Molecular data and ecological niche modelling reveal the Pleistocene history of a semi-aquatic bug (Microvelia douglasi douglasi) in East Asia

Zhen Ye, Gengping Zhu, Pingping Chen, Danli Zhang, Wenjun Bu

MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 3080–3096 doi: 10.1111/mec.12797

Molecular data and ecological niche modelling reveal the Pleistocene history of a semi-aquatic bug (*Microvelia douglasi douglasi*) in East Asia

ZHEN YE,*[a] GENGPING ZHU,†[a] PINGPING CHEN,‡ DANLI ZHANG* and WENJUN BU*

*Institute of Entomology, College of Life Sciences, Nankai University, 94 Weijin Road, Tianjin 300071, China, †Tianjin Key Laboratory of Animal and Plant Resistance, College of Life Sciences, Tianjin Normal University, 393 Binshui West Road, Tianjin 300387, China, ‡Netherlands Biodiversity Centre – Naturalis, 2300 RA Leiden, The Netherlands

Abstract

This study investigated the Pleistocene history of a semi-aquatic bug, *Microvelia douglasi douglasi* Scott, 1874 (Hemiptera: Veliidae) in East Asia. We used *M. douglasi douglasi* as a model species to explore the effects of historical climatic fluctuations on montane semi-aquatic invertebrate species. Two hypotheses were developed using ecological niche models (ENMs). First, we hypothesized that *M. douglasi douglasi* persisted in suitable habitats in southern Guizhou, southern Yunnan, Hainan, Taiwan and southeast China during the LIG. After that, the populations expanded (Hypothesis 1). As the spatial prediction in the LGM was significantly larger than in the LIG, we then hypothesized that the population expanded during the LIG to LGM transition (Hypothesis 2). We tested these hypotheses using mitochondrial data (COI+COII) and nuclear data (ITS1 + 5.8S+ITS2). Young lineages, relatively deep splits, lineage differentiation among mountain ranges in central, south and southwest China and high genetic diversities were observed in these suitable habitats. Evidence of mismatch distributions and neutrality tests indicate that a population expansion occurred in the late Pleistocene. The Bayesian skyline plot (BSP) revealed an unusual population expansion that likely happened during the cooling transition between LIG and LGM. The results of genetic data were mostly consistent with the spatial predictions from ENM, a finding that can profoundly improve phylogeographic research. The ecological requirements of *M. douglasi douglasi*, together with the geographical heterogeneity and climatic fluctuations of Pleistocene in East Asia, could have shaped this unusual demographic history. Our study contributes to our knowledge of semi-aquatic bug/invertebrate responses to Pleistocene climatic fluctuations in East Asia.

Keywords: East Asia, ecological niche modelling, invertebrate, *Microvelia douglasi douglasi*, phylogeography, semi-aquatic bug

Received 22 November 2013; revision received 14 May 2014; accepted 14 May 2014

Introduction

Intraspecific phylogenies provide valuable insights on how populations are shaped by historical or contemporary processes (Avise 2000). In particular, the pattern of sequence differences in a gene genealogy contains information on a species' demographic history. These data can be used to reveal colonization events or genetic barriers (Emerson *et al.* 2001). Many factors (e.g. geological, ecological and dispersal) interact to determine species distribution and genetic diversity (Byrne 2008). Environmental changes during the Pleistocene glacial–interglacial cycles played an important role in shaping the temporal distribution and genetic diversity of species in Europe and North America (Avise 2000; Hewitt 2000, 2004). For example, phylogeography and paleontology have shown that European and North American terrestrial species expanded from southern refugia (e.g.

Correspondence: Wenjun Bu, Fax: +86-22-23408957;
E-mail: wenjunbu@nankai.edu.cn

[a]The first two authors contributed equally.

Mediterranean refugia scenarios) after the Last Glacial Maximum (LGM), and genetic diversity was usually geographically structured and lower in the recolonized northern areas (Hewitt 1999, 2000). Unlike terrestrial species, many recent phylogeographic studies show that montane aquatic insects survived in the Central European periglacial, that is on the periphery of Pleistocene glaciers, and in extra-Mediterranean mountain systems (e.g. Pauls *et al.* 2006; Lehrian *et al.* 2010), contrary to classical terrestrial Mediterranean refugia scenarios (Schmitt 2007). Aquatic species exhibit different life histories than terrestrial species because their physical environment is fundamentally different (Danks 2007). As a small number of permanently running water cannot cool much below 0 °C, aquatic organisms would have occupied the remaining habitats and been subject to much less severe temperature decreases than their terrestrial counterparts (Pauls *et al.* 2006).

In East Asia, including China, Japan, Korea and Mongolia, most areas were not covered by ice sheets during the Pleistocene, unlike North America and Europe (Rost 1994) (Fig. S5, Supporting information). East Asia was a mosaic of mountains lower than 2000 m characterized by a relatively mild Pleistocene climate (Weaver *et al.* 1998; Pinot *et al.* 1999; Ju *et al.* 2007), potentially hosting microclimatic zones and supporting a variety of habitats in relative stability (Qian & Ricklefs 2000; Li *et al.* 2009; Porretta *et al.* 2012). Glacial refugia may therefore have been available for East Asian species throughout their entire ranges, instead of being limited to the south as in Europe and North America. Several recent studies of East Asian terrestrial species, the Chinese Hwamei bird, *Leucodioptron canorum canorum* (Li *et al.* 2009), the Chinese bamboo partridge *Bambusicola thoracica thoracica* (Huang *et al.* 2010) and the Elliot's laughing thrush *Garrulax elliotii* (Qu *et al.* 2011) revealed an unusual pre-LGM expansion, with demographic expansion during the Last Interglacial (LIG). This result followed the general consensus on cold-adapted species in European biogeography and phylogeography (Hewitt 2004) where cold-adapted species may have wider distributions during glacial periods and experience range contractions during interglacials, the opposite pattern to temperate species.

To date, the influence of Late Pleistocene fluctuations on East Asian semi-aquatic invertebrate species remains poorly understood. We used a small semi-aquatic bug, *Microvelia douglasi douglasi* Scott 1874 (Fig. 1), which belongs to the family Veliidae of Hemiptera, as a model specie, and investigated their Pleistocene history in East Asia. Semi-aquatic bugs are a special organism group that lives in humid environments, usually on the water surface (Andersen 1982) throughout their lives. The alternating cooler (stadial) and warmer (interstadial) episodes during the late Pleistocene might have easily affected the physiological activity of semi-aquatic bugs and also restricted their habitat requirement for bodies of water, which might freeze during cool episodes. *M. douglasi douglasi* is considered as a representative of the East Asian semi-aquatic microhabitat biome. It is widely distributed in East and Southeast Asia (Miyamoto & Lee 1963; Chen & Andersen 1993; Chen *et al.* 2005). In China, it is distributed in a southern area extending from the southwestern mountain region to the eastern flatlands. *M. douglasi douglasi* usually lives on quiet water surfaces of nearshore areas, including rain pools, ditches, swampy ground, the margin of ponds or paddy fields, where it is a natural predator feeding on rice pests (Nakasuji & Dyck 1984) or mosquito larvae (Kurihara 1974).

Phylogeographic analysis is a powerful way to gain insights into the historical processes that have shaped temporal distribution and genetic variation on the species level (Avise 2000). A major issue that can confound the usefulness of genetic markers (e.g. mitochondrial versus nuclear data) for historical inference is the possibility of incongruous results (Zink & Barrowclough 2008). An alternative scenario is derived from ecological niche modelling (ENM). ENM has become a popular tool in modern phylogeography, especially for reconstructing the paleodistribution of species for which fossil data are not available (Carstens & Richards 2007), as is the case for most insects. Using the environmental variables detected in species occurrence records, ENM seeks to characterize environmental conditions that were suitable for the species, and then identify where suitable environments were distributed spatially (Pearson 2007). Based on niche conservatism, the characterized niche can be projected to identify areas of historical potential distribution. This approach to generating a historical distribution is spatially explicit and independent from phylogeographic inference (Collevatti *et al.* 2012). Richards *et al.* (2007) suggested that the coupled use of ENM to develop alternative phylogeographic hypotheses that can be tested with genetic approaches can profoundly improve phylogeographic research. Several recent studies have successfully applied this approach (e.g. Jakob *et al.* 2009; Buckley *et al.* 2010).

In the present study, we first generated potential distribution maps of *M. douglasi douglasi* in the present, the LIG and the LGM. These hypotheses were then validated with mitochondrial and nuclear sequences using a phylogeographic approach. Based on ENM, we derived two hypotheses concerning the suitable habitat and demographic history of *M. douglasi douglasi*. ENM projected a narrowed (disconnected) suitable habitat in southern Yunnan, Hainan and Taiwan, and a relatively

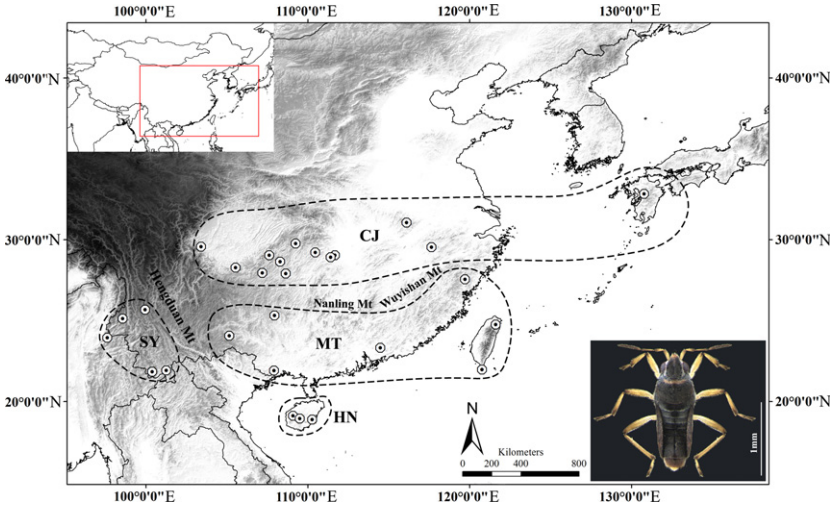

Fig. 1 Map of subtropical China showing sampling localities of *Microvelia douglasi douglasi*. Dotted lines suggesting four subregions defined by SAMOVA based on mitochondrial data (i.e. CJ, Central China and Southern Japan region; MT, Min-Guang and Taiwan subregion; SY, Southern Yunnan mountain subregion; HN, Hainan island subregion).

large expansion of suitable habitat in southeast China during the LIG (Fig. 9). We thus hypothesized that *M. douglasi douglasi* persisted in these suitable habitats during the LIG, after which the populations expanded (Hypothesis 1). In this scenario, we would expect relatively distinct mtDNA (or nuclear data) lineages or genetic differentiation among mountain ranges in central, south and southwest China, and high genetic diversities in these suitable habitats. Further, The ENM showed that unsuitable LIG climate conditions changed to suitable climate conditions during the LGM in East Asia (Fig. 8). We hypothesized that the population expansion of *M. douglasi douglasi* occurred during the LIG to LGM transition (Hypothesis 2). Under this scenario, we expected analyses of demographic history (mtDNA) in the Bayesian skyline plot to also show that population expansion occurred during that time.

