Acta Botanica Brasilica, 2022, 36: e2021abb0255
doi: 10.1590/0102-33062021abb0255

Original article

Reconstruction of the evolutionary biogeography reveals the origins of *Alpinia* Roxb. (Zingiberaceae): A case of "out-of-Asia" migration to the Southern Hemisphere

Chi Toan Le[1*], Thi Bich Do[1], Thi Minh Anh Pham[1], Van Du Nguyen[2, 3], Sinh Khang Nguyen[2], Van Hoach Nguyen[4], Phi Bang Cao[4] and Wyckliffe Omondi Omollo[5, 6]

Received: August 09, 2021
Accepted: January 20, 2022

ABSTRACT

Alpinia is considered the largest genus of Zingiberaceae with ca. 250 species occurring in tropical and subtropical Asia, Australia, and Pacific Islands. The historical biogeography of *Alpinia* was conducted to explain where *Alpinia* originated and how it migrated to other regions. The phylogeny, divergence times and ancestral area reconstruction of *Alpinia* were performed by using the molecular data based on the comprehensive taxon sampling. Our results provide an objective approach to understand the historical biogeography of *Alpinia*. The genus originated in Asia during the Late Cretaceous ca. 69 Ma and started to diverge after the K–Pg boundary during the early Paleocene with the presence and development of the tropical rainforest and a warm, moist climate. *Alpinia* migrated to Malesia and then dispersed to Australasia. The molecular analyses supported the diversification of *Alpinia* in Asia and Malesia. Additionally, the Indian *Alpinia* has likely a common ancestor with *Renealmia* and *Aframomum*, and it is possible that after originating in Asia, *Alpinia* migrated from Asia to India then to Africa during the early period of collision between the Indian subcontinent and Eurasia to form the common ancestor of Indian *Alpinia*, *Renealmia*, and *Aframomum*. Our phylogeny provides a framework for studies in biogeography, comparative ecology, and evolution.

Keywords: Zingiberaceae, *Alpinia*, molecular phylogeny, historical biogeography, Asia, origin.

Introduction

Zingiberaceae (the gingers family) is the largest in the order Zingiberales with 53 genera and over 1200 species in the world (Kress *et al*. 2002). In the Zingiberaceae family, *Alpinia* is considered the largest genus with ca. 250 species. *Alpinia* is distributed in tropical and subtropical Asia, Australia, and Pacific Islands (Larsen *et al*. 1998; Wu & Larsen 2000; Kress *et al*. 2005). Kress *et al*. (2005) suggested that *Alpinia* plays an important ecological role in understanding the understory vegetation of tropical

---

1 Hanoi Pedagogical University 2, 32 Nguyen Van Linh, Xuanhoa, Phucyen, Vinhphuc, Vietnam
2 Institute of Ecology and Biological Resources - Vietnam Academy of Science and Technology, Hanoi, Vietnam
3 Graduate University of Science and Technology - Vietnam Academy of Science and Technology, Hanoi, Vietnam
4 Faculty of Natural Sciences, Hung Vuong University, Phutho, Vietnam
5 State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China
6 University of Chinese Academy of Sciences, Beijing, 100049, China

*Corresponding author: Lechitoan@hpu2.edu.vn

and subtropical forests. Several species of *Alpinia*, such as *A. officinarum*, *A. galanga*, *A. coriandriodora*, *A. kwangsiensis* are used as important economical, ornamental, food, or medicinal plants (Wu & Larsen 2000).

The molecular phylogenetic relationships of *Alpinia* have been conducted by several studies (Rangsiruji *et al*. 2000; Kress *et al*. 2002; 2005; Le 2021). These studies used molecular data from both ITS and *mat*K regions. The results of molecular analyses from the studies showed that *Alpinia* is polyphyletic; six clades within *Alpinia* were recognized. Kress *et al*. (2005) based on dataset from 72 species of *Alpinia* suggested that the molecular results are associated with morphological features of clades within *Alpinia*, and further study with comprehensive sampling is necessary.

de Boer *et al*. (2018) and Poulsen *et al*. (2018) constructed the phylogeny of *Amomum* and *Elettaria* respectively based on molecular data including several genera of the subfamily Alpinioideae. The two studies supported the non-monophyly of *Alpinia*, *Elettaria*, and *Amomum*, in which several species of these three genera are nested. The misidentification of some Cardamom species in the *Alpinia zerumbet* clade could be a reason for this complication (Poulsen *et al*. 2018). Further studies and specific sampling will be needed to solve the polyphyly of some clades in Alpinioideae.

Le (2021) conducted an update for phylogeny of *Alpinia* based on 110 species of *Alpinia*. *Alpinia* is shown to consist of six clades, clade I and clade II placed together with *Aframomum*, *Renealmia*, *Amomum*, *Elettariopsis* and *Paramomum*; while clades III, IV, V and VI formed together with *Nicolaia*, *Etlingera*, *Amomum* and *Hornstedtia*. Additionally, the study further noted the need to clarify the biogeography of the genus based on current distribution.

Kress & Specht (2006) conducted an investigation of evolution, biogeography, and diversification of Zingiberales based on all the families in the order and three gene regions (including *rbc*L, *atp*B, and ISS), but without any sample of *Alpinia*. The results indicated that the order Zingiberales originated around 124 million years ago (Ma), with major family-level lineages becoming established in the Late Cretaceous (80-110 Ma) and crown lineages within each family beginning to diversify during the early to Mid-Tertiary (29-64 Ma). In this study, Zingiberaceae was estimated to have originated at 105 ± 4.2 Ma. However, the evolutionary history of lineages (including *Alpinia*) in Zingiberaceae was not clarified.

The historical biogeography of *Alpinia* is an interesting question given by the current distribution in distinct areas of the genus; however, it has not been fully studied. Kress *et al*. (2005) conducted phylogenetic analyses to understand the evolutionary patterns of *Alpinia*. The study suggested that the understanding of the patterns of evolution of various morphological and ecological characteristics of taxa can determine the phylogenetic history. However, the study sampled only 72 species of *Alpinia* and did not perform the molecular dating and ancestral area reconstruction analyses, thus the historical biogeography of *Alpinia* was not investigated and remains unclear.

In order to investigate these global and continental-scale biogeographies, it is necessary to formulate a hypothesis for the relationships and divergence times of *Alpinia*. Here we conducted phylogenetic and biogeographical analyses using molecular data from DNA regions based on intensive taxon sampling (including 110 of the 250 species of *Alpinia*). Our aims are to reconstruct the biogeographical history of *Alpinia* reconciled with present-day distributions, fossil data, and geological information to specifically: (1) address the time of origin for *Alpinia* and (2) predict hypothetical distributions, the original center, and probable dispersal routes of *Alpinia* species.

