Molecular phylogeny, biogeography and character evolution of the montane genus Incarvillea Juss. (Bignoniaceae)

Santosh Kumar Rana, Dong Luo, Hum Kala Rana, Shaotian Chen, Hang Sun

Abstract

The complex orogeny of the Himalaya and the Qinghai-Tibet Plateau (QTP) fosters habitat fragmentation that drives morphological differentiation of mountain plant species. Consequently, determining phylogenetic relationships between plant subgenera using morphological characters is unreliable. Therefore, we used both molecular phylogeny and historical biogeographic analysis to infer the ancestral states of several vegetative and reproductive characters of the montane genus Incarvillea. We determined the taxonomic position of the genus Incarvillea within its family and inferred the biogeographical origin of taxa through Bayesian inference (BI), maximum likelihood (ML) and maximum parsimony (MP) analyses using three molecular data sets (trnL-trnF sequences, nr ITS sequences, and a data set of combined sequences) derived from 81% of the total species of the genus Incarvillea. Within the genus-level phylogenetic framework, we examined the character evolution of 10 key morphological characters, and inferred the ancestral area and biogeographical history of the genus. Our analyses revealed that the genus Incarvillea is monophyletic and originated in Central Asia during mid-Oligocene ca. 29.42 Ma. The earliest diverging lineages were subsequently split into the Western Himalaya and Sino-Himalaya during the early Miocene ca. 21.12 Ma. These lineages resulted in five re-circumscribed subgenera (Amphicome, Olgaea, Niedzwedzchia, Incarvillea, and Pteroscleris). Moreover, character mapping revealed the ancestral character states of the genus Incarvillea (e.g., suffruticose habit, cylindrical capsule shape, subligneous capsule texture, absence of capsule wing, and loculicidal capsule dehiscence) that are retained at the earliest diverging ancestral nodes across the genus. Our phylogenetic tree of the genus Incarvillea differs from previously proposed phylogenies, thereby recommending the placement of the subgenus Niedzwedzchia close to the subgenus Incarvillea and maintaining two main divergent lineages.

Keywords: Biodiversity hotspots, Biogeography, Incarvillea, Molecular phylogeny, Phytools, Stochastic character mapping

1. Introduction

Understanding the origin and evolution of endemic species has long intrigued biogeographers, evolutionary biologists, and molecular ecologists (Gaston, 2000; Lohmann et al., 2013). Accordingly, researchers have increasingly examined the flora of the mountains of Central Asia, including the Himalayas and the Hengduan Mountains, which are unique biodiversity hotspots (Myers et al., 2000; Mittermeier et al., 2011). These mountains surround the Qinghai-Tibet Plateau (QTP), the world’s highest and most extensive plateau (Zhang et al., 2002). The ongoing uplift of the plateau, which started 25 million years ago (Ma), has had a major influence on plant diversity in the region (Harrison et al., 1992; Li et al., 1995; Guo et al., 2002; Spicer et al., 2003). Specifically, the QTP uplift has created a large elevational gradient that fragments the habitat of alpine plant species, creating barriers to gene flow that increase divergent evolution and speciation (Filatov et al., 2016). As a result, the uplift of the QTP has created young endemic ‘plant cradles’ in the Hengduan Mountains, which act as an ‘evolutionary front’ in China (Lopez-Pujol et al., 2011). However, the origin and evolution of many montane species (e.g., Incarvillea spp.) and the biogeographical patterns (e.g., dispersal routes) between the mountainous hotspots of Central Asia have...
yet to be fully studied; therefore, understanding the causes and consequences of these patterns remains a challenge (e.g., Nauheimer et al., 2012).

To date, few studies have investigated the connections between the QTP and adjoining biodiversity hotspots in Asia. The similarity of floristic elements and geographic distribution patterns of plants that occurs throughout Asia may reflect evidence of dispersal of montane plants to Southeast Asia from the Himalaya region, as well as Central Asia, or the reverse.

This unusual biogeographical connection may be explained by long-distance dispersal that uses SE Asia as a ‘stepping stone’ (Liu et al., 2002). The QTP has acted as both a source and a sink for plant taxa that occur in the Himalaya and adjoining regions, the New and Old World, the Middle East, as well Indochina (Hajra and Rao, 1990; Zhu and Roos, 2004; Xu et al., 2014). Surprisingly, research has shown that the ‘highlands of Central Asia’ (Li et al., 2014) and the ‘Beringia’ land bridge connecting Eastern Asia with Northern America (Nie et al., 2006; Deng et al., 2018) have acted as corridors for plant genera that originated in the QTP (Xu et al., 2010; Zhang et al., 2014). However, evidence suggests that plant taxa may have originated outside the QTP and diversified after their ancestors arrived to the QTP region (Liu et al., 2002; Yue et al., 2009; Tu et al., 2016). In addition, the QTP might serve as a ‘refugium’ for Tertiary relict floras (Milne and Abbott, 2002).

Incarvillea Juss. is a temperate and herbaceous montane genus that comprises 16 species distributed throughout the western, southern and eastern fringe of the QTP (Grierson, 1961; Zhao, 1985; Zhang and Santisuk, 1998; Chen et al., 2005). Recent molecular and taxonomic work by Chen et al. (2005, 2006) revealed five major subgenera, including the subgenus Olgaea (Table 1). Of these subgenera, the subgenus Incarvillea is distributed mainly in East Asia (I. sinensis Lam.: from SW to NE China) and Mongolia (I. potaninii Batalin), whereas the remaining four subgenera are distributed at various elevations (500–5500 m a.s.l.) in the three major biodiversity hotspots of Asia, i.e., the Mountains of Central Asia (subgenus Olgaea and Niedzwiedzki), the Himalaya (subgenus Ampicome and Pteroscleris), and Mountains of SW China (subgenus Pteroscleris) (Chen et al., 2005) (Fig. 1a). Previous research on Incarvillea has focused on molecular diversity and species richness assessments (genus Incarvillea; Chen et al., 2010), population genetic analysis (I. younghusbandi Sprague; Zhu et al., 2009: I. sinensis; Chen et al., 2012: I. arguta (Royle) Royle; Rana et al., 2019), pollen dispensing mechanisms (I. arguta; Han et al., 2008) and reproductive strategies of a pollinator-limited species (I. mairei (H.Lev) Grierson; Ai et al., 2013). However, few studies have examined how, when and where the species of this genus originated and what stochastic characters led to the evolution of the different subgenera.