Materials and methods

Sample collection and DNA sequencing

A total of 28 *M. douglasi douglasi* populations were collected across the natural distribution area in subtropical China and southern Japan. Of these, one population was from Japan, three from Hainan Island, two from Taiwan Island and the rest from mainland China. In each location, one to ten mature males or females were collected, preserved in 95% ethanol and stored in a freezer at −20 °C until DNA extraction. The latitude and longitude of each collection site were recorded using a handheld GPS unit. All the material was deposited in the College of Life Sciences at Nankai University (Tianjin, China). Because of the small size of *M. douglasi douglasi* (macropterous morph: body length: 1.85–2.00 mm; apterous morph: body length: 1.58–1.62 mm), genomic DNA was extracted from the entire body excluding the wings, posterior part of the abdomen and genitalia. Total genomic DNA was isolated using either the CTAB-based method (Reineke *et al.* 1998) or a General AllGen Kit. The precipitated DNA was resuspended in ddH_2O for later use. Polymerase chain reactions (PCRs) were performed using specific primers designed in the present study (Table S1, Supporting information). The PCR procedure for COI, COII and nuclear markers (ITS1+5.8S+ITS2) included an initial denaturation at 94 °C for 2 min, followed by 31–33 cycles of 30 s at 92 °C, 30 s at 48–52 °C and 1 min at 72 °C, ending with a final extension at 72 °C for 8 min. All fragments were sequenced in both directions using the HiSeq 2000 sequencing system. Sequences were visually proofread and aligned in BIOEDIT software (version: 7.1, Hall 2012).

Genetic analysis

Genetic polymorphism and structure. Genetic diversity was estimated by the number of polymorphic sites (*S*), haplotype distribution (Hap), number of haplotypes (Nhap), haplotype diversity (*Hd*) and nucleotide diversity (π), which were all calculated in DNASP 4.0 (Rozas *et al.* 2003). The program SAMOVA 1.0 (Dupanloup *et al.* 2002) was used to define groups of populations using genetic (rather than geographical) criteria. SAMOVA was run using 100 simulated annealing processes for $K = 2$ to 10. We then estimated the genetic differentiation among the defined groups based on pairwise F_{ST} values in ARLEQUIN 3.5 (Excoffier & Lischer 2010). For testing an isolation-by-distance model, we used MEGA 5.01 (Tamura *et al.* 2011) to calculate the genetic distance matrix, and GENALEX 6.5 (Peakall & Smouse 2012) to calculate the geographical distance matrix. The Mantel test was

performed in IBDWS (Jensen *et al.* 2005) using 1000 randomizations.

Phylogeographic analysis and haplotype network reconstruction

The PERMUT 1.0 program was used to compute parameters of differentiation (i.e. G_{ST} and N_{ST}). G_{ST} is a population differentiation estimate based solely on haplotype frequencies, whereas N_{ST} is a parameter that considers both haplotype frequencies and their genetic divergence (Pons & Petit 1996; Burban *et al.* 1999). Phylogeographical signals can be inferred by testing $N_{ST} > G_{ST}$. If N_{ST} is significantly higher than G_{ST}, then genealogically closely related haplotypes tend to occur together within populations, providing evidence for phylogeographical structure (Pons & Petit 1996). Our comparison of G_{ST} and N_{ST} was conducted based on 1000 random permutations of haplotypes across geographic populations. The potential geographical zones associated with genetic discontinuities across the entire sample region were further investigated using Monmonier's algorithm implemented in BARRIER 2.2 (Manni *et al.* 2004). A median-joining (MJ) haplotype network (Bandelt *et al.* 1999) was constructed with default settings in NETWORK 4.6.1.2 (Fluxus Technology, Suffolk, UK). We then colour coded the origin of each specimen carrying a given haplotype to illustrate haplotype distribution.

Historical demographic changes based on mitochondrial data

We used a mismatch distribution approach to reveal the demographic history of *M. douglasi douglasi* in East Asia. Mismatch distribution defined groups of populations by SAMOVA. The growth/decline statistical charts within all sample individuals were calculated based on a population growth-decline model in DNASP 4.0 (Rozas *et al.* 2003). In addition, three neutrality tests, the Tajima's *D* (Tajima 1989), Fu and Li's *D** (Fu & Li 1993) and Fu's Fs (Fu 1997) tests, implemented in DNASP (Rozas *et al.* 2003) or ARLEQUIN 3.5 (Excoffier & Lischer 2010), were used to detect departures from the mutation–drift equilibrium that would be indicative of changes in historical demography and natural selection. We inferred population expansion over time using a Bayesian coalescent-based method (Bayesian skyline plot, Drummond *et al.* 2005), a widely used module implemented in BEAST 1.6.1 (Drummond & Rambaut 2007). We ran the chains for 100 million generations, checked to make sure the ESSs was more than 200 and then discarded the first 10% as 'burn-in'. The best fit nucleotide substitution model HKY + G + I was estimated by Modeltest 3.7 (Posada & Crandall 1998). A relaxed uncorrelated lognormal molecular clock was applied with the mutation rate 0.4–0.8%/Ma for *M. douglasi douglasi* (Andersen *et al.* 2000; Damgaard & Zettel 2003). Divergence time among the four groups was estimated using a Yule process model, with the chains run for 100 million generations sampling every 1000 generations. Demographic history through time was reconstructed in TRACER 1.4 (Drummond & Rambaut 2007). TREEANNOTATOR 1.6.1 (in BEAST package) was used to summarize trees with 'Mean height', after we discarded the first 25% as burn-in.

Paleoclimate niche modelling reconstruction

A total of 97 occurrence records were obtained for niche modelling, including 28 records from the alcohol preserved specimens and 69 records from pinned specimens in Nankai University or the published literature. To avoid overemphasis on sampled areas, we selected points for model calibration using a subsampling regime to reduce sampling bias and spatial autocorrelation. We first generated models using all available occurrence points and measured spatial autocorrelation among model pseudo-residuals (1 – probability of occurrence generated by model) by calculating Moran's *I* at multiple distance classes using SAM v4.0 (Rangel *et al.* 2006). Significance was determined using permutation tests. A minimum distance of 150 km was detected. We therefore created a grid with cell dimensions of 1.5 × 1.5° and selected the occurrence point closest to the centroid of each grid cell. This procedure reduced the number of native occurrences to 69 points used for niche modelling, leaving the rest for model evaluation. Environmental dimensions in which to characterize the climate niche of *M. douglasi douglasi* were selected. We chose seven bioclimatic variables (Hijmans *et al.* 2005) that are most likely to restrict *M. douglasi douglasi* distribution: annual mean temperature (BIO1), mean diurnal temperature range (BIO2), maximum temperature of the warmest month (BIO5), minimum temperature of the coldest month (BIO6), annual mean precipitation (BIO12), precipitation of the wettest month (BIO13) and precipitation of the driest month (BIO14). All the variables were derived from the WorldClim data centre at a resolution of 5-arc. Prior to niche modelling, principle component analysis (PCA) was performed to identify the suitable environmental variables in reduced dimensions among the geographically separated populations.

We used maximum entropy implemented in MAXENT (version 3.3.3k; Phillips *et al.* 2006) to estimate niches in environmental dimensions. Analysis was run using default program conditions (cumulative output, convergence threshold (10^{-5}), maximum number of iterations (500)). Both the area under the curve (AUC) of the

receiver operating characteristic (ROC) plot and the binary omission rate were used for model evaluation. AUC is a composite measure of model performance and weights the omission error (predicted absence in areas of actual presence) and commission error (predicted presence in areas of actual absence) equally, ranging from 0 to 1, where 1 is a perfect fit. The binary omission rate was calculated by the proportion of test points that were not predicted at the 10th training presence threshold (see below).

For hindcasting the effect of Pleistocene climatic fluctuations, the current native niche models were calibrated using the above environmental variables and screened records and then transferred onto the reconstructed climatic conditions during the LGM and LIG periods, which were simulated by the Community Climate System Model 3.0 (Collins *et al.* 2006). Two general circulation models (GCM) for LGM conditions were used (the Community Climate System Model 3, CCSM 3 and Model for Interdisciplinary Research on Climate, MIROC 3.2). Model predictions based either on CCSM or MIROC were summed as the consensus for LGM prediction. The choice of threshold used to derive binary prediction can alter projections of species range under historical climate scenarios (Nenzén & Araújo 2011). Here, we overlaid possibility predictions with binary predictions to highlight predictions that might be masked by the latter. We transformed model predictions to produce binary maps by establishing the level at which 90% of the input occurrence points were included in the prediction (i.e. 10th training presence threshold).