Materials and methods

Taxon sampling

To fully understand the historical biogeography of *Alpinia*, we used the molecular data from the most recent phylogenetic study of the genus *Alpinia* by Le (2021) for our historical biogeography analyses (Tab. S1). In the molecular dataset, all clades in the phylogeny of *Alpinia* were included, thus representing the geographical and taxonomic diversity of the genus. In the molecular dataset, all the duplicates were removed, and we only kept one individual per species. In total, 110 species of *Alpinia* and 39 non-*Alpinia* species were included in the molecular dataset following Kress *et al*. (2002; 2005). Two markers *mat*K and the nuclear ribosomal internal transcribed spacer region ITS (including the 5.8S and flanking ITS1 and ITS2 spacers) were used for analyses. We checked the presence of paralogues in our ITS dataset by evaluation rates of evolution, GC content, and mutations between the 5.8S (conserved region) and the ITS1, ITS2 regions (fast evolving regions). Voucher information and GenBank accession numbers are presented in Tab. S1. Each sequence in the combined dataset was carefully checked by using BLAST (blast.ncbi.nlm.nih.gov) under default parameters. The two single datasets of two genes *mat*K and ITS were aligned by Geneious v.8.0.5 (Kearse *et al*. 2012). Finally, the combined dataset was partitioned into two subsets corresponding to two gene regions for molecular analyses.

Molecular phylogenetic analyses

The two methods maximum likelihood (ML) and Bayesian Inference (BI) were performed on the CIPRES Science Gateway Portal (Miller *et al*. 2010). The ML analyses were conducted in RAxML 8.2.12 (Stamatakis 2006) using the GTR+I+G standard nucleotide substitution model generated by jModeltest 2.1.6 (Darriba *et al*. 2012) for each DNA region and applying 1,000 bootstrap replicates. The BI analysis was performed on MrBayes 3.2.6 (Ronquist *et* *al*. 2012) on the CIPRES using the nucleotide substitution models that estimated separately each gene region by jModeltest 2.1.6 (Darriba *et al*. 2012) (*mat*K: GTR+G, ITS: GTR+I+G). Two independent runs, each comprising four Markov chain Monte Carlo (MCMC) chains, were conducted. We ran the Markov chain Monte Carlo (MCMC) for 10 million generations, and trees were sampled every 1000 generations. To check the effective sample sizes (ESSs) of all relevant parameters (>200) we used Tracer v.1.6 (Rambaut *et al*. 2014). Majority-rule consensus tree and Bayesian posterior probabilities (PP) were obtained after removing the first 25 % of sampled generations as burn-in. The obtained trees were visualised using FigTree v.1.4.0 (Rambaut 2009).

Divergence time estimation

The combined molecular dataset and the uncorrelated lognormal Bayesian method in BEAST v.1.8.0 (Drummond *et al*. 2012) were used to estimate the divergence times of *Alpinia*. We partitioned the combined molecular dataset using the "unlink substitution model" option, and applied a substitution model from the results of jModelTest v.2.1.6 to each partition. The Yule process tree prior was run for all dating. Condamine *et al.* (2015) noted that the Yule prior was recommended in the manual use of BEAST, and most molecular dating analyses used the Yule prior. Although Condamine *et al.* (2015) suggested that the birth-death process had a better fit than the Yule process in some situations, they also suggest that some recent studies have used both the Yule and birth-death process but they did not find any differences in age estimates between the two models. The birth-death process will not necessarily always better fit the data just because it seems more biologically realistic than a simpler Yule model, especially for shallower divergence times. The uncorrelated lognormal clock model is more likely to yield accurate estimates than the uncorrelated relaxed clock model that assumes an exponential distribution of the evolutionary rates.

We used lognormal distribution for the calibration point, and normal distribution was used for the divergence age following results from the previous studies (Lu *et al*. 2013; Deng *et al*. 2014; Liu *et al*. 2018). We ran the two separate MCMC for 150 million generations with trees taken every 15,000 generations. Tracer v.1.6 (Rambaut *et al*. 2014) was used to check if ESSs for all relevant parameters were well above 200 and that stationarity had been reached. A maximum clade credibility tree was then built by TreeAnnotator v.1.8.0 (Rambaut & Drummond 2010) with the initial 25 % of trees discarded as burn-in. The final result was visualized in Figtree v.1.4.0 (Rambaut 2009).

Hickey & Peterson (1978) noted that the fossil leaf of genus *Zingiberopsis* is closely related to the extant genus *Alpinia*. Wilf *et al*. (2000) further recognized that the nearest living relative to *Zingiberopsis* is the Asian genus *Alpinia*. The fossil genus *Zingiberopsis* was known from Late Cretaceous through the earliest Oligocene (Hickey & Peterson 1978; Wilf *et al*. 2000). Moreover, Kress & Specht (2006) used *Zingiberopsis* at 65 Ma for estimating the divergence time of Zingiberales. In addition, Forest (2009) suggested that the age of a fossil is generally treated as a minimum constraint in calibration procedures for the true age of the clade which the fossil is assigned. Thus, we accepted *Zingiberopsis* to be related to *Alpinia* and used it to calibrate the crown node of *Alpinia* at 65 Ma (95 % HPD of 63.1 - 69.1 Ma) in this study. *Alpinia* is non-monophyletic, some non-monophyletic clades of *Aframomum*, *Renealmia*, *Amomum*, *Elettariopsis*, *Paramomum*, *Nicolaia*, *Etlingera*, *Hornstedtia*, *Leptosolena*, *Vanoverberghia*, and *Plagiostachys* were nested within *Alpinia*, however this could be as a result of misidentifications (Poulsen *et al*. 2018). Therefore, we considered the crown node of *Alpinia* including all the nested species within this genus.

Kress & Specht (2006) estimated the divergence time of Zingiberaceae at 105 ± 4.2 Ma, hence, we consider 105 Ma as the divergence age of Zingiberaceae and calibrated the stem node of Zingiberaceae at 105 Ma (95 % HPD of 100.1 - 109.9 Ma). Additionally, the two extinct zingiberalean species *Spirematospermum chandlerae* and *Spirematospermum wetzleri* were respectively described by Friis (1988) and Li *et al*. (2018) as fossils of Zingiberaceae. However, Magallón *et al*. (2015) suggested that, although *Spirematospermum* is usually considered to Zingiberaceae, the genus likely has a closer relationship to Musaceae based on the presence of a chalazal member, rudimentary micropylar collar and hilar cavity. These features do not occur in extant Zingiberaceae. *Spirematospermum* has been used to calibrate the stem nodes of Zingiberales and Commelinales (Magallón *et al*. 2015). Thus to determine the position of the *Spirematospermum* in divergence time estimation of *Alpinia* or even Zingiberaceae is uncertain. The divergence age of Zingiberaceae from Magallón *et al*. (2015) is likely too young to be used as a *Zingiberopsis* fossil. Therefore, we do not consider *Spirematospermum* as the calibration point of Zingiberaceae from Magallón *et al*. (2015) for divergence time estimation of *Alpinia*.

Ancestral area reconstruction

Bayesian approach to dispersal-vicariance analysis (Bayes-DIVA; Nylander *et al*. 2008) and BioGeoBEARS (Matzke 2013) were applied to reconstruct the ancestral area of *Alpinia*.

