Biogeographic patterns and diversification dynamics of the genus Cardiodactylus Saussure (Orthoptera, Grylloidea, Eneopterinae) in Southeast Asia

Jiajia Dong, Gael Kergoat, Natállia Vicente, Cahyo Rahmadi, Shengquan Xu, Tony Robillard

Title

Biogeographic patterns and diversification dynamics of the genus *Cardiodactylus* Saussure (Orthoptera, Grylloidea, Eneopterinae) in Southeast Asia

Authors

Jiajia Dong[a,b,*], Gael J. Kergoat[c], Natállia Vicente[d], Cahyo Rahmadi[e], Shengquan Xu[b], Tony Robillard[a]

Authors' affiliations

[a] *Institut de Systématique, Evolution et Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, 57 rue Cuvier, CP 50, 75231 Paris Cedex 05, France*

[b] *College of Life Science, Shaanxi Normal University, 710119, Xi'an, Shaanxi, P.R. China*

[c] *CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France*

[d] *Universidade Federal de Viçosa, Programa de pós-graduação em Ecologia. Av. PH Rolfs s/n. Viçosa, Minas Gerais, CEP 36570-900, Brazil*

[e] *Division of Zoology, RC–Biology, Cibinong Science Center – Indonesian Institute of Sciences (LIPI), Cibinong Science Center, Jl. Raya Jakarta- Bogor Km. 46, Cibinong, 16911 Indonesia*

*****Corresponding author: jia_jia_dong@hotmail.com

Running title: Biogeography of *Cardiodactylus* crickets

Abstract

Southeast Asia harbors an extraordinary species richness and endemism. While only covering 4% of the Earth's landmass, this region includes four of the planet's 34 biodiversity hotspots. Its complex geological history generated a megadiverse and highly endemic biota, attracting a lot of attention, especially in the field of island biogeography. Here we used the cricket genus *Cardiodactylus* as a model system to study biogeographic patterns in Southeast Asia. We carried out molecular analyses to: (1) infer phylogenetic relationships based on five mitochondrial and four nuclear markers, (2) estimate divergence times and infer biogeographical ancestral areas, (3) depict colonization routes, and summarize emigration and immigration events, as well as *in situ* diversification, and (4) determine whether shifts in species diversification occurred during the course of *Cardiodactylus* evolution. Our results support the monophyly of the genus and of one of its species groups. Dating and biogeographical analyses suggest that *Cardiodactylus* originated in the Southwest Pacific during the Middle Eocene. Our reconstructions indicate that Southeast Asia was independently colonized twice during the Early Miocene (*ca.* 19-16 Million years ago), and once during the Middle Miocene (*ca.* 13 Million years ago), with New Guinea acting as a corridor allowing westward dispersal through four different passageways: Sulawesi, the Philippines, Java and the Lesser Sunda Islands. Sulawesi also served as a diversification hub for *Cardiodactylus* through a combination of high immigration and *in situ* diversification events, which can be accounted for by the complex geological history of the Wallacea region.

Keywords

Biogeography; climate change; crickets; diversification analyses; geological changes; molecular dating

1. Introduction

Southeast Asia (SEA) is considered as one of the most geologically dynamic regions of the planet (Lohman et al., 2011). It aggregates four major biodiversity hotspots (de Bruyn et al., 2014; Mittermeier et al., 2004; Zachos and Habel, 2011), which are divided by sharp, yet porous, biogeographic boundaries (Hearty et al., 2007; Morley et al., 2012; Wallace, 1860). The complex geological history of SEA generated a megadiverse and highly endemic biota that has attracted a lot of attention, especially in the field of island biogeography (Condamine et al., 2013, 2015; Hall, 2009a; Lohman et al., 2011; Metcalfe, 2006; Woodruff, 2010). In SEA, dynamic geological and climatic histories acted together to generate the world's largest island complex, in which the spatial distribution of terrestrial habitats has been altered extensively through time (Bird et al., 2005; Hall, 1996, 2009a, 2011; Hall and Sevastjanova, 2012; Heaney, 1991; Metcalfe, 2006; Voris, 2000; Zahirovic et al., 2014). Volcanic uplifts and repeated sea-level fluctuations also promoted species diversification by providing countless opportunities for allopatric speciation (Heaney, 2000; Steppan et al., 2003; Jansa et al., 2006; Outlaw and Voelker, 2008). The interplay between the movement of tectonic plates, oscillations in the Earth's orbit, and the variable configuration of landmasses and ocean currents also affected the climate of SEA (Morley, 2012). As a result, the vegetation and the distribution of habitats of SEA were profoundly modified through time (Heaney, 1991). Tropical rainforest in SEA greatly expanded during the Paleogene Period (66.0-23.0 Million years ago (Ma)) and the beginning of the Neogene Period, due to warmer temperatures and more humid climates. Tropical rainforests reached a peak during the mid-Miocene Climatic Optimum (MCO; 17-15 Ma), when their distribution extended further north to southern China and Japan (Heaney, 1991; Morley, 2012; Zachos et al., 2001). It is not until the Late Miocene and Early Pliocene that tropical rainforests started to decrease in range as a result of global cooling (Hall, 2012; Zachos et al., 2001). Lower sea levels, cooler temperatures, and modified rainfall patterns related with the development of continental glaciers during the Pleistocene (2.6-0.01 Ma) also had a great impact on SEA vegetation setting (Heaney, 1991). Studies of fossil pollen records from the Last Glacial Maximum (20,000 years ago)

indicate that SEA vegetation differed substantially from that of today, being characterized by an increase in the extent of montane vegetation and savannah, and by the presence of a continuous belt of evergreen rainforests across Sundaland (Bird et al., 2005). During the Last Glacial Period (110,000-11,700 years ago), short warmer interglacial periods also resulted in raises of the global sea level, which led to the fragmentation of some parts of the Sundaland into islands and peninsulas; for instance, the Thai-Malay Peninsula, Borneo, Sumatra and Java were separated from each other every time the sea level rose.

The complex geological and climate history and the fragmented island habitats have raised many questions on the origins of the current biota of SEA. In a comprehensive study, de Bruyn et al. (2014) reviewed previous biogeographical works by conducting meta-analyses of geological, climatic and biological data. Based on inferences of colonization routes and ancestral areas derived from meta-analyses of phylogenetic data, they inferred that Borneo and Indochina acted as major evolutionary hotspots during the Cenozoic. One of the reasons invoked is the length of their emergent histories (more than 80 million years), which surpasses by far those of other regions in SEA. De Bruyn et al. (2014) also found out that within-area diversification (*in situ* diversification) and subsequent emigration played a major role in Borneo and Indochina, especially during the course of the Neogene and Quaternary sea-level fluctuations.

To investigate biogeographic and species diversification patterns in SEA, we use a species-rich clade of crickets from the subfamily Eneopterinae, the genus *Cardiodactylus* Saussure, as a model system. Eneopterinae crickets are mostly distributed in tropical areas (Vicente et al., 2017) and are characterized by a diversity of traits in relation with acoustic communication (*e.g.*, Robillard et al., 2007, 2013; ter Hofstede et al., 2015). With 82 known species, *Cardiodactylus* is currently the most speciose eneopterine genus. It is also the most widespread one, as its distribution runs from Japan (northern Ryukyu Islands), to the coasts of Southeast Asia, Northern Australia and the archipelagos of the Western Pacific across the Caroline Islands, Fiji, Indonesia, New Caledonia, Palau, Papua New Guinea, the Philippines, the Samoan Islands, the Solomon Islands, the Mariana Islands, Tonga, Vanuatu and Wallis

and Futuna (Otte, 2007; Robillard and Ichikawa, 2009; Robillard et al., 2014). The genus is subdivided in two species groups, *Efordi* (22 known species) and *Novaeguineae* (60 known species), which show contrasted patterns of geographical distributions. Only one species of the *Efordi* group is distributed in SEA; the other species are distributed in New Guinea and in archipelagoes of the Southwest Pacific. In contrast, the *Novaeguineae* species group has a wider distribution, and is well diversified in SEA (35 species are found in SEA: Dong and Robillard, 2016; Robillard et al., 2014). Interestingly, the distribution map of the *Novaeguineae* group indicates that species richness is inconsistent with landmass sizes (see details in Figure 1). For example, only one species occurs in the northern part of Indochina (*Cardiodactylus guttulus,* distributed from Japan to Northern Vietnam), while four species are endemic to Sulawesi. Some species are not endemic to a single island; for example, *C. empagatao* is found in the Philippines and Sulawesi, and *C. borneoe* is widely distributed in Borneo and in the Thai-Malay Peninsula. Disentangling the origin of such diverse distribution patterns thus calls for the need of a thorough biogeographical study of the genus.

According to the study of Vicente et al. (2017) on Eneopterinae the genus *Cardiodactylus* did not originate in SEA but in the Southwest Pacific; however, this inference was potentially biased by the settings of the analysis in terms of time slices, and by the fact that it relied on a sparse sampling for the genus (only six *Cardiodactylus* species were sampled). Here, we use a more comprehensive sampling (41 *Cardiodactylus* species) for the genus and more appropriate time slices to infer the biogeographic patterns and diversification dynamics of the genus using a time-calibrated species tree. This allows us to assess with more precision the dynamics of diversification of the genus, and to determine whether Borneo and Indochina played a major role in its diversification.

2. Materials and methods

2.1. Taxon sampling

The taxon sampling of *Cardiodactylus* is focused on species distributed in SEA, but also includes representatives of the whole distribution and taxonomic diversity of the genus. In

this study, we sampled 113 individuals representing 41 species of *Cardiodactylus* crickets, including 28 species from SEA and 13 species from New Guinea and Southwest Pacific archipelagoes (New Caledonia, Solomon Islands and Vanuatu). For outgroup selection, we referred to a previous study (Anso et al., 2016), and selected 22 species representing all five tribes of the Eneopterinae subfamily, as well as two more distant species belonging to the subfamily Gryllinae. Detailed information on specimens is presented in Appendix S1.

Most specimens were collected during recent fieldworks and through collaborative works in several countries in SEA. In addition, some specimens used in the current study were obtained from the following institutions: the 'Muséum d'Histoire Naturelle de Genève' (MHNG), the 'Muséum National d'Histoire Naturelle, Paris' (MNHN), the 'Museum Zoologicum Bogoriese, Bogor' (MZB), the 'Nationaal Natuurhistorisch Museum, Leiden' (RMNH), the 'Natural History Museum, London' (BMNH), the 'Naturhistorisches Museum, Vienna' (NHMW), the 'Royal Belgian Institute of Natural sciences, Brussels' (RBINS), and the 'Zoological Institute, Russian Academy of Sciences, Saint Petersburg' (ZIN).

2.2. DNA extraction, amplification and sequencing

The molecular work was performed at the 'Service de Systématique Moléculaire' of the MNHN. Whole genomic DNA was extracted from the median/hind femora of dried or alcohol-preserved specimens using the NucleoSpin® 96 Tissue Kit (Macherey-Nagel, Germany) and an automatic pipetting robot Eppendorf epMotion® 5075 TMX. Five mitochondrial markers and four nuclear markers used in previous phylogenetic studies on crickets (Anso et al., 2016; Chintauan-Marquier et al., 2016; Nattier et al., 2011, 2012; Robillard and Desutter-Grandcolas, 2006) were amplified and sequenced. Mitochondrial markers were partial fragments of two non-protein coding(12S ribosomal RNA (12S), 16S ribosomal RNA (16S)), and three protein-coding, (cytochrome c oxidase subunit I (COI), cytochrome c oxidase subunit II (COII) and cytochrome b (Cytb)) genes. Nuclear markers included partial fragments of elongation factor 1 alpha (EF1a), histone H3 (H3), 18S ribosomal subunit (18S) and 28S ribosomal subunit (28S). The amplified fragment of EF1a consisted of both protein-

coding and non-protein coding regions whereas the amplified fragment of H3 only consisted of protein-coding regions. Whenever possible, we obtained molecular data from the same previous voucher specimens. We used the primers and settings detailed in Anso et al. (2016), Nattier et al. (2011) and Chintauan-Marquier et al. (2016). All sequences obtained in this study were cleaned, checked for sequencing errors and pseudogenes (through BLAST searches, DNA sequences and amino acid translation alignments and inference of gene trees), and deposited in GenBank (see details in Appendix S1). The COI sequences for all sequenced specimens of *Cardiodactylus borneoe* were discarded because of suspicion of pseudogenes. The complete combined dataset consisted of 4,111 aligned base pairs (bp) for 145 terminals: 528 bp for 16S, 419 bp for 12S, 706 bp for COI, 340 bp for COII, 345 bp for Cytb, 653 bp for 18S, 375 bp for 28S, 327 bp for H3 and 405 bp for EF1a.

