

RESEARCH ARTICLE

Open Access



Genetic and geographical structure of boreal plants in their southern range: phylogeography of *Hippuris vulgaris* in China

Qixiang Lu[†], Jinning Zhu[†], Dan Yu^{*} and Xinwei Xu^{*}

Abstract

Background: Our current understanding of the evolutionary history of boreal and arctic-alpine plants in their southern range in Asia remains relatively poor. Using three cpDNA non-coding regions and nine nuclear microsatellite (nSSR) loci, we examine the phylogeographic pattern in a broad geographic sampling of the boreal plant *Hippuris vulgaris* to infer its dispersal and diversification in China. In addition, the species distributions at the Last Glacial Maximum (LGM) and at present were predicted using ecological niche modeling (ENM).

Results: The cpDNA results revealed two distinct lineages, A and B. A is restricted to Northeast China; B is distributed in Northwest China, the Qinghai-Tibet Plateau (QTP) and North and Northeast (NNE) China; and A and B diverged ca. 1.36 Ma. The nSSR data revealed two genetic clusters corresponding to the two cpDNA lineages and nonreciprocal hybridization with lineage A as the maternal lineage in Northeast China. Cluster B further divided into three subclusters: I, mainly in NNE China and the northeastern border of the QTP; II, in Northwest China and the QTP; and III, on the QTP. ENM predicted a marked range shift on the QTP at the LGM, retreating from the platform to the northeast and southeast edges.

Conclusions: *Hippuris vulgaris* probably diverged into lineages A and B in high latitudes and then immigrated into Northeast China and Northwest China, respectively. Lineage A persisted and diversified in Northeast China. Lineage B reached the QTP during the mid-Pleistocene, diversified in that region due to the influence of climatic oscillations, migrated into Northeast China and subsequently hybridized with lineage A. Our findings give empirical evidence that boreal plants display complex evolutionary history in their southern range in Asia and provide new insights into the evolution of boreal and arctic-alpine plants.

Keywords: Boreal plants, China, Chloroplast DNA, Ecological niche modeling, *Hippuris vulgaris*, Microsatellites, Phylogeography

Background

Climatic oscillations in the Quaternary have greatly affected the distribution and abundance of boreal and arctic species [1]. These species have undergone major range shifts with the advances and retreats of ice sheets. In the past two decades, numerous studies have focused on plant species in boreal regions and/or arctic latitudes and southern

mountain ranges, and different phylogeographic patterns amongst plants have been revealed, especially in Europe and North America [2–5]. Some of these studies included samples from mountains in Central Asia have revealed that Central Asia is an important refugium for boreal and arctic plants and colonization from Central Asia to Europe or to Beringia are conceivable [6–10]. However, colonization routes of these species between Central Asia and mountains in low latitudes in Asia need to be further explored. At present, our understanding of the southern range dynamics of boreal and arctic-alpine plants in Asia is based

* Correspondence: lakeyd@163.com; xuxw@whu.edu.cn

[†]Equal contributors

National Field Station of Freshwater Ecosystem of Liangzi Lake, College of Life Sciences, Wuhan University, Wuhan, P. R. China



mostly on data from Japan [11–13]. Phylogeographic studies with a broader geographic sampling in Asia are necessary to better elucidate evolutionary histories of boreal and arctic-alpine plants.

China provides suitable habitats for many boreal and arctic-alpine plant species in their southern range, including the Qinghai-Tibetan Plateau (QTP) and mountain ranges in Northwest China and Northeast China [14]. Two putative routes have been suggested for the migration of the flora between the QTP and the arctic regions: one is the “Central Asiatic Highland Corridor” pathway via mountain ranges in Northwest China [15], and another is the “Himalayan-Hengduan Mountain-Qinling-Northeast China” route [16]. We thus hypothesized that similar colonization routes might also be present in a single species of this plant group. In addition, the QTP is considered a center of the development and differentiation of the elements from boreal or Arctic-Tertiary flora since the late Tertiary and Quaternary glacial periods [17]. Similarly, infra-specific lineage divergence in boreal and arctic-alpine plants might also occur in this region.

To infer refugial areas and colonization routes of boreal and arctic-alpine plants in China, we should conduct phylogeographic studies on species with a widespread range covering the QTP and mountains in northeast and northwest China. In this study, we focused on *Hippuris vulgaris* (Plantaginaceae), an aquatic plant of circumboreal distribution. China is the southern range of this species in Asia, and *H. vulgaris* is mainly distributed in the north, northeast, northwest and southwest of China [18]. The plant is rhizomatous perennial and often grows in fresh water at the edges of ponds, lakes and rivers. A recent chloroplast DNA (cpDNA) phylogeography of *H. vulgaris* in the QTP and adjacent areas revealed a mid-to-late Pleistocene split of two lineages on the QTP and possible colonization between the QTP and Northwest China [19], whereas no specific refugia areas were deduced and no samples from Northeast China were included. We hypothesized that colonization between the QTP and Northeast China was likely existed in *H. vulgaris*. In the present study, we used both cpDNA sequences and nuclear microsatellite (nSSR) data to examine the phylogeographic pattern of *H. vulgaris* based on a large-scale sampling covering its distribution range in China. In addition, we inferred the glacial range dynamics of *H. vulgaris* by using ecological niche modeling (ENM; [20]). Our aims were to (1) test whether lineage divergence on the QTP was detected by both cpDNA and nSSR, and (2) deduced potential refugia and putative colonization routes of *H. vulgaris* in China. Investigating these questions was conducive to (3) elucidating the evolutionary history of *H. vulgaris* in its southern range in Asia.

Methods

Plant materials

A total of 637 individuals were collected at 91 sites throughout the distribution range of *H. vulgaris* in China, including Northwest China, the QTP, and North and Northeast (NNE) China (Additional file 1). At each site, 1–16 individuals were sampled at intervals of at least 5 m according to the population size. The leaf material of each individual was collected and immediately stored in silica gel. Voucher specimens were deposited in the herbarium of Wuhan University (WH).