The Bayes-DIVA method can minimize phylogenetic uncertainty by utilizing the posterior distribution of trees resulting from BEAST and generate credibility support values for alternative phylogenetic relationships (Nylander *et al*. 2008; Yu *et al*. 2011). For the Bayes-DIVA analysis, the results from the BEAST analysis were used as input data. The final representative tree was computed after the first 25 % of sampled generations discarded as burn-in.

The BioGeoBEARS was implemented in R (R Development Core Team 2016). Two likelihood-based models under the

BioGeoBEARS method were implemented: Dispersal-Extinction-Cladogenesis (DEC; Ree & Smith 2008), and the likelihood version of dispersal-vicariance (DIVA; Ronquist 1997; herein DIVALIKE). In each model, an additional "*j*" parameter (founder event/jump speciation) was added, which allows descendant lineages to have a different area from the direct ancestor (Matzke 2013; Vasconcelos *et al*. 2017). Thus, a total of four models resulted. Species distribution data were obtained from field and herbarium studies (E, U, L, HN, PE, KUN, HAST), references (Larsen *et al*. 1998; Wu & Larsen 2000). Additional data were downloaded from the Global Biodiversity Facility (http://www.gbif.org). Atkins *et al.* (2020) noted that the sampling density can also impact the historical biogeography analyses, thus a comparison of our sampling rates against current species estimates to evaluate the impact of sampling density on the results is presented in Tab. S2.

Six main regions were defined based on the floristic regions and the extant distribution of *Alpinia* and outgroups: A, Asia (including East Asia, Indochina and Myanmar); B, India (including India and Sri Lanka); C, Malesia (including Malaysia, Indonesia, Philippines, and Singapore); D, Australasia (including Australia, New Zealand, New Guinea, and Pacific Islands); E, America; F, Africa. The floristic regions named as America (E) and Africa (F) were set for outgroups taxa only. Thus, we set six as the maximum number of areas for both the two methods in ancestral area reconstruction analyses.

Results

Phylogeny of Alpinia

The aligned positions of *mat*K and ITS data sets were 2994 and 837 base pairs, respectively. The combined matrix of the two markers consisted of 3831 base pairs (Tab. S3). We found that most ITS sequences are pseudogenes, only 21/144 sequences (14.5 %) are functional genes. The functional sequences are found in all the six clades of *Alpinia*. Bailey *et al.* (2003) suggested that it should not be assumed that functional and non-functional sequences included in phylogenetic analyses of taxa are necessarily paralogous to one another or that functional copies are necessarily orthologous, while exclusion of putative pseudogenes a priori from phylogenetic analyses may not be necessary (Razafimandimbison *et al.* 2004). There is only one individual per species in our molecular data (including the ITS dataset) and no indication in our data that the observed polymorphisms predate the divergence of these taxa. Therefore, it is reasonable for our ITS dataset to be used to conduct the phylogenetic and historical biogeographic analyses of *Alpinia* (Razafimandimbison *et al.* 2004).

The topologies from ML and BI analyses of the combined dataset were congruent, with the differences having low support. Thus, the molecular phylogeny of *Alpinia* is presented in Fig. 1 based on the Bayesian tree with BS and PP values. *Alpinia* was supported to be non-monophyletic and six major clades were recognized within *Alpinia* (Fig. 1). These molecular results were congruent with the previous studies (Kress *et al*. 2002; 2005; Le 2021). However, the relationships among these clades are not well resolved and showed a similar level of resolution to previous results (Kress *et al*. 2002; 2005; Le 2021).

Clade I was weakly supported as close relative to *Renealmia* and *Aframomum* (Fig. 1), while the remaining clades were well supported as separate clades by molecular data except clade IV (Fig. 1). Some non-monophyletic clades of *Nicolaia*, *Etlingera*, *Amomum*, *Hornstedtia*, and *Leptosolena* are closely related to clade V but weakly supported, while *Vanoverberghia* is nested within clade V (Fig. 1). *Plagiostachys* is nested within clade IV. Nevertheless, the systematic position of the clades is still uncertain.

Age estimates

The divergence time estimates for *Alpinia* are presented in Fig. 2. The stem node of *Alpinia* was estimated to be 69.06 Ma (95 % HPD: 64.86-75.06 Ma) (node 0, Fig. 2), while, the crown age of *Alpinia* was estimated to be 65.06 Ma (95 % HPD: 64.03-69.39 Ma) (node 1, Fig. 2). Clade VI splitted from its ancestor at 62.96 Ma (95 % HPD: 57.31-68.01 Ma) (node 3, Fig. 2). Clades III splitted from a common ancestor with clade V together with *Etlingera*, *Hornstedtia*, and *Nicolaia* at 56.86 Ma (95 % HPD: 49.66-63.22) (node 5, Fig. 2). The largest clade - clade IV diverged at 53.05 Ma (95 % HPD: 45.47-59.97 Ma) (node 9, Fig. 2).

Ancestral area reconstruction

The comparison of sampling rates indicated that our sampling density is better than Kress *et al.* (2005) in all geographic regions for *Alpinia* with the highest in Asia (164 %) and lowest in Australasia (100 %) with the same number of species distributed in this region (Tab. S2). Thus, the results from this study are more advanced. The reconstructions obtained for the major clades of interest using the BioGeoBEARS and Bayes-DIVA were highly congruent (Tab. 1). However, the results from BioGeoBEARS show higher resolution for more nodes than Bayes-DIVA. Among the four models, the results indicated higher log-likelihood values for three parameters in comparison to two parameters (Tab. 2), hence, indicating jump speciation (*i.e.*, dispersal between non-adjacent areas) as an important pattern in range variation of *Alpinia*. Moreover, the model DEC+*j* in the BioGeoBEARS analyses was the best-fit biogeographical model. Therefore, we only present the reconstruction of BioGeoBEARS under the DEC+*j* model (Fig. 3). The result of ancestral area reconstructions from Bayes-DIVA is shown in Fig. S1.