2.3. Phylogenetic analyses

To check for possible contaminations and artifacts, preliminary phylogenetic analyses were carried out for each marker using the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/; Trifinopoulos et al., 2016) (see the resulting gene trees in Appendix S5). Then the aligned sequences of all nine markers were concatenated in Geneious R8.1.4 (Biomatters Ltd., New Zealand, www. geneious.com).

The concatenated dataset was then analyzed using Bayesian inference (BI) and maximum likelihood (ML). For both BI and ML we used PartitionFinder V2.1.1 (Lanfear et al., 2017) to determine best-fit partitioning schemes and the associated substitution models. One partition was specified for each of the non-protein coding genes (12S, 16S, 18S and 28S) and one partition per codon for the protein-coding genes (COI, COII, Cytb and H3), with the exception of EF1a which included several protein-coding and non-protein coding regions. For the latter we only used one partition to avoid specifying too many small partitions for the protein-coding and non-protein coding regions. PartitionFinder analyses were further carried out using the default '*greedy'* algorithm option and either the '*mrbayes'* or '*raxml'* set of models (for BI and ML analyses, respectively); we also used the '*linked branch lengths'*

option, in order to limit the risk of overparameterization by lowering the number of inferred partitions. The Bayesian information criterion (BIC) was preferentially used to compare partitioning schemes and substitution models following Ripplinger and Sullivan (2008).

Bayesian inference analyses were performed with MrBayes 3.2.6 (Ronquist et al., 2012) whereas ML analyses were performed with RAxML 8.2.8 (Stamatakis, 2014) and IQ-TREE 1.6.2 (Nguyen et al., 2015). All corresponding analyses were performed using the CIPRES Science Gateway 3.3 (Miller et al., 2015).

For ML analyses best-scoring trees were obtained using heuristic searches relying on 100 random-addition replicates. Clade support was assessed using non-parametric bootstrap (for both RAxML and IQ-TREE analyses); for each analysis 1,000 bootstrap replicates were conducted. Nodes supported by bootstrap support values (BS) $\geq$ 70% were considered strongly supported following Hillis and Bull (1993).

For BI analyses we conducted two independent runs with eight Markov chain Monte Carlo (MCMC): one cold and seven incrementally heated that ran for 50 million generations with trees sampled every 1,000 generations. We used a conservative burn-in of 12.5 million generations per run after checking for stability on the log-likelihood curves and the split-frequencies of the runs in Tracer v.1.7 (Rambaut et al., 2014). Support of nodes for MrBayes analyses was provided by clade posterior probabilities (PP) as directly estimated from the majority-rule consensus topology. A clade with a PP value higher than 0.95 was considered as well supported following Erixon et al. (2003). Whenever a *Cardiodactylus* species (for which multiple specimens were sequenced) or a specific taxonomic group of interest was recovered paraphyletic, we used Bayes factors (B_F) to assess whether there was statistical support for their non-monophyly. To do so, specific analyses (in which taxa of interest are constrained to be monophyletic) were carried out using MrBayes.

2.4. Divergence time estimation

We generated a species-level dataset (with one specimen per sampled *Cardiodactylus* species; see details in Appendix S1) to properly conduct dating and all subsequent analyses

(*i.e.* historical biogeography and diversification analyses). Divergence times were estimated using Bayesian relaxed clocks as implemented in BEAST 1.8.4 (Drummond et al., 2012). The partitions/clocks and substitution models were selected under PartitionFinder 2.1.1 following the settings presented above but with the '*beast'* set of models (see details in Table 1). We were not able to fully implement a two steps strategy recommended by several authors (*e.g.,* see Foster and Ho, 2017). The first step corresponds to a PartitionFinder analysis with the partition-rich '*linked branch lengths*' option, to be followed by the use of the 'ClockstaR' R package (Duchêne et al., 2014), which requires having no missing gene fragments in the dataset; the latter step is expected to reduce the number of clocks in molecular dating procedures (hence limiting the risk of overparameterization). For comparison purpose, we nonetheless decided to conduct BEAST analyses with both options (number of partitions/clocks inferred with either the '*linked branch lengths*' or the '*unlinked branch lengths*' option). BEAST analyses were performed on the CIPRES Science Gateway using BEAGLE to improve and speed up the likelihood calculation (Ayres et al., 2012; Miller et al., 2015). For each clock model/partitioning scheme an uncorrelated lognormal relaxed clock was implemented. The *Tree Model* was set to a birth-death speciation process (Gernhard, 2008) to better account for extinct and missing lineages. The *ucld.mean* prior of each clock model was set to an uninformative interval (0.0001-1.0) with a uniform prior distribution.

Based on the available worldwide biogeographical framework of Eneopterinae crickets (Vicente et al., 2017), three secondary calibration points were enforced using normal distributions centered on previously estimated median ages: (1) the node for the most recent common ancestor (MRCA) of *Cardiodactylus* was assigned an interval of 33.60-55.19 Ma (*normalPrior mean*="42.6" *stdev*="6.0"), (2) the node for the MRCA of the *Novaeguineae* species group was assigned an interval of 23.32-39.73 Ma (*normalPrior mean*="30.0" *stdev*="6.0"), and (3) the node for the MRCA of *C. guttulus*, *C. oeroe*, *C. tankara* and *C. singapura* was assigned an interval of 17.83-31.17 Ma (*normalPrior mean*="23.4" *stdev*="6.0").

BEAST analyses consisted of 50 million generations of MCMC with the parameters and trees sampled every 1,000 generations. A burn-in of 25% was applied after checking the log-likelihood curves. Trees obtained from distinct analyses were combined using LogCombiner v1.8.4 (http://beast.bio.ed.ac.uk/). The maximum credibility tree, median ages and their 95% highest posterior density (HPD) were generated with TreeAnnotator v1.8.4 (https://github.com/beast-dev/beast-mcmc/releases/tag/v1.8.4).

2.5. Ancestral area estimation

The package BioGeoBEARS (Matzke, 2014) implemented in R was used to infer the biogeographical history of *Cardiodactylus* crickets across SEA. The dispersal-extinction-cladogenesis (DEC; Ree and Smith, 2008), dispersal-vicariance analysis (DIVA; Ronquist, 1997; herein DIVALIKE) and Bayesian Analysis of Biogeography (BayArea, Landis et al., 2013; herein BAYAREALIKE) models were used to estimate ancestral areas. Because of concerns with its statistical validity (Ree and Sanmartín, 2018) we did not use the +J model of Matzke (2014) in our analyses. The analyses were conducted with the maximum clade credibility (MCC) tree inferred under BEAST. The geographical distribution of each species was estimated based on fieldwork data, extant literature, and information provided by museum specimens (see details in Appendix S1). Twelve operational areas were defined (see details in Appendix S2) based on the distribution ranges of extant *Cardiodactylus* crickets (Robillard et al., 2014): Indochina (A), Thai-Malay Peninsula (B), Sumatra (C), the Philippines (minus Palawan) (D), Palawan (E), Borneo (F), Java (G), Sulawesi (H), Lesser Sunda Islands (I), Moluccas (J), New Guinea (K) and the archipelagoes in the Southwest Pacific (L).

In previous studies (de Bruyn et al., 2014; Hall, 2011; Morley, 2012), it has been shown that three major Cenozoic collision events had a great influence on the current archipelago setting: (1) the India-Asia collision at *ca*. 45 Ma facilitated the isolation of West Sulawesi from Sundaland, (2) the Australia-Asia collision at *ca.* 25 Ma generated the Bird's Head region of New Guinea, and (3) the collision of the Sula Spur promontory with the

Southeast Asian margin in Sulawesi *ca.* 15 Ma formed the region now known as Wallacea (Hall, 2009a). To account for these major geological events, three distinct time slices were enforced in our time-stratified biogeographical model; the first time slice runs from 45.0 to 25.0 Ma, the second one from 25.0 to 15.0 Ma and the last one from 15.0 Ma to the present (Hall, 2013; Hall and Sevastjanova, 2012; see more details in Appendix S2). For each time slice, we constructed a basal matrix of scaling factors (multipliers) (0, 0.01, 0.1, 0.25, 0.50, 0.75, 1.0) for the dispersal rates between areas according to their geographical position and changes in sea level. Alternative and more simplified sets of multipliers were also tested (0, 0.01, 0.1, 0.25, 1.0; 0, 0.01, 0.1, 0.50, 1.0; and 0, 0.01, 0.1, 0.75, 1.0). Analyses were repeated with two, three or four maximum ancestral areas (parameter *max_range_size* in BioGeoBEARS). Therefore, a total of 36 processes (four distinct multipliers sets, three speciation models (DEC, DIVALIKE and BAYAREALIKE), three distinct maximum range sizes), were implemented to estimate ancestral areas (the details of these processes are listed in Appendix S2 and S3).

Finally, to ensure that the time framework imposed by the secondary calibration does not constrain the biogeographical results through strong constraint on allowed areas and dispersal multipliers, we implemented nine additional processes without time slices (unstratified analyses) using the basal set of multipliers with three speciation models and three distinct maximum ancestral areas (see Appendix S3).

2.6. Colonization events and in situ *diversification*

As mentioned above, only one member of the *Efordi* species group is distributed in SEA; in contrast, the *Novaeguineae* species group is more widely distributed, including in SEA, New Guinea and the Southwest Pacific. Therefore, we study colonization events and *in situ* diversification events of *Cardiodactylus* in SEA by focusing on the *Novaeguineae* species group. Within this clade, major colonization routes were inferred for each time slice. Emigration and immigration events for each area were analyzed by summarizing colonization

events using SPPS Statistics 22 (SPSS Inc.). We also followed de Bruyn et al. (2014) to estimate *in situ* diversification events using SPPS Statistics v.22 (SPSS Inc.).

2.7. Diversification analyses

Bayesian Analysis of Macroevolutionary Mixtures 2.5 (BAMM; Rabosky et al., 2013) was used: (1) to estimate rates of speciation (λ), extinction (μ), and net diversification (γ) either for all *Cardiodactylus* species or only for the SEA species belonging to the *Novaeguineae* species group (which constitutes a monophyletic group), (2) to conduct rate-through-time analysis of these rates, and (3) to identify and visualize shifts in species rates across the *Cardiodactylus* phylogeny. BAMM accounts for non-random and incomplete taxon sampling in the phylogenetic trees by allowing all non-sampled species to be associated with a particular tip or more inclusive clade. Species numbers were obtained from the 'Orthoptera species file online' website (Cigliano et al., 2017) and published sources.

Priors for BAMM were generated using the R package BAMMtools v.2.5.0 (Rabosky et al., 2014a,b) by providing the BEAST maximum clade credibility tree and the total species numbers across *Cardiodactylus* (see Appendix S4). We used a gradient of prior values ranging from 0.1 to 1.0 to test the sensitivity to the prior, to account for the recent debate on the reliability of BAMM estimates (Moore et al., 2016, but see Rabosky et al., 2017). We ran BAMM by setting four independent MCMC running for 50 million generations and sampled every 50,000 generations; convergence was assessed by computing ESS of log likelihoods. After removing 10% of trees as burn-in, the BAMM output were analyzed with employing BAMMtools package in R and the 95% credible rate shift configurations was estimated using Bayes factors. The best shift configuration with the highest maximum a posteriori probability was estimated in this analysis. Rates-through-time plots were generated for speciation (λ), extinction (μ) and net diversification (γ), either for all *Cardiodactylus* species or only for the SEA species. All visualization was performed using R and C code available through the R package BAMMtools.

