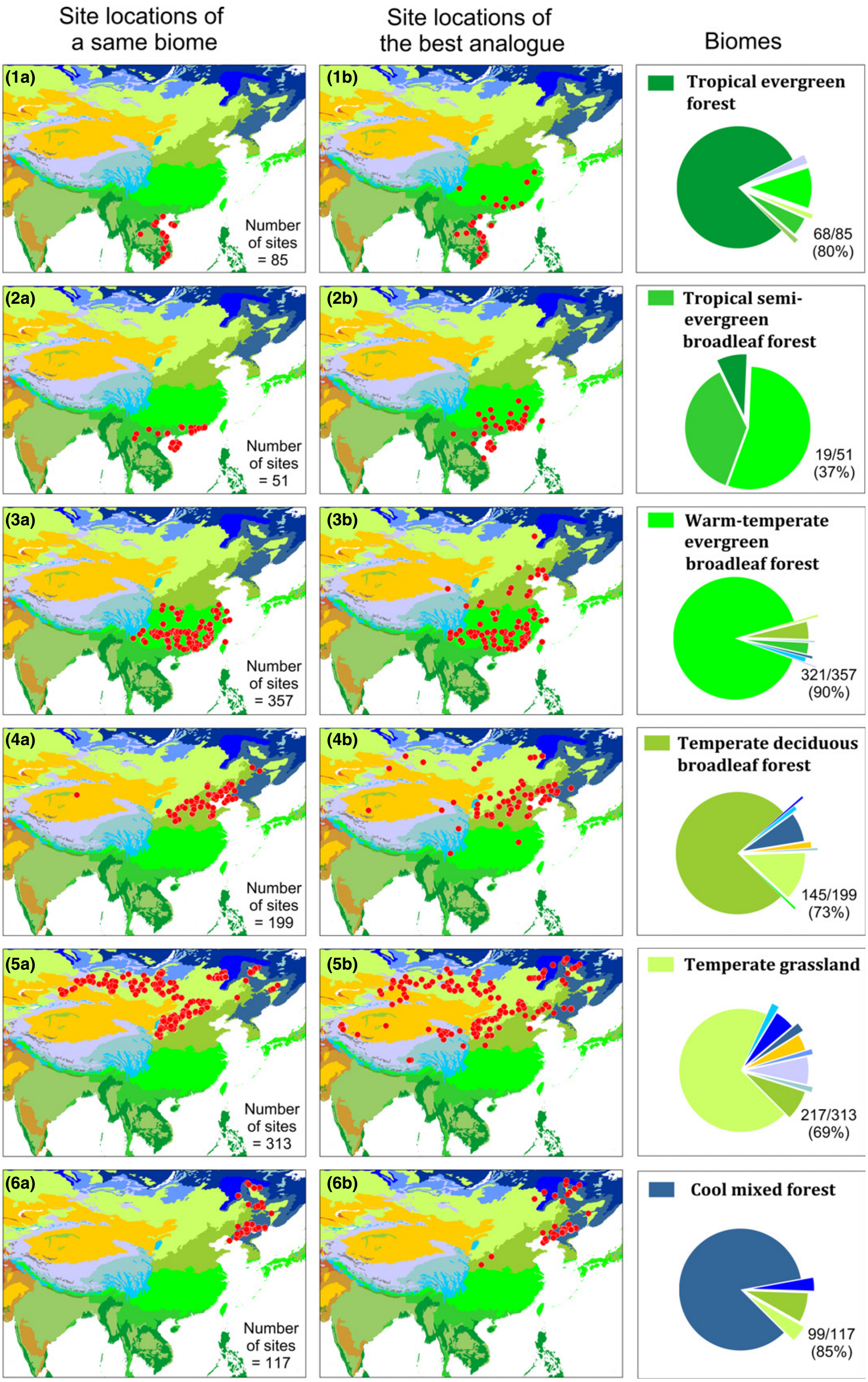

Figure 3 Biome comparison between the original assignments and the analogues for the 1748 sites from the East Asian Pollen Database (EAPD). The left column (a) displays the biome assigned to the actual location of the pollen sample sites. The middle column (b) shows the location of the sites computed as best analogues, several of which are situated outside the original biome. The number of sites in each biome and the number of analogues are shown on the right.