Results

Genetic polymorphism and structure

For the mitochondrial data, 1405 bp of protein-coding regions of mitochondrial genome were obtained from 221 individuals, including sections of the COI (726 bp) and COII (679 bp) genes. 72 unique haplotypes were derived among all individuals. The 81 polymorphic sites included 52 singleton variable and 29 parsimony informative sites. The nucleotide diversities and haplotype diversities ranged from 0 to 0.004 with an average of 0.003, and 0 to 1.000 with an average of 0.858, respectively (Table 1). Among 28 *M. douglasi douglasi* populations, five (GDBL, GZFJ, HBXF, JXML and JP) contained a single haplotype (Table 1). Both nucleotide and haplotype diversities decreased with increasing latitude (Fig. 2). SAMOVA showed a distinct increase in F_{CT} from $K = 2$ to 4, and a slight increase of F_{CT} values from $K = 5$ to 10 (Fig. 3). When the $K \geq 5$ the grouping structure started to disappear, with each group composing a single population, we used $K = 4$ as the grouping scheme. The four groups corresponded well to four geographic regions (Fig. 1). The first group (CJ) is from Japan and northern China (AHYX, GZDZ, GZFJ, GZYH, GZSY, HBXF, HNSP, HNXX, HNZJ, JXML, SCEM, SCXY, JP), the second (MT) from southern China (GDBL, GZLB, GXSS, TWKD, TWYL, YNBM, ZJTS), the third (SY) from western China (YNBS, YNDL, YNNY, YNMN, YNRL) and the fourth (HN) from Hainan Island (HNBW, HNJL, HNMY). Pairwise F_{ST} values among the four defined regions ranged from 0.603 to 0.869 and were significant ($P < 0.05$, Fig. S1, Supporting information) except for the MT and HN pair ($F_{ST} = 0.433$, $P > 0.05$, Table 2), which suggested that the four defined regions are differentiated, whereas the MT and HN have existent gene flow. A Mantel test revealed a significant positive relationship between genetic and geographical distances ($r = 0.3353$, $P < 0.001$), indicating that gene flow is restricted by distance (Fig. 4).

For the nuclear sequences data, 1039 bp fragments were successfully obtained from 199 individuals, including ITS1, 5.8S and ITS2 genes. 36 unique haplotypes were derived among all individuals. The 44 polymorphic sites included 28 singleton variable and 16 parsimony informative sites. The nucleotide diversities and haplotype diversities ranged from 0 to 0.00169 with an average of 0.00325, and 0.222 to 0.900 with an average of 0.863, respectively (Table S3, Supporting information). SAMOVA showed a slight increase in F_{CT} from $K = 2$ to 3, and a slight decrease of F_{CT} values from $K = 4$ to 10 (Fig. 3). Then we used $K = 3$ as the grouping scheme. The first group is from Japan and northern China (AHYX, GZDZ, GZFJ, GZYH, GZSY, HBXF, HNSP, HNXX, HNZJ, JXML, JP), the second from south and southwest China (SCEM, SCXY, GDBL, GZLB, GXSS, TWKD, TWYL, YNBM, ZJTS, YNBS, YNDL, YNNY, YNMN, YNRL), the third also from Hainan Island (HNBW, HNJL, HNMY). The first and third groups are similar to the result of mitochondrial data except for excluding the sample of SCEM, SCXY. However, the mtDNA structure of MT and SY in *M. douglasi douglasi* is not apparent in the nuclear data. The incongruence of mitochondrial versus nuclear data is a common issue in phylogeographic studies (Zink & Barrowclough 2008). Here, this could indicate that mtDNA has relatively high variability and a rapid evolutionary rate relative to the nuclear gene, or potentially female-biased dispersal.

Phylogeography analysis and haplotype network

For the mitochondrial data, parameters of differentiation ($N_{ST} = 0.673$; $G_{ST} = 0.365$; $P < 0.001$) were computed.

Table 1 Nucleotide polymorphism in each geographic population

COI+COII	Lat.	Long.	Sample size	*S*	Hap	*Hd*	π
CJ region							
AHYX	31°3′36″	116°6′36″	4	1	**h1**(3), h2	0.500	0.00036
GZDZ	29°2′24″	107°37′12″	10	2	**h1**(9), h6	0.200	0.00028
GZFJ	27°54′36″	108°38′24″	4	0	**h1**(4)	0.000	0.00000
GZYH	28°38′24″	108°17′24″	10	1	**h1**(9), h11	0.200	0.00014
GZSY	27°56′61″	107°11′24″	6	1	**h1**(5), h12	0.333	0.00024
HBXF	29°45′36″	109°15′2″	2	0	**h1**(2)	0.000	0.00000
HNSP	29°2′24″	111°40′48″	10	1	**h1**(7), **h29**(3)	0.467	0.00033
HNXX	28°55′12″	111°24′36″	5	3	**h1**(2), h30, h31, h32	0.900	0.00085
HNZJ	29°12′36″	110°27′36″	1	—	**h1**	—	—
JXML	29°32′24″	117°38′24″	10	0	**h1**(10)	0.000	0.00000
SCEM	29°34′12″	103°24′36″	10	7	**h1**(4), h33, h34, h35, **h36**(2), h37	0.844	0.00111
SCXY	28°16′12″	105°32′24″	10	7	**h1**(5), h38, h39, h40, h41, h42	0.778	0.00100
JP	32°49′48″	130°47′24″	5	0	**h1**(5)	0.000	0.00000
MT region							
GDBL	23°19′12″	114°28′48″	9	0	**h3**(9)	0.000	0.00000
GZLB	25°18′36″	107°56′24″	10	5	**h3**(5), **h7**(2), h8, h9, h10	0.756	0.00093
GXSS	21°54′36″	107°54′36″	9	2	**h3**(7), h4, h5	0.417	0.00032
TWKD	21°58′12″	120°46′12″	10	5	**h3**(5), h43, h44, h45, h46, h47	0.778	0.00071
TWYL	24°45′1″	121°36′36″	10	4	**h3**(7), h48, h49, h50	0.533	0.00057
YNBM	24°3′36″	105°9′36″	9	3	**h3**(7), h60, h61	0.417	0.00047
ZJTS	27°33′36″	119°43′12″	7	6	**h3**(4), h70, h71, h72	0.714	0.00149
HN region							
HNBW	19°6′36″	109°5′24″	4	8	**h3**, **h13**, h14, **h15**	1.000	0.00308
HNJL	18°52′48″	110°16′12″	10	15	**h3**, **h13**(2), **h15**, h16, h17, **h18**, h19, h20, h21	0.978	0.00354
HNMY	18°56′24″	109°30′36″	10	22	**h15**(2), **h18**, h22, h23, h24, h25, h26, h27, h28	0.978	0.00444
SYregion							
YNBS	25°7′12″	98°33′36″	10	2	h51, **h52**(9)	0.200	0.00028
YNDL	25°40′12″	99°57′36″	10	11	**h52**, h53, h54, h55, h56, h57, **h58**(2), **h59**(2)	0.956	0.00263
YNNY	21°50′24″	100°23′24″	10	3	**h52**(6), **h59**, **h62**(2), h63	0.644	0.00054
YNMN	21°55′12″	101°15′36″	6	5	**h52**(2), **h59**, **h62**, h64, h65	0.933	0.00133
YNRL	23°56′24″	97°36′36″	10	10	**h52**(4), **h69**(2), h66, h67, h68, h69	0.844	0.00166

S, number of segregating sites; Hap, haplotype distribution; *Hd*, haplotype diversity; π, nucleotide diversity (haplotypes in bold were found in two or more localities; the number in brackets indicating how many times a haplotype was observed at a particular locality).

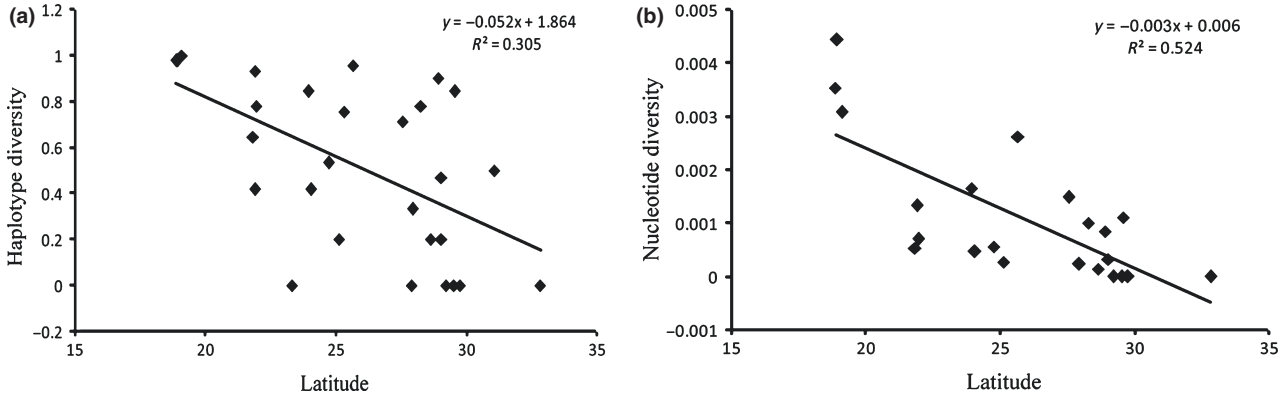

Fig. 2 Scatter plot showing the relationship between latitude and haplotype diversity (a) and nucleotide diversity (b) based on mitochondrial data.

A significant difference of N_{ST} and G_{ST} was observed ($N_{ST} > G_{ST}$; $P < 0.001$), indicating a strong phylogeographical structure existing in haplotype distributions. Zones of genetic discontinuities identified by Barrier 2.2 showed the potential geographical barriers associated with the genetic abruption (Fig. S2a, Supporting

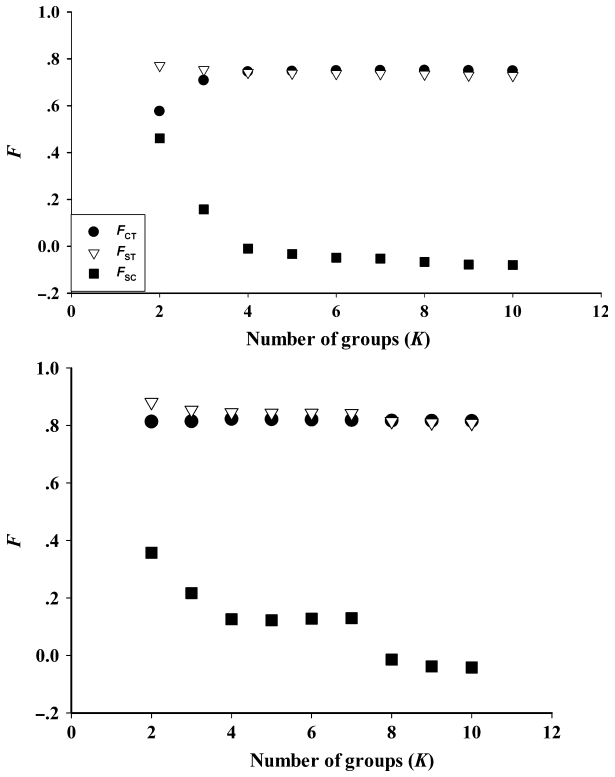

Fig. 3 Values of fixation indices (*F*) as a function of number of groups (*K*). F_{CT} is the differentiation between groups, F_{ST} is the differentiation between populations among groups and F_{SC} is the differentiation between populations within groups. Top panel is based on mitochondrial data. Lower panel is based on nuclear data.