3. Results

3.1. Data collection and phylogenetic analyses

The best-fit partition schemes and substitution models used in BI and ML analyses of the combined dataset are showed in Table 1. Both BI and ML phylogenetic analyses yielded robust and largely congruent topologies (Figure 2; see also Appendix S6 for original outputs of both BI and ML analyses); it is especially the case for the results of RAxML analyses which are presented in Figure 2 (90% of nodes within *Cardiodactylus* supported by BV $\geq$ 70%). In all analyses the subfamily Eneopterinae and the genus *Cardiodactylus* are recovered as monophyletic with a high support (BS_{RAxML} of 100%, PP of 1.0 and $BS_{IQ\text{-}TREE}$ of 100% and 76%, for Eneopterinae and *Cardiodactylus*, respectively). All *Cardiodactylus* species are also recovered as monophyletic, with the exception of *C. floresiensis*. However, additional analyses carried out under BI do not support the paraphyly of *C. floresiensis*: MCMC runs where *C. floresiensis* representatives are constrained to be monophyletic yield a harmonic mean estimate of -50151.29 versus -50150.18 for the unconstrained MCMC runs; hence the difference corresponds to a non-statistically significant B_F of 2.22 ($B_F < 10$; see Kass and Raftery, 1995).

Within *Cardiodactylus*, the monophyly of the *Novaeguineae* species group is strongly supported in both analyses ($BS_{RAxML/IQ\text{-}TREE}$ of 100%, PP of 1.0). By contrast, the *Efordi* species group is recovered as paraphyletic in all analyses due to the placement of *C. javarere* and *C. enkraussi*. *Cardiodactylus javarere* is found as sister to the *Novaeguineae* species group, forming a well-supported clade (BS_{RAxML} of 97%, $BS_{IQ\text{-}TREE}$ of 78%, PP of 1.0). Both RAxML and MrBayes analyses recover *Cardiodactylus enkraussi* as sister to the clade (*C. javarere* + *Novaeguineae* species group), but its position is less robust (BS_{RAxML} of 85%, PP of 0.73); with IQ-TREE *C. enkraussi* is grouped with *C. niugini*, this clade being the sister group of *C. javarere* + *Novaeguineae* species group, with a very weak support ($BS_{IQ\text{-}TREE}$ of 34% and 24% for the two corresponding nodes). Both RAxML and MrBayes analyses indicate that the remaining species of the *Efordi* species group constitute a well-supported clade (BS_{RAxML} of 78% and PP of 0.96) whereas IQ-TREE analyses infer two distinct

lineages, due to the placement of *C. enkraussi* (see above). Additional analyses carried out under BI support the paraphyly of the *Efordi* species group: MCMC runs where members of the *Efordi* species group are constrained to be monophyletic yield a harmonic mean estimate of -50162.07 versus -50150.18 for the unconstrained MCMC runs. Hence the difference corresponds to a statistically significant B_F of 21.68 (B_F > 10; see Kass and Raftery, 1995).

Within the *Novaeguineae* species group, *C. haddocki* and *C. novaeguineae* form a robust clade ($BS_{RAxML/IQ\text{-}TREE}$ of 100%, PP of 1.0), sister to another well-supported clade (BS_{RAxML} of 99%, $BS_{IQ\text{-}TREE}$ of 66%, PP of 1.0) comprising all the remaining species of the group. Within this clade, three species are the first lineages to branch off, sister to the remaining species of the group; these three species correspond to two species distributed in New Guinea (*C. lucus* and *C. maai*) and one in the Western Pacific (*C. tankara*). The remaining members of the species group form a generally well-supported clade (BS_{RAxML} of 76%, PP of 0.97; but $BS_{IQ\text{-}TREE}$ of only 18%), which splits into two main clades (hereby referred as clade 1 and clade 2 on Figure 2). Clade 1 is generally well supported (BS of 84%, PP of 1.0; but $BS_{IQ\text{-}TREE}$ of only 40%) and consists of lineages that are also well supported (see Figure 2 for details). Clade 2 is only moderately supported, with low BS (BS_{RAxML} of 40%, $BS_{IQ\text{-}TREE}$ of 19%) and a moderate support under BI (PP of 0.91). Within clade 2, two main clades (clade 3 and clade 4) can be distinguished. All species belonging to the clade 4 are distributed in SEA, whereas only two species (*C. manus* and *C. quatei*) are not distributed in SEA in clade 3.

3.2. Divergence time estimation

BEAST analyses relying on the 11 partitions/clocks selected through the use of the '*linked branch lengths*' option did not converge, as underlined by low ESS values (below 200) for 53 parameters. On the contrary BEAST analyses relying on the two partitions/clocks selected through the use of the '*unlinked branch lengths*' option converged quickly, with all parameters showing ESS values $\geq$ 200. As a result, we only present and discuss the results of the latter analyses in our study. The resulting median ages are presented in Figure 3 and

the original output results (including those of the analyses with low ESS values) in Appendix S7. Age estimates for the genus *Cardiodactylus* suggest an origin in the Middle Eocene *ca*. 43.2 Ma (95% HPD: 33.47-52.93 Ma). The members of the *Efordi* species group form a monophyletic group, except *C. javarere*, which is recovered as sister to the *Novaeguineae* species group. Members of the *Efordi* species group apparently started their diversification in the Late Eocene about 41.0 Ma (95% HPD: 29.22-48.83 Ma). The lineage leading to *C. javarere* then diverged from the *Novaeguineae* species group *ca*. 39.1 Ma, while members of the *Novaeguineae* species group started diversifying during the Late Oligocene about 27.0 Ma (95% HPD: 20.39-33.67 Ma). The divergence time for the MRCA of *C. haddocki* and *C. novaeguineae*, which are sister to the remaining members of the *Novaeguineae* species group, is inferred at *ca*. 8.4 Ma (95% HPD: 5.18-11.84 Ma), at the end of the Miocene. Most speciation events within the *Novaeguineae* species group happened during the Miocene (*ca.* 23.0-5.3 Ma). Only a few lineages appear to have diverged recently (less than 4 Ma).

3.3. Ancestral area estimation

DIVALIKE model shows higher statistical support over the DEC and BAYAREA models for all tested dispersal multiplier matrices (see more details in Appendix S3). Furthermore, the matrix of scaling factors involving five distinct rates (0, 0.01, 0.25, 0.5, 0.75, 1.0) implemented with a DIVALIKE model and a maximum range size of two areas was supported as the best-fit model over more simplified sets of multipliers and more allowed areas; the corresponding ancestral areas estimation is presented in Figure 3. Analyses without time slices (unstratified analyses) result in the same biogeographical pattern (see more details in Appendix S3) in SEA *Novaeguineae* species group, but it inferred an origin in New Guinea for the genus *Cardiodactylus* and for the *Efordi* species group. However, according to Hall's work (2009, 2013), New Guinea only began to emerge about 25 Ma; therefore, a New Guinea origin of *Cardiodactylus ca.* 43 Ma is not compatible with geological evidences. The latter illustrates the importance of using stratified analyses to account for geological changes through time; hence for our study we only focus on the results inferred with the stratified

417 analysis relying on the inferred best-fit model (DIVALIKE model with five distinct rates and a
418 maximum range size of two areas).

419 Ancestral areas estimation suggests an origin of *Cardiodactylus* in the Southwest
420 Pacific during the Middle Eocene, followed by a complex and dynamic biogeographical
421 history. Lineages belonging to the *Efordi* species group diverged earlier (*ca.* 40 Ma) than the
422 *Novaeguineae* species group (*ca.* 27 Ma) and colonized New Guinea several times, twice
423 independently, after 37.1 Myrs (*C. niugini*) and 39.1 Myrs (*C. javarere*), and twice
424 independently after 22.5 Myrs (*C. busu* and *C. nobilis*).

425 The members of the *Novaeguineae* species group colonized New Guinea four times
426 from the Southwest Pacific: twice at the end of the Oligocene (once after 24.0 Myrs (*C. lucus*)
427 and once *ca.* 22.1 Ma (main clade in SEA), and twice more recently, *ca*. 8.4 Ma for the clade
428 made of (*C. haddocki* + *C. novaeguineae*). From New Guinea, *Cardiodactylus* independently
429 colonized SEA three times: the first one linked New Guinea to SEA through the Philippines
430 *ca.* 19.5 Ma; the second one through Sulawesi *ca.* 16.4 Ma; a third colonization of SEA from
431 New Guinea occurred more recently, during the Middle Miocene, through Java and the
432 Lesser Sunda Islands *ca.* 13 Ma.

433 Through the first passageway, *Cardiodactylus* colonized Palawan after 19 Myrs
434 through the Philippines, then the Thai-Malay Peninsula and Borneo. *Cardiodactylus*
435 colonized Indochina twice recently, once from the Thai-Malay Peninsula, *ca.* 4 Ma (*C.*
436 *thailandia*) and once from the Philippines *ca.* 5 Ma (*C. guttulus*). Meanwhile a lineage from
437 the Thai-Malay Peninsula dispersed to Borneo after 3.4 Myrs.

438 From the Philippines, *Cardiodactylus* also colonized the Moluccas (*C. halmahera*)
439 after 17 Myrs, and recently dispersed to Sulawesi (*C. empagatao*). After colonizing Borneo
440 *ca.* 15 Ma, the genus colonized Sumatra *ca.* 12 Ma and reached Java through Sumatra *ca.* 8
441 Ma.

442

443 *3.4. Colonization routes, emigration / immigration events and* in situ *diversification*

444 Colonization routes of the *Novaeguineae* species group are presented in Figure 4a. Before
445 the Miocene, as deduced from our biogeographical analyses, no colonization event
446 happened. During the Miocene, 10 distinct colonization routes are inferred; interestingly New
447 Guinea is recovered as a major biogeographical crossroad (four different routes are inferred,
448 to Java, Sulawesi, the Lesser Sunda Island and the Philippines), followed by the Philippines
449 (three routes), the Southwest Pacific (two routes), and Borneo and Sumatra (one route each).
450 Along these routes, colonization events from the Southwest Pacific to New Guinea happened
451 twice. During the Plio-Pleistocene eight colonization routes are identified: three depart from
452 the Philippines; two depart from New Guinea, and two from the Thai-Malay; and one goes
453 from the Southwest Pacific to New Guinea.
454 Estimated emigration and immigration events through time are presented in Figure 4b.
455 When summing the total numbers of events for each area, six areas only had immigration
456 events inferred (Indochina, Java, the Lesser Sunda Islands, Palawan, Sulawesi and the
457 Moluccas). The total number of inferred emigration events is higher than that of immigration
458 events for New Guinea, the Philippines, the Southwest Pacific and the Thai-Malay Peninsula,
459 while the opposite (more immigration than emigration events) is inferred only for Borneo.
460 Moreover, the total number of emigration and immigration events is at equilibrium only for
461 Sumatra.
462 *In situ* diversification events through time are presented in Figure 4c. When summing
463 the total numbers of events for each area, three areas (New Guinea, Sulawesi and the
464 Philippines) are associated with the highest number of *in situ* diversification events (four
465 events), followed by the Lesser Sunda Islands and the Thai-Malay Peninsula (three events),
466 Sumatra and the Southwest Pacific (two events) and Borneo (one event).
467
468 *3.5. Diversification analyses*
469 A *poissonRatePrior* of 0.5 was recovered as the best-fit prior by BAMMtools (see more
470 details in Appendix S10). Convergence of the MCMC chains in the BAMM analyses was
471 observed after discarding the burn-in period (ESS > 900 for both the number of shifts and log

likelihoods). The 95% credible set of rate shift configurations sampled with BAMM provides more support (probability of 0.36) for a scenario without shifts in diversification rates within *Cardiodactylus* (Figure 5); less supported scenarios (probabilities of 0.16, 0.17 and 0.17) inferred a single shift in diversification rates near the base of the *Novaeguinae* species group (the placement of the shift differs, as illustrated in Figure 5). BAMM analyses strongly support a diversity-dependent speciation process across *Cardiodactylus* with the following rates: (1) net diversification rate of 0.06 species/Myr (95% quantile=0.023-0.071), (2) speciation rate of 0.12 species/Myr (95% quantile=0.075-0.202), and (3) extinction rate of 0.06 species/Myr (95% quantile=0.003-0.170). Rate-through-time plots for the three corresponding rates are presented in Figure 6, for the whole genus and within the *Novaeguineae* species group. The resulting plots indicate that *Cardiodactylus* speciation and net diversification rates have decreased through time, with the exception of an increase in speciation and net diversification rates occurring from 20.0 to 15.0 Ma. By contrast, the inferred extinction rates are quite constant and relatively low. A similar trend was recovered in the *Novaeguineae* species group. For the whole genus, speciation and net diversification rates rised slowly with several fluctuations between the Late Oligocene (*ca.* 26 Ma) and the Middle Miocene (*ca.* 15 Ma) (Figure 6a). These tendencies are similar for the *Novaeguineae* species group, with a sharper peak during the Early Miocene (*ca.* 18-23 Ma) (Figure 6b).