DNA extraction, amplification and sequencing

Total genomic DNA was extracted from silica-dried leaves using the DNA secure Plant Kit (Tiangen Biotech, Beijing, China). Six cpDNA non-coding regions were examined in preliminary experiment in 2012 and three regions, *trnH-psbA* [21], *trnQ-rps16* and *rps16-trnK* [22], were chosen due to their high polymorphism (Additional file 2). Polymerase chain reaction (PCR) was performed in a volume of 25 μ L containing 10–30 ng of genomic DNA, 0.1 μ M of each primer, 0.2 mM of each dNTP, 2 mM $MgCl_2$ and 0.6 U of ExTaq DNA polymerase (TaKaRa, Dalian, China) under the following conditions: 3 min at 95 $^{\circ}C$, followed by 35 cycles of 30 s at 95 $^{\circ}C$, 30 s at 55 $^{\circ}C$, and 90 s at 72 $^{\circ}C$, and then a final 5 min extension at 72 $^{\circ}C$. Amplifications were conducted in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Foster City, USA). The PCR products were purified and sequenced by the Beijing Genomic Institute in Wuhan, China.

Microsatellite loci development and genotyping

We developed nSSR primers using the microsatellite-enriched library method. Detailed methods and processes followed Wu *et al.* [23]. A total of nine nSSR loci were developed for subsequent analysis (Additional file 3). Microsatellites were amplified under the following conditions: 3 min at 95 $^{\circ}C$; 35 cycles of 30 s at 95 $^{\circ}C$, 30 s at 49–57 $^{\circ}C$ (Additional file 3), and 1 min at 72 $^{\circ}C$; and a final extension at 72 $^{\circ}C$ for 5 min. PCR products were analyzed on an ABI 3730XL, and genotyping was performed using GeneMapper version 4.0 software (Applied Biosystems).

Population genetic structure analyses

All cpDNA sequences were aligned using the program MAFFT 6.7 [24]. The three chloroplast fragments were concatenated for subsequent analyses. Identical sequences were collapsed into a single haplotype using DNASP 5.10 [25], and all sequences of different haplotypes were deposited in GenBank (Accession Nos. KT921227–KT921258). A statistical parsimony network was constructed using TCS 1.21 [26], treating sequence gaps as the fifth

character state. In the network analysis, gaps of two or more base pairs were re-coded as single-base-pair indels.

Regarding the nSSR data, genetically distinct clusters of 91 populations were identified using an individual-based assignment approach as implemented in STRUCTURE 2.3.4 [27]. Ten independent runs were performed for each K value (K = 1 to 16) with a burn-in period of 20,000 iterations and 100,000 Markov chain Monte Carlo (MCMC) iterations under the admixture model. The best-fit number of clusters was determined based on the ΔK method [28]. The geographical distribution of different cpDNA haplotypes and nSSR clusters identified was mapped using ArcGIS 9.0 (Esri, Redlands, CA, USA). For both cpDNA and nSSR data, analyses of molecular variance (AMOVA) were used to partition genetic variation among and within populations, as implemented in ARLEQUIN 3.1 [29]. The individuals with the same nSSR genotype were removed and only one was kept within a population in genetic analyses.

Phylogenetic reconstruction and divergence time estimation

The phylogenetic relationships of cpDNA haplotypes were inferred using maximum likelihood (ML) analysis implemented in GARLI 0.951 [30], starting from random trees and using 10,000,000 generations per search. The ML bootstrap support was estimated from 1000 bootstrap replicates in GARLI. Bayesian inference (BI) implemented in MrBayes 3.1.2 [31] was also used for phylogenetic reconstruction. Two independent MCMC analysis runs were conducted simultaneously, beginning with a random tree, with each run including four chains (one cold and three hot). Two million generations were run, with sampling at every 1,000 generations. Chain convergence was checked using Tracer 1.5 [32], and the first 25 % of samples were discarded as burn-in. The K81uf + I substitution model for the ML and BI analyses was identified using the Akaike information criterion (AIC) implemented in ModelTest 3.7 [33]. The divergence times of cpDNA lineages were estimated by using a Bayesian method implemented in BEAST 1.4.7 [34]. An evolutionary rate of 1.52×10^{-9} s/s/y [35] for the three combined cpDNA non-coding regions was used to estimate divergence time under an uncorrelated lognormal relaxed clock model. MCMC analysis of 50,000,000 generations was implemented, in which every 1,000 generations were sampled. The first 10 % of generations were discarded as burn-in, and the parameters were checked using Tracer 1.5.

Demographic analyses

Regarding the cpDNA data, we examined pairwise mismatch distributions to detect historical demographic expansions using ARLEQUIN. The sum of squared deviations

(SSD) and Harpending's raggedness index (Rag) [36] were used to test the goodness of fit under a sudden-expansion model. Their significances were tested using a parametric bootstrap approach with 1,000 replicates [37]. We also performed two powerful tests, Fu's F_S test [38] and the R_2 test [39], to detect potential population growth using DNASP.