Molecular phylogenetic analysis of the genus Incarvillea based on ITS and trnL-trnF sequences has provided strong evidence that species within subgenus Pteroscleris might have undergone a recent radiation, perhaps related to the uplift of the Himalaya-Hengduan Mountains (Chen et al., 2005). Thus, members of the genus Incarvillea provide an excellent opportunity to use biogeography to examine species adaptation to recent diversification and rapid radiation. Furthermore, because the five subgenera of Incarvillea clearly differ in their habit, stamens, capsule texture, and seeds (Chen et al., 2004), morphological character evolution may help understand the phylogenetic placement of subgenera.

In this study, we aim to identify the historical biogeography and ancestral characters that have resulted in present-day patterns of Incarvillea diversification. Specifically, we asked when, where and how the genus Incarvillea originated and what conspicuous ancestral characters led to the evolution of the five subgenera of Incarvillea. To answer these questions, we reconstructed a phylogenetic framework of the genus Incarvillea. Our molecular tree incorporated plastid trnL-trnF, nuclear ribosomal Internal Transcribed Spacer (nrITS), and combined sequences from the genus Incarvillea, along with large trnL-trnF sequences from the Bignoniaceae family (except Bignoniae). We then combined information from our time-calibrated tree derived from our molecular data with that of biogeographical considerations, morphological character evolution, and the distribution of representatives of extant Incarvillea species to discuss the events that underlie the evolutionary history of the genus Incarvillea.

### 2. Material and methods

#### 2.1. Taxon sampling and DNA sequencing

We sampled ca. 81% of extant species of the genus Incarvillea, including representative species from five Incarvillea subgenera. Specifically, we used 18 GenBank accessions representing 13 species of the genus Incarvillea (except Incarvillea altissima Forrest, I. forrestii Fletcher, and I. potaninii). We assume that the generic coverage of the genus Incarvillea represents the overall distribution range (Fig. 1a), which is likely to lead the reliable reconstruction of the biogeographic history.

New sequences generated in this study include Incarvillea arguta nrITS sequences (MT533886–MT533889) from four individuals representing the entire geographic range of the species (Fig. 1a). Voucher specimens were deposited in the National Herbarium and Plant Laboratories (KATH), Nepal and Kunming Institute of Botany (KUN), China.

Outgroups comprise the tribes or clades within the family Bignoniaceae (except clade Bignoineae) (Grose and Olmstead.

### Table 1

Morphological comparison of five major clades (i.e., subgenera of the genus Incarvillea); modified from Chen et al. (2005).

| Characteristics | Incarvillea subgenus | I. subgenus Ampicome | I. subgenus Incarvillea | I. subgenus Pteroscleris | I. subgenus Olgaea |
|-----------------|----------------------|----------------------|------------------------|-------------------------------|----------------------|
| Habit           | Suffruticose         | Suffruticose         | Suffruticose           | Herbaceous                      | Suffruticose         |
| Leaf arrangement| Alternate            | Alternate            | Alternate or radial     | Not swollen                     | Not swollen           |
| Calyx teeth base| Not swollen           | Not swollen           | Swollen                | Racemose                        | Glabrous             |
| Inflorescence   | Racemose             | Racemose             | Racemose               | Glabrous                        | Glabrous             |
| Anther texture  | Glabrous             | Glabrous             | Glabrous               | Cylindrical                     | Cylindrical          |
| Capsule shape   | Ovate                | Elongate, linear cylindrical | Fibrous               | Coriaceous                      | Coriaceous           |
| Capsule texture | Subligneous          | No                   | No                     | Subligneous                     | No                   |
| Capsule wing    | Six wings, longitudinal | No               | No                     | Loculical                       | Loculical            |
| Capsule dehiscence | Septifragal         | Loculical            | Loculical              | Hyaline wing                    | Opaque wing          |
| Seed wing       | Opaque wing          | Coma at ends         | Hyaline wing           | Opaque wing                     | Hyaline wing         |
2007; Olmstead et al., 2009; Refulio-Rodriguez and Olmstead, 2014): Tecomea tribe (9 genera), Catalpaeae tribe (3 genera), Tabebuia alliance (9 genera), Plaeotropical clade (13 genera) Delostoma D. Don (2 species), Oroxyloae (1 genus), Taurrettia Foug. (2 species), Argylia D. Don (2 species), and Jacaranda Juss. (4 species) (Table S1). To determine the position of the genus Incarvillea within the family Bignoniaceae, we also used Pedalum murex L. and Sesamum indicum L. (only for genus-level phylogeny reconstruction) from the family Pedaliaceae (Gormley et al., 2015).

Total genomic DNA was extracted from dried leaves of four Incarvillea arguta specimens following the manufacturer’s protocol for the DNAsecure Plant Kit (Tiangen Biotechnology Co. Ltd., Beijing, China). Plastid trnL-trnF sequences were obtained from Rana et al. (2019), whereas nrITS sequences were amplified using primers ITS1 and ITS4 (White et al., 1990). PCR reactions were performed in 30 μl volume containing 20–40 ng of sample DNA, 15 μl 2× Taq Plus Master Mix with dye (Tiangen Biotech), 12 μl of double distilled H2O and 1 μl of 10 μM stock of each primer. The reaction was performed under the following thermal conditions: 94 °C for 3 min, followed by 35 cycles at 94 °C for 30 s, 53 °C for 30 s and 72 °C for 7 min, with reactions held at 10 °C until further processing. PCR products were purified and sequenced by Tsingke Biological Technology Co. (Beijing, China).