4. Discussion

4.1. Phylogenetic relationships

Both BI and ML analyses recovered a similar placement of the genus *Cardiodactylus* within the subfamily Eneopterinae in the tribe Lebinthini (Figure 2), as in previous studies (Anso et al., 2016; Nattier et al., 2011; Robillard and Desutter-Grandcolas, 2004, 2006; Vicente et al., 2017). Within *Cardiodactylus*, we were able to assess the status of the two species groups defined by Otte (2007) on morphological grounds.

Our results clearly support the monophyly of the species group *Novaeguineae* and the paraphyly of the *Efordi* species group (-50150.18 versus -50162.0, B_F of 21.78; B_F>10). The paraphyly of the latter can be accounted for by the placement of two species: *C. javarere* and *C. enkraussi*. When using B_F we found out that the placement of *C. enkraussi* outside of the *Efordi* species group is not statistically supported (harmonic mean estimate of -50150.18 versus -50151.98, B_F of 3.6; B_F<10); it is also the case for *C. javarere* (harmonic mean estimate of -50150.18 versus -50149.44; B_F of 1.48, B_F<10). Further molecular analyses with a denser sampling will be necessary to reach a stable conclusion on the status of the *Efordi* species group. Within the *Novaeguineae* species group, our results strongly support a clade grouping *C. novaeguineae* and *C. haddocki* as the sister group of all remaining species of the group. When comparing results of BI and ML analyses, similar relationships are inferred within the species group, except for the positions of *C. manus* and *C. quatei*, perhaps as a consequence of missing taxa from New Guinea and the Southwest Pacific. Nevertheless, the species distributed in SEA are clearly nested within a clade including species from New Guinea and the Southwest Pacific.

4.2. Origin of Cardiodactylus *in the Southwest Pacific*

Our study estimated the divergence time and ancestral areas of *Cardiodactylus* by implementing secondary calibrations based on the study of Vicente et al. (2017), which inferred that *Cardiodactylus* diverged from its sister clade within the tribe Lebinthini in SEA. As a result, similar age estimates for the MRCA of *Cardiodactylus* were recovered: 43.2 Ma in our study vs. 42.57 Ma in the study of Vicente et al. (2017). Secondary calibrations may be a source of bias in dating procedures (Graur and Martin, 2004; Schenk, 2016); however, it is worth highlighting that another study (relying on a larger phylogenetic context) using a different set of secondary calibrations gave a similar age range for *Cardiodactylus* (Anso et al., 2016).

All DIVALIKE stratified analyses infer an origin for *Cardiodactylus* in the Southwest Pacific; it is also the case for all DEC stratified analyses and some of the BAYAREA stratified analyses. This result seems to conflict with the widely accepted assertion that less diverse island communities are easier to invade (Bellemain and Ricklefs, 2008). However, the global biogeographical history of the tribe Lebinthini, as inferred by Vicente et al. (2017), suggests that *Cardiodactylus* diverged from its sister clade in SEA before colonizing the islands of the Southwest Pacific during the Middle Eocene. During that time, the species richness in SEA was likely much higher than that of the Southwest Pacific. The genus probably first occurred in islands now under water, since many islands in this region had not emerged yet (*e.g*., the Solomon and Fiji Islands) or were just coral reef formations (*i.e.* New Caledonia; Neall and Trewick, 2008). In fact, the origin of *Cardiodactylus* in the Southwest Pacific and its recent recolonization of SEA support the "reverse colonization" hypothesis (from small Pacific islands back to continent) as presented in Bellemain and Ricklefs (2008). This hypothesis was recently supported by several studies of other insect clades, such as in *Polyura* butterflies (Toussaint and Balke, 2016) and *Camponotus* ants (Clouse et al., 2015), while it was not recovered for other clades (Economo et al., 2015). Our results consequently bring some new insights into the discussion of Pacific island biogeography and of the role of Pacific clades in contribution to the diversity of surrounding archipelagoes (*e.g*., Claridge et al., 2017). In the case of *Cardiodactylus*, additional studies with a denser sampling of the species distributed in the archipelagoes of the Southwest Pacific will be necessary to precise their origin in the Southwest Pacific. As a reminder, our conclusions are based on *ca.* 50% of the species of *Cardiodactylus* only. Consequenly, some conclusions are likely to change as our knowledge on this genus will continue improving and be further implemented in future biogeographical studies.

4.3. New Guinea as a corridor between the Southwest Pacific and SEA

During the period ranging from its origin in the Middle Eocene to the collision between Australia and Asia (*ca.* 25 Ma), there is no evidence that *Cardiodactylus* colonized any area

552 other than New Guinea (Figure 3). Our results inferred colonization events from the
553 Southwest Pacific to New Guinea after the emergence of the Bird's Head region of New
554 Guinea (*ca.* 25 Ma). New Guinea thus appeared to have been a significant corridor between
555 the Southwest Pacific and SEA, even if this aspect of the biogeography of *Cardiodactylus* is
556 likely not yet addressed with a sufficient taxonomic sampling.

557 Two independent colonization events of SEA from New Guinea are inferred before
558 the emergence of the Wallacea (*ca.* 15 Ma), and a third colonization event occurred
559 afterward immediately (*ca.* 13 Ma). This result is consistent with the dispersal route
560 documented in flightless beetles, which crossed Lydekker's line from New Guinea and
561 reached the Moluccas *ca.* 5 Ma (Tanzler et al., 2016). *Cardiodactylus* initially departed from
562 New Guinea and colonized SEA by crossing the Wallace line, which is a good example of
563 faunal boundaries permeation between Wallace's line and Lydekker's line (Lohman et al.,
564 2011). It was also found in other insect groups: Müller et al. (2013) infered that colonization
565 events happened several times between Wallacea and New Guinea in Pieridae butterflies;
566 similarly, Kalkman et al. (2018) found that the Odonate family Argiolestidae colonized
567 Sulawesi from the north of Australia.

568

569 *4.4. Colonization routes from east to west in SEA and priority of adjacent area in colonization*
570 *routes*

571 *Cardiodactylus* colonized SEA three times independently from east to west across New
572 Guinea, through four different passageways: the Philippines, Sulawesi, the Lesser Sunda
573 Islands and Java.

574 The colonization of SEA through the Philippines occurred during the early Miocene (*ca.*
575 19 Ma), when the Philippines (minus Palawan), a true oceanic island archipelago, began to
576 emerge, initially scattered along the margin of the Philippine Sea Plate (Yumul et al., 2004).
577 From the Philippines, the genus rapidly dispersed to the rest of the region. It also colonized
578 Sulawesi twice, first *ca*. 16.4 Ma, then recently as a range expansion of *C. empagatao*.

579 The colonization of SEA through Sulawesi *ca.* 16.4 Ma was followed by subsequent *in*
580 *situ* diversification within this island. This is consistent with the "predominantly tectonic
581 dispersal origin" of 20 taxa in Sulawesi, as shown by the review of Stelbrink et al. (2012)
582 based on 27 animal datasets: in this study, the authors inferred that speciation on Sulawesi
583 did not occur before the Miocene, which is consistent with geological evidence indicating an
584 increase of landmasses along with more heterogeneous landscapes (Hall, 2009b).
585 During the Late Miocene, colonization events occurred less frequently and were only
586 confined to adjacent areas: the Thai-Malay Peninsula / Indochina, Java / Sumatra, the
587 Philippines / Indochina (Figure 3 and Figure 4a). However, the frequent transgressing events
588 along the Wallace's line (or the modified Huxley's line) documented in the beetle genus
589 *Rhantus* and the avian family Campephagidae (Lohman et al., 2011), were not found in
590 *Cardiodactylus*. These results are consistent with the conclusions on flightless weevils
591 (Tänzler et al., 2016), which showed that the Sunda Arc (region including Borneo, Java and
592 Sumatra) could be a potential dispersal corridor between mainland SEA and Melanesia
593 through its continuous chain of islands.
594 The study of de Bruyn et al. (2014) inferred that colonization events happened more
595 frequently between adjacent areas than between distant ones. This conclusion is verified in
596 our study, but it could however be tempered by the fact that the probabilities of transition
597 amongst our areas were set proportionally to the distance between them.
598

599 *4.5. Weak influence of recent sea level changes on* Cardiodactylus *diversification dynamics*

600 The drastic changes of sea level during the Plio-Pleistocene have been considered of greater
601 influence on species diversification in SEA, through the fragmentation of areas when the sea
602 level rose and the increasing landmass area when the sea level decreased (Guo et al., 2015;
603 Lohman et al., 2011). Interestingly, according to the results of BAMM analyses, net
604 diversification rates of *Cardiodactylus* rose rapidly during the Miocene, and then decreased
605 before the Plio-Pleistocene (Figure 6). Moreover, inferred colonization events during the Plio-

606 Pleistocene in SEA are far less common than expected. These recent events only happened
607 from the Philippines to the Moluccas, Palawan and Sulawesi, and from the Thai-Malay
608 Peninsula to Borneo and Indochina, but they did not generate any detectable diversification
609 (Figures 3 and 4a). It may indicate that the recent changes of sea level and their
610 consequences in terms of fragmentation of habitats were not sufficient to establish stable
611 barriers between the populations, or that the flight abilities of the species prevent speciation
612 to occur.

613 These results are consistent with the Thai-Malay Peninsula position as a
614 biogeographic crossroad between Indochina and the Sundaic faunal regions (Borneo, Java
615 and Sumatra; Lohman et al., 2011). One interesting point in our results is the absence of
616 inferred immigration of *Cardiodactylus* in Java and Sumatra (Figure 4b in Plio-Pleistocene),
617 which differs from the conclusion of de Bruyn et al. (2014), where these two areas are
618 characterized by higher levels of immigration than emigration. This is related with the low
619 diversity of the genus in these islands, which can in turn be linked with missing knowledge of
620 the species distributed in large areas of Sumatra and Java, which could not be sampled.

621

622 *4.6. Dynamics of diversification through the history of* Cardiodactylus

623 Our diversification analysis indicates that there was no distinct rate shift within
624 *Cardiodactylus* and that the speciation rate of the genus changed slowly through time
625 (Figures 5 and 6). However, when we focus on the diversity in each area of SEA, each
626 lineage of *Cardiodactylus* shows a contrasted history.

627 According to the meta-analysis implemented by de Bruyn et al. (2014), the diversity of
628 SEA biota may have arisen through the accumulation of immigrants, by *in situ* diversification,
629 or by a combination of the two. In our results (Figures 4b and 4c), the highest level of
630 immigration occurred in Borneo, Indochina, Java, Moluccas and Sulawesi, and to a lesser
631 extend in the remaining areas.