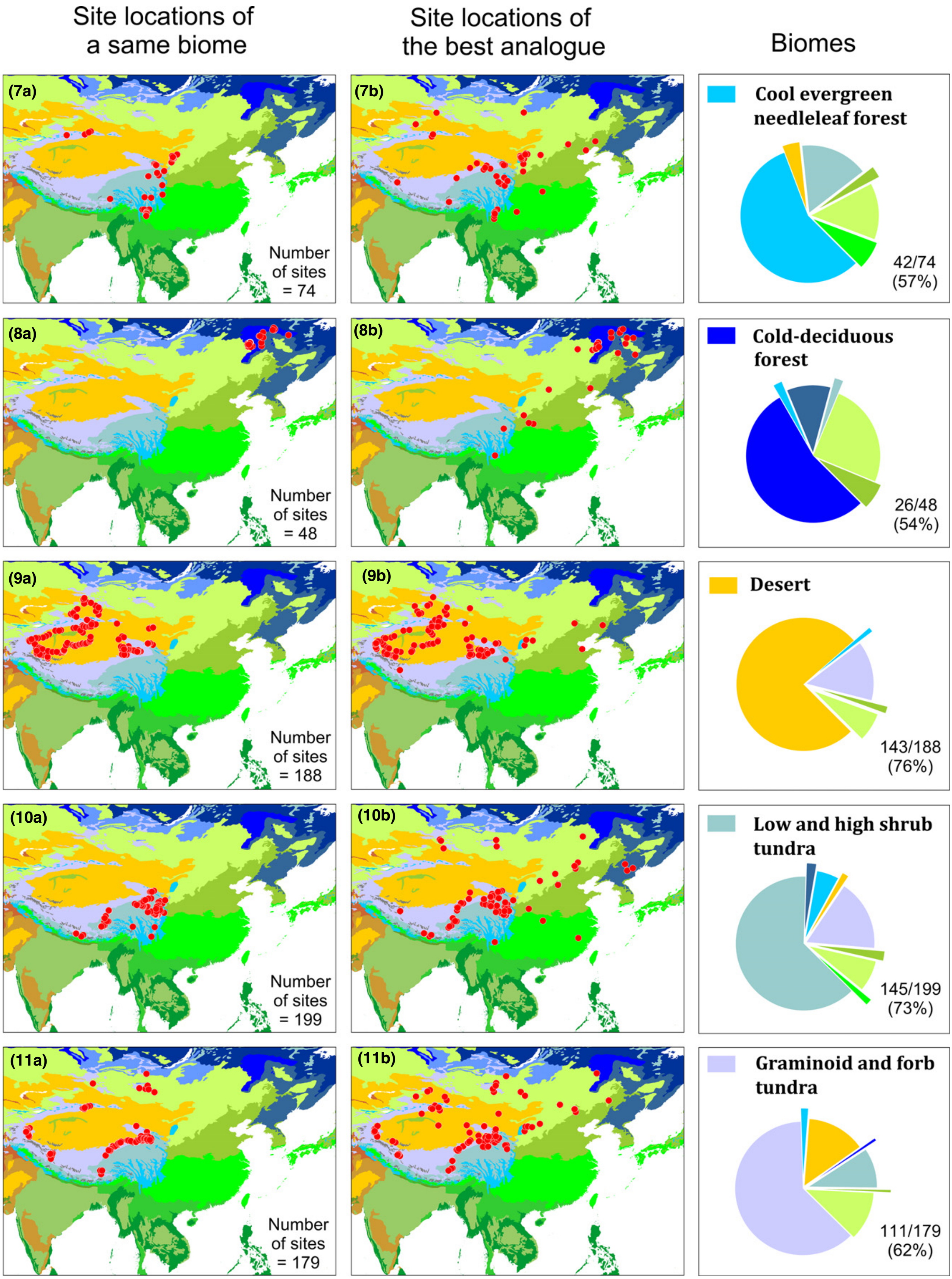

Figure 3 Continued.