Table 2 Pairwise F_{ST} values for the four defined regions of *Microvelia douglasi douglasi* based on mitochondrial data

	CJ	MT	HN	SY
CJ	0.00000			
MT	0.86979*	0.00000		
HN	0.69327*	0.43340	0.00000	
SY	0.76112*	0.81727*	0.60386*	0.00000

*$P < 0.05$.

information). The bold lines a, c, d, e and f (Fig. S2, Supporting information) divided the entire sampling region into four subregions, with significant genetic/geographical isolation between them. Isolations between pairs of SY/MT, CJ/MT and HN/MT (Table 2) were consistent with the results of SAMOVA. Interestingly, the bold line b divided the HN populations (HNBW, HNJL, HNMY) from each other, which indicated that different lineages existed in the Hainan region. Three haplogroups were recognized in the MJ haplotype network of COI+COII (Fig. 5a), which could be used for inferring

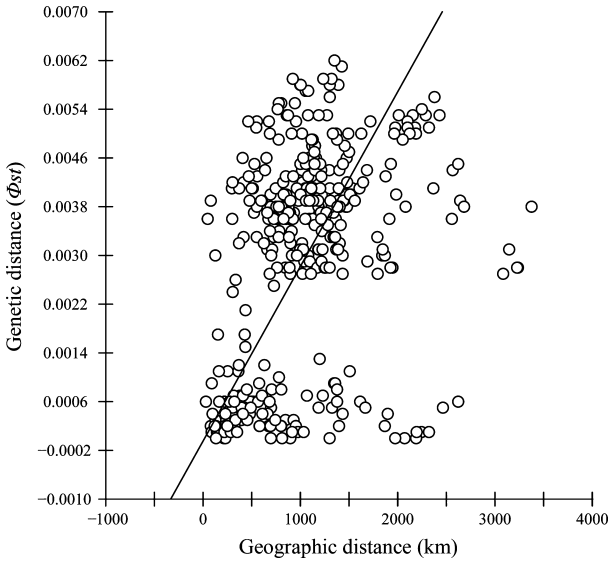

Fig. 4 Scatter plot showing the relationship between genetic distances (*Φst*) and geographical distances (km) based on mitochondrial data.

population history because the most ancient haplotypes should be located at the centre of the gene tree and be geographically widespread, whereas the most recent haplotypes should be at the tips of the gene tree and be localized geographically (Schaal *et al.* 1998). Several distributional patterns were observed in the haplotype network. First, the haplotype tended to be closer to those sampled from the same or nearby regions than to those from more geographically distant populations (Fig. 5a), consistent with a pattern of restricted gene flow under a model of isolation by distance. Second, h1, h3 and h52 were three most frequent haplotypes, characterizing 29.9%, 20.8% and 10.1% individuals, respectively. The h1 and h52 haplotypes were found in the CJ and SY regions, whereas h3 was the ancestral haplotype found in the MT and HN regions (Table 1). Third, the haplogroup of the HN region has a long branch length in network, suggesting that the Hainan populations remained stable after colonization. Three starlike shapes in the network (Fig. 5a) suggest *M. douglasi douglasi* experienced at least three population expansion events (Slatkin & Hudson 1991).

For the nuclear sequences data, Barrier 2.2 showed the bold lines a, b, c (Fig. S2b, Supporting information) dividing the entire sampling region into three subregions. Two haplogroups were recognized in the MJ haplotype network (Fig. 5b). The haplogroup of Japan and northern China (except SCEM, SCXY) has five base pair changes, reflecting significant genetic isolation. The starlike shapes in the network (Fig. 5b) also suggest *M. douglasi douglasi* experienced population expansion events.

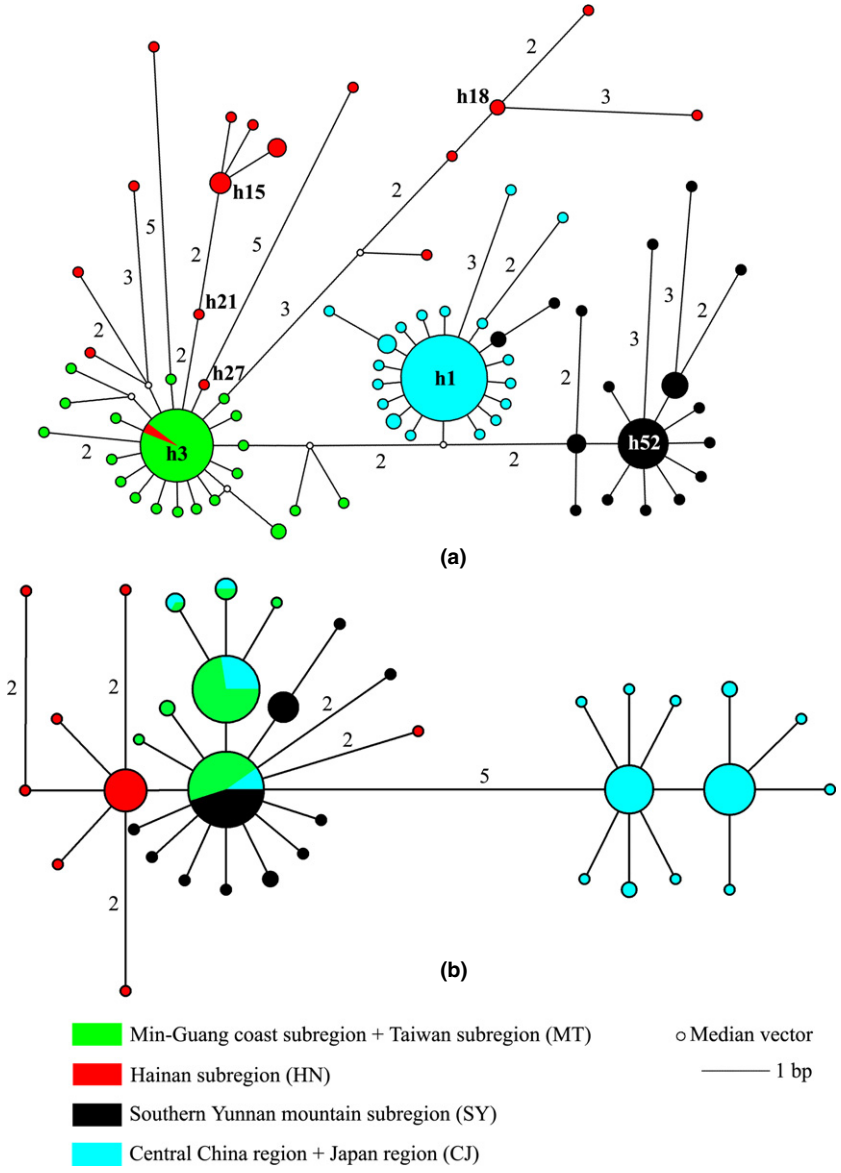

Fig. 5 Median-joining haplotype network constructed using Network. Haplotype circle size denotes the number of sampled individuals. Colours correspond to different regions. Numbers of base pair changes (no number = 1 bp) are given: (a) based on mitochondrial data. (b) based on nuclear data.

Historical demographic changes from the mitochondrial data

Multimodal mismatch distributions are assumed to characterize old populations of constant size, whereas expanding populations are thought to be unimodal (Harpending 1994). The former was observed in HN and the entire sample (Fig. 6d,e), whereas the latter was observed in the CJ, MT and SY populations (Fig. 6a–c; $P > 0.05$). The negative values of Tajima's *D*, Fu and Li's *D**, and the significance of Fu's *F*s (Table 3) suggested a scenario of demographic expansion in the CJ, MT and SY populations. Combining the results of neutrality tests, the negative values of *D*, *D**, and non-significance of *F*s (Table 3) all indicated a model of relative constant population size in the Hainan population. For the whole data set, the negative values of *D*, *D**, together with the significance of *F*s (Table 3) suggested an overall past population expansion. In the Bayesian skyline plot, the advent of a phase of demographic growth after a prolonged phase of substantial demographic stability was observed, starting about 50 000 years ago. The demographic trend then plateaued about 25 000 years ago (i.e. population expansion occurred during the transition from LIG to LGM, Fig. 7). The divergence time between the CJ+SY and MT+HN clades was estimated to be 0.38 million years ago (Ma) (95% HPD: 0.18–0.53 Ma), and the divergence of the CJ/SY and MT/HN subclades occurred 0.29 Ma (95% HPD: 0.13–0.40 Ma) and 0.30 Ma (95% HPD: 0.14–0.42 Ma), respectively (Fig. S3b, Supporting information), all