The areas are however characterized by significantly different levels of *in situ* diversification. There was no event of *in situ* diversification in Indochina, Palawan, Java and the Moluccas, and only one in Borneo. Sulawesi and the Philippines show the highest level of *in situ* diversification, followed by the Thai-Malay Peninsula and the Lesser Sunda Islands. The early colonization of Sulawesi, the Philippine and the Thai-Malay Peninsula from New Guinea and the fragmentation of these areas from the Miocene to the present day may explain that the levels of *in situ* diversification are higher compared with other areas. This result is consistent with the conclusions of de Bruyn et al. (2014), who showed that the level of *in situ* diversification in ancient areas (originating area) was higher than in others. However, the higher level of *in situ* diversification found in the Lesser Sunda Islands contrasts with this conclusion: this group of islands is relatively small and recent, since they emerged during the Middle Miocene. They partly come from microcontinental fragments sliced from Java or/and Australia, while some parts of these islands truly emerged from the ocean. During their history, connections and disconnections between the Lesser Sunda Islands were very common (Hall, 2009a) and this may have resulted in higher speciation rate. Linking the complex geological history in the Lesser Sunda Islands easily explains the species diversity found in this area, despite its recent age and small size. These results suggest that higher levels of species diversification can be driven by frequent connection and disconnection between two areas, at least in some parts of the genus distribution. However, we cannot exclude that this level of *in situ* diversification could partly be linked to a positive sampling bias, since extensive field work has been done recently in this archipelago, while other regions of SEA remain less sampled (Robillard et al., 2014). Similarly, the low level of *in situ* diversification estimated in large islands colonized almost as early as Sulawesi, such as Java and Borneo, could be explained by biases in taxonomic sampling in these areas.

To summarize, the species diversity of *Cardiodactylus* in SEA seemingly results from: (1) the accumulation of colonizers in Indochina, Palawan, Java and Moluccas, and (2) both the accumulation of colonizers and *in situ* diversification in the Thai-Malay Peninsula, Sumatra, the Philippines, Borneo, Sulawesi and the Lesser Sunda Islands.

5. Conclusion

In summary, the historical biogeography of *Cardiodactylus* appears to be linked to the early geological history of SEA. After originating from the Southwest Pacific (diverging from its sister group among the Lebinthini tribe), *Cardiodactylus* lineages crossed the Lydekker's line multiple times, but more rarely transgressed the Wallace's line. SEA was colonized three times independently by members of the *Novaeguineae* species group and once by the *Efordi* species group. In addition, the New Guinea acted as a major colonization source of *Cardiodactylus* in SEA from east to west, while the Thai-Malay Peninsula served as an important corridor between Indochina and Borneo. In addition, Sulawesi served as a diversification hub for *Cardiodactylus* through a combination of high immigration and *in situ* diversification events.

6. Acknowledgements

We thank Alfried Vogler and two anonymous reviewers for numerous valuable comments and suggestions on a previous version of the manuscript. This work was conducted in the context of the PhD thesis of JD, which was funded by China Scholarship Council (CSC), the Innovation Funds of Graduate Programs of Shaanxi Normal University [2012CXB019], the National Science Foundation of Chian [Grant No. 31402006] and a supporting grant from "La Société des Amis du Muséum" and the Institut de Systematique, Evolution, Biodiversité (ISYEB). NV was supported by the post-doc funding "CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico [Processo: 164922/2017-2]" in Brazil. Laboratory access and assistance was provided by the "Service de Systematique Moleculaire" of the Muséum national d'Histoire naturelle, Paris (CNRS UMS 2700). The molecular work was partly supported by agreement no. 2005/67 between the Genoscope (Evry, France) and the MNHN project 'Macrophylogeny of life'; sequencing was also undertaken in the project @SpeedID proposed by F-BoL, the French Barcode of life initiative and the network

687 ‘Bibliothèque du vivant’ funded by the CNRS, INRA and MNHN; and throught several MNHN
688 grants (ATM “Génomique et collections” (Sarah Samadi and Régis Debruyne; ATM blanche
689 2016). Fieldwork was supported by the following projects and institutions: the Museum
690 national d’Histoire naturelle, Paris (grants from ATM “Biodiversite actuelle et fossile”,
691 Stéphane Peigné and Philippe Janvier; ATM “Formes possibles, formes réalisées”, Vincent
692 Bels & Pierre-Henry Gouyon); the Project Lengguru 2014 (www.lengguru.org), conducted by
693 the French Institut de Recherche pour le Développement (IRD), the Indonesian Institute of
694 Sciences (LIPI), the University of Papua (UNIPA), the University of Cendrawasih (UNCEN),
695 the University of Musamus (UNMUS) and the Sorong Fisheries Academy (APSOR) with
696 corporate sponsorship from COLAS Group, Veolia Water and the Total Foundation; the
697 project “Our Planet Reviewed Papua-New-Guinea 2012-2013” set up by Pro-Natura
698 International, MNHN (France), IRD (France) in partnership with the Royal Belgian Institute of
699 Natural Sciences, the New Guinea Binatang Research Center, the University of Papua New
700 Guinea, and the Divine Word University of Madang and with core funding of Prince Albert II
701 of Monaco Foundation, the Stavros Niarchos Foundation, the Total Foundation, the
702 Fondation d’entreprise EDF, the Fonds Pacifique, Spiecapag, Entrepose Contracting, the
703 New-Caledonia Government, the Reef Foundation and the Belgian National Lottery. We
704 thank Jérôme Constant and Carole Paleco (RBINS) for their help during the study of the
705 collections of the Royal Belgian Institute of Natural sciences, Brussels, funded by the
706 SYNTHESYS European program (BE-TAF-6640), Judith Marshall and George Beccaloni
707 (BMNH) for their help during the study in the Natural History Museum, London, funded by the
708 SYNTHESYS European program (GB-TAF-531); Andrej Gorochov (ZIN) for providing
709 *Cardiodactylus* samples from the collections of Saint Petersburg; Rob de Vries and Caroline
710 Pepermans (RMNH) for their help during the study in Leiden collections, funded by a grant
711 from the PPF "État et structure phylogénétique de la biodiversité actuelle et fossile", MNHN
712 (Philippe Janvier); Peter Schwendinger and John Hollier (MHNG) for their help during the
713 study of Genève collections; and Oscar Effendi and Erni Ernawati (MZB, Indonesia) for their

help during the study of eneopterine crickets in Cibinong, Indonesia (MZB), and Shepherd Myers and Franck Howarth (BPBM) for loaning crickets from the collections of Honolulu.

7. References

Anso, J., Barrabe, L., Desutter-Grandcolas, L., Jourdan, H., Grandcolas, P., Dong, J., Robillard, T., 2016. Old Lineage on an old Island: *Pixibinthus*, a new cricket genus endemic to New Caledonia shed light on gryllid diversification in a hotspot of biodiversity. PLoS One 11, e0150920.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P., Rambaut, A., Suchart, M.A. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst. Biol. 61, 170–173.

Bellemain, E., Ricklefs, R.E., 2008. Are islands the end of the colonization road? Trends Ecol. Evol. 23, 461–468.

Bird, M.I., Taylor, D., Hunt, C., 2005. Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? Quat. Sci. Rev. 24, 2228–2242.

Chintauan-Marquier, I.C., Legendre, F., Hugel, S., Robillard, T., Grandcolas, P., Nel, A., Zuccon, D., Desutter-Grandcolas, L., 2016. Laying the foundations of evolutionary and systematic studies in crickets (Insecta, Orthoptera): a multilocus phylogenetic analysis. Cladistics 32, 54–81.

Cigliano, M.M., Braun, H., Eades, D.C., Otte, D., 2017. Orthoptera Species File. Version 5.0/5.0. [01/09/2017]. <http://Orthoptera.SpeciesFile.org>

Claridge, E.M., Gillespie, R.G., Brewer, M.S., Roderick, G.K., 2017. Stepping-stones across space and time: repeated radiation of Pacific flightless broad-nosed weevils (Coleoptera: Curculionidae: Entiminae: *Rhyncogonus*). J. Biogeogr. 44, 784–796.

Clouse, R.M., Janda, M., Blanchard, B., Sharma, P., Hoffmann, D.H., Andersen, A.N., Czekanski-Moir, J.E., Krushelnycky, P., Rabeling, C., Wilson, E.O., Economo, E.P.,

742 Sarnat, E.M., General, D.M., Alpert, G.D., Wheeler, W.C., 2014. Molecular phylogeny
743 of Indo-Pacific carpenter ants (Hymenoptera: Formicidae, *Camponotus*) reveals waves
744 of dispersal and colonization from diverse source areas. Cladistics 31, 1–14.
745 Condamine, F.L., Toussaint, E.F.A., Clamens, A.-L., Genson, G., Sperling, F.A.H., Kergoat,
746 G.J., 2015. Deciphering the evolution of birdwing butterflies 150 years after Alfred
747 Russell Wallace. Sci. Rep. 5, 11860.
748 Condamine, F.L., Toussaint, E.F.A., Cotton, A.M., Sperling, F.A.H., Genson, G., Kergoat,
749 G.J., 2013. Fine-scale biogeographical and temporal diversification processes of
750 peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian
751 Archipelago. Cladistics 29, 88–111.
752 De Bruyn, M., Stelbrink, B., Morley, R.J., Hall, R., Carvalho, G.R., Cannon, C.H., Van den
753 Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L., Maiorano, L., Shoup, R., Von Rintelen,
754 T., 2014. Borneo and Indochina are major evolutionary hotspots for Southeast Asian
755 biodiversity. Syst. Biol. 63, 879–901.
756 Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with
757 BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
758 Duchêne, S., Molak, M., Ho, S.Y.W., 2014. ClockstaR: choosing the number of relaxed-clock
759 models in molecular phylogenetic analysis. Bioinformatics 30, 1017–1019.
760 Economo, E.P., Sarnat, E.M., Janda, M., Clouse, R., Klimov, P.B., Fischer, G., Blanchard,
761 B.D., Ramirez, L.N., Andersen, A.N., Berman, M., Guénard, B., Lucky, A., Rabeling, C.,
762 Wilson, E.O., Knowles, L.L., 2015. Breaking out of biogeographical modules: range
763 expansion and taxon cycles in the hyperdiverse ant genus *Pheidole*. J. Biogeogr. 42,
764 2289–2301.
765 Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior
766 probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52, 665–673.
767 Foster, C.S.P., Ho, S.Y.W., 2017. Strategies for partitioning clock models in phylogenomic
768 dating: application to the angiosperm evolutionary timescale. Genome Biol. Evol. 9,
769 2752–2763.

770 Gernhard, T., 2008. The conditioned reconstructed process. J. Theor. Biol. 253, 769–778.
771 Graur, D., Martin, W., 2004. Reading the entrails of chickens: molecular timescales of
772 evolution and the illusion of precision. Trends Genet. 20, 81–86.
773 Guo, Y.Y., Luo, Y.B., Liu, Z.J., Wang, X.Q., 2015. Reticulate evolution and sea-level
774 fluctuations together drove species diversification of slipper orchids (*Paphiopedilum*) in
775 South-East Asia. Mol. Ecol. 24, 2838–2855.
776 Hall, R., 1996. Reconstructing Cenozoic SE Asia, in: Hall, R. and Blundell, D (Eds), Tectonic
777 Evolution of Southeat Asia, Geological Society, London, pp. 153–184.
778 Hall, R., 2009. Southeast Asia's changing palaeogeography. Blumea. 54, 148–161.
779 Hall, R., 2009b. Continetal grouth at the Indonesian margins of southeast Asia. Arizona
780 Geol.Soc. Digest. 22, 245–258.
781 Hall, R., 2011. Australia-SE Asia collision: plate tectonics and crustal flow, in: Hall, R.,
782 Cottam, M. A., Wilson, M. E. J. (Eds), The SE Asian Gateway: History and Tectonics of
783 the Australia-Asia Collision. Geological Society, London, Special Publications, 355, 75–
784 109.
785 Hall, R., 2012. Sundaland and Wallacea: geology, plate tectonics and palaeogeography, in:
786 Gower, D.J., Richardson, J.E., Rosen, B.R., Rüber, L., Williams, S.T. (Eds.), Biotic
787 evolution and environmental change in Southeast Asia, Cambridge University Press,
788 Cambridge, pp. 32–78.
789 Hall, R., 2013. The palaeogeography of Sundaland and Wallacea since the Late Jurassic. J.
790 Limnol. 72, 1–17.
791 Hall, R., Sevastjanova, I., 2012. Australian crust in Indonesia. Aus. J. Earth Sci. 59, 827–844.
792 Hearty, P.J., Hollin, J.T., Neumann, C.A., O'Leary, M.J., McCulloch, M., 2007. Global sea-
793 level fluctuations during the last Interglaciation (MIS 5e). Quat. Sci. Rev. 26, 2090–
794 2112.
795 Heaney, L.R., 1991. A synopsis of climatic and vegetational change in Southeast Asia. Clim.
796 Change. 19, 53–61.