2.2. Molecular phylogeny of the montane genus Incarvillea

Phylogenetic relationships of the genus Incarvillea within the family (family-level) and subgenera within the genus (genus-level) were inferred by using plastid trnL-trnF sequences. Phylogenetic estimates were based on Bayesian inference (BI), maximum likelihood (ML) and maximum parsimony (MP) approaches using MrBayes v.3.2.6 (Ronquist et al., 2012), RAxML (Stamatakis, 2014) and PAUP* v.4.0a (Swofford, 2003) respectively. Furthermore, all three approaches were applied to reconstruct the subgeneric phylogenetic position within the genus Incarvillea using nrITS and combined (trnL-trnF and nrITS) sequences. Indels were treated as missing data, by default. We used GTR + G as the best fitting substitution model based on the Akaike information criterion (AIC) using ModelTest 2.16 (Posada and Crandall, 1998). Bayesian analysis was carried out in online CIPRES Science Gateway v.3.3 (Miller et al., 2010; https://www.phylo.org) using 5 × 107 generations, four chains with sampling every 1000th generations and first 20% was discarded as burn-in. ML analyses were performed using the graphical front-end raxmlGUIv1.5b2 (Silvestro and Michalak, 2012) for RAxML v.8.2.x with 1000 rapid bootstraps (Stamatakis, 2014). MP analyses were performed using heuristic searches with 1000 replicates of random addition, 10 trees held in each step during stepwise addition, tree bisection reconnection (TBR) swapping and 1000 bootstrapping replications with 10,000 maximum number of trees. The bootstrap results are summarized in a 50% majority-rule consensus cladogram. The clade support on the tree shows the Bayesian posterior probabilities (PP), maximum likelihood bootstrap values and maximum parsimony bootstrap values estimated by BI, ML and MP approaches respectively. The degree of phylogenetic incongruence between trnL-trnF and ITS partitions as a combined sequence was assessed using the congruence-length difference (ILD) test (Farris et al., 1994, 1995) in PAUP* v.4.0a, with 1000 replications using a heuristic tree search and 1000 additional sequence replicates. Furthermore, only parsimony-informative sites were used to ensure the accuracy of the P-value in the ILD test.

2.3. Estimation of divergence time

We calculated the divergence time based on the fossil and secondary calibration points using Bayesian statistics in BEAST v.1.8.4 (Drummond and Rambaut, 2007; Drummond et al., 2012). The Bayesian relationship of plastid trnL-trnF datasets were conducted using a TVM + G substitution model selected by the AIC in jModelTest 2.16 (Posada and Crandall, 1998) and uncorrelated lognormal relaxed clock (P < 0.05, a likelihood-ratio test in PAUP* v.4.0a; Swofford, 2003). A yule process was specified as tree prior. Three calibration points were used to construct nodes (nodes 1, 2, and 3). First, based on meta-calibration work on flowering plants (Magallon et al., 2015), the divergence time between Bignoniaceae and Pedaliaceae (Pedalum murex) was set to 47.03 Ma as our prior for the root node 1. We assumed a normal distribution with a mean of 47.03 Ma, and a standard deviation of 5.8, giving a 95% HPD of 37.49–56.57 Ma. Several other fossil records reveal that the age of the family is ca. 50 million years old (e.g. Wehr and Hopkins, 1994; Wehr, 1995; Wehr and Manchester, 1996; Wilf, 1997; Pigg and Wehr, 2002). Lohmann et al. (2013) applied these fossil records using a normal prior with a mean of 49 Ma and a standard deviation of 3.0 to calibrate Bignoniaceae. Secondly, fossil records for the Catalpa—Campsis clade, which are estimated as 38.8 Ma, sets as the minimum age for the clade (Meyer and Manchester, 1997; Manchester, 2000) to calibrate the crown node 2 of the Catalpa—Campsis clade. We used a lognormal distribution with the minimum age constraints 38.8 Ma taken as zero offset, and both the lognormal mean and the standard deviation set to 1 (95% HPD: 38.94–52.52 Ma). The fossil from the seed of Catalpa Scop. from the John Day Formation, Wheeler County, Oregon, USA (Meyer and Manchester, 1997), is estimated to be from the Late Eocene. The fossil-bearing sediments are estimated to be ca. 38.8 Ma, based on the dating of the overlying tuff (38.4 ± 0.7 Ma), and the underlying ignimbrite (39.17 ± 0.15 Ma; Manchester, 2000). Finally, the secondary calibration point estimated as 19.906 Ma was set as the divergence of the genus Incarvillea lineages (Rana et al., 2019). We assumed a normal distribution with a mean of 19.906 Ma, and a standard deviation of 3.89, giving a 95% HPD of 13.51–26.3 Ma as our prior for the crown node.

BEAST used a MCMC chain length of 5 × 107 generations, sampling every 1000th generations, following a burn-in of the initial 20% cycles. MCMC samples were checked in TRACER to confirm sampling adequacy and convergence of the chains to a stationary distribution. We started the analysis independently to detect the effective sample size (ESS > 200) for all parameters of each run (Rambaut and Drummond, 2007). TREEANNOTATOR v.1.8.4 (Drummond et al., 2012) and FigTree v.1.4.0 (Rambaut, 2009) were used to summarize and display the sampled trees, respectively.