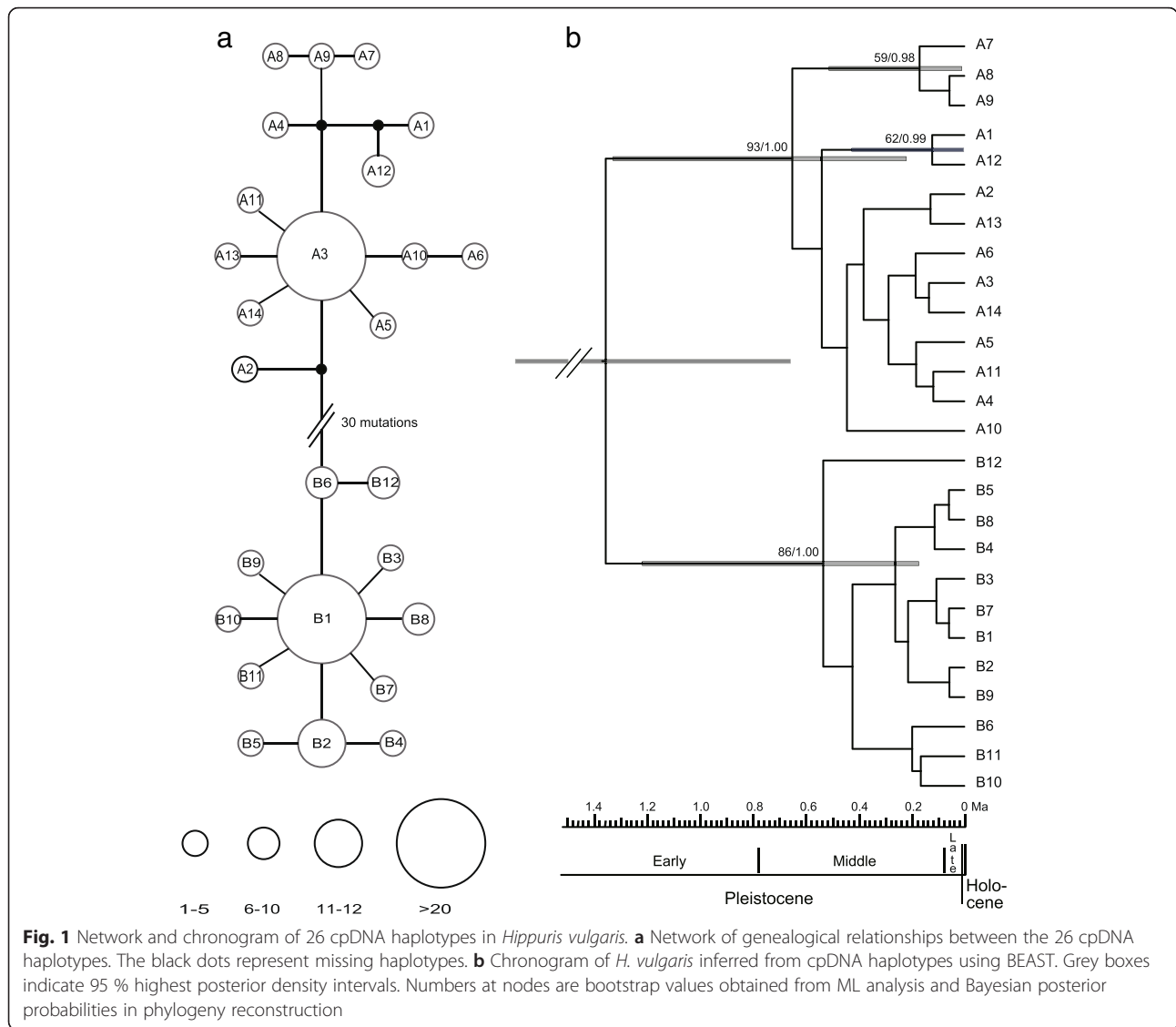
Present and past distribution modeling

ENM was used to predict the suitable past (Last Glacial Maximum, LGM) and present climate envelopes for each lineage (see results), as implemented in MAXENT 3.3.3 k [20, 40]. In addition to the 91 sites in this study (Additional file 1), 47 sites from Chen *et al.* [19] and 68 collection records from the Chinese Virtual Herbarium (www.cvh.org.cn) were included. Based on this total of 206 records, we obtained an environmental data set comprising 19 BioClim variables and altitudes from the WORLDCLIM database with resolutions of 2.5' ([41]; www.worldclim.org). We examined pairwise correlations among the 20 variables and retained seven environmental variables (annual mean temperature, temperature seasonality, temperature annual range, mean temperature of warmest quarter and coldest quarter, precipitation of warmest quarter and coldest quarter) with pairwise Pearson correlation coefficients of $r < 0.7$ to minimize biased fitting of niche models. A current distribution model was developed using these seven bioclimatic data layers. This model was then projected onto the LGM dataset simulated by the Community Climate System Model v3.0 [42]. We used the default parameters of MAXENT to construct ENMs. The accuracy of each model prediction was evaluated using the area under the (receiver operating characteristic) curve (AUC).

Results

Chloroplast DNA haplotype network

The lengths of the aligned sequences for *trnH-psbA*, *trnQ-rps16* and *rps16-trnK* were 436, 970 and 829 bp, respectively. A total of 56 polymorphic sites, including 21 indels and 35 base substitutions, were observed in the concatenated sequences. The sequences were collapsed into 26 haplotypes (A1-A14 and B1-B12). A haplotype network was constructed with two distinct groups: A and B (Fig. 1a). Group A was restricted to mountainous areas in Northeast China. Group A was found in 28 populations, with haplotype A3 widespread and present in 21 populations, haplotypes A4 and A12 each found in two populations, and other haplotypes each restricted to a single population (Fig. 2a). Group B occurred in all three regions and was found in 64 populations. The most common haplotype, B1, was present in 59 populations; among the remaining 11 haplotypes, only B2 was shared between the QTP and NNE China. Northwest China, the QTP and NNE China contained one, three, and six private