Heaney, L.R., 2000. Dynamic disequilibrium: a long-term, large-scale perspective on the equilibrium model of island biogeography. Glob. Ecol. Biogeogr. 9, 59–74.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.

Jansa, S.A., Barker, F.K., Heaney, L.R., 2006. The Pattern and Timing of Diversification of Philippine Endemic Rodents Evidence from Mitochondrial and Nuclear Gene Sequences. Syst. Biol. 55, 73–88.

Kalkman, V.J., Dijkstra, D.B, Dow, R.A., Stokvis, F.R., van Tol, J., 2018. Out of Australia: the Argiolestidae reveal the Melanesian Arc System and East Papua Composite Terrane as possible ancient dispersal routes to the Indo-Australian Archipelago (Odonata: Argiolestidae). Int. J. Odonat. 21, 1–14.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Landis, M., Matzke, N.J., Moore, B.R., Huelsenbeck, J.P., 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62, 789–804.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.

Lohman, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hal, R., Ng, P.K.L., Shih, H.-T., Carvalho, G.R., von Rintelen, T., 2011. Biogeography of the Indo-Australian Archipelago. Annu. Rev. Ecol. Evol. Syst. 42, 205–226.

Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63, 951–970.

Metcalfe, I., 2006. Paleozoic and Mesozoic tectonic evolution and palaeogeography of East Asian crustal fragments: The Korean Peninsula in context. Gondwana Res. 9, 24–46.

Miller, M.A., Schwartz, T., Pickett, B.E., He, S., Klem, E.B., Scheuermann, R.H., Passarotti, M., Kaufman, S., O'Leary, M.A., 2015. A RESTful API for access to phylogenetic tools via the CIPRES Science Gateway. Evol. Bioinform. 11, 43–48.

824 Mittermeier, R.A., Gil, P.R., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C.G.,
825 Lamoreux, J., Da Fonseca, G.A.B., 2004. Hotspots Revisited - Earth's biologically
826 richest and most endangered terrestrial ecoregions. The University of Chicago Press,
827 Chicago.
828 Moore, B.R., Höhna, S., May, M.R., Rannala, B., Huelsenbeck, J.P. 2016. Critically
829 evaluating the theory and performance of Bayesian analysis of macroevolutionary
830 mixtures. Proc. Natl. Acad. Sci. USA 113, 9569–9574.
831 Morley, R.J., 2012. A review of the Cenozoic palaeoclimate history of Southeast Asia, in:
832 Gower, D., Johnson, K., Richardson, J., Rosen, B., Rüber, L., Williams, S. (Eds), Biotic
833 evolution and environmental change in Southeast Asia, Cambridge University Press,
834 Cambridge, pp. 79–114.
835 Müller, C.J., Matos-Maravi, P.E., Beheregaray, L.B., 2013. Delving into *Delias* Hübner
836 (Lepidoptera: Pieridae): fine-scale biogeography, phylogenetics and systematics of the
837 world's largest butterfly genus. J. Biogeogr. 40, 881–893.
838 Nattier, R., Grandcolas, P., Elias, M., Desutter-Grandcolas, L., Jourdan, H., Couloux, A.,
839 Robillard, T., 2012. Secondary sympatry caused by range expansion informs on the
840 dynamics of microendemism in a biodiversity hotspot. PLoS One 7, e48047.
841 Nattier, R., Robillard, T., Desutter-Grandcolas, L., Couloux, A., Grandcolas, P., 2011. Older
842 than New Caledonia emergence? A molecular phylogenetic study of the eneopterine
843 crickets (Orthoptera: Grylloidea). J. Biogeogr. 38, 2195–2209.
844 Neall, V.E., Trewick, S.A., 2008. The age and origin of the Pacific islands: a geological
845 overview. Phil. Trans. R. Soc. B. 363, 3293–3308.
846 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQTREE: A fast and
847 effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol.
848 Evol. 32, 268–274.
849 Otte, D., 2007. New species of *Cardiodactylus* from the western Pacific region (Gryllidae:
850 Eneopterinae). Proc. Acad. Nat. Sci. Phila. 156, 341–400.

851 Outlaw D.C., Voelker, G., 2008. Pliocene climatic change in insular Southeast Asia as an
852 engine of diversification in *Ficedula* flycatchers. J. Biogeogr. 35, 739–752.
853 Rabosky, D.L., Donnellan, S.C., Grundler, M., Lovette, J.J., 2014a. Analysis and visualization
854 of complex macroevolutionary dynamics: an example from Australian scincid lizards.
855 Syst. Biol. 63, 610–627.
856 Rabosky, D.L., Grundler, M., Anderson, C., Title, P., Shi, J.J., Brown, J.W., Huang, H.,
857 Larson, J.G., Kembel, S., 2014b. BAMMtools: an R package for the analysis of
858 evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol. 5, 701–707.
859 Rabosky, D.L., Mitchell, J.S., Chang, J., 2017. Is BAMM flawed? Theoretical and practical
860 concerns in the analysis of multi-rate diversification models. Syst. Biol. 66, 477–498.
861 Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J., Alfaro, M.E.,
862 2013. Rates of speciation and morphological evolution are correlated across the
863 largest vertebrate radiation. Nat. Comm. 4, 1958.
864 Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6, Available from
865 http://beast.bio.ed.ac.uk/Tracer.
866 Ree, R.H., Sanmartín, I., 2018. Conceptual and statistical problems with the DEC+J model of
867 founder-event speciation and its comparison with DEC via model selection. J.
868 Biogeogr. 45, 741–749.
869 Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution
870 by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.
871 Ripplinger, J., Sullivan, J., 2008. Does Choice in Model Selection Affect Maximum Likelihood
872 Analysis? Syst. Biol. 57, 76–85.
873 Robillard, T., Desutter-Grandcolas, L., 2004. Phylogeny and the modalities of acoustic
874 diversification in extant Eneopterinae (Insecta, Orthoptera, Grylloidea, Eneopteridae).
875 Cladistics 20, 271-293.
876 Robillard, T., Desutter-Grandcolas, L., 2006. Phylogeny of the cricket subfamily
877 Eneopterinae (Orthoptera, Grylloidea, Eneopteidae) based on four molecular loci and
878 morphology. Mol. Phylogenet. Evol. 40, 643–661.

879 Robillard, T., Grandcolas, P., Desutter-Grandcolas, P., 2007. A shift toward harmonics for
880 high-frequency calling shown with phylogenetic study of frequency spectra in
881 Eneopterinae crickets (Orthoptera, Grylloidea, Eneopteridae). Can. J. Zool. 85, 1264–
882 1274.
883 Robillard, T., Ichikawa, A., 2009. Redescription of two *Cardiodactylus* species (Orthoptera,
884 Grylloidea, Eneopterinae): the supposedly well-known *C. novaeguineae* (Haan, 1842),
885 and the semi-forgotten *C. guttulus* (Matsumura, 1913) from Japan. Zool. Sci. 26, 878–
886 891.
887 Robillard, T., Montealegre-Z, F., Desutter-Grandcolas, L., Grandcolas, P., Robert, D., 2013.
888 Mechanisms of high-frequency song generation in brachypterous crickets and the role
889 of ghost frequencies. J. Exp. Biol. 216, 2001–2011.
890 Robillard, T., Dong, J.J., 2016. The *Cardiodactylus* crickets from Eastern New Guinea, with
891 description of five new species (Orthoptera: Gryllidae: Eneopterinae), in: Robillard, T.,
892 Legendre, F., Villemant, C., Leponce, M. (Eds.), Insects of Mount Wilhelm, Papua New
893 Guinea, Mémoires du Muséum National d'Histoire Naturelle 209, Muséum national
894 d'Histoire Naturelle, Paris, pp. 203–258.
895 Robillard, T., Gorochov, A.V., Poulain, S., Suhardjono, Y.R., 2014. Revision of the cricket
896 genus *Cardiodactylus* (Orthoptera, Eneopterinae, Lebinthini): the species from both
897 sides of the Wallace line, with description of 25 new species. Zootaxa 3854, 1–104.
898 Ronquist, F., 1997. Dispersal-Vicariance Analysis: A new approach to the quantification of
899 historical biogeography. Syst. Biol. 46, 195–203.
900 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B.,
901 Liu, L., Suchard, M. A, Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian
902 phylogenetic inference and model choice across a large model space. Syst. Biol. 61,
903 539–542.
904 Schenk, J.J., 2016. Consequences of secondary calibrations on divergence time estimates.
905 PLoS One 11, e0148288.

906 Stelbrink, B., Albrecht, C., Hall, R., von Rintelen, T., 2012. The biogeography of Sulawesi
907 revisited: is there evidence for a vicariant origin of taxa on Wallace's "anomalous
908 island"? Evolution 66, 2252–2271.
909 Steppan, S.J., Zawadzki, C., Heaney, L.R., 2003. Molecular phylogeny of the endemic
910 Philippine rodent *Apomys* (Muridae) and the dynamics of diversification in an oceanic
911 archipelago. Biol. J. Linn. Soc. 80, 699–715.
912 Tanzler, R., Van Dam, M.H., Toussaint, E.F.A., Suhardjono, Y.R., Balke, M., Riedel, A.,
913 2016. Macroevolution of hyperdiverse flightless beetles reflects the complex geological
914 history of the Sunda Arc. Sci. Rep. 6, 18793.
915 Ter Hofstede, H.M., Schöneich, S., Robillard, T., Hedwig, B., 2015. Evolution of a
916 communication system by sensory exploitation of startle behavior. Curr. Biol., 25,
917 3245–3252.
918 Toussaint, E.F.A., Balke, M., 2016. Historical biogeography of *Polyura* butterflies in the
919 oriental Palaeotropics: trans-archipelagic routes and South Pacific island hopping. J.
920 Biogeogr. 43, 1560–1572.
921 Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., Minh, B.Q., 2016. W-IQ-TREE: a fast online
922 phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44, W232–
923 W235.
924 Vicente, N., Kergoat, G.J., Dong, J., Yotoko, K., Legendre, F., Nattier, R., Robillard, T., 2017.
925 In and out of the Neotropics: historical biogeography of Eneopterinae crickets. J.
926 Biogeogr. 44, 2199–2210.
927 Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeat Asia: shorelines, river systems
928 and time durations. J. Biogeogr. 27, 1153–1167.
929 Wallace, A.R., 1860. On the zoological geography of the Malay archipelago. Zool. J. Linn.
930 Soc. 4, 172–184.
931 Woodruff, D.S., 2010. Biogeography and conservation in Southeast Asia: how 2.7 million
932 years of repeated environmental fluctuations affect today's patterns and the future of
933 the remaining refugial-phase biodiversity. Biodivers. Conserv. 19, 919–941.

934 Yumul, G.P., Dimalanta, C.B., Tamayo, R.A., Maury, R.C., Bellon, H., Polvé, M.,
935 Maglambayan, V.B., Querubin, C.L., Cotten, J., 2004. Geology of the Zamboanga
936 Peninsula, Mindanao, Philippines: an enigmatic South China continental fragment?
937 Geol. Soc. London Spec. Publ. 266, 289–312.
938 Zachos, F.E., Habel, J.C., 2011. Biodiversity hotspots: Distribution and protection of
939 conservation priority areas. Springer-Verlag, Berlin.
940 Zachos, J., Pagani, M., Slloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms and
941 aberrations in global climate 65 Ma to present. Science 292, 686–693.
942 Zahirovic, S., Seton, M., Müller, R.D., 2014. The Cretaceous and Cenozoic tectonic evolution
943 of Southeast Asia. Solid Earth 5, 227–273.
944

945 **Figure titles**

946

947 **Figure 1**. Distribution map of *Cardiodactylus* in Southeast Asia. In this study, we sampled 28
948 out of 35 *Cardiodactylus* species distributed in Southeast Asia. The sampled species
949 are written in black (collecting localities in black circles), and species missing in the
950 sampling are shown in grey.. Map modified after Robillard et al. (2014).