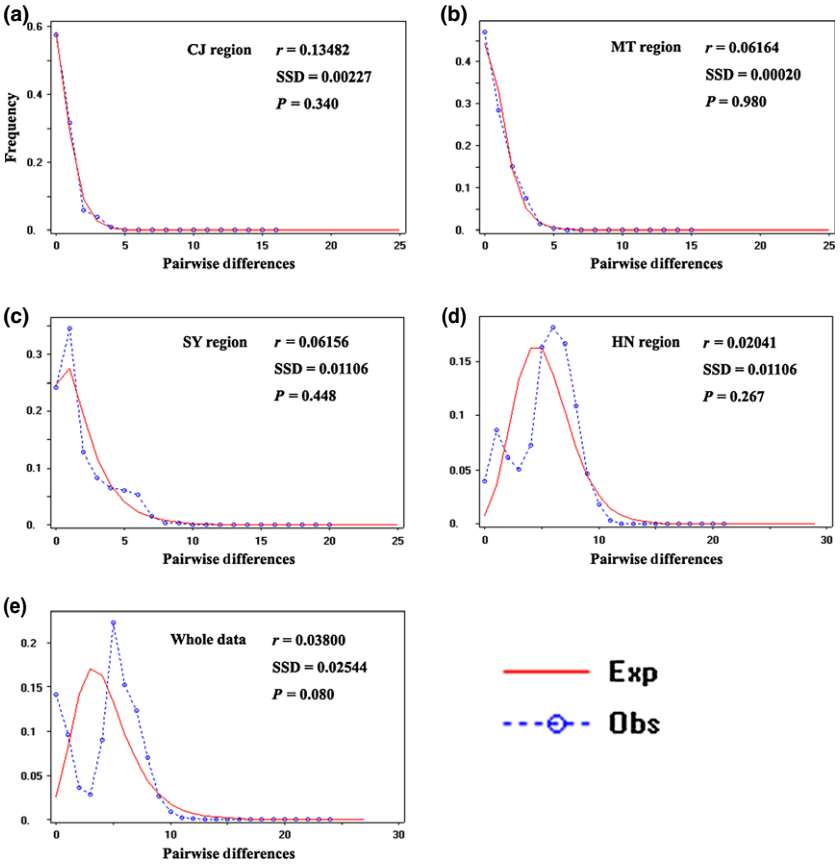

Fig. 6 Mismatch distributions for the subregional groups and for the entire sample based on mitochondrial data. Curves represent the frequency distribution of pairwise differences: observed (Obs), expected (Exp).

Table 3 Nucleotide polymorphism and neutrality tests in defined groups and the whole data set based on mitochondrial data

Parameter	MT	HN	SY	CJ	Whole set
Sample size	64	24	46	87	221
S	21	30	25	21	81
Nhap	20	17	17	19	72
Hd	0.530	0.960	0.758	0.426	0.858
π	0.00064	0.00373	0.00136	0.00042	0.00315
Tajima's *D*	−2.46832**	−1.31326	−2.19850**	−2.54431*	−2.04795*
Fu's Fs	−24.48984***	−7.30086	−10.92220***	−26.36854***	−25.37206***
Fu and Li's *D**	−4.54455**	−2.13758	−3.98216**	−5.73486**	−7.56331**

*$P < 0.05$; **$P < 0.02$; ***$P < 0.001$.

during the transition from the mid-Pleistocene to the late Pleistocene.

Principle component analysis and ecological niche modelling

Principle component analysis of seven bioclimatic variables associated with *M. douglasi douglasi* occurrences revealed reduced dimensions that account for the observed distribution (Table S2; Fig. S4, Supporting information). The first three components were significant and together explained 82.93% of the variance. The first component (PC-1) was associated with annual precipitation and the minimum temperature of the coldest month. The second (PC-2) and the third (PC-3) were less clearly associated with a single dimension (Table S2, Supporting information). In the reduced dimensions, the CJ, HN and MT occurrences were clustered, whereas the SY occurrences showed some departure from the cluster. Principal component analysis provided a representation of the species' climate space across populations; the clustering of occurrences suggested

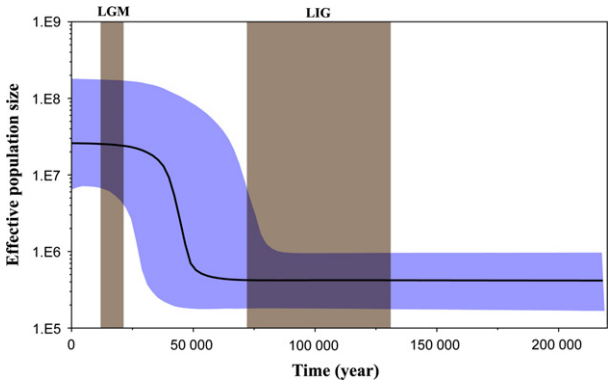

Fig. 7 Historical demographic trends of the entire sample represented by Bayesian skyline plot (BSP) based on mitochondrial data. X-axis is the timescale before present, and Y-axis is the estimated effective population size. Estimates of means are joined by a solid line, while the shaded range delineates the 95% HPD limits. LGM represents Last Glacial Maximum, and LIG represents the Last Interglacial.

climate niche conservatism while separation signified a potential divergence from niche conservation in a strict sense (Warren *et al.* 2008; Peterson 2011).

Niche models showed good performance in capturing the remaining records in native prediction (AUC = 0.95; omission = 0), when calibrating a niche model using the trimmed records. Highly suitable areas were observed in southern China, western Vietnam, northeastern Burma, Philippines, Malaysia and Indonesia (Fig. 8). Transferring the current niche model into the simulated LGM and LIG climate conditions, a great expansion was observed during the transition from LIG to LGM in both the binary and possibility prediction (Fig. 8). In the LIG prediction, highly suitable areas were identified in Hainan, Taiwan, Fujian and southern Yunnan in China, and in western Vietnam in Southeast Asia. These areas were most likely to be the suitable habitat for *M. douglasi douglasi* during the LIG. In the LGM period, the suitable areas significantly increased and the range expanded greatly. A slight range expansion towards the north was also observed after the LGM in China, South Korea and southern Japan.

Discussion

The main aim of this study was to investigate the influence of late Pleistocene fluctuations on a semi-aquatic invertebrate species, *M. douglasi douglasi*, in East Asia. We applied ENM to develop two hypotheses about the population demography and tested the hypotheses using two genetic data sets. This methodology was promoted by Richards *et al.* (2007) and is regarded as a key innovation in modern phylogeography (Hickerson *et al.* 2010). Below, we discuss the two hypotheses in the light of our a priori expectations and the observed results.

Phylogeographic structure

Combining the mtDNA COI and COII sequences, four groups (CJ, MT, SY and HN) were defined by SAMOVA across the distributional range of *M. douglasi douglasi* in subtropical China and southern Japan. In China, the four groups were located in four geographically distinct regions agreeing with the fauna division of China (Zhang 1999): the CJ group in the central China and southern Japan subregion, and the MT, SY and HN groups in the Min-Guang and Taiwan subregions, Southern Yunnan mountain subregion, and Hainan subregion, respectively. Significant genetic/geographical isolation in CJ/MT, SY/MT and HN/MT was observed (Figs. 1, S2a, Supporting information), suggesting that landscape barriers or climate ecological factors (e.g. temperature, water) might have blocked gene flow in CJ/MT, SY/MT and HN/MT and also might be due to relatively low dispersal capabilities of *M. douglasi douglasi*: individuals in most populations being apterous. The Hengduan, Wuyi, Nanling Mountains and Qiongzhou Strait might have acted as a substantial geographical barrier, which has been observed for the Chinese yew tree (*Taxus wallichiana*; Gao *et al.* 2007), Chinese Canopy tree (*Eurycorymbus cavaleriei*; Wang *et al.* 2009) and small freshwater fishes (*Squalidus argentatus*; Yang *et al.* 2012). According to the latest Wallace's Zoogeographic regions of the World (Holt *et al.* 2013), the CJ region is located in the Sino-Japanese realm and the MT region is located in the Oriental realm. PCA analysis also indicated that temperature and water, particularly minimum temperature of the coldest month (BIO6) and annual precipitation (BIO12), are the most important measured variables controlling the geographic distribution of *M. dougalsi douglasi*, so specific climate ecological factors in different zoogeographic realms might also contribute to the restriction of gene flow between CJ and MT. In aquatic insects, low geneflow rates are usually explained by limited adult lateral dispersal capabilities or philopatry restricting movement between habitats at larger geographic scales (Pauls *et al.* 2009; Taubmann *et al.* 2011). The nuclear sequences data also showed a distinctly phylogeographic structure. Two haplogroups were recognized in the MJ haplotype network (Fig. 5b). The significant split first occurred between the CJ (except SCEM, SCXY) and the other one (Fig. S2b, Supporting information). The relatively distinct mtDNA/nuclear data lineages and genetic differentiation among mountain ranges in central, south and southwest China support hypothesis 1, the survival of *M. douglasi douglasi* in these suitable habitats during the LIG. Endemic

