Evolution of biogeographic disjunction between eastern Asia and North America in *Chamaecyparis*: Insights from ecological niche models

Ping Liu a, b, Jun Wen c, **, Tingshuang Yi a, *

a Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
b Kunming College of Life Sciences, University of Chinese Academy of Sciences, Kunming, Yunnan 650201, China
c Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States

**Corresponding author.**

**Corresponding author.**

E-mail addresses: wenj@si.edu (J. Wen), tingshuangyi@mail.kib.ac.cn (T. Yi).

1. Introduction

The biogeographic disjunction between eastern Asia (EA) and North America (NA) is one of the most fascinating biogeographic patterns (Graham, 1972; Wen, 1999, 2001; Xiang et al., 2015). However, most plant taxa in these two regions show an eastern Asian and eastern North American disjunction (Dengler, 1972; Tiffney, 1985; Wen et al., 1996, 2002; Wen, 1999, 2001; Wen and Shi, 1999; Xiang et al., 2004). Fewer plant groups are disjunctly distributed between EA and western North America (WNA), even though EA and WNA are closer geographically and were connected via the Bering land bridge throughout most of the Cenozoic Era between the Paleogene (65.5 Ma) and Pliocene (5.5–4.8 Ma) (Donoghue and Smith, 2004; Milne, 2006; Wen et al., 2016). EA and eastern North America (ENA) also show higher similarities in floristic composition (Li, 1952; Wen, 1999). The formation and historical processes of this biogeographic pattern have been debated for over 150 years (Gray, 1846, 1859; Darwin, 1859; Li, 1952; Wen, 1999, 2001). Two hypotheses have been proposed to explain why more lineages show disjunctions between EA and WNA (Wen, 2001). The first hypothesis is that temperate forest elements were initially widely distributed in the Northern Hemisphere during the mid-Tertiary, but were subsequently extirpated in WNA and western Europe in response to climatic cooling during the late Neogene and Quaternary. The second hypothesis is that the current intercontinental disjunct distribution of *Chamaecyparis* species in WNA experienced a significant niche shift.
et al., 2016); the second hypothesis is that ENA and EA have more similar habitats, and therefore taxa that migrated between the two regions might have survived and adapted in the new geographic area more easily (Wen, 2001). These two hypotheses need to be further tested via evidence based on phylogeny, ecology and paleontology.

Ecological niche models (ENMs) have been increasingly used to answer the fundamental questions about niche diversification and biogeography within a clade (Warren et al., 2008; Wen et al., 2013). The niche overlap between two taxa can be quantified with Schoener's D (Warren et al., 2008; Broennimann et al., 2012; Li et al., 2014), varying from 0 (no overlap) to 1 (all overlap). This approach has been considered as an accurate method for evaluating niche dynamics (Broennimann et al., 2012; Li et al., 2014). Following previous studies (Petitpierre et al., 2012; Li et al., 2014; Di Cola et al., 2017), we applied three more specific components to reveal niche differences between NA range and EA range of Chamaecyparis: niche stability, shared niche of both ranges; niche expansion in peculiar to one range; and niche unfilling in peculiar to the other range. These parameters may be applied to measure the niche changes between closely related taxa that show disjunct distributions.

Chamaecyparis consists of six species disjunctly distributed in EA, WNA, and ENA, with C. obtusa and C. pisifera from Japan, C. taiwanensis and C. formosensis from Taiwan, C. thyoides from ENA and C. Lawsoniana from WNA (Wang et al., 2003; Liao et al., 2010). A few phylogenetic studies have been carried out on this genus, but the nuclear and plastid data produced incongruent topologies (cf. Li et al., 2003; Wang et al., 2003; Liao et al., 2010; Mao et al., 2012). High homoplasy and incomplete concerted evolution of the ITS fragment may have contributed to incorrect inferences of phylogenetic relationships of Chamaecyparis (Liao et al., 2010). We thus used the tree based on plastid markers (Liao et al., 2010; Mao et al., 2012), in which Chamaecyparis was resolved into a C. thyoides (C. formosensis + C. pisifera) clade and a C. Lawsoniana (C. obtusa + C. taiwanensis) clade. In addition, this genus has abundant Tertiary fossil records from Europe, Asia and NA (Liu et al., 2009). Chamaecyparis is thus a good model to test the previous hypotheses concerning the formation of the eastern Asian and North American biogeographic disjunction.