951

952 **Figure 2**. Phylogeny of the genus *Cardiodactylus* inferred from Maximum likelihood (ML) and
953 Bayesian inference (BI). Values on nodes indicate branch support; the first two values
954 correspond to non-parametric bootstrap values (BV from RAxML analyses first, then
955 BV from IQ-TREE analyses) whereas the third value corresponds to BI posterior
956 probabilities (PP). Asterisks are used to indicate maximum support (100% for BV and
957 1.0 for PP). Signs '-' indicate topological incongruences between analyses, the
958 topology inferred with RAxML being represented. The bold red circles with letters
959 indicate the secondary calibration points used in the dating analyses; the circled
960 numbers highlight clades discussed in the text.

961

962 **Figure 3.** Reconstruction of historical biogeography for *Cardiodactylus* using a stratified
963 dispersal-vicariance analysis (DIVALIKE) model. The left panel and map represent the
964 12 areas implemented in the biogeographical model. Present-day distributions of each
965 species are given at the tips by colored circles corresponding to colored areas on map;
966 a colored square represents the inferred area(s) with the highest relative probability in
967 the DIVALIKE analysis; corner positions represent geographic ranges immediately after
968 a cladogenesis event. The red numbers near each node correspond to the median age
969 inferred in Beast analysis and the gray lines indicate its 95% height posterior
970 distribution (HPD). Bayesian posterior probabilities below 0.50 are not shown.

971

Figure 4. Colonization routes, numbers of emigration and immigration events and *in situ* diversification events inferred from ancestral area estimation for lineages of the *Novaeguineae* species group during the pre-Miocene, Miocene and Plio-Pleistocene. a) Colonization routes; line colors correspond to inferred area of origin (same color code as in simplified map): multiple lines of the same color correspond to multiple colonization events. b) Number of emigration (positive bars) and immigration (negative bars) events inferred from colonization events (Figure 4a) for each area in the pre-Miocene, Miocene and Plio-Pleistocene; column colors correspond to inferred areas as above. c) *In situ* diversification events inferred from ancestral area estimation for each area in the pre-Miocene, Miocene and Plio-Pleistocene.

Figure 5. Rate shift configurations with the four highest posterior probabilities from the 95% credible set within *Cardiodactylus*. The tree topology referred to Figure 3. The colored histogram indicates the speciation (λ), the color gradient corresponds the speciation rate.

Figure 6. Rates-through-time analysis of speciation (λ), extinction (μ) and net diversification (γ) in *Cardiodactylus* (a) and within *Novaeguineae* species group (b).

992 **Table**

993 **Table** . Best-fit models of sequence evolution and partitioning schemes selected with
994 PartitionFinder (‘*unlinked branch lengths’* option) for phylogenetic reconstructions using
995 Bayesian Inference (MrBayes/BEAST) and Maximum Likelihood (RAxML/IQ-TREE).
996 Codon position is denoted by pos1, pos2 and pos3. Subsets are denoted by p1 and p2.

Analyses	Partitions	Models
Bayesian inference (BI) analyses	p1: 16S, COI_pos2, Cytb_pos2, 18S, 28S, H3_pos1, H3_pos2, H3_pos3, EF1a;	GTR+I+G
	p2: 12S, COI_pos1, COI_pos3, COII_pos1, COII_pos2, COII_pos3, Cytb_pos1, Cytb_pos3;	GTR+I+G
Maximum likelihood (ML) analyses	p1: 16S, COI_pos2, Cytb_pos2, 18S, 28S, H3_pos1, H3_pos2, H3_pos3, EF1a;	GTR+I+G
	p2: 12S, COI_pos1, COI_pos3, COII_pos1, COII_pos2, COII_pos3, Cytb_pos1, Cytb_pos3;	GTR+I+G
Bayesian evolutionary analyses by sampling trees (BEAST)	p1: 16S, COI_pos2, Cytb_pos2, 18S, 28S, H3_pos1, H3_pos2, H3_pos3, EF1a;	GTR+I+G
	p2: 12S, COI_pos1, COI_pos3, COII_pos1, COII_pos2, COII_pos3, Cytb_pos1, Cytb_pos3;	GTR+I+G

997
998

Supporting Information

Appendix S1. Taxon sampling.

Appendix S2. Time-stratified biogeographical standard implemented in BioGeoBEARS analyses.

Appendix S3. Summary of ancestral area estimation with different models and dispersal rate multipliers with time slices (Table S3a) and without time slices (Table S3b). The gray line highlights the model with the highest statistical support.

Appendix S4. Sampling of species groups in BAMM analysis.

Appendix S5. Gene trees inferred for each marker: 16S (a), 12S (b), COI (c), COII (d), Cytb (e), 18S (f), 28S (g), EF1a (h), H3 (i).

Appendix S6. Original output results of Bayesian inference (BI) and maximum likelihood (ML) analyses for the concatenated dataset.

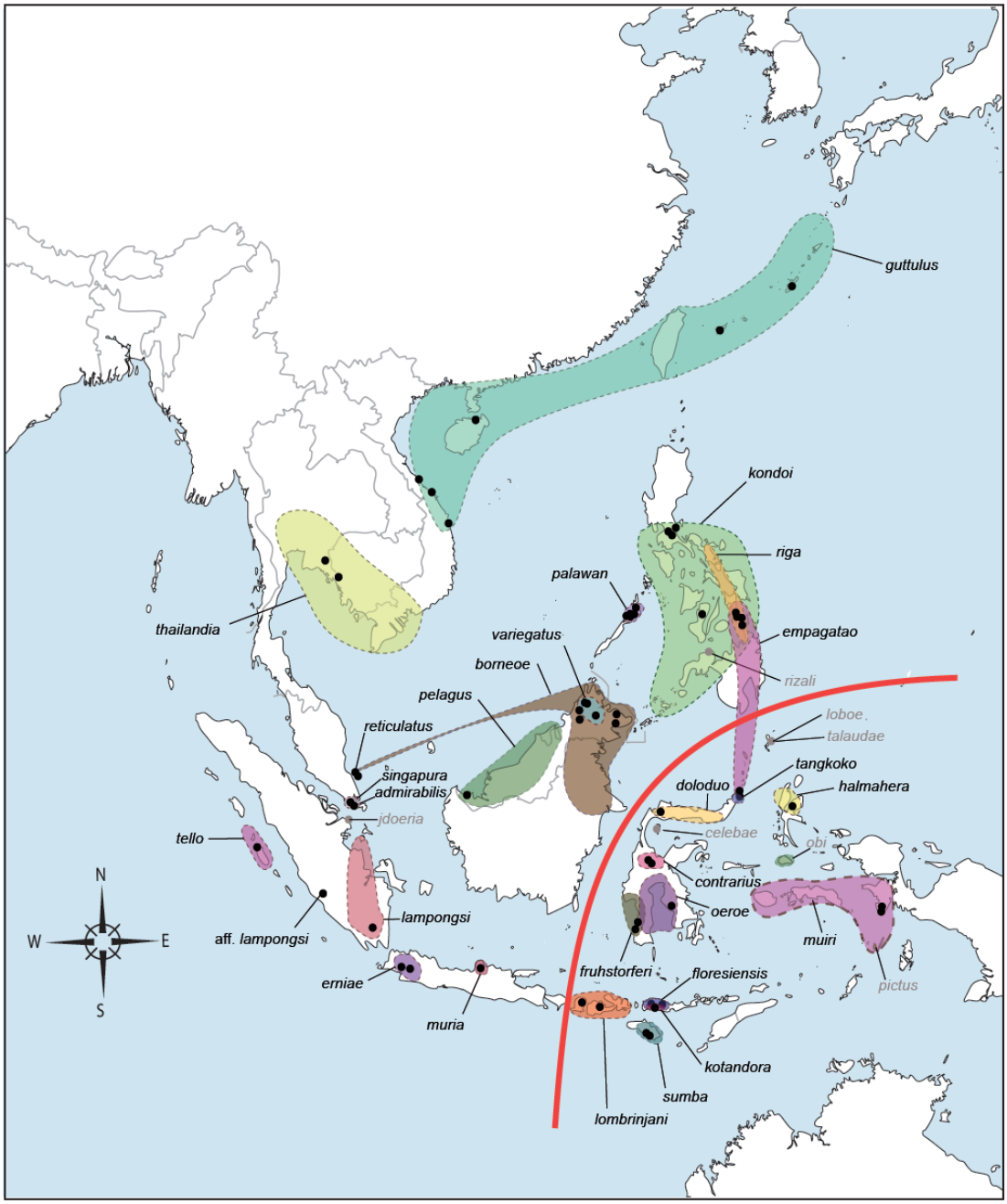
Appendix S7. Original outputs for dating analyses.

Appendix S8. Results of ancestral area estimation in BioGeoBEARS based on the most complex set of dispersal rate multipliers with three time slices, two maximum ancestral areas and DIVALIKE speciation model.

Appendix S9. Results of ancestral area estimation in BioGeoBEARS based on the most complex set of dispersal rate multipliers without time slices, two maximum ancestral areas and DIVALIKE speciation model.

Appendix S10. Summary of diversification rate under BAMM analysis with gradient of prior values ranging from 0.1 to 1.0. The gray line highlights the model with the highest statistical support.