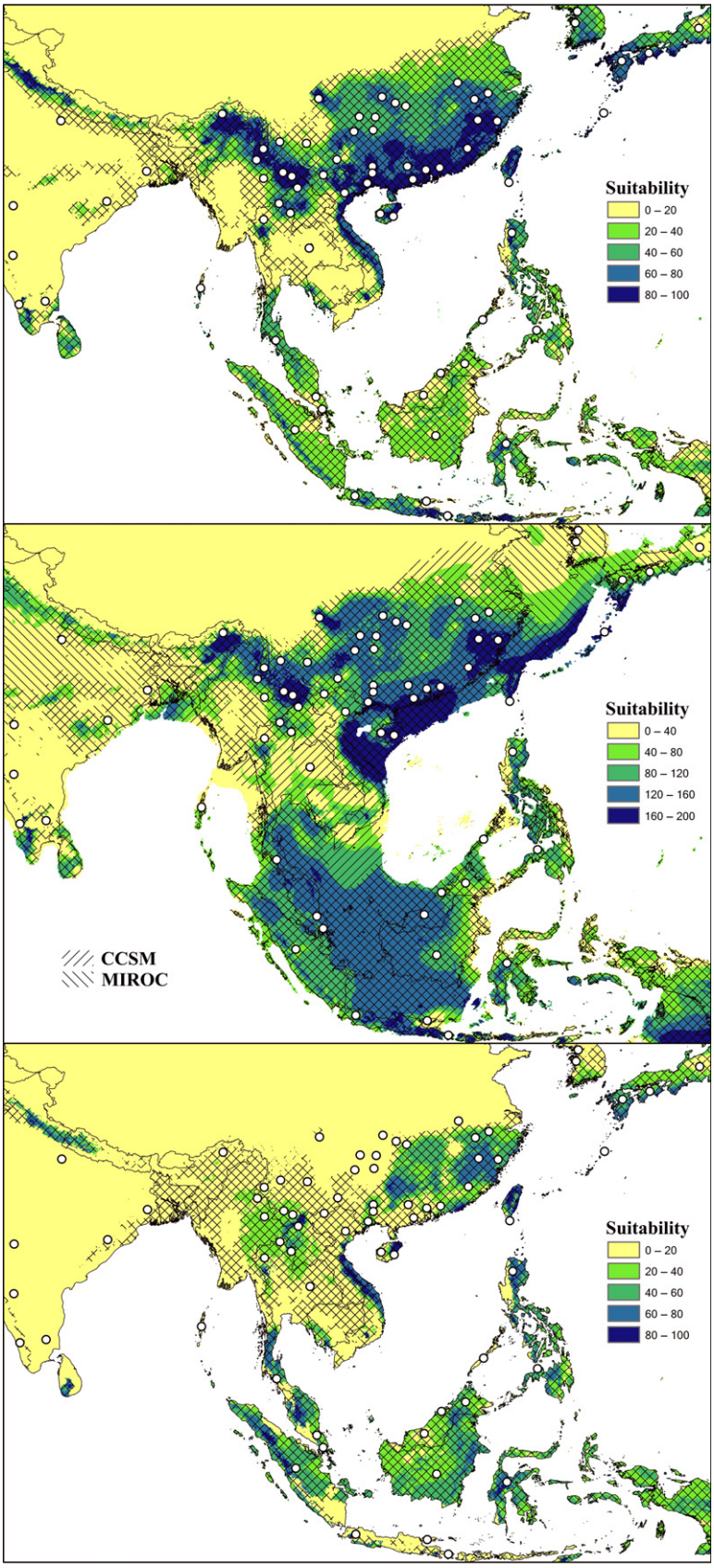

Fig. 8 Hindcasting the current niche model (up panel) onto the LGM (middle panel) and LIG (lower panel) periods in East Asia. White dots indicate occurrence records used to fit the niche model. Slash areas suggest model predictions approved by the 10th training presence threshold.

haplotypes dominate the genetic structure of *M. douglasi douglasi*. The differentiation between mountain ranges indicates restricted gene flow and relatively long-term isolation among these groups. This is a common pattern in montane aquatic insects in particular (e.g. Pauls *et al.* 2009; Lehrian *et al.* 2010).

The suitable habitats usually harbour high genetic diversity relative to those in the descendant populations, and the reduction of diversity would increase with the distance from a suitable habitat (Schaal *et al.* 1998; Petit *et al.* 2003). ENM projected relatively large expanses of suitable habitat in the MT region during the LIG, mainly located in Taiwan, Fujian and southern Guizhou (Fig. 9). Our genetic data indicated that the populations of TWKD, ZJTS and GZLB in the MT region have relatively high genetic diversity (TWKD: $Hd = 0.778$ $\pi = 0.00071$; ZJTS: $Hd = 0.714$ $\pi = 0.00149$; GZLB: $Hd = 0.756$ $\pi = 0.00093$). Individual specimens were collected in the Maolan Karst Forest, Wuyi Mountain and southern Taiwan, respectively (Fig. 9). These three regions also usually act as LGM refugia in southeast China (Cheng *et al.* 2005; Fang *et al.* 2013; Wu *et al.* 2013), and this genetic result is also consistent with the suitable habitat in the potential distribution maps generated by ENM (Fig. 9).

The small genetic distances among the wide Central China and Japan (CJ) haplotypes indicate that in this region, populations would have expanded from a shared Pleistocene suitable habitat during the LIG. ENM projected a suitable habitat in the CJ region during the LIG mainly located in eastern China and southern Japan (Fig. 9). However, our genetic data indicated that the population genetic diversity in eastern China and southern Japan was very low (e.g. JXML: Hd=0.000 $\pi = 0.00000$; JP: Hd=0.000 $\pi = 0.00000$). This pattern is consistent with a founder effect caused by a relatively few individuals dispersing over distances to recolonize eastern China and southern Japan, which showed that the eastern China and southern Japan populations were likely recolonized from the other source, which contradicts the results of ENM. Our genetic data also show that we can reject a putative recolonization of the eastern China and southern Japan populations from the MT, SY and HN source. The CJ lineage appears totally independent of the sampled MT, SY and HN populations. This is evident from the lack of shared haplotypes and the high degree of genetic differentiation between CJ and the other populations based on mitochondrial and nuclear sequences (Figs 5 and S2, Supporting information). We speculated that the suitable habitat in the CJ region during the LIG was mainly located in the western area within the CJ, which was supported by our genetic data. The genetic diversities of the SCEM, SCXY and HNXX populations located in the Eastern Himalaya range were much higher (SCEM: $Hd = 0.844$ $\pi = 0.00111$; SCXY: $Hd = 0.778$ $\pi = 0.00100$; HNXX: $Hd = 0.900$ $\pi = 0.00085$). The Eastern Himalaya is considered to be one of the biologically richest areas of the planet and also was an important LGM refugia in East Asia (Myers *et al.* 2000). An alternative scenario is that the CJ populations probably recolonized from some other source that was not accounted for in our sampling, for example Vietnam, northern India or Malaysia as projected by ENM (Figs 8 and 9).

Hainan Island was formed during the late Tertiary and early Quaternary periods by tectonic activity (Xing *et al.* 1995). Rising and falling sea levels as a result of glacial cycles repeatedly connected Hainan Island to continental China in the late Pleistocene. The bimodal curve in mismatch distributions, nonsignificant negative value of neutrality test and the long branch of haplogroups suggest that the HN populations might have a long evolutionary history of being maintained at a large stable size after colonization. ENM projected a suitable habitat in the HN region during the LIG (Fig. 9), which is consistent with our genetic data (genetic diversity: *Hd*

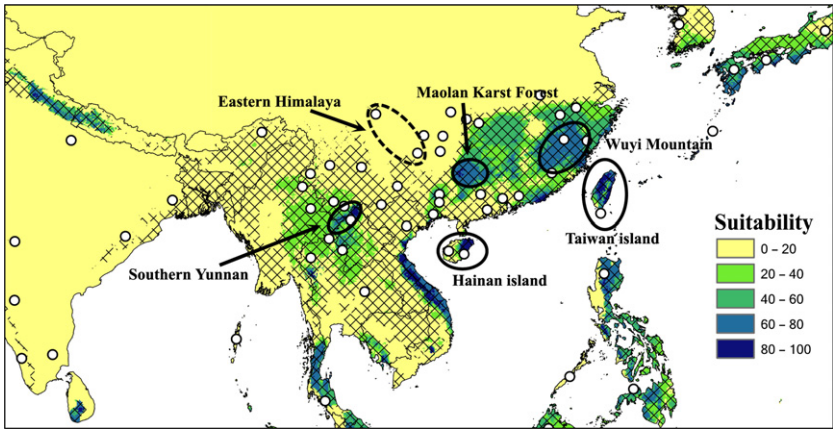

Fig. 9 The localization of putative suitable habitat for *Microvelia douglasi douglasi* during the LIG in East Asia. Solid circles indicate suitable habitat based on genetic data and ENM projection. Broken circles indicate suitable habitat based only on the genetic data.

= 0.960 π = 0.00373). When Hainan Island was surrounded by sea, it might have retained running water during the dry period in the late Pleistocene (Gathorne-Hardy *et al.* 2002), providing suitable conditions for accumulating genetic diversity since the late Pleistocene. However, we observed a lack of ancient haplotypes (Fig. 5a) and relative differentiation within Hainan populations (Fig. S2a, Supporting information). The results indicated that Hainan was recolonized by more than one lineage before the LIG, and multiple and independent genetic lineages from other regions resulted in high genetic diversity. The ancestral haplotype h3 in the MT was also found in the HNJL population. Some haplotypes (h21, h27, Fig. 5a) have a close relationship with the MT haplogroup, but the other haplotypes have distinct genetic barriers with the MT haplogroup (h15, h18, Fig. 5a), which suggested that at least one lineage in the HN region would have been recolonized from a population from the MT region. Other lineages were not accounted for in our sampling (Vietnam, northern India or Malaysia as projected by ENM) (Figs 8 and 9).

The SY region is located in the Hengduan Mountains, a dramatic series of north–south trending valleys and ridges extending from 1000 to over 6000 m in elevation (Peng *et al.* 2000). The Hengduan region has been found to be a centre of speciation for many organisms as a biodiversity hot spot (Myers *et al.* 2000). It was a refugium for many animals and plants during Pleistocene glacial periods (Wang & Liu 1994; Li *et al.* 2005; Long *et al.* 2006; Zhang *et al.* 2010). ENM projected a suitable habitat in the SY region during the LIG, mainly located in southern Yunnan (Fig. 9). High nucleotide and haplotype diversities (*Hd* = 0.758 π = 0.00136) were observed in the SY haplogroup, with the highest diversity observed in the YNDL population (*Hd* = 0.956 π = 0.00263). We therefore hypothesized that the complicated topology of the Hengduan region might account for the observed high genetic diversity of SY populations. Areas of sheltered topography could provide suitable stable microclimates, and species could experience frequent movement of populations between different altitudes and survived there during climatic oscillations (Rull 2009). A resumption of demographic expansion by southern Yunnan populations within the SY populations therefore might have occurred after the LIG.

Divergence times and unusual demographic expansion

Divergence times derived from the analysis of the mitochondrial sequence indicated that diversification within *M. douglasi douglasi* was relatively recent (Fig. S3, Supporting information). All diversification probably occurred within 0.13–0.53 million years, and all appeared during the transition from the mid-Pleistocene to the late Pleistocene. It has been suggested that our planet experienced a dramatic climate shift during this period, termed the 'mid-Pleistocene revolution' (Ruddiman *et al.* 1989). The mid-Pleistocene climate transition (MPT) was characterized by a gradual change in the dominant climate periodicity from 40 to 100 ka centred on marine isotope stage (MIS) 23/22 boundary at about 0.9 million years (Berger *et al.* 1993). The dramatic climate shift cycles were characterized by increasing severity and duration of cold stages, which had profound impacts on the biota, especially in the Northern Hemisphere. A major faunal turnover has been detected in East Asia (Wang *et al.* 2000) around the time of the climate transition. It is possible that the drier and colder environment in the later Pleistocene induced the population divisions within *M. douglasi douglasi*.