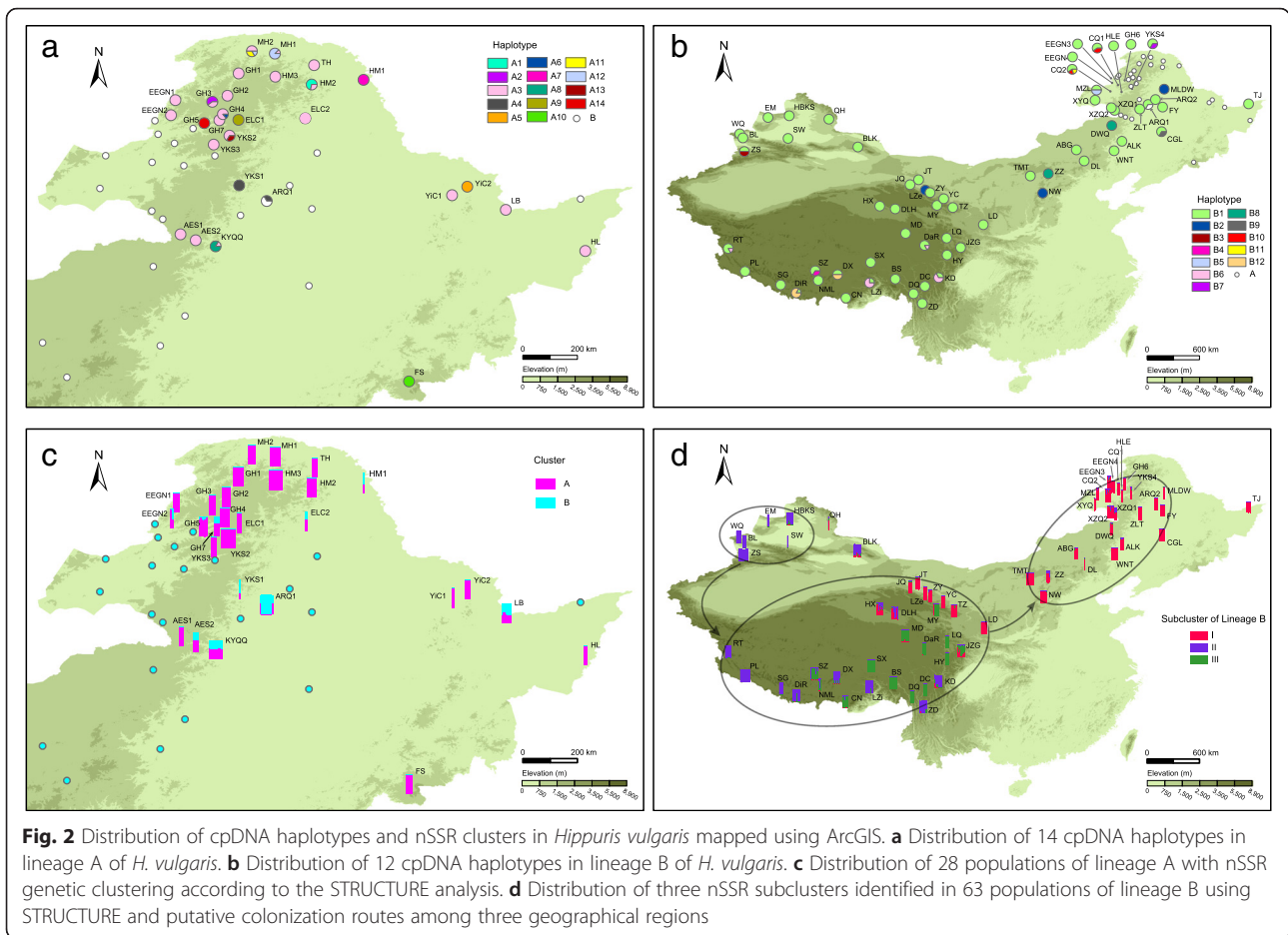
haplotypes, respectively (Fig. 2b). Only one population, ARQ1, contained haplotypes from both group A (A4) and group B (B1) (Fig. 2a, b).

Phylogenetic lineages and divergence time

The trees generated using ML and BI analyses showed the same topology, in which two reciprocally monophyletic lineages (A and B) were recovered with robust support (Fig. 1b). These two lineages corresponded to the two groups in the haplotype network. The divergence time of the two lineages was estimated at 1.36 Ma ago (Ma) (with a 95 % highest posterior density [HPD] interval of 0.66 - 2.52 Ma), which yields time estimates in the early Pleistocene. The diversification of lineages A and B was estimated at 0.65 Ma (95 % HPD: 0.22 - 1.33 Ma) and 0.53 Ma (95 % HPD: 0.17 - 1.22 Ma), respectively (Fig. 1b), both of which are in the mid-Pleistocene.

Population genetic structure

The STRUCTURE analysis using SSR data suggested K = 2 as the optimal number of genetic clusters based on the calculation of ΔK (Additional file 4). We defined these two clusters as A and B. All individuals of 18 and 63 populations were assigned to cluster A and B with a posterior probability higher than 0.95 (Additional file 5) and had cpDNA lineages A and B, respectively (Fig. 2c, d). Most or all individuals of nine populations were admixed with a probability range from 0.326 to 0.890 (Additional file 5) and had cpDNA lineage A (Fig. 2c). In population ARQ1, two individuals with admixed assignment had cpDNA lineage A, and other individuals with cluster B assignment had cpDNA lineage B (Additional file 5). All of the results suggested that hybridization occurred between lineage A and B and that hybrids were produced only when lineage A was the ovulate parent. Further STRUCTURE analysis was



also performed for cluster A and cluster B, and the optimal solution was $K = 2$ and $K = 3$, respectively (Additional file 4). In cluster A, all individuals were admixed and no subclusters were identified (Additional file 5). In cluster B, most individuals were assigned to three subclusters with high posterior probability (Additional file 5). These three subclusters had different geographic distributions: subcluster I mainly occurred in NNE China and Northeast QTP edge, subcluster II was mostly found in Northwest China and the QTP, and subcluster III was restricted to the QTP (Fig. 2d).

Hierarchical AMOVA revealed that most of this species' total variation (92.78 % in cpDNA and 64.02 % in nSSR) occurred among lineages (Table 1), indicating high genetic differentiation between lineage A and B. Nonhierarchical AMOVA for lineage A revealed that most of the genetic variation in both cpDNA and nSSR was distributed within populations (Table 1). For lineage B, the AMOVA results revealed that most of the genetic variation in both cpDNA and nSSR was found within populations, with a low percentage occurring among three regions (Northwest China, the QTP, and NNE China), indicating frequent gene flow among populations and regions (Table 1).

Demographic analyses

Mismatch distributions for lineage A and lineage B were inconsistent with the unimodal curve expected for populations under conditions of demographic expansion (figures not shown). The demographic expansion of lineage A was supported by nonsignificant H_{Rag} value but not supported by significant SSD value and Fu's F_S and R_2 tests. Similarly, the demographic expansion of lineage B was supported by nonsignificant H_{Rag} and SSD values but not supported by Fu's F_S and R_2 tests (Table 2). These findings suggest that recent population expansion is not present in either of the two lineages.