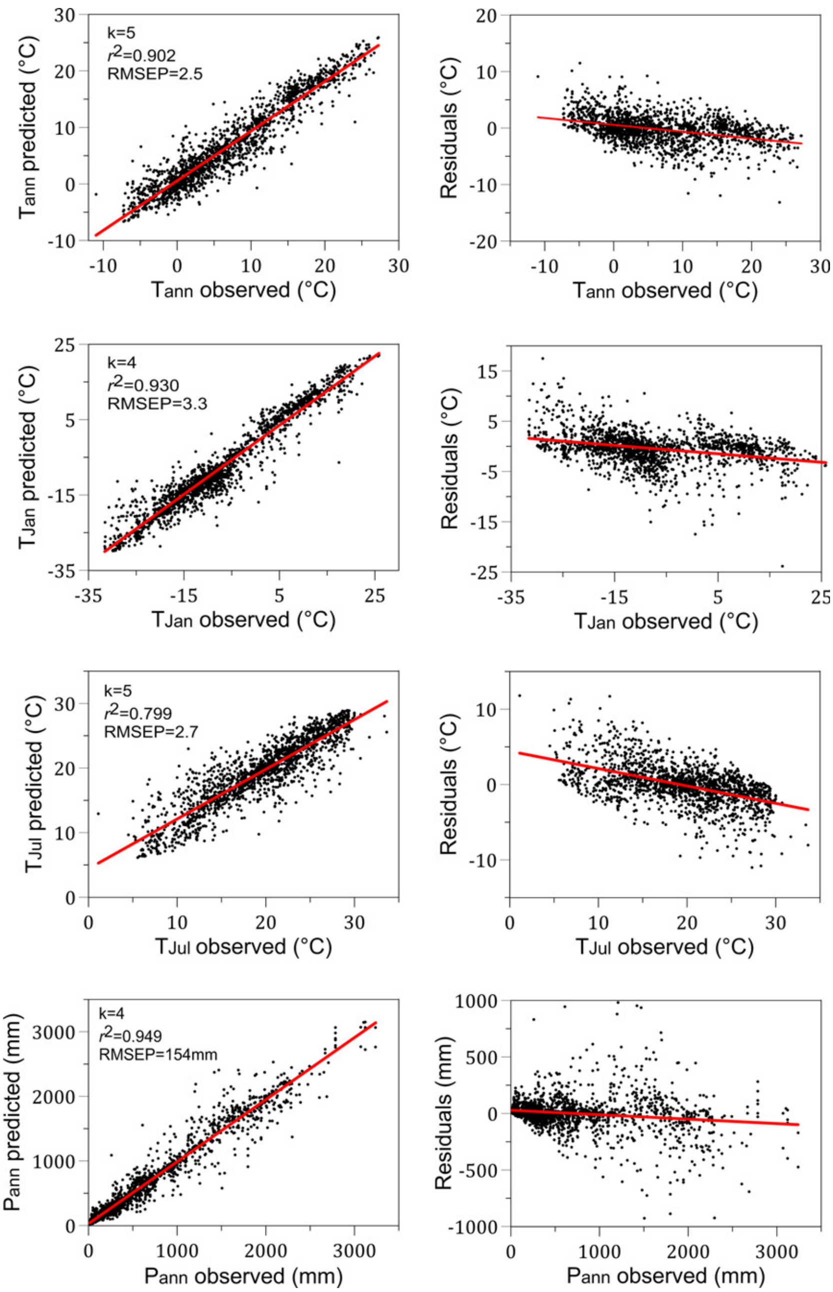

Figure 4 Pollen-based climate reconstruction using the modern analogue technique (MAT) in East Asia. The left column shows correlation coefficients between the observed climates and predicted values based on the East Asian Pollen Database (EAPD) dataset; the right column is the scatter plot of residuals from pollen-based climate reconstructions (predicted values) versus corresponding observed climate values of the sites. The plots show high correlations between the observed and the predicted values. The mean temperature of July (T_{Jul}) shows a slightly more dispersed correlation plot than the other parameters because the thermal gradient of July is the lowest. For the mean annual precipitation (P_{ann}), the plot of residuals between the observed and the predicted values shows a more dispersed distribution towards higher precipitations (in particular higher than 1000 mm). RMSEP, root mean square error of prediction; k, the number of analogues.

pollen percentages has a southern limit at approximately 22° N, while that of *Picea* is located at approximately 33° N. These two taxa range from near sea level in the northern part of the region to as high as 4000 m in the low latitudes (see Fig. S2 in Appendix S1). *Larix* occurs in the cold temperate zone and in the high mountains of the warm temperate area, and it typifies the deciduous conifer forest in the East Asia region. *Pinus* pollen occurs in more than 70% of the samples and is abundant in Southeast Asia. *Pinus* percentage isolines greater than 20% align well with the 400 mm precipitation contour, which delineates the boundary between the humid monsoon zone and the western arid zone. *Pinus* shows little correlation with the latitudinal temperature gradient. Among tropical gymnosperms, *Dacrydium* is limited to Indo-China and Hainan Island, and *Podocarpus* is limited to the southern subtropical regions.

The utility of pollen data for reconstructing past climates is dependent upon the establishment of quantitative relationships between pollen abundances and climate variables. Scatter plots of pollen percentages against climate variables indicate that most pollen types, e.g. *Abies, Betula,* Amaranthaceae, etc. have a well-defined climate envelope (Fig. 2). A few taxa that comprise a limited number of species in restricted geographical localities, such as *Abies* and *Betula*, have more restricted climate ranges. Amaranthaceae, with a great number of species, has broader climate ranges. This family (Amaranthaceae–Chenopodiaceae) is considered as an indicator of dry climates, with pollen values peaking

(> 40%) when P_{ann} is lower than 400–500 mm and P_{Jul} is approximately 150 mm. Although Cyperaceae is identifiable only to the family level, it has a relatively narrow envelope with T_{ann} from −7 °C to 4 °C and P_{ann} < 800 mm. In contrast, Poaceae, which includes a great number of genera and species that are widely distributed in East Asia, has a broader climatic range. *Pinus*, identified to the genus level, also has a wide climate range, with more than 80% of the sites falling between −8 °C and 25 °C for T_{ann} and between 400 mm and 2000 mm for P_{ann} (see Fig. S3 in Appendix S1).