Evidence from mismatch distributions (Fig. 6), neutrality tests (Table 3) and the Bayesian skyline plot (BSP) (Fig. 7) all indicates a population expansion in the late Pleistocene. The population expansion occurred during the transition from the LIG to LGM, around 50 000 years ago (based on a mutation rate of 0.4–0.8%/Myr), and later plateaued about 25 000 years ago (Fig. 7). This result was in line with the spatial prediction of ecological niche models (Hypothesis 2), which is independent from the hypothesis generated by phylogeography (Collevatti *et al.* 2012). Niche modelling showed a significant suitable range contraction and expansion in the LIG and LGM (Fig. 8). The two independent hindcasting methods (phylogeographic analysis and ENM) achieved a similar pattern, consolidating the results of range expansion estimation for *M. douglasi douglasi* during the LIG to LGM transition. We speculate that the restricted distribution or low population during the LIG may be due to increasing temperatures, as temperatures during the LIG were about 2–5 °C warmer than the present (Zhao *et al.* 2011). The high temperature not only affected the bug's physiological activity (Chown & Terblanche 2007), but also destroyed suitable conditions for *M. douglasi douglasi* at low altitudes in coastal areas, as the Pacific sea level reached approximately 6 m above the present (Camoin *et al.* 2001), reducing the number of habitats on the mainland (including small rain pools, ditches and swampy ground) along with excessive evaporation in low-lying areas. These two reasons might account for the contracted distribution of *M. douglasi douglasi* during the LIG.

As the climate cooled during the coming of the LGM, our coupled phylogeography and ecological niche modelling approaches both revealed a rapid demographic expansion. It started approximately 50 000 years ago, and plateaued about 25 000 years ago, during Marine

Isotope Stage (MIS) 3 of the Quaternary Glaciations in China (Li *et al.* 2009; Zhao *et al.* 2011). We hypothesized that under relatively mild climate conditions in the East Asia (Li *et al.* 2009), relatively low temperatures and rich rainfall increased the number of suitable habitats (small rain pools, ditches, swampy ground) for *M. douglasi douglasi* along with low evaporation and water storage during the MIS 3 (Zhao *et al.* 2007; Wang 2010). In MIS 3, strong monsoon rainfall marked this region with intensive erosion, exposing more and more topological surface (Schaefer *et al.* 2008; Chen *et al.* 2011). We speculate that the formation of complicated topological surfaces increased the number of suitable habitats for *M. douglasi douglasi*. In this scenario, *M. douglasi douglasi* might have dispersed into suitable habitats and established populations there. In addition, *M. douglasi douglasi* is capable of high fecundity and long dispersal ability in its macropterous form (Nakasuji & Dyck 1984). The increase in suitable habitats available to *M. douglasi douglasi*, together with wide ecological adaptability, paved the way for demographic and distributional expansion during the LIG to LGM transition.

This scenario has also been detected for two eastern Asian birds (Dai *et al.* 2011; Wang *et al.* 2012) and a brown tree frog (Lin *et al.* 2012), and is different from the classic post-LGM expansion scenario observed in many other organisms (Hewitt 1999). Schmitt (2007) proposed two general biogeographical arctic–alpine distribution patterns: (i) disconnected distribution during glacial phases, resulting in comparably deeper/older genetic splits and (ii) widely distributed species throughout the glacial period, with interruption of gene flow during recent climate warming, resulting in relatively young genetic splits. Interestingly, the biogeographical distribution patterns of *M. douglasi douglasi* in East Asia are very similar to the latter. Further study of this species, for instance a comprehensive sample from southeast Asia, should enhance our biogeographical knowledge of semi-aquatic insect.

Conclusions

Coupled phylogeography and ecological niche modelling unveiled the Pleistocene history of a semi-aquatic bug (*Microvelia douglasi douglasi*) in East Asia. Molecular data reveal subtle genetic differentiation among mountain ranges in central, south and southwest China, and high genetic diversities in these suitable habitats. The unusual population expansion is inconsistent with that of most species, and its range expanded rather than contracted during the glacial periods and contracted rather than expanded during the interglacial periods. The results of genetic data are mostly consistent with the spatial predictions from ENM. The ecological requirements of *M. douglasi douglasi*, together with the geographical heterogeneity and mild Pleistocene climate of East Asia, could have shaped this unusual demographic history. Further studies are needed to obtain more insights into the effects of Pleistocene climatic fluctuations of semi-aquatic insect/invertebrate species in East Asia.

Acknowledgements

The authors wish to thank Prof. Jin-Tang Dong (Nankai University, China) and Prof. Qiang Xie (Nankai University, China) for reviewing the manuscript and providing very helpful advice. We thank Dr. Jose Ricardo (Universidade Federal do Pampa, Brazil) for providing specimens from Japan. Funding for this research was provided by the National Natural Science Foundation of China (No. 31071959, J1210005), Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20100031110026), Special Project of Technologically Basic Work (No. 2011FY120200), and a talent introduction programme award to Gengping Zhu at Tianjin Normal University (5RL127).

References

Andersen NM (1982) The semiaquatic bugs (Hemiptera, Gerromorpha). Phylogeny, adaptations, biogeography and classification. *Entomonograph*, **3**, 1–455.

Andersen NM, Cheng L, Damgaard J *et al.* (2000) Mitochondrial DNA sequence variation and phylogeography of oceanic insects (Hemiptera: Gerridae: *Halobates* spp.). *Marine Biology*, **136**, 421–430.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Berger WH, Bickert T, Jansen E, Wefer G, Yashuda M (1993) The central mystery of the Quaternary Ice Age. *Oceanus*, **36**, 53–56.

Buckley TR, Marske K, Attanayake D (2010) Phylogeography and ecological niche modelling of the New Zealand stick insect *Clitarchus hookeri* (White) support survival in multiple coastal refugia. *Journal of Biogeography*, **37**, 682–695.

Burban C, Petit RJ, Carcreff E *et al.* (1999) Rangewide variation of the maritime pine bast scale *Matsococcus feytauti* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Molecular Ecology*, **8**, 1593–1602.

Byrne M (2008) Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quaternary Science Reviews*, **27**, 2576–2585.

Camoin GF, Ebren P, Eisenhauer A *et al.* (2001) A 300 000-yr coral reef record of sea level changes, Mururoa atoll (Tuamotu archipelago, French Polynesia). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **175**, 325–341.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Chen PP, Andersen NM (1993) A checklist of Gerromorpha from China (Hemiptera). *Chinese Journal of Entomology*, **13**, 69–75.

Chen PP, Nieser N, Zettel H (2005) *The Aquatic and Semi-Aquatic Bugs (Heteroptera: Nepomorpha & Gerromorpha) of Malesia*. Fauna Malesiana Handbook 5, p. 546. Brill, Leiden, Boston.

Chen YX, Li YK, Zhang M *et al.* (2011) Terrestrial cosmogenic nuclide ^{10}Be exposure ages of the samples from Wangkun till in the Kunlun Pass. *Journal of Glaciology and Geocryology*, **33**, 101–109.

Cheng YP, Hwang SY, Lin TP (2005) Potential refugia in Taiwan revealed by the phylogeographical study of Castanopsis carlesii Hayata (Fagaceae). *Molecular Ecology*, **14**, 2075–2085, Blackwell Publishing, Ltd.

Chown SL, Terblanche JS (2007) Physiological diversity in insects: ecological and evolutionary contexts. *Advances in Insect Physiology*, **33**, 50–152.

Collevatti RG, Terribile LC, Lima-Ribeiro MS *et al.* (2012) A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Molecular Ecology*, **21**, 5845–5863.

Collins WD, Bitz CM, Blackmon ML *et al.* (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, **19**, 2122–2143.

Dai C, Zhao N, Wang W *et al.* (2011) Profound climatic effects on two East Asian black-throated tits (Aves: Aegithalidae), revealed by ecological niche models and phylogeographic analysis. *PLoS ONE*, **6**, e29329.

Damgaard J, Zettel H (2003) Genetic diversity, species phylogeny and historical biogeography of the *Aquarius paludum* group (Heteroptera: Gerridae). *Insect Systematics and Evolution*, **34**, 313–328.

Danks HV (2007) How aquatic insects live in cold climates. *Canadian Entomology*, **139**, 443–471.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Rambaut A, Shapiro B *et al.* (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Emerson BC, Paradis E, Thebaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution*, **16**, 707–716.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fang F, Sun HY, Zhao Q *et al.* (2013) Patterns of diversity, areas of endemism, and multiple glacial refuges for freshwater crabs of the genus *Sinopotamon* in China (Decapoda: Brachyura: Potamidae). *PLoS ONE*, **8**, e53143.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Gao LM, Möller M, Zhang XM *et al.* (2007) High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. *Molecular Ecology*, **16**, 4684–4698.

Gathorne-Hardy FJ, Syaukani, Davies RG *et al.* (2002) Quaternary rainforest refugia in south-east Asia: using termites (Isoptera) as indicators. *Biological Journal of the Linnean Society*, **75**, 453–466.

Hall T (2012) BioEdit version 7.1.7. Available form: http://www.mbio.ncsu.edu/bioedit/bioedit.html. (accessed 25 November 2012).

Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial-DNA mismatch distribution. *Human Biology*, **66**, 591–600.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean society*, **68**, 87–112.

Hewitt GM (2000) The genetic legacy of the quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 183–195.

Hickerson MJ, Carstens BC, Cavender-Bares J *et al.* (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.

Hijmans RJ, Cameron SE, Parra JL *et al.* (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Holt BG, Lessard JP, Beorregaard MK *et al.* (2013) An update of Wallace's zoogeographic regions of the World. *Science*, **339**, 74–77.

Huang Z, Liu N, Liang W *et al.* (2010) Phylogeography of Chinese bamboo partridge, *Bambusicola thoracica thoracica* (Aves: Galliformes) in south China: inference from mitochondrial DNA control-region sequences. *Molecular Phylogenetics and Evolution*, **56**, 273–280.