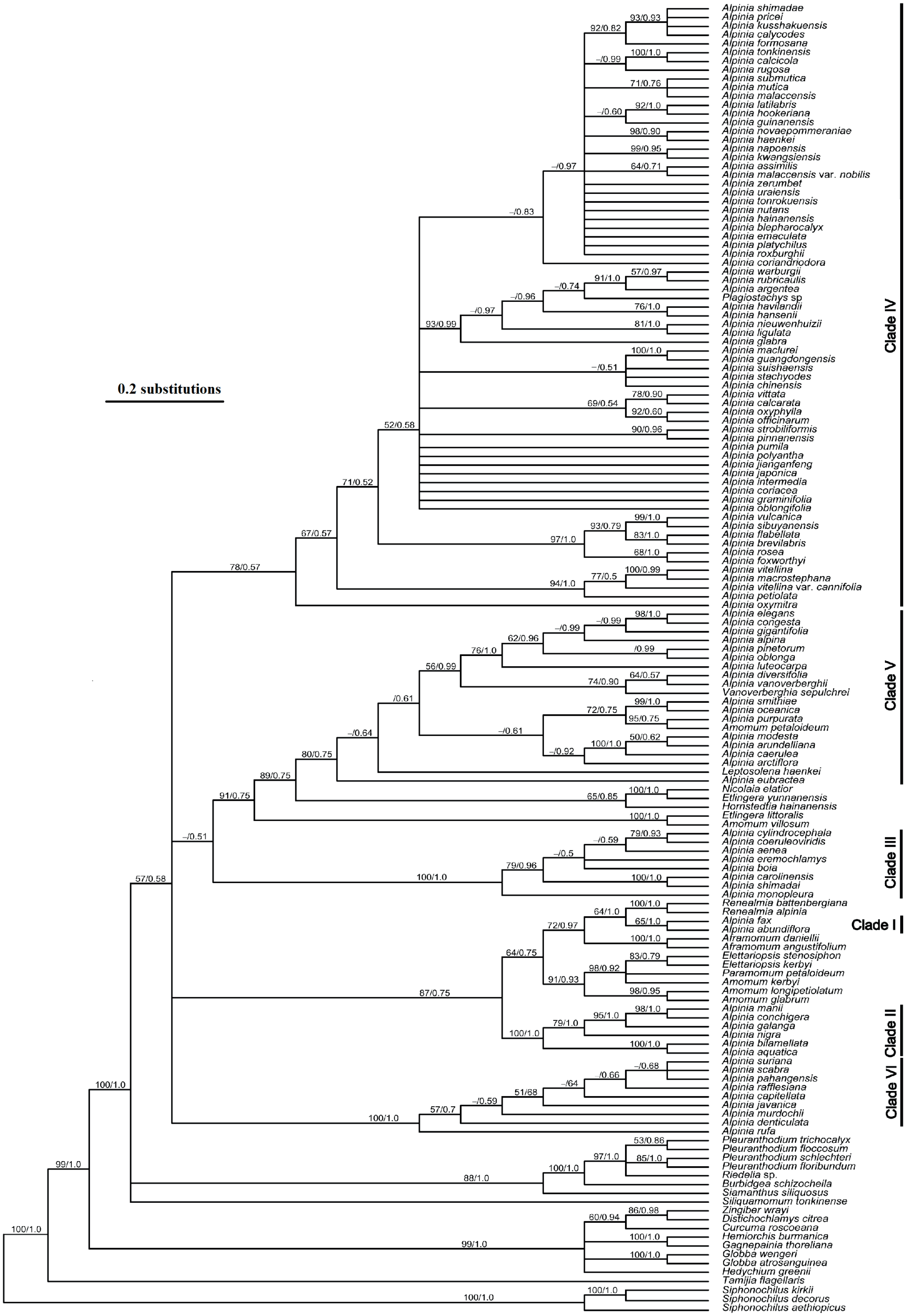

Figure 1. Majority rule consensus tree of *Alpinia* from the Bayesian inference based on the combined datasets of two DNA regions (*mat*K *and* ITS). ML bootstrap values and posterior probabilities (PP) of the BI analysis are presented above the branches. "–" indicates the support values less than 50 %. The circumscription of sections follows the classification of Kress *et al*. (2005).

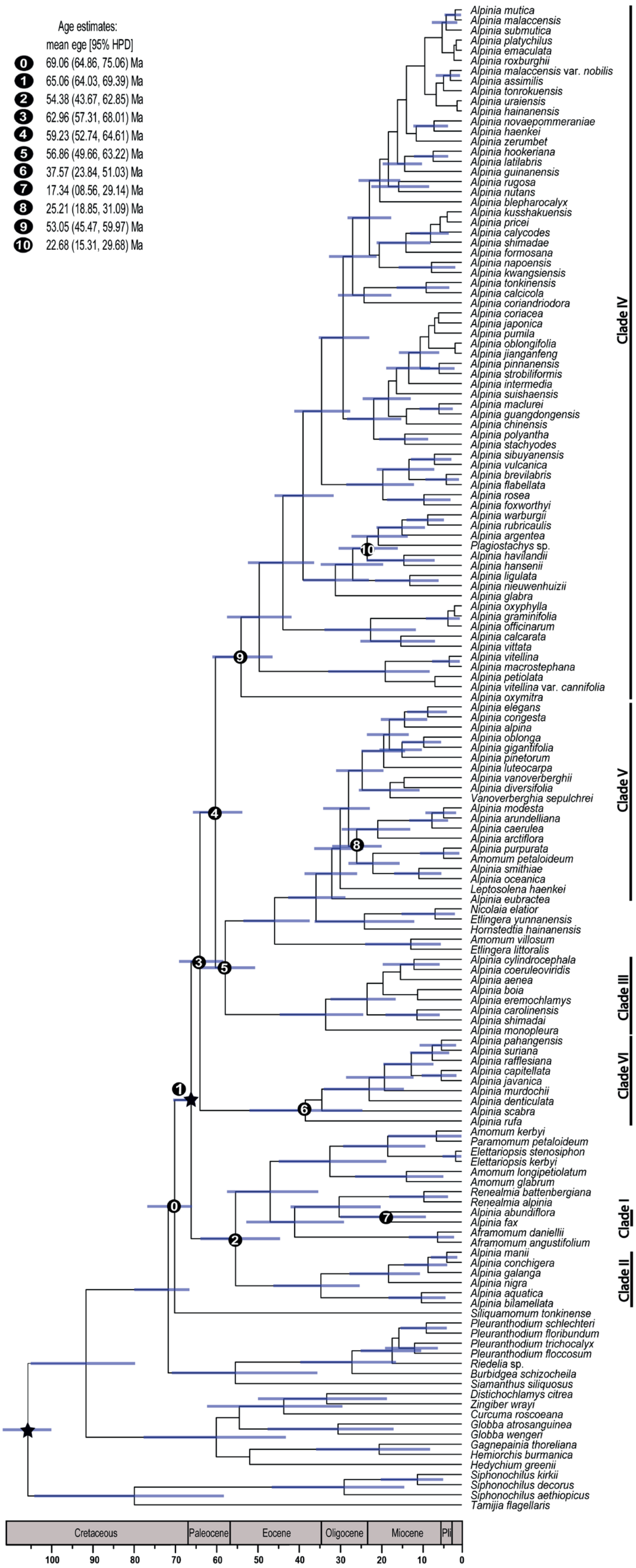

Figure 2. Maximum clade credibility tree inferred from BEAST based on the combined datasets of two DNA regions. The bars around node ages indicate 95 % highest posterior density intervals. Node constraints are indicated with stars. Nodes of interests were marked as 0–10.