A number of taxa are restricted to the dry end of the climate range. *Ephedra*, a shrub characteristic of arid and semi-arid desert, is limited to areas with P_{ann} and P_{Jul} lower than 200 mm and 40 mm, respectively. *Ephedra* occurs over a broad temperature range, with T_{ann} and T_{Jul} ranging from −2 to 12 °C and from 12 to 28 °C (Fig. S3 in Appendix S1), respectively.

Comparisons of climate thresholds between plant species (Huang *et al.*, 2008) and pollen taxa show that most pollen–climate ranges match the actual climate envelope of the source plants. For example, the superposition of the observed climate envelopes of 10 dominant species of *Castanopsis* and the scatter plot of *Castanopsis* pollen shows substantial overlap in both temperature (*c.* 12–22 °C T_{ann}) and precipitation (*c.* 1000–2100 mm P_{ann}). The *Castanopsis* pollen percentages higher than 20% also coincide with the observed climate envelope (see Fig. S4 in Appendix S1).

Predicted biomes based on pollen spectra

To find the best-matched sample, each pollen assemblage was compared with all other spectra using the MAT method. The MAT analysis revealed that for more than 71% of all sites, the 'best analogue' occurred within the same East Asian biome. Approximately 90% of the 321 pollen samples from the WTEF have their best analogue within their own biome (Fig. 3). However, we obtained a lower match percentage for other biomes. Biomes were assigned to sites according to their geographical coordinates, regardless of their elevation. One should, therefore, be cautious when analysing mountainous areas, especially the high Plateau of Qinghai-Tibet, where elevation-driven vegetation belts contribute to rapid changes in plant species composition. Better matches could be obtained in the future if more consideration were given to the effect of elevation on species composition and finer-scale biome mappings.

The main biomes in the tropical zones are the TEBF and the TSBF. For the TEBF, 80% of the samples have best analogues assigned within their observed biomes. In contrast, only 37% of the TSBF sites have a TSBF best analogue; for the remaining samples, the best analogue occurs in the adjacent WTEF. The TSBF biome sites in our dataset are located near the northern limit of the tropical zone, where the vegetation grades into the WTEF, and the mapped boundary between these two biomes does not capture this transition. Furthermore, a number of TSBF sites are located in high-elevation mountains, which are more temperate than the continent-scale biomes would indicate. Thus, the WTEF best analogues for many of the TSBF sites are not surprising.

In the high-latitude biomes, 85% of the best analogues for cool mixed forest (CMF) were from the same biome, while only 54% of the cold deciduous forest (CDDF) samples matched the observed biome. The reason for the low matching percentage is that forests similar to the CDDF occur along the mountain chains to the south, in Daxinganling, Taihang, Qingling and eastern Tibet. Consequently, the CDDF analogues were partly from these mountain chains, which belong to the cold evergreen needleleaf forest and CMF biomes when considered at the continental scale.

In the arid and semi-arid regions, the desert demonstrated the best match (76%) between the pollen-based reconstructed biome and the observed biome. Samples from ecotones had fewer analogues within their own biomes. For temperate grassland (TG), many of the best-analogue samples were from either farther north or from alpine meadows farther south, in the high Tibetan Plateau.

Climate reconstruction using MAT

In the validation test of MAT for climate reconstruction, observed and predicted climate values were highly correlated. The correlation coefficient (r^2) of the temperature variables ranged from 0.799 to 0.930 (Fig. 4). To measure calibration precision, we used the root mean square error of prediction (RMSEP) to evaluate the difference between the predicted and the reference values. The RMSEP for T_{ann} and T_{Jul} were 2.51 and 2.65 °C, respectively, but that of T_{Jan} was 3.30 °C (Table 2). The correlation coefficient for T_{Jan} was the highest (r^2 = 0.930) compared with the other temperature parameters. The r^2 for the mean annual precipitation was 0.949, but the data are more variable when precipitation exceeds 1000 mm. Although the RMSEP of the mean annual precipitation was as high as 154 mm, it was still very low if calculated as a percentage, especially in the regions subject to monsoon rainfall. The ratio of the residuals of the predicted values to the observed values was lower than 0.2 for more than 90% of the data points. The spatial map for the taxa-based climate calibration shows that the observed and predicted climate values are comparable (Fig. 5). These strong

Table 2 Statistic test of the pollen-based climate reconstruction: bias, correlation coefficient (r^2) and root mean square error of prediction (RMSEP) validation of the modern climate reconstruction across 1748 surface pollen samples in East Asia using the modern analogue technique (MAT).