2.4. Ancestral area reconstruction

To limit methodological biases from influencing subsequent inferences, we used multiple approaches to reconstruct the ancestral area of the genus Incarvillea. A dispersal-extinction-cladogenesis model (DEC-Lagrange) and Bayesian Binary Method (BBM) were implemented in LAGRANGE v.20130526 (Rie and Smith, 2008) and RASP v.4.0 (Reconstruct Ancestral State in Phylogenies; Yu et al., 2015), respectively, using all post-burn-in trees from the Beast analysis. We excluded all the outgroups of BEAST MCMC trees from chronograms to represent only the biogeographic range of the genus Incarvillea. Biogeographic data for species within the genus Incarvillea were compiled from the distribution information described in the literature, online flora of China and herbarium specimens. We divided the entire geographic range of the genus Incarvillea (Fig. 1a)
into six areas corresponding to the floristic divisions of East Asia (Wu and Wu, 1996; Sun, 2013): (A) the Sino-Japanese region and Mongolian steppe region (SMJ), (B) the Sino-Himalaya region (SH), (C) the QTP region, (D) the Western Himalaya (WH), (E) the Far-West Himalaya (FWH), and (F) Central Asia (CA). The number of maximum areas occupied at each node was set to four. For these analyses, range constraints were restricted to the region ‘E’ and ‘F’ from ‘A’, ‘B’ and vice-versa. The dispersal rate between areas was set to 0.5 (intermittent barriers; from A to D; B to E and vice versa) and 0.01 (with barriers; from A–B to F and vice versa) (Zhou et al., 2019). The BBM was run with the fixed state frequencies model (Jukes-Cantor) with equal among-site rate variation for 10 independent runs of 50,000,000 generations of the MCMC chains, with sampling every 1000th generations; the first 20% was discarded as burn-in in BBM analyses.

### 2.5. Stochastic character mapping

We created a matrix of 10 morphological characters that included habit, leaf arrangement, calyx teeth base, inflorescence, anther texture, capsule shape, capsule texture, capsule wing, capsule dehiscence, and seed wing (Table S2). The characters were selected based on Chen et al. (2005, 2006), the published literature, flora of China and examination of herbarium specimens, photographs (Table 1). We scored these characters for each species included in the plastid trnL–trnF data set, except for the outgroup taxa. Maximum likelihood reconstruction of ancestral states as a function of stochastic character mapping (SCM) (Hulesenbeck et al., 2003) was implemented to reconstruct the evolution of morphological characters of the genus Incarvillea. The package phytools v.0.6–99 (Revell, 2019) in R (R Core Team, 2019) was used to project the habitats onto the maximum clade credibility tree from a BEAST analysis. The best-fitting model for each character was selected by likelihood ratio tests, and the set of adequately fitting models was found by comparing corrected Akaike information criterion (AICc) scores. Out of the three transition models (ER, equal rates model; SYM, symmetrical model; ARD, all-rates-different model), we used ER following model selection via the corrected AICc (Table 3). The AIC is the Akaike information criterion for a finite sample size (Burnham and Anderson (2002) and vice versa) and 0.01 (with barriers; from A–B to F and vice versa) (Zhou et al., 2019). The BBM was run with the fixed state frequencies model (Jukes-Cantor) with equal among-site rate variation for 10 independent runs of 50,000,000 generations of the MCMC chains, with sampling every 1000th generations; the first 20% was discarded as burn-in in BBM analyses.

### Table 2

| SN | Major clades and numbered nodes | Age estimates (Ma) | Ancestral area reconstructions | DEC-Lagrange (Prob.) |
|----|---------------------------------|-------------------|------------------------------|---------------------|
| 1  | Node 4, Stem node of the genus Incarvillea | 29.42 | 23.34–35.78 | NA |
| 2  | Node 5, subgenera Niedzwedzki–Incarvillea–Pteroscleris clade | 18.38 | 13.08–24.06 | F | BJ (25%) |
| 3  | Node 6, subgenera Olgaea–Amphiomica clade | 16.99 | 9.87–23.85 | F | DJF (28.5%) |
| 4  | Node 7, subgenera Incarvillea–Pteroscleris clade | 15.29 | 9.94–21.2 | B | BJ (54%) |
| 5  | Node 8, subgen. Pteroscleris clade | 7.65 | 5.36–12.85 | BC | BJ (37.1%) |
| 6  | Tribe Tecomae clade | 33.43 | 27.06–39.17 | NA |

BBM, Bayesian binary Markov chain Monte Carlo; HPD, highest posterior density; DEC, Dispersal–Extinction–Cladogenesis; Prob., relative probability.

### Table 3

| Character | Model | Int. | AICc | dAICc | wAICc |
|-----------|-------|------|------|------|-------|
| Habit     | ARD   | 3.5849 | 11.169853 | 1.61901 | 0.18203 |
|          | SYM   | 3.7754 | 9.550848 | 0 | 0.40898 |
|          | ER    | 3.7754 | 9.550848 | 0 | 0.40898 |
| Leaf arrangement | ARD | 3.9169 | 30.27383 | 6.54656 | 0.02992 |
|          | SYM   | 10.341 | 26.6812 | 2.93593 | 0.18033 |
|          | ER    | 10.864 | 23.7277 | 0 | 0.78576 |
| Calyx teeth base | ARD | 2.7632 | 9.526412 | 1.14393 | 0.22087 |
|          | SYM   | 3.1957 | 8.391486 | 0 | 0.38957 |
|          | ER    | 3.1957 | 8.391486 | 0 | 0.38957 |
| Inflorescence | ARD | 5.585 | 16.9922 | 0.22183 | 0.18203 |
|          | SYM   | 10.041 | 8.082168 | 0 | 0.38752 |
|          | ER    | 3.0411 | 8.082168 | 0 | 0.38752 |
| Anther texture | ARD | 3.9023 | 11.804559 | 1.8105 | 0.16821 |
|          | SYM   | 3.9977 | 9.994056 | 0 | 0.4159 |
|          | ER    | 3.9977 | 9.994056 | 0 | 0.4159 |
| Capsule shape | ARD | 2.5665 | 9.133063 | 1.19576 | 0.21568 |
|          | SYM   | 2.9687 | 9.737036 | 0 | 0.39216 |
|          | ER    | 2.9687 | 9.737036 | 0 | 0.39216 |
| Capsule wing | ARD | 3.9023 | 11.804559 | 1.8105 | 0.16821 |
|          | SYM   | 2.9687 | 9.737036 | 0 | 0.39216 |
|          | ER    | 2.9687 | 9.737036 | 0 | 0.39216 |
| Capsule dehiscence | ARD | 5.585 | 16.9922 | 0.22183 | 0.18203 |
|          | SYM   | 2.9687 | 9.737036 | 0 | 0.39216 |
|          | ER    | 2.9687 | 9.737036 | 0 | 0.39216 |
| Seed wing | ARD   | 8.0014 | 28.0270 | 8.84777 | 0.01005 |
|          | SYM   | 8.2908 | 22.5815 | 3.42661 | 0.1512 |
|          | ER    | 8.5775 | 15.1849 | 0 | 0.83875 |