In this study, we first applied reciprocal Maxent models between EA species and NA (both ENA and WNA) species to test if the species in EA and ENA share more similar habitats. We then applied Maxent models to predict the suitable area of Chamaecyparis in the Last Glacial Maximum (LGM) period using the Northern Hemisphere as a background to test if extirpations during LGM played a role in the formation of EA and NA disjunct distribution of this genus. Finally, we calculated niche diversification among the species in EA, ENA, and WNA to test if the species in EA and ENA share more similar niches.

2. Materials and methods

2.1. Occurrence data

All six species of Chamaecyparis were included in this study. We put all four species of Chamaecyparis in EA into one group (the EA group); and we designated C. taiwanensis and C. obtusa as EA group 1 and C. formosensis and C. pisifera as EA group 2. We firstly compiled geo-referenced presence records by querying the Global Biodiversity Information Facility Data Portal (data.gbif.org; accessed 1/9/16) and TROPICOS (tropicals.org; accessed 1/9/16). Additional eastern Asian occurrences were obtained from the database Plants of TAIWAN (http://taiz2.ntu.edu.tw; accessed 1/10/16). Duplicates and occurrences within 5 km of each other were removed to minimize spatial autocorrelation. The final dataset included 832 occurrences (Table S1).

2.2. Climate data

We downloaded 19 bioclimatic layers (Table S2) from WorldClim (worldclim.org; accessed 1/9/16) at 2.5 arc-minutes (ca. 5 km) resolution (Hijmans et al., 2005), both from current and LGM periods. CCSM4 model was used for LGM period. Climate data were extracted at each occurrence using ArcGIS v10.2 (ESRI, 2011). In order to minimize redundancy among climate variables by removing highly correlated variables, we calculated Pairwise Pearson correlations between the 19 factors. When a pair had a Pearson correlation >0.8, one of the two variables was removed (Tables S3–S5). The variables showing low jackknife training gain values by generating an ENM in Maxent v3.3.3k (Phillips et al., 2006) were finally removed. This process resulted in a dataset of eight climatic variables for the EA group: bio2, bio6, bio7, bio8, bio9, bio12, bio14 and bio15; seven climatic variables for C. thyoides: bio1, bio2, bio8, bio9, bio14, bio15 and bio18; and seven climatic variables for C. Lawsoniana: bio3, bio4, bio9, bio11, bio15, bio18 and bio19.

2.3. Maxent modeling

Reciprocal ENMs (Peterson et al., 1999) created by Maxent 3.3.3 k were applied to examine niches in geographic space. ENMs were separately calibrated with the native occurrence and background data of the EA group, EA group 1, and EA group 2, which were projected onto the NA range; and those of the eastern North American C. Thyoides and the western North American C. Lawsoniana, which were projected onto the EA range. The last analysis was calibrated with Chamaecyparis in current period and projected onto LGM period of the Northern Hemisphere (10°N–75°N). We delineated EA across 90°E–180°E, 10°N–75°N and NA across 60°W–170°W, 10°N–75°N following previous studies (Qian, 2002; Ricklefs et al., 2004) (Fig. 1). Maxent was run using the 10,000 background points generated from the kernel density maps, with default settings, jackkni ング and logistic output. We measured climate variable importance by comparing jackknife of training gain values when models were made with individual variables. An area under the curve (AUC) of the receiver operating characteristic plot can be a poor indicator of model accuracy when using pseudo-absences, as with Maxent.

2.4. Quantifying the niche dynamics

The extent of the study area has important effects on niche comparisons given its current distribution and the timescale considered in the study (Guisan et al., 2014; Li et al., 2014). We used Japanese and Taiwanese administrative boundaries to define the range of the EA group (Li et al., 2014); and we used the biomes in which a species occurs to define the ranges of C. Thyoides and C. Lawsoniana (Olson et al., 2001; Guisan et al., 2014).

Species data were projected onto the first two axes of a principal components analysis (PCA), depicting a multivariate climatic space calculated with the remaining climatic variables used in our study (Fig. S1). Following previous studies (Petitpierre et al., 2012; Broennimann et al., 2012), we did not include additional axes because the first two explained a large proportion of the total climate variation (Table 1; Fig. S1). The PCA was calibrated on climate factors distributed in both extents (referred to as PCAexp in Broennimann et al., 2012). Species occurrences were then transformed into species density using a kernel smoother in the gridded PCA environmental space (at a resolution of 100 × 100 cells).
(Broennimann et al., 2012). This approach reduces the risk that a difference between the numbers of two species records would cause an analytical bias in our results. This approach allowed species occupancy to be defined by correcting species densities through incorporating differences in environmental availability among Chamaecyparis ranges (Broennimann et al., 2012).