Variables	T_{ann} (°C)	T_{Jul} (°C)	T_{Jan} (°C)	P_{ann} (mm)
Range	−6.7–29.9	6.2–29.0	−29.8–22.1	20–3152
Average bias	−0.27	−0.04	−0.43	−2
RMSEP	2.51	2.65	3.30	154
r^2	0.90	0.80	0.93	0.95

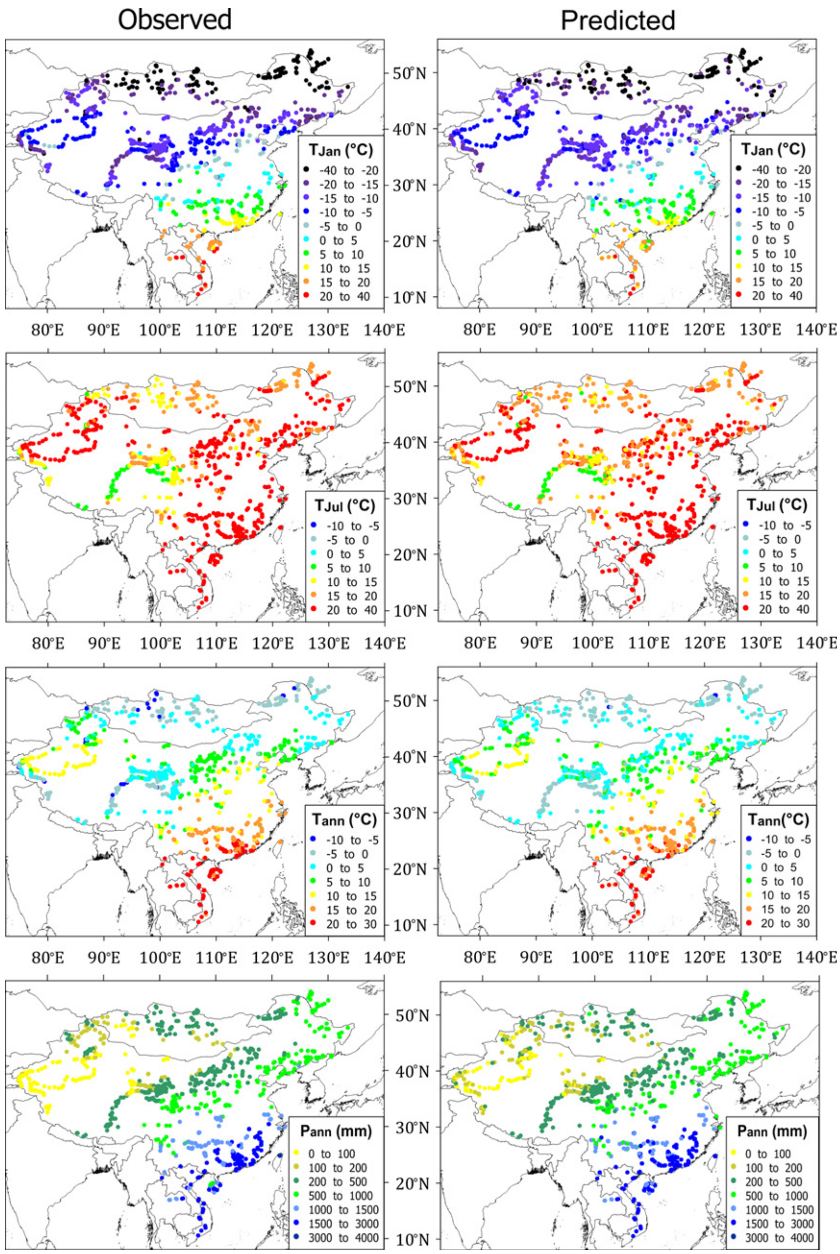

Figure 5 Evaluation of pollen-based climate reconstruction by spatial comparison between observed (left column) and predicted (right column) climate parameters (°C for temperature and mm for precipitation) in East Asia.

correlations illustrate the usefulness of the EAPD for quantitative palaeovegetation and palaeoclimate reconstructions and for data–model comparisons.

CONCLUSIONS

This paper describes a newly established standardized surface pollen database for East Asia that is suitable for reconstructing past vegetation and climate. We assembled 1748 pollen spectra from 2858 sites in the EAPD to map taxa distributions and reconstruct biome and climate variables. This work presents the first comprehensive mapping of several major plant taxa across East Asia. The geographical distribution of 120 selected taxa illustrates the potential for biome reconstructions at the sub-continental scale. The modelled ranges of most of the taxa were strongly correlated with the observed vegetation zones and related bioclimatic variables. For example, Amaranthaceae and *Ephedra* characterized desert regions, and *Artemisia* was a temperate grassland indicator in semi-arid regions. The spatial distributions of most of the analysed tree taxa accurately represented the distribution of the source plant species. Several pollen types identifiable at the family level are widely distributed, however, and derived from many species that might have diverse ecological requirements and tolerances.