Present and past (LGM) distribution of lineages

For lineages A and B, the AUC values for the ENM were 0.970 and 0.911, respectively, indicating that model prediction performed well. The predicted distribution of lineage A under the current conditions was similar to its actual distribution (Fig. 3a). At the LGM, the potential distribution of lineage A mainly occurred in lowland areas in Northeast China, differing from the present (Fig. 3b). For lineage B, the potential range under the current conditions was generally similar to its actual distribution

Table 1 Analyses of molecular variance based on cpDNA and nSSR data for populations of *Hippuris vulgaris*

Source of variation	cpDNA data		Microsatellite data	
	df	Percentage of total variance (%)	df	Percentage of total variance (%)
All populations				
Among lineages	1	92.78***	1	64.02***
Among populations within lineages	79	2.75***	79	13.33***
Within populations	131	4.47***	523	22.65***
Lineage A				
Among populations	17	36.90*	17	5.76
Within populations	25	63.10***	84	94.24***
Lineage B				
Among populations	62	38.35***	62	40.01***
Within populations	106	61.65***	439	59.99***
Lineage B				
Among regions	2	0.79	2	12.79***
Among populations within regions	60	37.75***	60	30.37***
Within populations	106	61.46***	439	56.84***

Note: * $P < 0.05$, *** $P < 0.001$

(Fig. 3c). The predicted distribution at the LGM differed greatly on the QTP, retreating entirely from the platform to the northeast edge and southeast edge (Fig. 3d).

Discussion

Both cpDNA and nSSR results show that *H. vulgaris* in China is composed of two distinct lineages A and B with different geographic distributions. The distribution ranges of three other species in the genus *Hippuris* are to the north of 53° north latitude, which are in the north of the range of *H. vulgaris* [43, 44]. In addition, the fossil records of *H. vulgaris*, e.g., one of Plio-Pleistocene age from North Greenland [45] and another of middle and upper Pleistocene from central England [46], were also found in the north area of its present distribution. Both species distribution and fossil record suggested that the ancestral range of *H. vulgaris* was in the north of its present range and probably to the north of 53° north latitude. Therefore, lineages A and B had likely split in high latitudes before they migrated into China. The occurrence of different infra-specific lineages in high latitudes of Asia has been revealed in some arctic-alpine plants [3, 5, 6, 8, 9, 47]. Our molecular dating

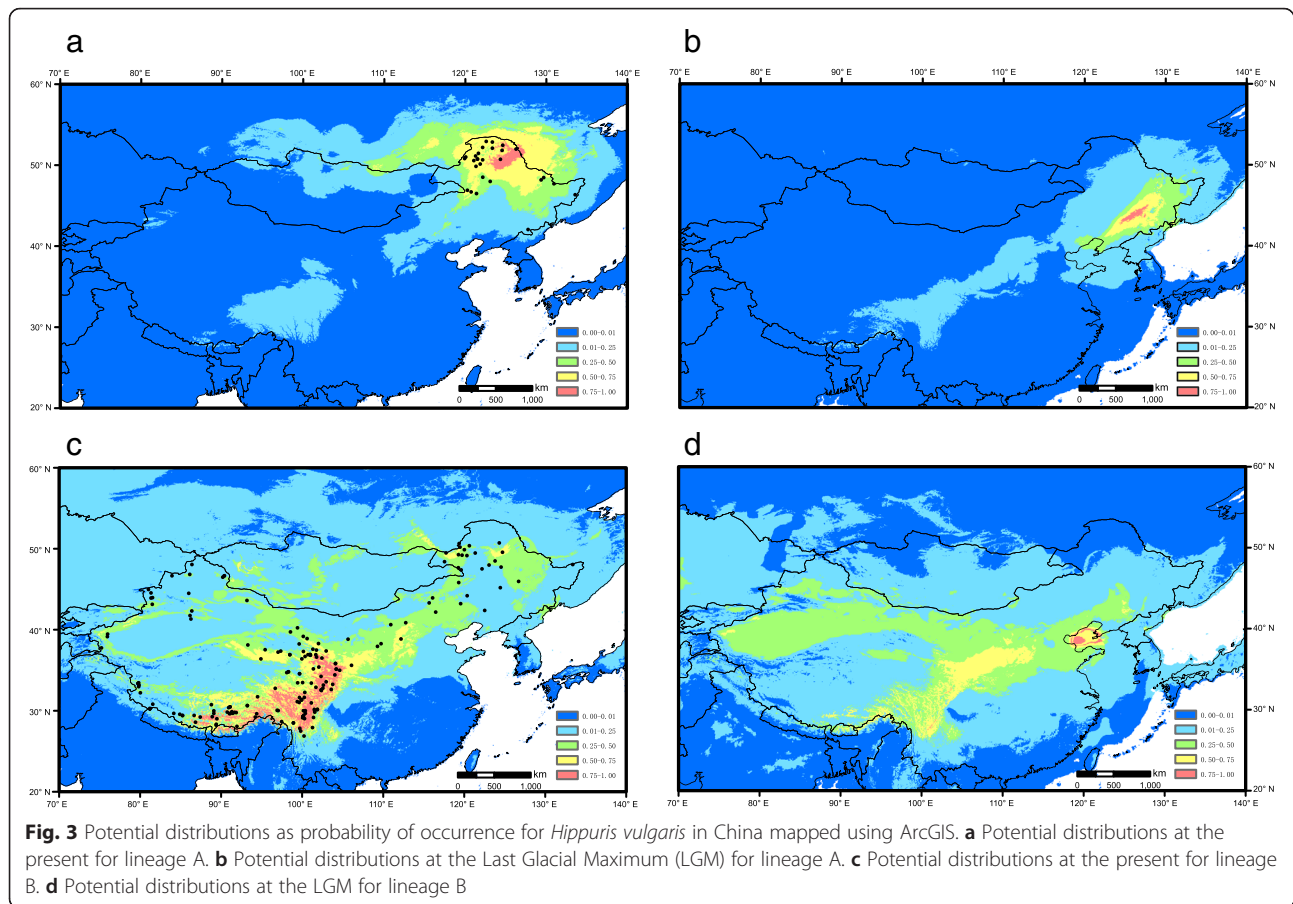
Table 2 Summary of mismatch distribution parameters and neutrality test for two lineages of *Hippuris vulgaris*

	SSD	P	H_{Rag}	P	Fu's F_S	P	R_2	P
Lineage A	0.257	0.001	0.235	0.983	2.108	0.832	0.053	0.069
Lineage B	0.053	0.067	0.457	0.572	6.791	0.949	0.019	0.104

indicates that the divergence of lineages A and B occurred during the early Pleistocene (*ca.* 1.36 Ma), a period displaying minor glacial/interglacial variability before the mid-Pleistocene revolution followed with higher-amplitude fluctuations [48], suggesting that the latitudinal range shifts during the early Pleistocene may not have been sufficient to push *H. vulgaris* southwards into China. The mid-Pleistocene stem ages of lineage A and lineage B (Fig. 1b) also suggest that the two lineages migrated into China driven by the mid-Pleistocene glaciations and thereafter diversified separately. This scenario needs to be tested by further studies including samples from high latitudes of Asia.