Int, log-likelihood of the model; AICc, corrected Akaike information criterion; dAICc, difference between the AICc of the model and the best model; wAICc, Akaike weight of the model. The dAICc, AICc<sub>K</sub>—min AICc follows the following criteria: dAICc<sub>K</sub> ≤ 1, substantial support for a model with AICc<sub>K</sub> ≤ dAICc<sub>K</sub> ≥ 4, low support for a model with AICc<sub>K</sub>; dAICc<sub>K</sub> > 10, no support for a model with AICc<sub>K</sub>. The AICc weights indicate the probability that the model is the best among the whole set of candidate models (Burnham and Anderson, 2002; Sivalls and Joyce, 2005). ER, equal rates model; SYM, symmetrical model; ARD, all-rates-different model. Bold font in the Model columns indicates the selected model by the likelihood ratio test.
3. Results

3.1. Sequence data and phylogenetic relationships

The family-level phylogeny was represented by 75 trnL-trnF accessions (18 accessions of 13 species as ingroup and 57 accessions as outgroups). The aligned matrix was 1083 characters long with 208 phylogenetically parsimony-informative sites and 144 parsimony-uninformative sites. For both family and genus-level phylogenetic relationships, we compared the most likely tree obtained from the BI, ML and MP analyses, along with their support values. We considered weak support (0.5 < PP < 0.7), moderate support (0.7 < PP < 0.95) and strong support (PP > 0.95) for Bayesian posterior probabilities, whereas good support (70 < values < 95) and strong support (values > 95) for likelihood and parsimony analyses.

The phylogenetic trees generated by BI, ML and MP analyses were somewhat congruent, differing only in the support values of the nodes (Fig. 2). BI, ML and MP analyses strongly support the genus Incarvillea being a monophyletic clade within Tecomeae (PP = 1, values = 100) (Figs. 2–4 and S1–6). The phylogenetic framework of the genus Incarvillea based on plastid trnL-trnF in this study is consistent with a previous molecular study by Olmstead et al. (2009) (Fig. 2). The five clades within the genus Incarvillea are equivalent to the five subgenera (Amphicome, Olgaea, Niedzwedzkia, Incarvillea and Pteroscleris) as proposed by Grierson (1961) and Chen et al. (2005, 2006). Of the five major clades, three clades (subgenera Niedzwedzkia, Incarvillea, Pteroscleris) are rooted as one with strong support (Figs. 2 and 3). It makes sense that the subgenus Niedzwedzkia (the Central Asian species) followed by subgenus Incarvillea is sister to subgenus Pteroscleris (East Asian species) clade (Figs. 2, 3a and S1), but with weak to moderate support. This topology was used to estimate divergence time, reconstruct the ancestral area and analyse character evolution.

Genus-level phylogenetic position was inferred separately using 22 accessions (18 accessions of 13 Incarvillea species and four outgroups) of each plastid trnL-trnF, nrITS and combined sequences. In the homogeneity test, the combined data sets provided a P value of 0.79: the acceptance of the null hypothesis. Therefore, the plastid trnL-trnF, nrITS and combined sequences were 979, 761 and 1740 characters long with 86, 144 and 230 phylogenetically parsimony-informative sites, and 48, 78 and 126 parsimony-uninformative sites respectively. The two data sets used in the genus-level phylogeny also indicate that the genus Incarvillea is a monophyletic group and is sister to Tecoma capensis (Thunb.) Spach (Fig. 3). These relationships were supported by 100% posterior probabilities and bootstrap values (Figs. 3a and S1), except when ML and MP analyses were used with the nrITS data set (PP = 0.70, bootstrap value = 84% by trnL-trnF) (Figs. 3b; S2 and S5). The five major clades equivalent to five subgenera as proposed by Chen et al. (2005) are resolved with moderate to high support values by trnL-trnF (Fig. 3). Of the five subgenera, subgenus Niedzwiédzka (Central Asian and semi-retschenskia) Griesmer) is rooted with East Asian subgenera Incarvillea and Pteroscleris with 80% support (Fig. 3a). The interrelationships of the five major sub-generic clades were only moderately resolved: for instance, the sister–group relationship between I. sinensis (subgenus Incarvillea) and members of the subgenus Pteroscleris was ambiguously supported by ML/MP analyses (bootstrap values = 50/71.5) (Figs. 3a; S1 and S4). In contrast, nrITS and combined data sets revealed a concordance topology of the subgenera, differing in the support values and placement of I. semi-retschenskia (subgenus Niedzwedzkia) (Figs. 3b and c; S2; S3; S4).

Fig. 2. Bayesian majority-rule consensus tree of the genus Incarvillea within Bignoniaceae (except Bignoniae) inferred from plastid trnL-trnF sequence. Bayesian posterior probabilities (PP > 0.50) are represented above the branches. Maximum likelihood bootstrap values/maximum parsimony bootstrap values are indicated below branches, with “—” indicating support values of less than 50% or collapse in analysis. Colour legend for the genus Incarvillea clade: purple, subgenus Pteroscleris; blue, subgenus Incarvillea; cyan, subgenus Niedzwedzkia; green, subgenus Amphicome; red, subgenus Olgaea.
The MCC tree based on \textit{trnL-trnF} coincides with the phylogenetic framework of BI, ML and MP analyses for the plastid DNA datasets at the family or genus-level (Figs. 2—4). Our molecular dating analyses (Fig. 4; Table 2) suggest that the genus \textit{Incarvillea} might have evolved before the Miocene, during the mid-Oligocene ca. 29.42 Ma (95% HPD: 23.34—35.78 Ma) and the earliest diverging lineages in the genus \textit{Incarvillea} arose during the early Miocene ca. 21.12 Ma (95% HPD: 15.96—26.4 Ma). Most of the major clades of subgenus \textit{Incarvillea} were established before the Quaternary period. The recent lineage diversification of subgenus \textit{Pteroscleris} occurred during the late Miocene ca. 7.65 Ma (95% HPD: 3.56—12.85 Ma). The tribe Tecomeae clade diverged during the early Oligocene ca. 33.43 Ma (95% HPD: 27.06—39.17 Ma) (Fig. 4; Table 2).