Many pollen-derived taxa have distinct climate envelopes, implying a high suitability for quantitative climate reconstruction. Comparisons of the observed and modelled climate envelopes confirmed that most of the modelled distributions matched the observed distributions of the studied taxa.

More than 71% of the modern pollen samples had a best analogue from their own biome. The warm temperate evergreen forest had the best matching rate, with 90% of the pollen samples having a best analogue from the same biome. The potential natural vegetation map at the sub-continental scale in East Asia suffers from inaccuracies and does not capture finer-scale variability. Many of the best analogue sites were not from the same biome. It is possible that this result is due to imprecise biome assignments derived from the coarse-scale map.

Observed and predicted climate variables were highly correlated, especially mean January temperatures and mean annual precipitation. The pollen-based climate reconstruction demonstrates the usefulness of the database, which provides modern pollen surface samples for quantitative reconstruction of climate in East Asia.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant nos 41230101, 40331011 and 41001118) and the Fundamental Research Funds for the Central Universities (10lgzd08 and 11lgjc13). We thank all contributors of the present database and all members of the China Quaternary Pollen Database (CPD) organized in 1998. Special thanks to Eric C. Grimm and two other anonymous referees for editing the text and giving many useful comments. The authors are also grateful to Qin Chaofeng, Li Zhen, Liao Wenbo and An Fangzhou for botanical and geological studies on the fieldwork.

REFERENCES

Anderson, P.M., Bartlein, P.J., Brubaker, L.B., Gajewski, K. & Ritchie, J.C. (1989) Modern analogues of late-Quaternary pollen spectra from the western interior of North America. *Journal of Biogeography*, **16**, 573–596.

Angiosperm Phylogeny Group (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society*, **141**, 399–436.

Anupama, K., Ramesh, B.R. & Bonnefille, R. (2000) Modern pollen rain from the Biligirirangan-Melagiri hills of southern Eastern Ghats, India. *Review of Palaeobotany and Palynology*, **108**, 175–196.

Barboni, D. & Bonnefille, R. (2001) Precipitation signal in pollen rain from tropical forests, South India. *Review of Palaenbotany and Palynology*, **114**, 239–258.

Bartlein, P.J., Webb, T., III & Flieri, E. (1984) Holocene climatic change in the northern Midwest: pollen-derived estimates. *Quaternary Research*, **222**, 361–74.

Cheddadi, R., Beaulieu, J.L., Jouzel, J., Andrieu-Ponel, V., Laurent, J.M., Reille, M., Raynaud, D. & Bar-Hen, A. (2005) Similarity of vegetation dynamics during interglacial periods. *Proceedings of the National Academy of Sciences USA*, **12**, 13939–13943.

Chen, Y., Ni, J. & Herzschuh, U. (2010) Quantifying modern biomes based on surface pollen data in China. *Global and Planetary Change*, **74**, 114–131.

Ding, Y.H. & Chan, J.C.L. (2005) The East Asian summer monsoon: an overview. *Meteorology and Atmospheric Physics*, **89**, 117–142.

Editorial Board of Vegetation Map of China, Chinese Academy of Sciences (2001) *1:1,000,000 Vegetation map of China*. Science Press, Beijing.

Fægri, K., Iversen, J. & Krzywinski, L. (1989) *Textbook of pollen analysis*, 4th edn. John Wiley & Sons, Chichester, UK.

FAO (2001) *FAOCLIM 2.0. World-wide agroclimatic database*. Food and Agriculture Organization of the United Nations, Rome.

Flenley, J.R. (1979) *The equatorial rain forest: a geological history*. Butterworth, London.

Flora of China Editorial Committee (1989–2013) *Flora of China*, Vols 1–25 (ed. by Z.Y. Wu, H.R. Peter and D.Y. Hong). Science Press, Beijing & Missouri Botanical Garden, St. Louis, MO.

Gajewski, K., Vance, R., Sawada, M., Fung, I., Gignac, L.D., Halsey, L., John, J., Maisongrande, P., Mandell, P., Mudie, P.J., Richard, P.J.H., Sherin, A.G., Soroko, J. & Vitt, D.H. (2000) The climate of North America and adjacent ocean waters ca. 6 ka. *Canadian Journal of Earth Sciences*, **37**, 661–681.

Gajewski, K., Lézine, A.M., Vincens, A., Delestan, A., Sawada, M. & African Pollen Database (2002) Modern climate–vegetation–pollen relations in Africa and adjacent areas. *Quaternary Science Reviews*, **21**, 1611–1631.

Gavin, D.G., Oswald, W.W., Wahl, E.R. & Williams, J.W. (2003) A statistical approach to evaluating distance metrics and analog assignments for pollen records. *Quaternary Research*, **60**, 356–367.

Gotanda, K., Nakagawa, T., Tarasov, P., Kitagawa, J., Inoue, Y. & Yasuda, Y. (2002) Biome classification from Japanese pollen data: application to modern-day and Late Quaternary samples. *Quaternary Science Reviews*, **21**, 647–657.