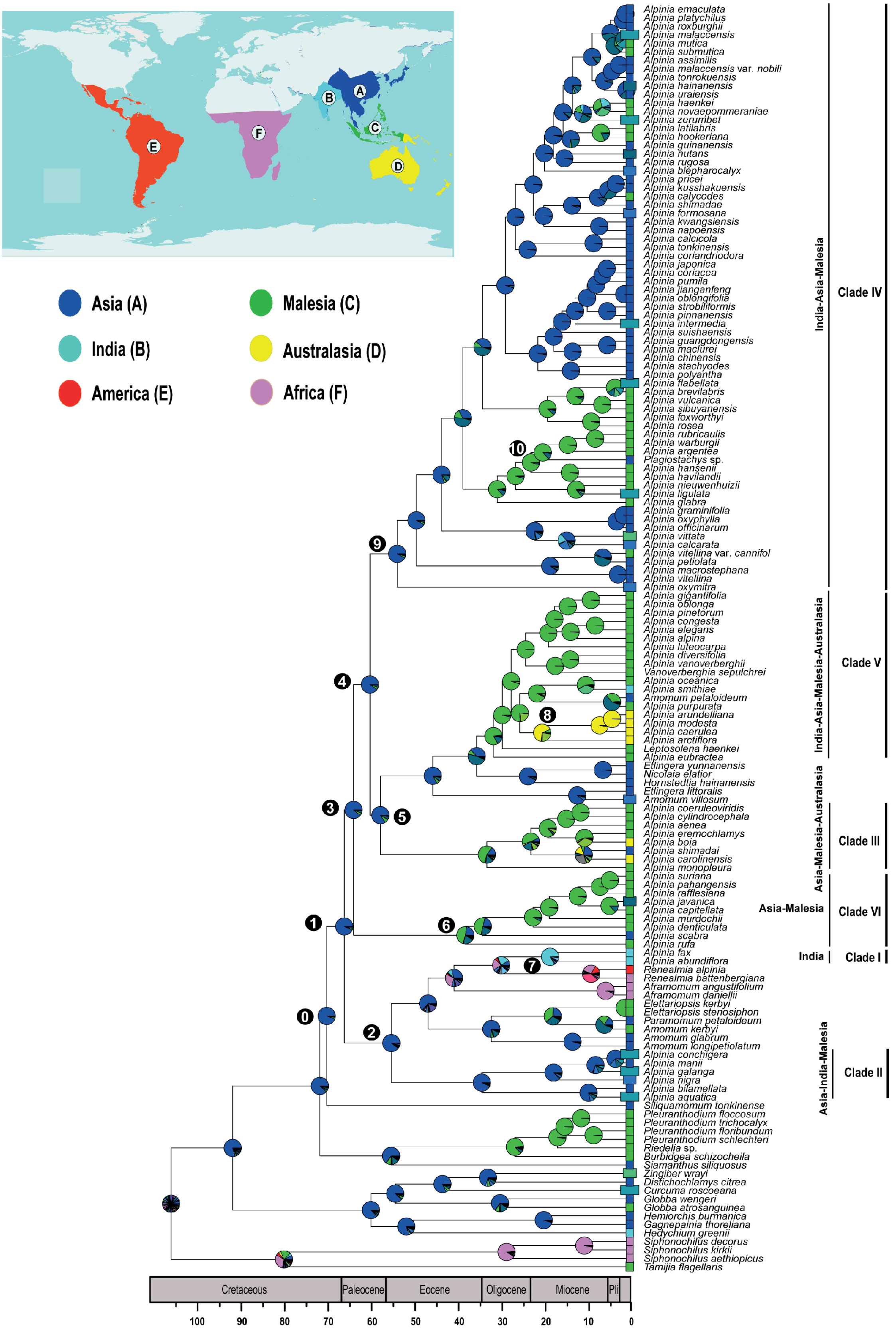

Figure 3. Ancestral area reconstruction of *Alpinia* by BioGeoBEARS ($j = 0.0028$, $\mathrm{LnL} = -250.18$). Geologic time scale is shown at the bottom. Area abbreviations are as follows: A, Asia (including East Asia, Indochina, and Myanmar); B, India (including India and Sri Lanka); C, Malesia (including Malaysia, Indonesia, Philippines, and Singapore); D, Australasia (including Australia, New Zealand, New Guinea, and Pacific Islands); E, America; F, Africa.

Our results suggest that *Alpinia* originated in Asia and then diverged to India, Malesia and Australasia (Fig. 3). The most recent common ancestor of *Alpinia* was likely located in Asia. Clade II of *Alpinia* originated in Asia and then widely spread in this area (Fig. 3). Similarly, clades III, IV, V and VI originated or split from ancestors in Asia, and then spread in Asia or dispersed to Malesia (Fig. 3), while two Indian species (*A. fax* and *A. abundiflora*) of clade I originated in India, but the geographical ancestral area of this clade is still not clear.

Discussion

*Origin of **Alpinia** in Asia*

Alpinia was supported as non-monophyletic by some studies (Kress *et al*. 2002; Kress *et al*. 2005; Le 2021; this study). Within *Alpinia*, several other genera were nested such as *Aframomum*, *Amomum*, *Renealmia*, *Elettariopsis*, *Paramomum*, *Leptosolena*, *Vanoverberghia*, *Plagiostachys*, *Etlingera*, *Hornstedtia*, and *Nicolaia* (Figs. 1, 2, 3) (Kress *et al*. 2002; Kress *et al*. 2005; Le 2021). Moreover, some of the above genera are also recognized as non-monophyletic (*Amomum*, *Etlingera*) (Figs. 1, 2, 3), therefore determining the geographical ancestral area of *Alpinia* is difficult. However, our results of ancestral area reconstruction analyses indicated that the main nodes in the backbone of the ancestral area reconstruction tree are in Asia (Fig. 3), which showed that the common ancestor of *Alpinia*, and its allies originated in Asia. This result is consistent with the only fossil (*Zingiberopsis*) of *Alpinia* dating to the Late Cretaceous which was discovered in Asia (Hickey & Peterson 1978; Wilf *et al*. 2000).

The biogeographical analyses and divergence time estimations suggest that the stem group of *Alpinia* originated in Asia ca. 69.06 Ma (95 % HPD: 64.86, 75.06 Ma) during the Late Cretaceous (node 0, Figs. 2, 3; Tab. 1), and soon after the initial origin *Alpinia* started to diverge in early Paleocene at 65.06 Ma (95 % HPD: 64.03, 69.39 Ma) (node 1, Figs. 2, 3; Tab. 1). During the Late Cretaceous, the Asian continent was a vast land dominated by forests (Lomolino *et al*. 2010), and the climate here was favorable (Zachos *et al*. 2001). The appearance of the K–Pg boundary (66 Ma) made a change for terrestrial plant ecosystems (Wolfe & Upchurch 1986; Vajda *et al*. 2001; McElwain & Punyasena 2007; Nichols & Johnson 2008; Schulte *et al*. 2010; Wang *et al*. 2012). The K–Pg boundary is associated with the Cretaceous–Paleogene extinction event, a mass extinction which destroyed the majority of the world's Mesozoic species (Fortey 1999). Up to 57 % of all plant species in North America disappeared (Wilf & Johnson 2004), 40 % disappeared in Central Africa and New Zealand (Morley 2000), and 70 % disappeared in tropical South America (Parra *et al*. 2007) at the end of the Cretaceous (Wang *et al*. 2012). However, the modern tropical rainforests began to appear after the K–Pg boundary (*e.g.* Upchurch & Wolfe 1987; Wing & Boucher 1998; Morley 2000).