Lineage A is restricted to Northeast China, indicating its immigration southwards into this region. The stem age of lineage A in the mid-Pleistocene (*ca.* 0.65 Ma) and the occurrence of its potential distribution in the lowland areas of Northeast China at the LGM (Fig. 3b) suggest that long-standing populations have persisted in Northeast China for several glacial/interglacial cycles. For lineage B, the migration route is more complex. Because the development of a wide arid belt or dryland in northern China and the Republic of Mongolia since *ca.* 2.6 Ma [49, 50] probably acted as a strong physical barrier for aquatic plant dispersal between Northeast China and Northwest China, lineage B should immigrate into China either from Northeast China or from Northwest China. Direct dispersal between the two regions was unlikely, therefore, lineage B reached its present distribution by dispersal either from Northeast China via the QTP to Northwest China or via the reverse route. The former route seems likely because NNE China had the largest number of haplotypes (Fig. 2b), whereas the latter route is supported by the fact that only subcluster I was found in NNE China (Fig. 2d). Combined with the exclusive occupation of the north range of Northeast China by lineage A (Fig. 2a) and the occurrence of inter-lineage hybrids only in Northeast China (Fig. 2c), we considered it more likely that lineage B immigrated southwards into Northwest China, dispersed to the QTP and subsequently migrated into NNE China (Fig. 2d).

The dispersal between highlands in Northwest China and the QTP via the “Central Asiatic Highland Corridor” pathway has been suggested [15] for the migration of the flora between the QTP and the arctic regions. This dispersal pathway was also suggested by Wen *et al.* [51] and verified through the results of a few recent phylogenetic and phylogeographic studies: the same or closely related lineages were present in the two regions [8, 52–54]. As for *H. vulgaris*, its dispersal between Northwest China and the QTP was supported by a shared cpDNA haplotype (Fig. 2b; [19]) and the same genetic subcluster II (Fig. 2d) present in the two regions. This colonization was also in accordance with the migratory route of waterbirds in China [55, 56], which is an important vector for long-distance dispersal of aquatic



plants [57, 58]. Furthermore, the fact that *H. vulgaris* is common in the diet of waterbirds at northern latitudes [59] indicates the possibility of long-distance dispersal by waterbirds in *H. vulgaris*. The dispersal between the QTP and Northeast China likely occurred through the “Himalayan-Hengduan Mountain-Qinling-Northeast China” route suggested by Wang [16], based on the geographic distribution of some plants. However, for *H. vulgaris*, another route between Northeast QTP edge and Northeast China seems more likely because the populations from Northeast QTP edge were assigned to genetic subcluster I, the same subcluster as populations from NNE China (Fig. 2d). This route is consistent with the dispersal pathway of a subspecies of sea buckthorn [60].

The QTP has biogeographic significance for alpine plant evolution [51]. Based on evidence from phylogenetic studies on several taxa, Wen *et al.* [51] suggested a pattern of Central Asian origin with subsequent diversification on the QTP. This pattern is suitable for lineage B of *H. vulgaris*. The stem age of lineage B in the mid-Pleistocene (*ca.* 0.53 Ma) is similar to the result of Chen *et al.* [19] (*ca.* 0.48 Ma), estimated based on another cpDNA dataset collected from samples on the QTP, suggesting that *H. vulgaris* reached the QTP during the mid-Pleistocene.

Two lineages were revealed on the QTP in Chen *et al.* [19], whereas no cpDNA lineage divergence except three private haplotypes was detected in the present study (Figs. 2b and 3), likely resulting from the use of different fragments. When nSSR data were considered, three subclusters were identified on the QTP: subcluster I in northeast edge, subcluster II from west to east, and subcluster III from middle to east (Fig. 2d). Chen *et al.* found that two cpDNA lineages was overlapped distribution and co-occurred in some populations due to repeated range expansions [19]. This conclusion is confirmed by our nSSR data because the same pattern was also revealed in our populations of subclusters II and III (Fig. 2d) and the sampling of Chen *et al.* [19] did not cover Northeast QTP edge. In addition, we tried to use ENM to infer potential glacial refugia of *H. vulgaris* based on BioClim variables. The distributions of aquatic plants species are correlated with climatic factors [61–63] and occurrences of different genetic lineages are associated with environmental dissimilarity or latitudinal difference in some widespread species [64, 65], indicating that eco-climatological niche modeling based on BioClim variables is applicable for aquatic plants. Based on the prediction of ENM, a marked range shift occurred in *H. vulgaris* at the LGM, retreating

from the platform to the northeast edge and southeast edge (Fig. 3c, d). This scenario and the locations of potential glacial refugia have been suggested by many phylogeographic studies on the QTP [66, 67]. The lineage divergence of *H. vulgaris* on the QTP was likely promoted by repeated climatic oscillations during mid-late Pleistocene.

Conclusions

Our study reveals that *Hippuris vulgaris* is highly structured genetically and geographically in China, with signs of immigration, dispersal, diversification and hybridization. This study provides empirical evidence that boreal plants display complex evolutionary history in their southern range in Asia. Comparative studies on other boreal and arctic-alpine plants would be valuable for better understanding the colonization and diversification in their southern range in Asia.

Availability of supporting data

The data set supporting the results of this article is included within the article and its additional files. The 26 cpDNA sequences supporting the results of this article are available in the National Center for Biotechnology Information (GenBank) under accession numbers KT921227-KT921258, <http://www.ncbi.nlm.nih.gov/Genbank/>. Moreover, the nSSR data and datasets for ecological niche modeling are deposited in LabArchives (<http://www.labarchives.com/bmc>) and are available via the DOI <http://dx.doi.org/10.6070/H4K64G3F>.