Gotanda, K., Nakagawa, T., Tarasov, P.E. & Yasuda, Y. (2008) Disturbed vegetation reconstruction using the biomization method from Japanese pollen data: modern and Late Quaternary samples. *Quaternary International*, **184**, 56–74.

Guiot, J. (1990) Methodology of the last climatic cycle reconstruction in France from pollen data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **80**, 49–69.

Haston, E., Richardson, J.E., Stevens, P.F., Chase, M.W. & Harris, D.J. (2009) The Linear Angiosperm Phylogeny Group (LAPG) III: a linear sequence of the families in APG III. *Botanical Journal of the Linnean Society*, **161**, 128–131.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Huang, K.Y., Zheng, Z. & Cheddadi, R. (2008) *Atlas of plants distributions and related climate in China*. China Review Academic Publishers, Guangzhou.

Huang, K.Y., Zheng, Z., Louis, F., Guang, D.S., Cheddadi, R., Zheng, Y.W., Wei, J.H. & Xu, Q.H. (2009) Bioclimate-based plant affinity groups and regional biome simulation of China. *Quaternary Sciences*, **29**, 1–12.

Jackson, S.T. & Williams, J.W. (2004) Modern analogs in Quaternary paleoecology: here today, gone yesterday, gone tomorrow? *Annual Review Earth and Planetary Sciences*, **32**, 495–537.

Jarvis, A., Reuter, H.I., Nelson, A. & Guevara, E. (2008) *Hole-filled SRTM for the globe*. Version 4. Available from the CGIAR-CSI SRTM 90 m Database: http://srtm.csi.cgiar.org/.

Kaplan, J.D. (2001) *Geophysical applications of vegetation modeling*. Lund University, Lund.

Markgraf, V., Webb, R.S., Anderson, K.H. & Anderson, L. (2002) Modern pollen/climate calibration for southern South America. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **181**, 375–397.

Maxwell, A.L. (2001) Holocene monsoon changes inferred from lake sediment pollen and carbonate records, north-eastern Cambodia. *Quaternary Research*, **56**, 390–400.

Members of China Quaternary Pollen Database (2000) Pollen based biome reconstruction at Middle Holocene (6 ka BP) and Last Glacial Maximum (18 ka BP) in China. *Acta Botanica Sinica*, **42**, 1201–1209.

Members of China Quaternary Pollen Database (2001) Simulation of China biome reconstruction based on pollen data from surface sediment samples. *Acta Botanica Sinica*, **43**, 201–209.

Nakagawa, T., Tarasov, P.E., Nishida, K., Gotanda, K. & Yasuda, Y. (2002) Quantitative pollen-based climate reconstruction in central Japan: application to surface and Late Quaternary spectra. *Quaternary Science Reviews*, **21**, 2099–2113.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. & Kassem, K.R. (2001) Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience*, **51**, 933–938.

Overpeck, J.T., Webb, T., III & Prentice, I.C. (1985) Quantitative interpretation of fossil pollen spectra: dissimilarity coefficients and the method of modern analogs. *Quaternary Research*, **23**, 87–108.

Penny, D. (2001) A 40,000 year palynological record from north-east Thailand; implications for biogeography and palaeo-environmental reconstruction. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **171**, 97–128.

Peterson, T.C. & Vose, R.S. (1997) An overview of the Global Historical Climatology Network temperature data base. *Bulletin of the American Meteorological Society*, **78**, 2837–2849.

Prentice, I.C., Guiot, J., Huntley, B., Jolly, D. & Cheddadi, R. (1996) Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Climate Dynamics*, **12**, 185–194.

R Core Team (2013) *R: a language and environment for statistical computing*. Version 2.15.2. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.r-project.org.

Sawada, M., Viau, A.E., Vettoretti, G., Peltier, W.R. & Gajewski, K. (2004) Comparison of North-American pollen-based temperature and global lake-status with CCCma AGCM2 output at 6 ka. *Quaternary Science Reviews*, **23**, 225–244.

Seppä, H., Birks, H.J.B., Odland, A., Poska, A. & Veski, S. (2004) A modern pollen–climate calibration set from northern Europe: developing and testing a tool for palaeoclimatological reconstructions. *Journal of Biogeography*, **31**, 251–267.

Simpson, G.L. (2007) Analogue methods in palaeoecology: using the analogue package. *Journal of Statistical Software*, **22**, 1–29.

Sun, X.J., Song, C.Q. & Chen, X.D. (1999) China Quaternary pollen database (CPD) and "Biome 6000" project. *Advance in Earth Sciences*, **14**, 407–411.

Tarasov, P.E., Webb, T., III, Andreev, A.A. *et al.* (1998) Present-day and mid-Holocene biomes reconstructed from pollen and plant macrofossil data from the former Soviet Union and Mongolia. *Journal of Biogeography*, **25**, 1029–1054.