3.3. Biogeographical analysis of the montane genus \textit{Incarvillea}

The result of DEC-Lagrange and BBM analyses is shown in Fig. 5 and interpreted in Table 2. Both analyses were congruent with most clades and indicated that the genus \textit{Incarvillea} originated in Central Asia (area F; Fig. 5, node 3; DEC-Lagrange 30.7% and BBM 89.5%). Furthermore, the analyses indicated that species of the Central Asian genus \textit{Incarvillea} may have dispersed to the Himalaya and the Mountains of SW China. For example, our analyses indicate that plants in the subgenus \textit{Pteroscleris} rapidly diversified in the Sino-Himalaya after colonizing the Sino-Japanese and Mongolian steppe regions, the Western Himalaya, and the Qinghai-Tibet Plateau. The split between the subgenera \textit{Pteroscleris} and \textit{Incarvillea} was accompanied by the colonization of the Sino-Himalaya region, and Sino-Japanese and Mongolian steppe regions, along with the early divergence of subgenus \textit{Niedzwedzka} (see clade 5 of subgenera \textit{Niedzwedzka—Incarvillea—Pteroscleris}). Likewise, the colonization of the Sino-Himalaya and Far/Western Himalaya gave rise to the subgenus \textit{Amphicome}, whose center of origin is the Western Himalaya (Rana et al., 2019) and is rooted with subgenus \textit{Olgaea} from Central Asia. In general, results from DEC-Lagrange and BBM are similar to what is known about the biogeography of the genus \textit{Incarvillea}. One notable difference, however, is that DEC-Lagrange suggested that lineages from the subgenus \textit{Pteroscleris} colonized from the Sino-Himalaya only, whereas BBM indicated that colonization occurred from both the QTP and Sino-Himalaya.

3.4. Stochastic character mapping as the function of character evolution

The likelihood ratio tests and the set of adequately fitting models obtained by comparing AICc scores indicated that ER and SYM were equally predictive for habit, calyx teeth base, inflorescence, anther texture, capsule shape, capsule wing, capsule dehiscence, whereas the remaining characters were favoured by ER model alone (Table 3). Therefore, to infer consistency of character mapping, we implemented the ER model as the best fitting model. The posterior probabilities inferred as the state frequencies calculated across 1000 stochastically mapped character histories for every 17 nodes (Table S2) were good enough to map characters using the ER model. Stochastic character mapping indicated that the ancestral characters of the genus \textit{Incarvillea} include a suffrutescent habit, alternate leaf arrangement, calyx teeth base not swollen, racemose inflorescence type, glabrous anther texture, cylindrical capsule shape, subgbose capsule texture, absence of capsule wing, loculicidal capsule dehiscence, and opaque seed wing (Figs. 1b—g and 6e). These characters are retained at the earliest diverging ancestral node across the genus. These character states persist through the backbone nodes with branches leading to several clades containing species of different subgenera. Most of the subgenus nodes share the ancestral characters except for another texture, which is most commonly pilose (Figs. 1e1—e4 and 6e). In addition, the subgenus \textit{Amphicome} has the highest likelihood of having fibrous capsules (Fig. 6g) and seed wing comas at the ends (Fig. 6j), but lack these ancestral characters.

4. Discussion

4.1. Monophyly of the genus \textit{Incarvillea}

The relationships between the five subgenera (\textit{Amphicome}, \textit{Incarvillea}, \textit{Pteroscleris}, \textit{Olgaea} and \textit{Niedzwedzka}) of the genus \textit{Incarvillea} have long remained unresolved. Previous studies that examined these relationships using either molecular data or morphological characters (habit, calyx, capsule, and seed) have generated inconsistent phylogenies (Grierson, 1961; Chen et al., 2005, 2006). In this study, phylogenetic analyses using \textit{trnL-trnF} data sets indicate that the genus \textit{Incarvillea} is monophyletic. Our phylogenies based on nrITS and combined data sets (Fig. 3b and c) are generally consistent with the generic classification of previous molecular studies by Chen et al. (2005). Our findings support the placement of subgenus \textit{Niedzwedzka} rooted within the subgenera \textit{Incarvillea} and \textit{Pteroscleris}, which contrasts with previous placements based on \textit{trnL-trnF} sequences (Chen et al., 2005). Moreover, our molecular phylogeny based on \textit{trnL-trnF} sequences revealed two major basal clades (i.e., \textit{Olgaea—Amphicome} and \textit{Niedzwedzka—Incarvillea—Pteroscleris}), which also differs from Chen et al.’s phylogeny (2005). Within the second basal clade, the phylogenetic arrangement of the basal subgenus \textit{Niedzwedzka} (Grierson, 1961; Chen et al., 2005) resulted in more resemblances with subgenus \textit{Incarvillea}. For example, the Central Asia endemic species \textit{I. semiretschenskia} has the East Asian habitat type (Illarionova, 2006) and stalked glands identical to members of subgenus \textit{Incarvillea}. The placement of taxa in subgenus \textit{Pteroscleris}, as revealed by ML and BI, might differ due to their respective support values. These differences should be considered carefully within a phylogenetic framework as they indicate the evolutionary significance of these taxonomic relationships. Importantly, the phylogenetic positions of subgenera should be resolved using multilocus data. Here, we provide the separate subgeneric rank of the genus \textit{Incarvillea} from two major basal clades revealed by \textit{trnL-trnF} based Bayesian and Likelihood chronograms as: (((\textit{Olgaea—Amphicome}), \textit{Niedzwedzka, (Incarvillea, Pteroscleris)})). Although our findings differ to some extent from Grierson (1961), Flora Reipublcae Sinicae (Wang et al., 1990), and Chen et al. (2005), we believe that the subgeneric positions are reasonable based on morphology and
biogeography. Therefore, correct placement of taxa requires that researchers consider morphological characters along with molecular data.