Additional files

Additional file 1: Geographic origins, sample sizes and cpDNA haplotypes of the 91 *Hippuris vulgaris* populations studied. (DOCX 28 kb)

Additional file 2: Polymorphisms in six cpDNA non-coding regions screened in preliminary experiment based on 8 individuals of *Hippuris vulgaris*. (DOCX 17 kb)

Additional file 3: Characteristics of nine nuclear microsatellite markers developed in *Hippuris vulgaris*. (DOC 38 kb)

Additional file 4: Modeling of the numbers of genetic clusters in *Hippuris vulgaris* for (a) all 91 populations, (b) the 18 populations in lineage A, and (c) the 63 populations in lineage B, respectively, using STRUCTURE. (DOC 140 kb)

Additional file 5: The bar plot depicts the STRUCTURE admixture coefficients for individuals of *Hippuris vulgaris*. (a) The bar plot for all populations when $K = 2$. (b) The bar plot for 18 populations of lineage A with no hybrids when $K = 2$. (c) The bar plot for the 63 populations of lineage B when $K = 3$. A single vertical bar displays the membership coefficient of each individual, with population names shown on the bottom. (DOCX 325 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XX and DY conceived the ideas; QL and JZ collected the data; and QL, JZ and XX analyzed the data and led the writing. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China to Xinwei Xu (31070190 and 31270265) and Dan Yu (30930011). We thank the members of Dan Yu's group for their advice and field assistance.

Received: 7 October 2015 Accepted: 28 January 2016

Published online: 09 February 2016

References

- Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. *Phil Trans R Soc B*. 2004;359:183–95.
- Abbott RJ, Brochmann C. History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Mol Ecol*. 2003;12:299–313.
- Schönswetter P, Popp M, Brochmann C. Central Asian origin of and strong genetic differentiation among populations of the rare and disjunct *Carex atrofusca* (Cyperaceae) in the Alps. *J Biogeogr*. 2006;33:948–56.
- Wróblewska A. The role of disjunction and postglacial population expansion on phylogeographical history and genetic diversity of the circumboreal plant *Chamaedaphne calyculata*. *Biol J Linn Soc*. 2012;105:761–75.
- Eidesen PB, Ehrich D, Bakkestuen V, Alsos IG, Gilg O, Taberlet P, et al. Genetic roadmap of the Arctic: plant dispersal highways, traffic barriers and capitals of diversity. *New Phytol*. 2013;200:898–910.
- Schönswetter P, Popp M, Brochmann C. Rare arctic-alpine plants of the European Alps have different immigration histories: the snow bed species *Minuartia biflora* and *Ranunculus pygmaeus*. *Mol Ecol*. 2006;15:709–20.
- Skrede I, Eidesen PB, Portela RP, Brochmann C. Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). *Mol Ecol*. 2006;15:1827–40.
- Allen GA, Marr KL, McCormick LJ, Hebda RJ. The impact of Pleistocene climate change on an ancient arctic-alpine plant: multiple lineages of disparate history in *Oxyria digyna*. *Ecol Evol*. 2012;2:649–65.
- Winkler M, Tribsch A, Schneeweiss GM, Brodbeck S, Gugerli F, Holderegger R, et al. Tales of the unexpected: Phylogeography of the arctic-alpine model plant *Saxifraga oppositifolia* (Saxifragaceae) revisited. *Mol Ecol*. 2012;21:4618–30.
- Allen GA, Marr KL, McCormick LJ, Hebda RJ. Geographical origins, migration patterns and refugia of *Sibbaldia procumbens*, an arctic-alpine plant with a fragmented range. *J Biogeogr*. 2015;42:1665–76.
- Ikeda H, Senni K, Fujii N, Setoguchi H. Survival and genetic divergence of an arctic-alpine plant, *Diapensia lapponica* subsp. *obovata* (Fr. Schm.) Hultén (Diapensiaceae), in the high mountains of central Japan during climatic oscillations. *Plant Syst Evol*. 2008;272:197–210.
- Ikeda H, Senni K, Fujii N, Setoguchi H. High mountains of the Japanese archipelago as refugia for arctic-alpine plants: phylogeography of *Loiseleuria procumbens* (L.) Desvaux (Ericaceae). *Biol J Linn Soc*. 2009;97:403–12.
- Ikeda H, Yoneta Y, Higashi H, Eidesen PB, Barkalov V, Yakubov V, et al. Persistent history of the bird-dispersed arctic-alpine plant *Vaccinium vitis-idaea* L. (Ericaceae) in Japan. *J Plant Res*. 2015;128:437–44.
- Wu ZY, Zhou ZK, Sun H, Li DZ, Peng H. *The Areal-Types of Seed Plants and Their Origin and Differentiation*. Kunming: Yunnan Science and Technology Press; 2006.
- Ohba H. *The Alpine Flora of the Nepal Himalayas: An Introductory Note*. Tokyo: University of Tokyo Press; 1988.
- Wang WT. On some distribution patterns and some migration routes found in the eastern Asiatic region. *Acta Phytotaxon Sin*. 1992;30:1–24.
- Sun H. Evolution of arctic-tertiary flora in Himalayan-Hengduan mountains. *Acta Bot Yunnan*. 2002;24:671–88.
- Chen JR, Funston AM. Hippuridaceae. In: Wu ZY, Raven PH, editors. *Flora of China*, vol. 13. Beijing: Science Press and St. Louis: Missouri Botanical Garden Press; 2007. p. 433.
- Chen JM, Du ZY, Sun SS, Gituru RW, Wang QF. Chloroplast DNA phylogeography reveals repeated range expansion in a widespread aquatic herb *Hippuris vulgaris* in the Qinghai-Tibetan Plateau and adjacent areas. *Plos One*. 2013;8, e60948.
- Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. 2006;190:231–59.

21. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, et al. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot.* 2005;92:142–66.
22. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot.* 2007;94:275–88.
23. Wu ZG, Yu D, Xu XW. Development of microsatellite markers in the hexaploid aquatic macrophyte, *Myriophyllum spicatum* (Haloragaceae). *Appl Plant Sci.* 2013;2, e1200230.
24. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* 2008;9:286–98.
25. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25:1451–2.
26. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. *Mol Ecol.* 2000;9:1657–9.
27. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945–59.
28. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005;14:2611–20.
29. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online.* 2005;1:47–50.
30. Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin; 2006.
31. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19:1572–4.
32. Rambaut A, Drummond AJ. Tracer: MCMC trace analysis tool, version 1.5. 2007. <http://tree.bio.ed.ac.uk/software/tracer>. Accessed 19 May 2014.
33. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 1998;14:817–8.
34. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7:214.
35. Yamane K, Yano K, Kawahara T. Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Res.* 2006;13:197–204.
36. Harpending HC. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol.* 1994;66:591–600.
37. Schneider S, Excoffier L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics.* 1999;152:1079–89.
38. Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics.* 1997;147:915–25.
39. Ramos-Onsins SE, Rozas J. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol.* 2002;19:2092–100.
40. Phillips SJ, Dudik M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography.* 2008;31:161–75.
41. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005;25:1965–78.
42. Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, et al. The community climate system model version 3 (CCSM3). *J Clim.* 2006;19:2122–43.
43. Elven R, Murray DF, Solstad H. *Hippuris* (Plantaginaceae). In: Flora of North America, Provisional Publication. Flora of North America Association. 2012. <http://floranorthamerica.org/files/Hippuris03h.CH> for Prov Pub.pdf. Accessed 15 May 2015.
44. Secretariat GBIF. GBIF Backbone Taxonomy. 2013. <http://www.gbif.org/species/3039572>. Accessed 29 November 2015.
45. Bennike L, Böcher J. Forest-tundra neighbouring the North Pole: plant and insect remains from the Plio-Pleistocene Kap København Formation, North Greenland. *Arctic.* 1990;43:331–8.
46. Kelly MR. Floras of middle and upper Pleistocene age from Brandon, Warwickshire. *Phil Trans R Soc B.* 1968;254:401–15.
47. Skrede I, Borgen L, Brochmann C. Genetic structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and high-arctic versus arctic-alpine distributions. *Heredity.* 2009;102:293–302.
48. Becquey S, Gersonde R. Past hydrographic and climatic changes in the Subantarctic Zone of the South Atlantic - The Pleistocene record from ODP Site 1090. *Palaeogeogr Palaeoclimatol.* 2002;182:221–39.
49. Lu H, Wang X, Li L. Aeolian sediment evidence that global cooling has driven late Cenozoic stepwise aridification in central Asia. In: Clift PD, Tada R, Zheng H, editors. *Monsoon Evolution and Tectonics - Climate Linkage in Asia*, vol. 243. London: Special Publications, Geological Society; 2010. p. 29–44.
50. Cai MT, Fang XM, Wu FL, Miao YF, Appel E. Pliocene-Pleistocene stepwise drying of Central Asia: evidence from paleomagnetism and sporopollen record of the deep borehole SG-3 in the western Qaidam Basin, NE Tibetan Plateau. *Global Planet Change.* 2012;94–95:72–81.
51. Wen J, Zhang JQ, Nie ZL, Zhong Y, Sun H. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front Genet.* 2014;5:4.
52. Yang FS, Li YF, Ding X, Wang XQ. Extensive population expansion of *Pedicularis longiflora* (Orobanchaceae) on the Qinghai-Tibetan Plateau and its correlation with the Quaternary climate change. *Mol Ecol.* 2008;17:5135–45.
53. Li GD, Kim C, Zha HG, Zhou Z, Nie ZL, Sun H. Molecular phylogeny and biogeography of the arctic-alpine genus *Lagotis* (Plantaginaceae). *Taxon.* 2014;63:103–15.
54. Wang Q, Liu JQ, Allen GA, Ma Y, Yue W, Marr KL, et al. Arctic plant origins and early formation of circumpolar distributions: a case study of the mountain sorrel, *Oxyria digyna*. *New Phytol.* 2016;209:343–53.
55. Sibley CG, Monroe BL. *Distribution and Taxonomy of Birds of the World*. New Haven: Yale University Press; 1990.
56. MacKinnon JR, Sha M, Cheung C, Carey G, Zhu X, Melville D. A Biodiversity Review of China. WWF International: Hong Kong; 1996.
57. Clausen P, Nolet BA, Fox AD, Klaassen M. Long-distance endozoochorous dispersal of submerged macrophyte seeds by migratory waterbirds in northern Europe—a critical review of possibilities and limitations. *Acta Oecol.* 2002;23:191–203.
58. Figuerola J, Green AJ. Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater Biol.* 2002;47:483–94.
59. Dessborn L, Brochet AL, Elmerg J, Legagneux P, Gauthier-Clerc M, Guillemain M. Geographical and temporal patterns in the diet of pintail *Anas acuta*, wigeon *Anas penelope*, mallard *Anas platyrhynchos* and teal *Anas crecca* in the Western Palearctic. *Eur J Wildl Res.* 2011;57:1119–29.
60. Jia DR, Abbott RJ, Liu TL, Mao KS, Bartish IV, Liu JQ. Out of the Qinghai-Tibet Plateau: evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophaë rhamnoides* (Elaeagnaceae). *New Phytol.* 2012;194:1123–33.
61. Stuckey RL. Phylogeographical outline of aquatic and wetland angiosperms in continental eastern North America. *Aquat Bot.* 1993;44:259–301.
62. Jacobs SWL, Wilson KL. A biogeographical analysis of the freshwater plants of Australasia. *Aust Syst Bot.* 1996;9:169–83.
63. Chambers PA, Lacoul P, Murphy KJ, Thomaz SM. Global diversity of aquatic macrophytes in freshwater. *Hydrobiologia.* 2008;595:9–26.
64. Zhu JN, Yu D, Xu XW. The phylogeographic structure of *Hydrilla verticillata* (Hydrocharitaceae) in China and its implications for the biogeographic history of this worldwide-distributed submerged macrophyte. *BMC Evol Biol.* 2015;15:95.
65. Wu ZG, Yu D, Li X, Xu XW. Influence of geography and environment on patterns of genetic differentiation in a widespread submerged macrophyte, Eurasian watermilfoil (*Myriophyllum spicatum* L., Haloragaceae). *Ecol Evol.* 2016;6:460–8.
66. Qiu YX, Fu CX, Comes HP. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of quaternary climate and environmental change in the world's most diverse temperate flora. *Mol Phylogenet Evol.* 2011;59:225–44.
67. Liu JQ, Sun YS, Ge XJ, Gao LM, Qiu YX. Phylogeographic studies of plants in China: advances in the past and directions in the future. *J Syst Evol.* 2012;50:267–75.