Table 1. Summary of supported clades, divergence age estimation (Ma) by BEAST and ancestral area reconstructions by Bayes-DIVA. Node numbers refer to Figs. 2, 3. Area abbreviations are as follows: A, Asia (including East Asia, Indochina and Myanmar); B, India (including India and Sri Lanka); C, Malesia (including Malaysia, Indonesia, Philippines, and Singapore); D, Australasia (including Australia, New Zealand, New Guinea, and Pacific Islands); E, America; F, Africa.

Node	Bayesian PP	Age estimates mean (95 % HPD) [Ma]	BioGeoBEARS (DEC+*j*)	Bayes-DIVA (maximum probability)
0	1.0	69.06 (64.86, 75.06)	A	A (100)
1	< 0.90	65.06 (64.03, 69.39)	A	A (98.43)
2	< 0.90	54.38 (43.67, 62.85)	A	A (84.67)
3	< 0.90	62.96 (57.31, 68.01)	A	AC (50.29) A (48.86)
4	< 0.90	59.23 (52.74, 64.61)	A	A (75.74)
5	< 0.90	56.86 (49.66, 63.22	A	AC (79.81)
6	< 0.90	37.57 (23.85, 51.04)	C	AC (100)
7	1.0	18.34 (08.56, 29.14)	AB	BEF (49.49) BF (49.49)
8	< 0.90	25.21 (18.85, 31.09)	C	C (100)
9	< 0.90	53.05 (45.47, 59.97)	A	A (100)
10	< 0.90	22.68 (15.30, 29.68)	C	AC (100)

Table 2. Comparison of the fit of different models of biogeographical range evolution and model specific estimates for different parameters (*d* = dispersal, *e* = extinction, *j* = weight of jump dispersal (founder speciation)).

Model	Parameter No	LnL	*D*	*e*	*j*	AIC	AIC weight
DEC	2	–250.58	5.5×10^{-3}	6.2×10^{-2}	0	496.3	0.468
DEC+*j*	3	–250.18	5.1×10^{-3}	6.3×10^{-2}	0.0028	495.5	0.258
DIVALIKE	2	–253.24	5.8×10^{-3}	6.8×10^{-2}	0	532.3	1.9×10^{-5}
DIVALIKE+*j*	3	–263.13	2.5×10^{-3}	1.0×10^{-8}	0.0143	510.5	1

Forests (including tropical forests, temperate forests and mountainous forests at high elevation) were present throughout the Northern Hemisphere during the early Cenozoic, which allowed intercontinental dispersal of plants (Wei *et al*. 2015; West *et al*. 2015). Morley (2000, 2003), Jablonski (2004), Schuettpelz & Pryer (2009), and Couvreur *et al*. (2011) suggested that the tropical forests were present during the Paleocene/Eocene, however, at that time this biome was located at the middle paleolatitudes where the climate was more favorable. Most *Alpinia* species prefer to grow in low- to mid-elevation forests and form clumps with stems from 1-3 m high except some species from east of Wallace's Line (*e.g.* the Moluccas, Fiji) which tend to grow much larger (*A. regia*, *A. boia*) or some species in New Guinea and Sulawesi which are found in montane forests up to 2000 m above sea level. Notably, the presence of *Alpinia* in tropical and subtropical forests is quite common (Kress *et al*. 2005).

Additionally, during the early Paleocene, the climate was warm before the thermal maximum in the late Paleocene (Zachos *et al*. 2001). Thus, the presence and development of the forest biomes could have provided potential habitats for the origin and divergence of *Alpinia* since it likes to grow in the understory, along forest margins, and in light gaps in the forests.

Many species of *Alpinia* are dominant in wetlands and along watercourses (Kress *et al*. 2005). The fruit of *Alpinia* is capsule and usually globose, thus the exocarp is able to protect the seeds for a long time even in a humid condition. *Alpinia* seems to be adapted to disperse by watercourses. However, the fruits of *Alpinia* have bright colors (orange and red), which can attract birds or small rodents, thus birds and small rodents can also act as dispersal agents for the seeds of *Alpinia* (Deng *et al*. 2015). Larsen *et al*. (1998) suggested that fruits of Zingiberaceae can be dispersed by birds, while seeds can be dispersed by ants, mice, squirrels or by rain and running water. Therefore, during the early evolution, *Alpinia* likely diverged and widely spread only in Asia (Fig. 3) creating endemic species in Asia, and later dispersed to Malesia and Australasia.

Migration to Malesia

Alpinia presents the richest diversity in humid tropical lowlands and mountain forests of Malesia and SE Asian monsoon regions (Larsen *et al*. 1998). Our biogeographical analyses supported that Asia and Malesia are the two richest-species regions where *Alpinia*'s ancestor was often dominant (Fig. 3). Results from ancestral area reconstruction analyses inferred that Asia was the ancestral area of Malesian *Alpinia*, and that *Alpinia* migrated from Asia to Malesia and back to Asia several times since 37.57 Ma during the Late Eocene (node 6, Fig. 3). The two areas, Asia and Malesia, are very close and even contiguous. In the earliest migration period, much of Sundaland was submerged (Hall 2012), however, many island chains between Asia and Australasia belonging to the Malesia region appeared during the late Eocene to early Oligocene (Holloway & Hall 1998; Meimberg *et al*. 2001; Morley 2003; Clayton *et al*. 2009; Hall 2012; Zahirovic *et al*. 2014; Liu *et al*. 2018). Some of the potential dispersers, *e.g.* fruit pigeons and large cuckoos, can travel at least several hundred kilometers (Higgins 1999) and the large seeds were occasionally passed within 5.5 h (Wotton *et al*. 2008; Liu *et al*. 2018). Moreover, all rivers in SE Asia flow into the South China sea. Thus, we hypothesize that following the watercourses, fruits of *Alpinia* could have migrated to the islands of the Malesia region. All the migrations between Asia and Malesia occurred since the late Eocene (less than 37.57 Ma, 95 % HPD: 23.85-51.04 Ma) (node 6, Fig. 3). In addition, the climate during this time was warm and moist in most of Southeast Asia (Zachos *et al*. 2001), which may have facilitated the migration of *Alpinia* through Asia to Malesia and then back to Asia.