4.2. Geographic origin, dispersal and colonization of the montane genus Incarvillea within the three major biodiversity hotspots

Our biogeographic reconstruction proposes that the genus *Incarvillea* originated in Central Asia before the Miocene (i.e., during mid-Oligocene ca. 29.42 Ma) and subsequently dispersed to the Western Himalaya and Sino-Himalaya during the early Miocene (21.12 Ma) (95% HPD: 15.96–26.4 Ma) (Figs. 4 and 5; Table 2). Although the genus may have originated in Central Asia, vicariance events in the Himalaya-Hengduan Mountains (HHM) have led to allopatric speciation, creating differentiation centres or centres of endemism (Chen et al., 2005; Sun et al., 2017) (Fig. 5). The dispersal of biota between the Western Himalaya and the Sino-Himalaya is one of the most noteworthy events related to uplift of the Himalaya-Hengduan Mountains. Thus, successive dispersal and vicariance events within the Himalaya may have driven additional species divergence and lineage diversification. The lack of differentiation centres in the Western Himalaya may be related to the influence of ‘Westertles’, prevailing winds from west to east, which enhance dispersal ability despite strong mountain barriers (Rana et al., 2019). This suggests that despite barriers to migration, weak dispersal of *Incarvillea* species might have been expedited by the Westertles, aiding migration from their region of origin over long distances in the early Pliocene (Chen et al., 2005; Qiong et al., 2017).

Despite gaps in taxon sampling and uneven geographical sampling, previous studies suggest that the southern part of the QTP may provide suitable landscape connectivity for species to migrate (Rana et al., 2019). Mountains do not form uniform barriers; river valleys cut deep into the Himalayas, crossing drainage divides, and forming dispersal corridors. For members of *Incarvillea*, geoclimatic factors are also likely to influence species that show weak dispersal abilities (Rana et al., 2019) as revealed by Rana et al. (2020a) in *Mirabilis himalaica* (Edgew.) Heimerl. Pollen dispersal in *Incarvillea* species is mediated by specialized floral structures with obvious herkogamy, often including a bilobed sensitive stigma and anther appendages (Cutting, 1921; Han et al., 2008; Ai et al., 2013). However, the reproductive strategies of alpine species of the genus *Incarvillea* endemic to the HHM region have not yet been fully characterized (Ai et al., 2013). Moreover, temperature shifts in mountain environments possess strong effects on the elevational distributions of species (Pearson and Dawson, 2003), limiting further suitable expansion. Therefore, plants must either adapt to existing environments and colonize longitudinally within the environment or suffer extinction (Rana et al., 2018, 2019, 2020a, b).

A floristic relationship between Central Asian and East Asian *Incarvillea* is likely to exist. The connection is supported by the distribution of the Central Asian endemic *I. semiretschenskia*, which is found outside of the East Asia type habitat in the eastern part of the ancient Chu-Li mountains (Illarionova, 2006). Such a connection can also be seen when the distribution range of the genus *Incarvillea* is projected under different climatic scenarios (paleoclimate, current, future), forming a ring of suitability around the QTP (Rana et al. unpublished manuscript). Our study may have found support for the biogeographic connection of the Central Asian species *Incarvillea semiretschenskia* within the subgenera *Incarvillea* and *Pteroscleris* if we had sampled *I. potaninii* from Mongolia. Nevertheless, the species *I. potaninii* of subgenus *Incarvillea* is still distributed in Mongolia (north of the QTP), which suggests that a floristic relationship between Central Asia and East Asia exists.

Alternatively, biogeographic analyses using *Incarvillea potaninii* (from Mongolia) may indicate that the dispersal route between Central and East Asia was direct rather than through the Mountains of SW China, which strongly supports the sister–group relationship of subgenus *Niedzwedzkia* with subgenera *Incarvillea* and *Pteroscleris*. The continuous distribution of the genus *Incarvillea* from the Mountains of Central Asia, the Himalaya-Hengduan Mountains to NE China and Mongolia, may have partly resulted from the Quaternary expansion as part of its range was glaciated (Chen et al., 2005).

Regardless, the intense dispersal of *Incarvillea* occurred from the Miocene onwards, while the colonization of the biodiversity hotspots and adjoining regions by representatives of *Incarvillea* subgenera started during the Quaternary period. The accelerated floristic exchange between the three major biodiversity hotspots (along with adjoining regions) might have been encouraged by Miocene cooling (Flower and Kennett, 1994; Zachos et al., 2001), which increased the occurrence of montane habitats in these regions.

4.3. Morphological characters evolution

We inferred ancestral states for several taxonomically important morphological stochastic characters to nodes deeply nested within the tree (Fig. 6). Our study strongly confirms that the most precise ancestral character that categorizes the genus *Incarvillea* is suffrutescose habit. Morphological data has traditionally been used to differentiate between *Incarvillea* subgenera. Of the characters we tested, capsule texture and seed wing, which had the highest AICc weights of the model, showed the greatest ability to differentiate between the five *Incarvillea* subgenera (Chen et al., 2004).

Additional specific ancestral characters distinguish *Incarvillea* subgenera. The large subgenus *Pteroscleris* is characterized by herbaceous habit and radical leaves, which are derived from ancestral suffrutescose habit and alternate leaves, respectively. Subgenus *Pteroscleris* also shares the following characters with subgenus *Incarvillea*: racemose inflorescence, glabrous anther, cylindrical capsule, absence of capsule wing, and loculicidal capsule dehiscence. Taken together, these findings indicate that subgenus *Pteroscleris* is advanced and is more closely related to subgenus *Incarvillea* (Grierson, 1961; Chen et al., 2005). The characters that typify the subgenus *Amphicome* are primitive (Grierson, 1961; Chen et al., 2005), except for the frequency with which they express the ancestral states of some characters (e.g., pilose anther texture, fibrous capsule texture, commas at the end of seed wings). These characteristics are consistent with Grierson’s views of the evolution of these three subgenera (Chen et al., 2005), suggesting that the subgenus *Amphicome* is ancestral.