Niche overlap was quantified with Schoener’s D (Warren et al., 2008; Broennimann et al., 2012), varying between 0 (no overlap) and 1 (total overlap). It also allows niche conservatism to be tested through a one-sided niche-similarity test based on D (Broennimann et al., 2012).

We also quantified three more specific components of niche changes: niche stability, niche expansion and niche unfilling (Petitpierre et al., 2012; Di Cola et al., 2017). In both the EA groups and C. thyoides extents, marginal climates with densities below 25% were not included to reduce the heterogeneity in climate availability between the EA groups and C. thyoides extents. But in both the EA groups and C. lawsoniana extents, analysis was performed on the whole environmental extent (EA groups and C. lawsoniana). The results were similar for different proportions of the intersection of the species densities (75%, 80%, 85%, 90%, 95%, 100% and NA; Fig. S2). The same R functions as in Petitpierre et al. (2012) and Di Cola et al. (2017) were used for the entire procedure.

In all, we quantified niche overlap, niche stability, niche expansion and niche unfilling among EA groups (the EA group, EA group 1 and EA group 2), the eastern North American C. thyoides and the western North American C. lawsoniana. Analyses were performed in R3.2.5 (RCoreTeam, 2016).

3. Results

3.1. Climate variables

All models performed well with AUC values >0.98 (Table S6, n = 10 replicate model runs), suggesting a high fit of the model (Phillips et al., 2006). The predicted distributions of Chamaecyparis in current period are consistent with the observed present distributions when we used the comprehensive variables, indicating that the distributions are restricted by environmental factors and the variables which were chosen in our study are reliable. A Jackknife of the regularized training gain (S1) revealed that bio14, bio17, bio19, bio12, bio6, bio4 and bio18 contributed the most to model development of EA group, EA group 1 and EA group 2. A Jackknife of the regularized training gain (S1) revealed that bio14, bio17, bio19, bio12, bio15, bio16, bio13 and bio1 contributed the most to model development of C. thyoides. A Jackknife of the regularized training gain (S1) revealed that bio19, bio3, bio4, bio16, bio13, bio11, bio9 and bio6 contributed the most to model development of C. lawsoniana.

3.2. Climatic niche in geographic space

The model trained with the EA group, the EA group 1 and the EA group 2 predicted the range of the eastern North American C. thyoides very well, but could not predict the range of the western North American C. lawsoniana (Fig. 2a–f). The model trained with the eastern North American C. thyoides could predict the range of the EA group (Fig. 2g and h). The model trained with the western North American C. lawsoniana could neither predict the range of the EA group nor the range of C. thyoides (Fig. 2i and j).

Most regions of the middle and western NA were inferred to have low suitable habitats for Chamaecyparis during both the LGM and current periods (Fig. 3a and b). However, southern Europe was inferred to have high suitable habitats for Chamaecyparis during the LGM period, and still possess low suitable habitats for C. thyoides during both the LGM and current periods (Fig. 3a and b). Interestingly, South China was inferred to have high suitable habitats for Chamaecyparis during the LGM period (Fig. 3b).

3.3. Niche dynamics among Chamaecyparis

Realized climatic niche conservatism (i.e., niche stability) was observed only from C. thyoides to C. lawsoniana at fine resolution (P < 0.05 for the similarity test) (Li et al., 2014) (Table 1). The D values and niche stabilities between EA group, EA group 1 or EA group 2 with C. thyoides were larger than those with C. lawsoniana.
Fig. 2. Probability of occurrence of Chamaecyparis in EA and NA, from 0 (blue) to 1 (red), obtained from different models: model trained on EA group (a) and projected onto NA (b); model trained on EA group 1 (c) and projected onto NA (d); model trained on EA group 2 (e) and projected onto NA (f); model trained on eastern North American C. thyoides (g) and projected onto EA (h); model trained on western North American C. lawsoniana (i) and projected onto EA (j) (MAXENT v3.3.3).
4. Discussion