Dispersal to Australasia

Our ancestral area reconstruction analyses indicated that all Australasia lineages of *Alpinia* originated in Malesia and dispersed to Australasia two times since the late Oligocene 25.21 Ma (95 % HPD: 19.19-31.09 Ma) (node 8, Fig. 3). During the late Oligocene, Malesia and Australasia were very close, thus, a hypothesis involving transoceanic long-distance dispersal should also be considered as an explanation to the disjunction of *Alpinia* between Malesia and Australasia. In the similarity of the migration to Malesia of *Alpinia*, a long-distance dispersal was likely mediated by watercourses or birds from Malesia to Australasia (Liu *et al*. 2018). This event is an example of "out-of-Asia" migration to the Southern Hemisphere.

Origin of Indian **Alpinia**

The two *Alpinia* species of clade I (*A. fax* and *A. abundiflora*) are endemic to India. However, the molecular data supported that the two species were closely related to *Renealmia* as well as *Aframomum* (Fig. 1). Therefore, they likely shared common ancestors. Our ancestral area reconstruction analyses inferred that the two species *A. fax* and *A. abundiflora* originated in India at 18.34 Ma (95 % HPD: 08.56-29.14 Ma) (node 7, Fig. 3) after the Indian subcontinent collided with Eurasia at 43-50 Ma. However, the origin of the two species is still not clear, because the common ancestors of the two species with *Renealmia* and *Aframomum* are not clearly determined (Fig. 3). Even so, our results showed that Asian *Alpinia* is likely the ancestor to the clade including Indian *Alpinia* (*A. fax* and *A. abundiflora*), *Renealmia* (*R. alpinia* and *R. battenbergiana*) and *Aframomum* (*A. angustifolium* and *A. daniellii*) (Fig. 3). Therefore, we hypothesize that after originating in Asia, *Alpinia* migrated from Asia to India then to Africa during the early period of collision between the Indian subcontinent and Eurasia to form the common ancestor of Indian *Alpinia*, *Renealmia* and *Aframomum*. In addition, the endemic species *A. smithiae* of India was derived from Malesia the closest area to India during the mid-Miocene when the climate was optimum (warm and wet) (Zachos *et al*. 2001) (Fig. 3).

Conclusion

The study based on molecular analyses of comprehensive taxon sampling confirms that Asia has been a major center of origin for *Alpinia* and its allies since the early Paleocene and then spread to Malesia and back to Asia several times since the late Eocene, likely by watercourses and birds. The lineages distributed in Australasia are younger members that dispersed from Malesia after the late Oligocene, whereas the origin of some Indian *Alpinia* share common ancestors with *Renealmia* and *Aframomum*. Thus, likely after originating in Asia, *Alpinia* migrated from Asia to India during the early period of collision between the Indian subcontinent and Eurasia to form the common ancestor of Indian *Alpinia*, *Renealmia*, and *Aframomum*. Our study adds an example of "out-of-Asia" migration in the biogeographic history to the Southern Hemisphere.

Acknowledgements

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant No. 106.03-2019.12.

References

Atkins HJ, Bramley GLC, Johnson MA, et al. 2020. A molecular phylogeny of Southeast Asian *Cyrtandra* (Gesneriaceae) supports an emerging paradigm for Malesian plant biogeography. Frontiers of Biogeography 12(1): 0-20.

Bailey CD, Carr TG, Harris SA, Hughes CE. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, andpseudogenes. Molecular Phylogenetics and Evolution 29: 435-455.

Clayton JW, Soltis PS, Soltis DE. 2009. Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, Sapindales). Systematic Biology 58: 395-410.

Condamine FL, Nagalingum NS, Marshall CR, Morlon H. 2015. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. BMC Evolutionary Biology 15: 1-18.

Couvreur TLP, Forest F, Baker WJ. 2011. Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. BMC Biology 9: 44.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9: 772.

De Boer H, Newman M, Poulsen AD, et al. 2018. Convergent morphology in Alpinieae (Zingiberaceae): Recircumscribing *Amomum* as a monophyletic genus. Taxon 67: 6-36.

Deng JB, Drew BT, Mavrodiev EV, Gitzendanner MA, Soltis PS, Soltis DE. 2014. Phylogeny, divergence times, and historical biogeography of the angiosperm family Saxifragaceae. Molecular Phylogenetics Evolution 83: 86-98.

Deng X, Deng W, Hughes AC, Mohandass D. 2015. Diverse nectar robbers on *Alpinia roxburghii* Sweet (Zingiberaceae). Journal of Asia-Pacific Biodiversity 8: 238-241.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Moleculer Biology Evolution 29: 1969-1973.

Forest F. 2009. Calibrating the tree of life: fossils, molecules and evolutionary timescales. Annals of Botany 104: 789-794.

Fortey R. 1999. Life: an unauthorized biography: a natural history of the first four thousand million years of life on earth. London, Flamingo.

Friis E. 1988. *Spirematospermum chandlerae* sp. nov., an extinct species of Zingiberaceae from the North American Cretaceous. Tertiary Research 9: 7-12.

Hall R. 2012. Late Jurassic-Cenozoic reconstructions of the Indonesian region and the Indian Ocean. Tectonophysics 570-571: 1-41.

Hickey LJ, Peterson RK. 1978. *Zingiberopsis*, a fossil genus of the ginger family from Late Cretaceous to early Eocene sediments of western interior North America. Canada Journal of Botany 56: 1136-1152.

Higgins PJ. 1999. Handbook of Australian, New Zealand and Antarctic Birds. Melbourne, Oxford University Press.

Holloway DJ, Hall R. 1998. SE Asian geology and biogeography: an introduction. In: Biogeography and Geological Evolution of SE Asia. Leiden, Backhuys Publishers. p. 1-23.

Jablonski NG. 2004. Primate homeland: forests and the evolution of primates during the Tertiary and Quaternary in Asia. Anthropology Science 113: 117-122.

Kearse M, Moir R, Wilson A, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.

Kress WJ, Liu AZ, Newman M, Li QJ. 2005. The molecular phylogeny of *Alpinia* (Zingiberaceae): A complex and polyphyletic genus of gingers. American Journal of Botany 92: 167-178.

Kress WJ, Prince LM, Williams KJ. 2002. The phylogeny and a new classification of the gingers (Zingiberaceae): Evidence from molecular data. American Journal of Botany: 89: 1682-1696.

Kress WJ, Specht CD. 2006. The evolutionary and biogeographic origin and diversification of the tropical monocot order Zingiberales. Aliso: A Journal of Systematic and Evolutionary Botany 22: 621-632.

Larsen K, Lock JM, Maas H, Maas PJM. 1998. Zingiberaceae. In: Kubitzki K. (ed.) The families and genera of vascular plants. Vol. 4. Berlin, Germany, Springer-Verlag. p. 474-495.

Le CT. 2021. A molecular phylogeny of *Alpinia* Roxb. (Zingiberaceae): An overview for phylogeny of Vietnamese *Alpinia*. TNU Journal of Science and Technology 226(05): 104-109.

Li Y, Yi TM, Li YZ, Li CS. 2018. *Spirematospermum wetzleri* (Heer) Chandler (Zingiberaceae) from the Miocene of Weichang, Hebei Province, North China and the phytogeographic history of the genus. Journal of Palaeogeography 7: 1-8.