Tarasov, P.E., Guiot, J., Cheddadi, R., Andreev, A.A., Bezusko, L.G., Blyakharchuk, T.A., Dorofeyuk, N.I., Filimonova, L.V., Volkova, V.S. & Zernitskay, V.P. (1999) Climate in northern Eurasia 6000 years ago reconstructed from pollen data. *Earth and Planetary Science Letters*, **171**, 635–645.

Tarasov, P.E., Nakagawa, T., Demske, D., Österle, H., Igarashi, Y., Kitagawa, J., Mokhova, L., Bazarova, V., Okuda, M., Gotanda, K., Miyoshi, N., Fujiki, T., Takemurak, K., Yonenobu, H. & Fleck, A. (2011) Progress in the reconstruction of Quaternary climate dynamics in the Northwest Pacific: a new modern analogue reference dataset and its application to the 430-kyr pollen record from Lake Biwa. *Earth-Science Reviews*, **108**, 64–79.

Whitmore, J., Gajewski, K., Sawada, M., Williams, J.W., Shuman, B., Bartlein, P.J., Minckley, T., Viau, A.E., Webb, T., III, Shafer, S., Anderson, P. & Brubaker, L. (2005) Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications. *Quaternary Science Reviews*, **24**, 1828–1848.

Williams, J.W. & Shuman, B. (2008) Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. *Quaternary Science Reviews*, **27**, 669–687.

Williams, J.W., Shuman, B. & Webb, T., III (2001) Dissimilarity analyses of late-Quaternary vegetation and climate in eastern North America. *Ecology*, **82**, 3346–3362.

WMO (1996) *Climatological normals (CLINO) for the period 1961–1990*. World Meteorological Organization, Geneva, Switzerland.

Wu, Z.Y. (1980) *Vegetation of China*. Science Press, Beijing.

Xu, Q.H., Li, Y.C., Yang, X.L. & Zheng, Z.H. (2005) Study on surface pollen of major steppe communities in northern China. *Geographical Research*, **24**, 934–402.

Yu, G., Prentice, I.C., Harrison, S.P. & Sun, X.J. (1998) Biome reconstructions for China at 0 and 6 ka. *Journal of Biogeography*, **25**, 1055–1069.

Yu, G., Ke, X.K., Xu, B. & Ni, J. (2004) The relationships between the surface arboreal pollen and the plants of the vegetation in China. *Review of Palaeobotany and Palynology*, **129**, 187–198.

Zheng, Z. & Guiot, J. (1999) A 400,000-year paleoclimate reconstruction in tropical region of China. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, **38**, 94–98.

Zheng, Z., Cour, P., Huang, C.X., Duzer, D., Robert, G., Calleja, M., Beaudouin, C., Deng, Y. & Huang, K.Y. (2007) Dust pollen distribution on a continental scale and its relation to present-day vegetation along north-south transects in east China. *Science in China Series D: Earth Sciences*, **50**, 236–246.

Zheng, Z., Huang, K.Y., Xu, Q.H., Lu, H.Y., Cheddadi, R., Luo, Y.L., Beaudouin, C., Luo, C.X., Zheng, Y.W., Li, C.H., Wei, J.H. & Du, C.B. (2008) Comparison of climatic threshold of geographical distribution between dominant plants and surface pollen in China. *Science in China Series D: Earth Sciences*, **8**, 107–1120.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Supplementary figures.

Appendix S2 List of the principal pollen types that are more than 0.001% of the total pollen count in the East Asian Pollen Database (EAPD).

Appendix S3 Pollen distribution maps for 120 taxa based on 1748 modern surface samples in East Asia.

DATA ACCESSIBILITY

The associated data and metadata for this study are available on the website of East Asian Pollen database (http://eapd.sysu.edu.cn/).

BIOSKETCH

Zhuo Zheng received his PhD at the University of Montpellier-II, France. He is a professor in the Department of Earth Sciences, Sun Yat-sen University, China, and works as a team leader of the Quaternary palaeoenvironment research group. His research experiences are mainly in the fields of palynology, pollen-based climate reconstruction, coastal environment, deep sea pollen analysis, variability of Asian monsoons, Holocene human impact and pre-historical agriculture. He is the project leader of a major programme of the National Natural Science Foundation of China (NSFC) (Spatial pollen distribution pattern in the East Asia), which supported the work described here.

Author contributions: Z.Z, J.H.W. and K.Y.H. are co-first authors with equal contribution; Q.H.X., H.Y.L., Y.L.L., C.H.L. and P.T. provided pollen data; and C.X.L., C.B., Y.D., A.D.P., Y.W.Z., S.X.Y. and H.H.P. analysed the data; all authors led the writing.

Editor: Jack Williams