---

Fig. 3. Bayesian majority-rule consensus tree of the genus *Incarvillea* inferred from (a) plastid trnL-trnF, (b) nrITS, and (c) combined (trnL-trnF and nrITS) sequences. Bayesian posterior probabilities (PP > 0.50) are represented below the branches. Maximum likelihood bootstrap values/maximum parsimony bootstrap values are indicated above branches, with “-” indicating support values of less than 50% or collapse in analysis. Colour legend for the genus *Incarvillea* clade: purple, subgenus *Pteroscleris*; blue, subgenus *Incarvillea*; cyan, subgenus *Niedzwedzkia*; green, subgenus *Amphicome*; red, subgenus *Olgara*.
An alternative interpretation is that the subgenus \textit{Amphicome} is primitive, subgenus \textit{Pteroscleris} is advanced, and subgenus \textit{Incarvillea} is an intermediate between these two genera (Chen et al., 2004, 2005). However, the two other subgenera (i.e., \textit{Niedzwedzkia} and \textit{Olgaea}) that form the two major lineages diverged 21.12 Ma, which would place the subgenus \textit{Amphicome} at the root with \textit{Olgaea}, and the subgenera \textit{Pteroscleris} and \textit{Incarvillea} at the root with subgenus \textit{Niedzwedzkia}. The two subgenera (\textit{Olgaea} and \textit{Niedzwedzkia}) that diverged early share ancestral character states. Specifically, the subgenus \textit{Niedzwedzkia}, which diverged first (18.38 Ma), possesses the most ancestral character states (e.g., suffruticosa habit, alternate leaves, a calyx teeth base that is not swollen, racemose in florrescence, glabrous anthers, subligneous capsule texture and opaque seed wings).

\textit{Incarvillea semiretschenscia} (subgenus \textit{Niedzwedzkia}) is the only species with a septifragal capsule (Chen et al., 2005) and possibly could have maintained the ancestral state in term of fruit dehiscence with a basal position in the genus. But, the formation of two main lineages in the genus could have resulted in the loculicidal capsule dehiscence as the ancestral state from subgenus \textit{Olgaea}, from which the septifragal dehiscence might be derived. It seems plausible that not only the capsule dehiscence but also the cylindrical capsule shape and absence of capsule wing from the subgenus \textit{Olgaea} are the ancestral state. The morphologically varied and distinct capsule characteristics (e.g., cylindric, winged and septifragal dehiscence) are peculiar evolutionarily derived character states. Hence, the subgenus \textit{Incarvillea} does not appear to form an intermediate between subgenera \textit{Pteroscleris} and \textit{Amphicome}, but instead forms a transition between subgenera \textit{Pteroscleris} and \textit{Niedzwedzkia}. Additional evidence from the East Asian \textit{I. semiretschenscia}, an endemic to Central Asia (Kazakhstan) with stalked glands, may support this hypothesis (Vassilczenko, 2000; Illarionova, 2006).
5. Conclusions

Our study re-examines the phylogenetic framework of the genus *Incarvillea* using *trnL-trnF* and *nrITS* (individually), and combined molecular data sets by BI, ML and MP. Furthermore, this study represents the first step towards an understanding of the biogeographical history of the genus *Incarvillea*. Our time-calibrated phylogenetic tree revealed that the genus *Incarvillea* is monophyletic and originated in Central Asia during mid-Oligocene ca. 29.42 Ma (95% HPD: 23.34–35.78 Ma) and the earliest diverging lineages arose during the early Miocene ca. 21.12 Ma (95% HPD: 15.96–26.4 Ma). Diversification of these lineages followed two simultaneous paths: dispersal to the Himalaya led to the *Olgaea*—*Amphicome* clade; dispersal to the Hengduan Mountains led to the *Niedzwedzkia*—*Incarvillea*—*Pteroscleris* clade. The subgenus *Pteroscleris* may have undergone a recent adaptive radiation, as it is a relatively new colonizer. Although our ability to make biogeographical inferences is limited due to low bootstrap support for nodes and limited taxon sampling, these analyses provide general insights into the complex history of the *Incarvillea* genus, which is characterized by repeated colonization of biogeographically defined areas. Stochastic character mapping provides strong evidence that there are two major lineages with ancestral character states; however, the phylogenetic placement of subgenus *Niedzwedzkia* must await further clarification. In the meantime, we expect that work on character evolution and the phylogeny of the genus *Incarvillea* will be well served by placing subgenus *Niedzwedzkia* close to subgenus *Incarvillea*. Finally, although base-line taxonomic work and additional sample collection are required, this study demonstrates the usefulness of using large secondary molecular data sets and biogeography to determine the evolutionary origins and diversification of plant taxa that occur in biodiversity hotspots.

Author contributions

SKR, SC and HS conceived the research design. SKR, HKR and DL performed the data analysis. SKR wrote the manuscript with the support of DL, HKR, and HS. All authors contributed to the final editing.

Declaration of competing interest

To the best of our knowledge, the named authors have no conflict of interest, financial or otherwise.

Acknowledgements

The authors thank Dr Zhuo Zhou for his help with the molecular dating analysis, Raymond Porter and Alexander Robert O’Neill (USA) for valuable inputs and English editing, and Dr Niu Yang, Ya-Zhou Zhang, Qia Wang for providing the plant photographs. This study was supported by the Second Tibetan Plateau Scientific Expedition and Research (STEP) Program (2019QZKK00502), the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20050203), NSFC-Yunnan joint fund to support key projects (U1802242), the Major Program of the National Natural Science Foundation of China (31590823) and the National Natural Science Foundation of China (31570203).