Jakob SS, Martinez-Meyer E, Blattner FR (2009) Phylogeographic analyses and paleodistribution modeling indicate Pleistocene *in situ* survival of *Hordeum* species (Poaceae) in southern Patagonia without genetic or spatial restriction. *Molecular Biology and Evolution*, **26**, 907–923.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics*, **6**, 13.

Ju L, Wang H, Jiang D (2007) Simulation of the Last Glacial Maximum climate over East Asia with a regional climate modelnested in a general circulation model. *Palaeogeography, Palaeoclimatology, and Palaeoecology*, **248**, 376–390.

Kurihara T (1974) Natural enemies of mosquito. *Seikatsu to Kankyou*, **19**, 21–26.

Lehrian S, Bálint M, Haase P, Pauls SU (2010) Genetic population structure of an autumn emerging caddisfly with inherently low dispersal capacity and insights into its phylogeography. *Journal of the North American Benthological Society*, **29**, 1100–1118.

Li M, Wei F, Goossens B *et al.* (2005) Mitochondrial phylogeography and subspecific variation in the red panda (*Ailurus fulgens*): implications for conservation. *Molecular Phylogenetics and Evolution*, **36**, 78–89.

Li SH, Yeung CKL, Feinstein J *et al.* (2009) Sailing through the Late Pleistocene: unusual historical demography of an East Asian endemic, the Chinese Hwamei (*Leucodioptron canorum canorum*), during the last glacial period. *Molecular Ecology*, **18**, 622–633.

Lin HD, Chen YR, Lin SM (2012) Strict consistency between genetic and topographic landscapes of the brown tree frog (*Buergeria robusta*) in Taiwan. *Molecular Phylogenetics and Evolution*, **62**, 251–262.

Long Y, Wan H, Yan FM *et al.* (2006) Glacial effects on sequence divergence of mitochondrial COII of *Polyura eudamippus* (Lepidoptera: Nymphalidae) in China. *Biochemical Genetics*, **44**, 361–377.

Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". *Human Biology*, **76**, 173–190.

Miyamoto S, Lee CE (1963) Water striders of Korea (Hemiptera, Heteroptera). *Kontyû*, **31**, 33–47.

Myers N, Mittermeier RA, Mittermeier CG *et al.* (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nakasuji F, Dyck VA (1984) Evaluation of the role of *Microvelia douglasi atrolineata* (Bergroth) (Heteroptera: Veliidae) as predator of the brown planthopper *Nilaparvata lugens* (Stål) (Homoptera: Delphacidae). *Researches on Population Ecology*, **26**, 134–149.

Nenzén HK, Araújo MB (2011) Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, **222**, 3346–3354.

Pauls SU, Lumbsch T, Haase P (2006) Phylogeography of the montane caddisfly Drusus discolor: evidence for multiple refugia and periglacial survival. *Molecular Ecology*, **15**, 2153–2169.

Pauls SU, Theissinger K, Ujvarosi L, Bálint M, Haase P (2009) Patterns of population structure in two closely related, partially sympatric caddisflies in Eastern Europe: historic introgression, limited dispersal, and cryptic diversity. *Journal of North American Benthological Society*, **28**, 517–536.

Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

Pearson RG (2007) Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History, Available at http://ncep.amnh.org.

Peng YN, Wan Y, Luo LS (2000) The study on the diversity and sustainable development in Hengduanshan Mountain of Yunnan. *Human Geography*, **15**, 50–53.

Peterson AT (2011) Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography*, **38**, 817–828.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pinot S, Ramstein G, Harrison SP *et al.* (1999) Tropical paleoclimates at the last glacial maximum: comparison of Paleoclimate Modeling Intercomparison Project (PMIP) simulations and paleodata. *Climate Dynamics*, **15**, 857–874.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237–1245.

Porretta D, Mastrantonio V, Bellini R *et al.* (2012) Glacial history of a modern invader: phylogeography and species distribution modelling of the Asian tiger mosquito *Aedes albopictus*. *PLoS ONE*, **7**, e44515.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Qian H, Ricklefs RE (2000) Large-scale processes and the Asian bias in temperate plant species diversity. *Nature*, **407**, 180–182.

Qu Y, Luo X, Zhang R *et al.* (2011) Lineage diversification and historical demography of a montane bird *Garrulax elliotii* – implications for the Pleistocene evolutionary history of the eastern Himalayas. *BMC Evolutionary Biology*, **11**, 174.

Rangel TF, Diniz-Filho JAF, Bini LM (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology and Biogeography*, **15**, 321–327.

Reineke A, Karlovsky P, Zebitz CPW (1998) Preparation and purification of DNA from insects for AFLP analysis. *Insect Molecular Biology*, **7**, 95–99.

Richards CL, Carstens BC, Knowles LL (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**, 1833–1845.

Rost KT (1994) Paleoclimatic field studies in and along the Qinling Shan (Central China). *GeoJournal*, **34**, 107–120.

Rozas J, Sanchez-DelBarrio JC, Messeguer X *et al.* (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Ruddiman WF, Raymo ME, Martinson DG *et al.* (1989) Pleistocene evolution: northern hemisphere ice sheets and North Atlantic Ocean. *Paleoceanography*, **4**, 353–412.

Rull V (2009) Microrefugia. *Journal of Biogeography*, **36**, 481–484.

Schaal BA, Hayworth DA, Oseni KM *et al.* (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**, 465–474.

Schaefer JM, Oberholzer P, Zhao ZZ *et al.* (2008) Cosmogenic beryllium-10 and neon-21 dating of late Pleistocene glaciations in Nyalam, monsoonal Himalayas. *Quaternary Science Reviews*, **27**, 295–311.

Schmitt T (2007) Molecular biogeography of Europe: pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Taubmann J, Theissinger K, Feldheim KA *et al.* (2011) Modelling range shifts and assessing genetic diversity distribution of the montane aquatic mayfly *Ameletus inopinatus* in Europe under climate change scenarios. *Conservation Genetics*, **12**, 503–515.

Wang J (2010) Glacial advance in the Qinghai-Xizang Plateau and peripheral mountains during the mid-MIS3. *Quaternary Sciences*, **30**, 1055–1065.

Wang XP, Liu YK (1994) *Theory and Practice of Biodiversity*. China Environmental Science Press, Beijing.

Wang R, Abelmann A, Li B *et al.* (2000) Abrupt variations of the radiolarian fauna at Mid-Pleistocene climate transition in the South China Sea. *Chinese Science Bulletin*, **45**, 952–955.

Wang J, Gao PX, Kang M *et al.* (2009) Refugia within refugia: the case study of a canopy tree (*Eurycorymbus cavaleriei*) in subtropical China. *Journal of Biogeography*, **36**, 2156–2164.

Wang W, McKay BD, Dai C *et al.* (2012) Glacial expansion and diversification of an East Asian montane bird, the green-backed tit (*Parus monticolus*). *Journal of Biogeography*, **40**, 1156–1169.

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.

Weaver AJ, Eby M, Fanning AF *et al.* (1998) Simulated influence of carbon dioxide, orbital forcing and ice sheets on the climate of the Last Glacial Maximum. *Nature*, **394**, 847–853.

Wu YK, Wang YZ, Jiang K *et al.* (2013) Significance of pre-Quaternary climate change for montane species diversity: insights from Asian salamanders (Salamandridae: *Pachytriton*). *Molecular Phylogenetics and Evolution*, **66**, 380–390.

Xing FW, Wu TL, Li ZX *et al.* (1995) Endemic plants of Hainan island. *Journal of Tropical and Subtropical Botany*, **3**, 1–12.

Yang JQ, Tang WQ, Liao TY *et al.* (2012) Phylogeographical analysis on *Squalidus argentatus* recapitulates historical landscapes and drainage evolution on the island of Taiwan and mainland China. *International Journal of Molecular Sciences*, **13**, 1405–1425.

Zhang RZ (1999) *Zoogeography of China*, pp. 1–502. Science Press, Beijing.

Zhang MW, Rao DQ, Yang JX *et al.* (2010) Molecular phylogeography and population structure of a midelevation montane frog *Leptobrachium ailaonicum* in a fragmented habitat of southwest China. *Molecular Phylogenetics and Evolution*, **54**, 47–58.

Zhao JD, Zhou SZ, Liu SY *et al.* (2007) A preliminary study of the glacier advance in MIS3b in the western regions of China. *Journal of Glaciology and Geocryology*, **29**, 233–241.

Zhao JD, Shi YF, Wang J (2011) Comparison between quaternary glaciations in china and the marine oxygen isotope stage (MIS): an improved schema. *Acta Geographica Sinica*, **66**, 867–884.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

Z.Y. analysed the molecular data and performed the research; G.P.Z. contributed to ecological niche modelling analysis; D.L.Z. performed the nuclear sequence data; P.P.C. and W.J.B. designed the research. The manuscript was written by Z.Y., G.P.Z., P.P.C. and W.J.B.

Data accessibility

DNA sequences have been deposited in GenBank under Accession nos KJ810577-KJ810577, KJ818903-KJ819539.

Details regarding individual samples are available in doi:10.5061/dryad.sd5 m9.

DNA sequence alignments, Bioclimate data and Maxent input files: doi:10.5061/dryad.sd5 m9.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Primer sequences used to amplify the mitochondrial and nuclear fragments.

Table S2 Principal component analysis (PCA) of bioclimatic variables associated with *M. douglasi douglasi* occurrence; eigenvalues for the most important variables (> 0.8) in PCA are in bold.

Table S3 Nucleotide polymorphism in each population with geographical coordinates.

Fig. S1 Pairwise F_{ST} values for the four defined subregions based on mitochondrial data.

Fig. S2 Zones of genetic discontinuities.

Fig. S3 Yule tree of the *M. douglasi douglasi* in East Asia and divergence time among the four haplogroups, constructed with the mitochondrial DNA.

Fig. S4 Principle component analysis of 7 bioclimatic variables associated with occurrence of *M. douglasi douglasi*.

Fig. S5 A map that compared the extent of Pleistocene ice sheets during the LGM in the North America, Europe and East Asia.