Our study indicated that both hypotheses may be jointly applied to explain the formation of the current intercontinental disjunct distribution of Chamaecyparis. Fossil records indicate the wide distribution of this genus in the Northern Hemisphere during the Tertiary (Liu et al., 2009). Climatic cooling at the end of the Neogene and the Quaternary significantly shaped the current distribution of Chamaecyparis (Graham, 1993; Manchester, 1999). Our ENMs indicated that there was suitable habitat in southern Europe during the LGM. The extinction of this genus in Europe may be attributed to their failure to migrate and survive in the southern refuges. Extinctions of many other taxa such as Aralia, Hamamelis, Liriodendron, and Nyssa in Europe have been proposed in the Cenozoic (Parks and Wendel, 1990; Wen, 1993, 1998, 1999; Tiffney and Manchester, 2001; Nie et al., 2008). Chamaecyparis also had a broader distribution in middle and western NA until the Pliocene, but only has a narrow current distribution from southwestern Oregon to northern California near the Pacific coast (Wang et al., 2003). Maxent modeling indicated that there were few suitable regions for Chamaecyparis in middle and western North America. The distribution of Chamaecyparis was greatly reduced in this region following the climate cooling from the end of the Neogene to the Quaternary. Our results indicated precipitation and temperature are two key factors to the distribution of Chamaecyparis. Greatly reduced precipitation and temperature during the LGM in the Northern Hemisphere might have greatly impacted the distribution of Chamaecyparis.

The results of reciprocal ENMs, integrating the four eastern Asian species as one group or separating them into two groups, showed that the model calibrated with the EA species could predict the distribution of eastern NA species well. However, it failed to predict the distribution of western NA species. Our results also indicated that the eastern Asian species have higher niche overlap with the eastern North American species. Both analyses suggest that similar habitat was shared by species in eastern Asia and eastern North America. The western North American species show low niche similarity with eastern Asian species, even when compared with its close eastern Asian relatives C. obtusa and C. taiwensis. Taxa migrating between EA and western NA should have experienced niche shifts to adapt to the new habitat. The similar habitats in the two regions may be the main reason that more plant groups show an eastern Asian and eastern North American disjunction.

In summary, our ecological niche models of Chamaecyparis support the hypothesis that eastern Asian and eastern North American disjunctions observed today may be due to both extinctions in western NA and Europe in response to the late Neogene and Quaternary climatic cooling (Graham, 1993; Manchester, 1999) and to more similar habitats between ENA and EA. However, more angiosperm and gymnosperm taxa showing the disjunct distribution between EA and NA need to be analyzed to further test this hypothesis.

Acknowledgments

This study was funded by grants from the Ministry of Science and Technology of China, Basic Research Project (No. 2013FY112600), and the Talent Project of Yunnan Province (No. 2011CIO42). We appreciate the help of Xiao-Jian Qu (Kunning Institute of Botany), Steven T. Callen (Saint Louis University), Qian-Long Liang (Sichuan University) and Xuan Liu (Institute of Zoology) for the study of ecological niche models. Kai Chen (Kunning Institute of Botany) and Ming-Cheng Wang (Sichuan University) for the study of ArcGIS. Jian-Jun Jin (Kunning Institute of Botany) for the collection of occurrence data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.pld.2017.04.001.

References

Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., et al., 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Global Ecol. Biogeogr. 21, 481–497.

Darwin, C.R., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.

Dengler, N.G., 1972. Ontogeny of the vegetative and floral apex of Calycanthus occidentalis. Can. J. Bot. 50, 1349–1356.

Di Cola, V., Broennimann, O., Petitpierre, B., et al., 2017. Ecospat: an R package to support spatial analyses and modeling of species niches and distributions. Ecography 40, 1–14.

Donoghue, M.J., Smith, S.A., 2004. Patterns in the assembly of temperate forests around the Northern Hemisphere. Philos. Trans. R. Soc. Lond. B 359, 1633–1644.

Graham, A., 1993. History of the vegetation: Cretaceous (Maastrichtian)-Tertiary. In: Flora of North America Editorial Committee. Flora of North America North of Mexico, vol. 1. Oxford Univ. Press, New York, pp. 57–70.

Gray, A., 1846. Analogy between the flora of Japan and that of the United States. Am. J. Sci. Arts 2, 135–136.

Gray, A., 1859. Diagnostic characters of phanerogamous plants, collected in Japan by Charles Wright, botanist of the U.S. North Pacific Exploring Expedition,