Liu B, Le CT, Barrett RL, et al. 2018. Historical biogeography of Loranthaceae (Santalales): Diversification agrees with emergence of tropical forests and radiation of songbird. Molecular Phylogenetics and Evolution 124: 199-212.

Lomolino MV, Riddle BR, Whittaker RJ, Brown JH. 2010. Biogeography, fourth ed. Sunderland, Massachusetts, USA, Sinauer Associates.

Lu LM, Wang W, Chen ZD, Wen J. 2013. Phylogeny of the non-monophyletic *Cayratia* Juss. Vitaceae and implications for character evolution and biogeography. Molecular Phylogenetics Evolution 68: 502-515.

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist 207: 437-453.

Matzke NJ. 2013. Probabilistic historical biogeography: new models for founder event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Frontiers of Biogeography 5: 242-248.

McElwain JC, Punyasena SW. 2007. Mass extinction events and the plant fossil record. Trends in Ecology and Evolution 22: 548-557.

Meimberg H, Wistuba A, Dittrich P, Heubl G. 2001. Molecular phylogeny of Nepenthaceae based on cladistic analysis of plastid trnK intron sequence data. Plant Biology 3: 164-175.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in: Proceedings of the gateway computing environments workshop (GCE). New Orleans, USA, Institute of Electrical and Electronics Engineers (IEEE). p: 1-8.

Morley RJ. 2000. Origin and evolution of tropical rain forests. New York, USA, John Wiley.

Morley RJ. 2003. Interplate dispersal paths for megathermal angiosperms. Perspectives in Plant Ecology, Evolution and Systematics 6: 5-20.

Nichols DJ, Johnson KR. 2008. Plants and the K-T boundary. Cambridge, UK, Cambridge University Press.

Nylander JA, Olsson U, Alstrom P, Sanmartin I. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal vicariance analysis of the thrushes (Aves: Turdus). Systematics Biology 57: 257-268.

Parra FDL, Jaramillo CA, Dilcher DL. 2007. Paleoecological changes of spore producing plants through the Cretaceous-Paleocene boundary in Colombia. Palynology 32: 258-259.

Poulsen AD, Mathisen HB, Newman MF, Ardiyani M, Lofthus Ø, Bjorå CS. 2018. *Sulettaria*: A new ginger genus disjunct from *Elettaria cardamomum*. Taxon 67: 725-738.

R Development Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria, R foundation for statistical computing.

Rambaut A. 2009. FigTree v.1.4. http://tree.bio.ed.ac.uk/software/ figtree/. 21 Dec. 2009.

Rambaut A, Drummond AJ. 2010. TreeAnnotator. Version 1.8, Distributed as Part of the BEAST Package. http://beast.community/installing/. 16 Oct. 2013.

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. http://tree.bio.ed.ac.uk/software/tracer/. 2 May 2014.

Rangsiruji A, Newman MF, Cronk QCB. 2000. A study of the infrageneric classification of *Alpinia* (Zingiberaceae) based on the ITS region of nuclear rDNA and the *trn*L-F spacer of chloroplast DNA. In: Wilson KL, Morrison DA. (eds.) Monocots-Systematics and evolution. Collingwood, Australia: CSIRO Publishing 2000. p. 728.

Razafimandimbison SG, Kellogg EA, Bremer B. 2004. Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: A Case Study from Naucleeae (Rubiaceae). Systematic Biology 53(2): 177-192.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57: 4-14.

Ronquist F, Teslenko M, Mark PVD, et al. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539-542.

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Systematic Biology 46: 195-203.

Schuettpelz E, Pryer KM. 2009. Evidence for a Cenozoic radiation of ferns in an an-giosperm-dominated canopy. Proceedings of the National Academy of Sciences USA 106: 11200-11205.

Schulte P, Alegret L, Arenillas I, et al. 2010. The chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. Science 327: 1214-1218.

Stamatakis A. 2006. RAxML-VI-HPC, maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.

Upchurch GR, Wolfe JA. 1987. Mid-Cretaceous to Early Tertiary vegetation and climate: evidence from fossil leaves and woods. In: Friis EM, Chaloner WG, Crane PH. (eds.) The origins of Angiosperms and their biological consequences. Cambridge, UK, Cambridge University Press. p. 75-105.

Vajda V, Raine JI, Hollis CJ. 2001. Indication of global deforestation at the Cretaceous-Tertiary boundary by New Zealand fern spike. Science 294: 1700-1702.

Vasconcelos TNC, Proença CEB, Ahmad B, et al. 2017. Myrteae phylogeny, calibration, biogeography and diversification patterns: increased understanding in the most species rich tribe of Myrtaceae. Molecular Phylogenetics Evolution 109: 113-137.

Wang W, Ortiz RDC, Jacques FMB, et al. 2012. Menispermaceae and the diversification of tropical rainforests near the Cretaceous-Paleogene boundary. New Phytologist 195: 470-478.

Wei R, Xiang QP, Schneider H, et al. 2015. Eurasian origin, boreotropical migration and transoceanic dis-persal in the pantropical fern genus *Diplazium* (Athyriaceae). Journal of Biogeography 42: 1809-1819.

West CK, Greenwood DR, Basinger JF. 2015. Was the Arctic Eocene ‘rainforest’ monsoonal? Estimates of seasonal precipitation from early Eocene megafloras from Ellesmere Island Nunavut. Earth Planetary Science Letters 427: 18-30.

Wilf P, Johnson KR. 2004. Land plant extinction at the end of the Cretaceous: a quantitative analysis of the North Dakota megafloral record. Paleobiology 30: 347-368.

Wilf P, Labandeira CC, Kress WJ, et al. 2000. Timing the radiations of leaf-beetles: Hispines on gingers from Latest Cretaceous to Recent. Science 289: 291-294.

Wing SL, Boucher LD. 1998. Ecological aspects of the Cretaceous flowering plant radiation. Annual Review of Earth and Planetary Sciences 26: 379-421.

Wolfe JA, Upchurch GR. 1986. Vegetation, climatic and floral changes at the Cretaceous-Tertiary boundary. Nature 324: 148-152.

Wotton DM, Clout MN, Kelly D. 2008. Seed retention times in the New Zealand pigeon (Hemiphaga novaezeelandiae novaeseelandiae). New Zealand Journal of Ecology 32: 1-6.

Wu TL, Larsen K. 2000. Zingiberaceae. In: Wu ZY, Raven PH, Hong DY. (eds.) Flora of China. Vol. 24. Beijing and St. Louis, Science Press and Missouri Botanical Garden Press. p. 322-377.

Yu Y, Harris AJ, He X. 2011. RASP (Reconstruct Ancestral State in Phylogenies) 2.0 Beta. http://mnh.scu.edu.cn/soft/blog/RASP. 19 Mar. 2012.

Zachos JC, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292: 686-693.

Zahirovic S, Seton M, Müller RD. 2014. The Cretaceous and Cenozoic tectonic evolution of Southeast Asia. Solid Earth 5: 227-273.