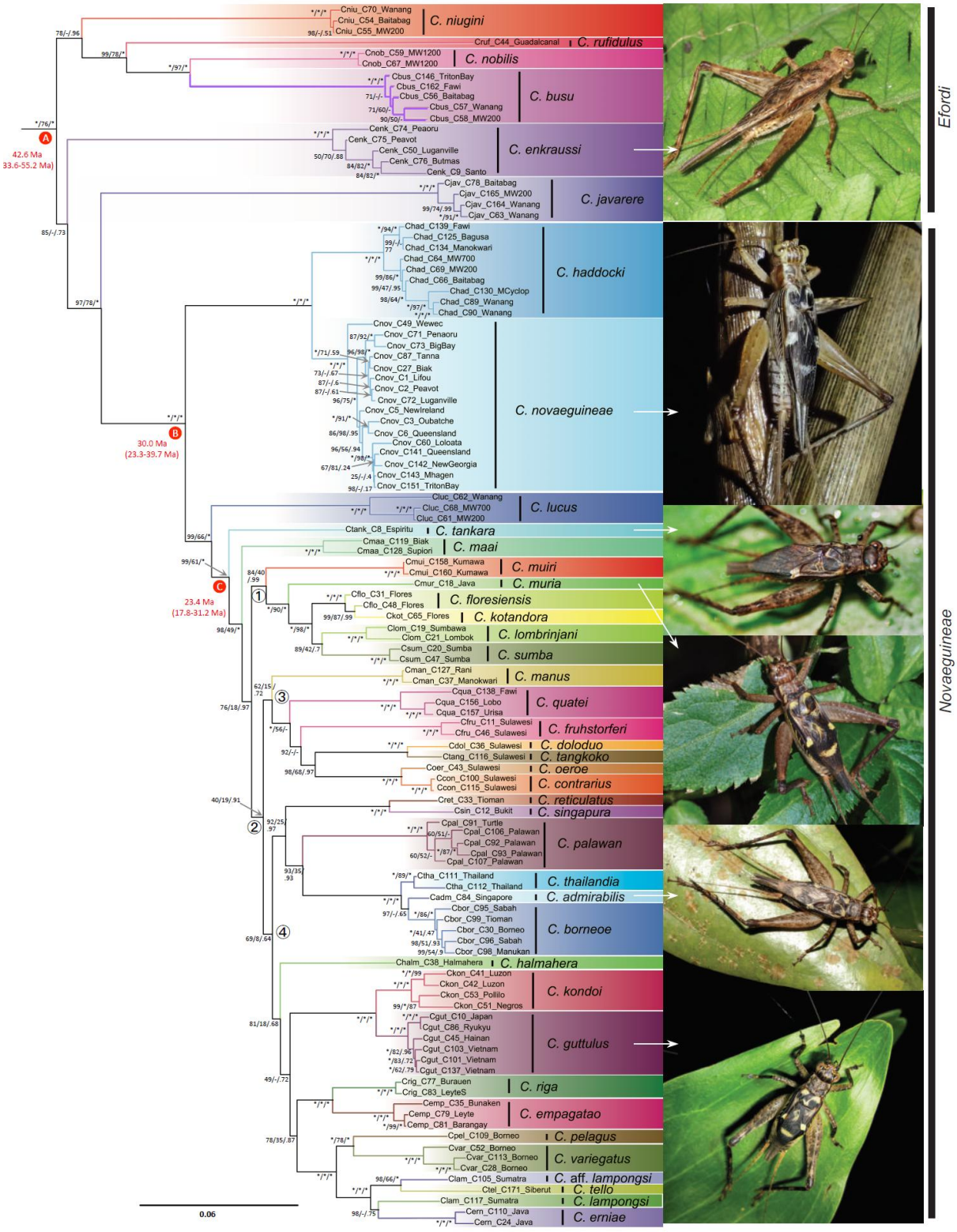
1024 Fig.1



1025

1026

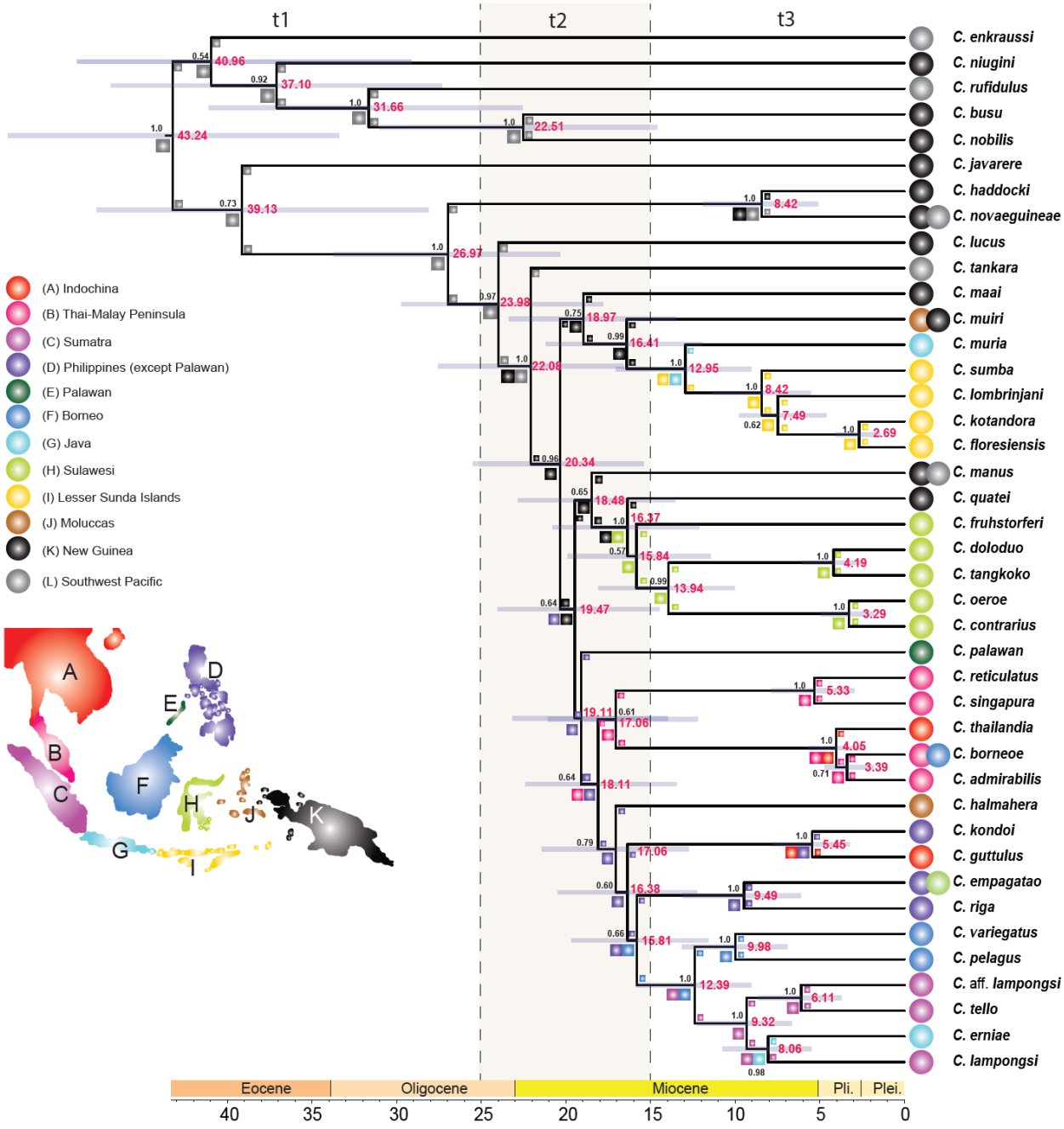
1027 Fig.2



1028

1029

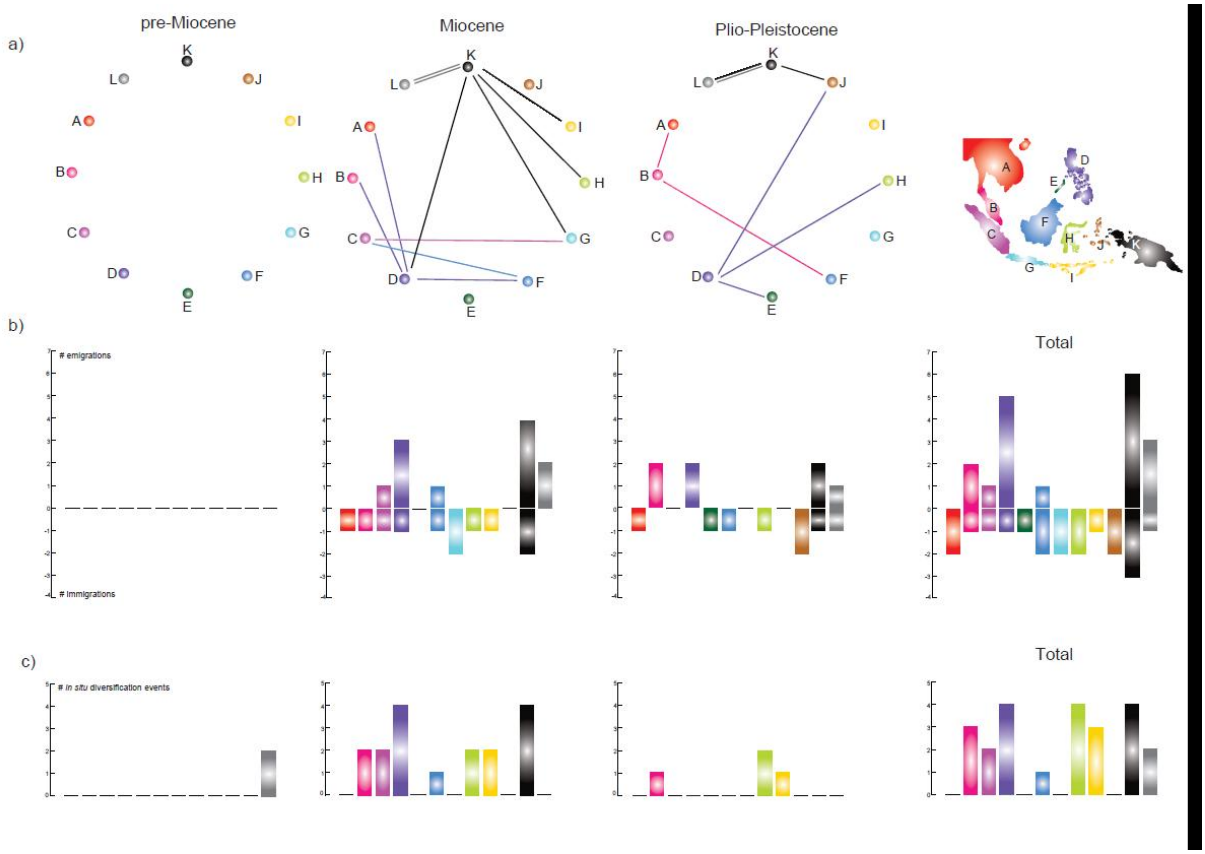
1030 Fig.3



1031

1032

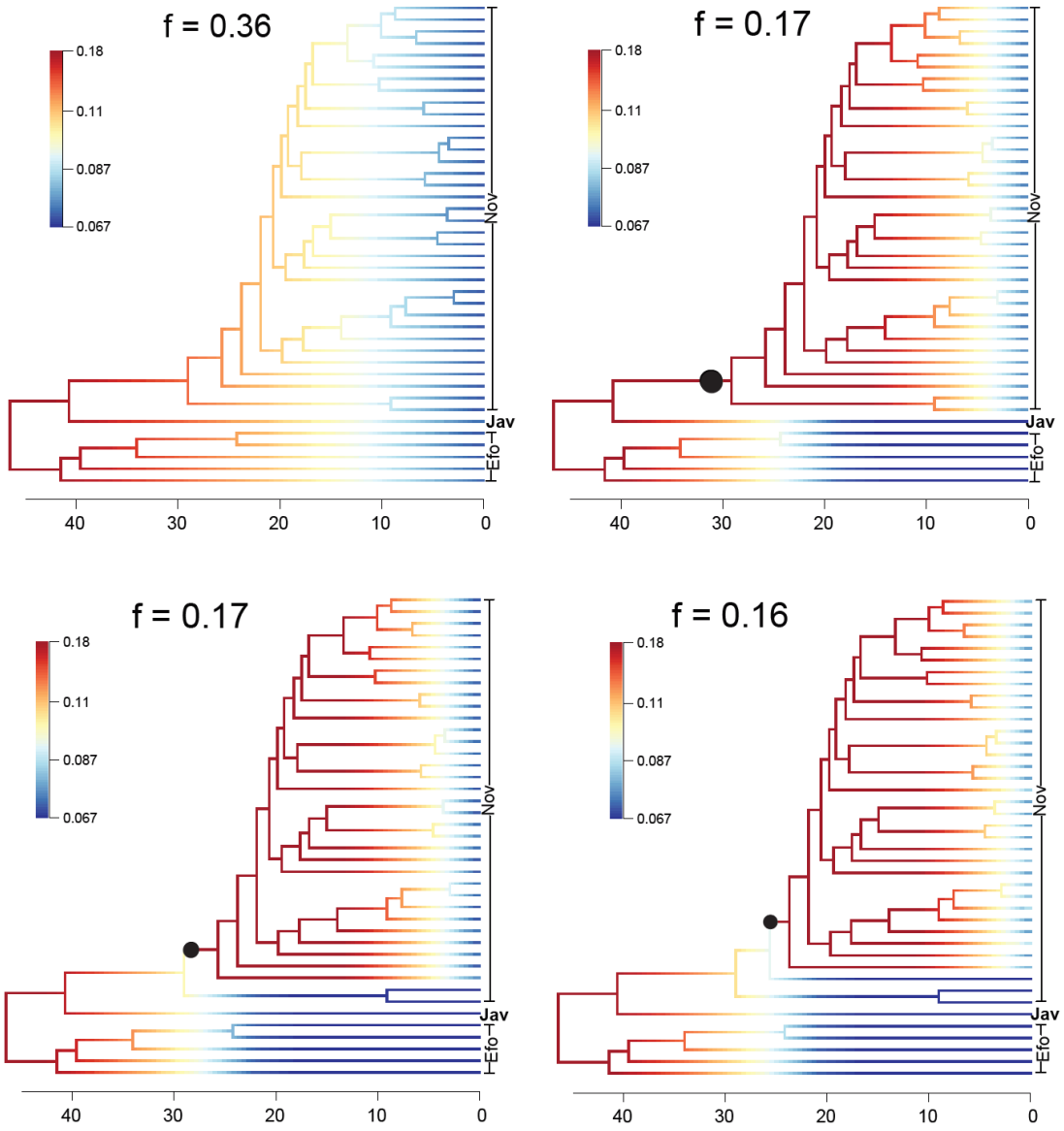
1033 Fig.4



1034

1035

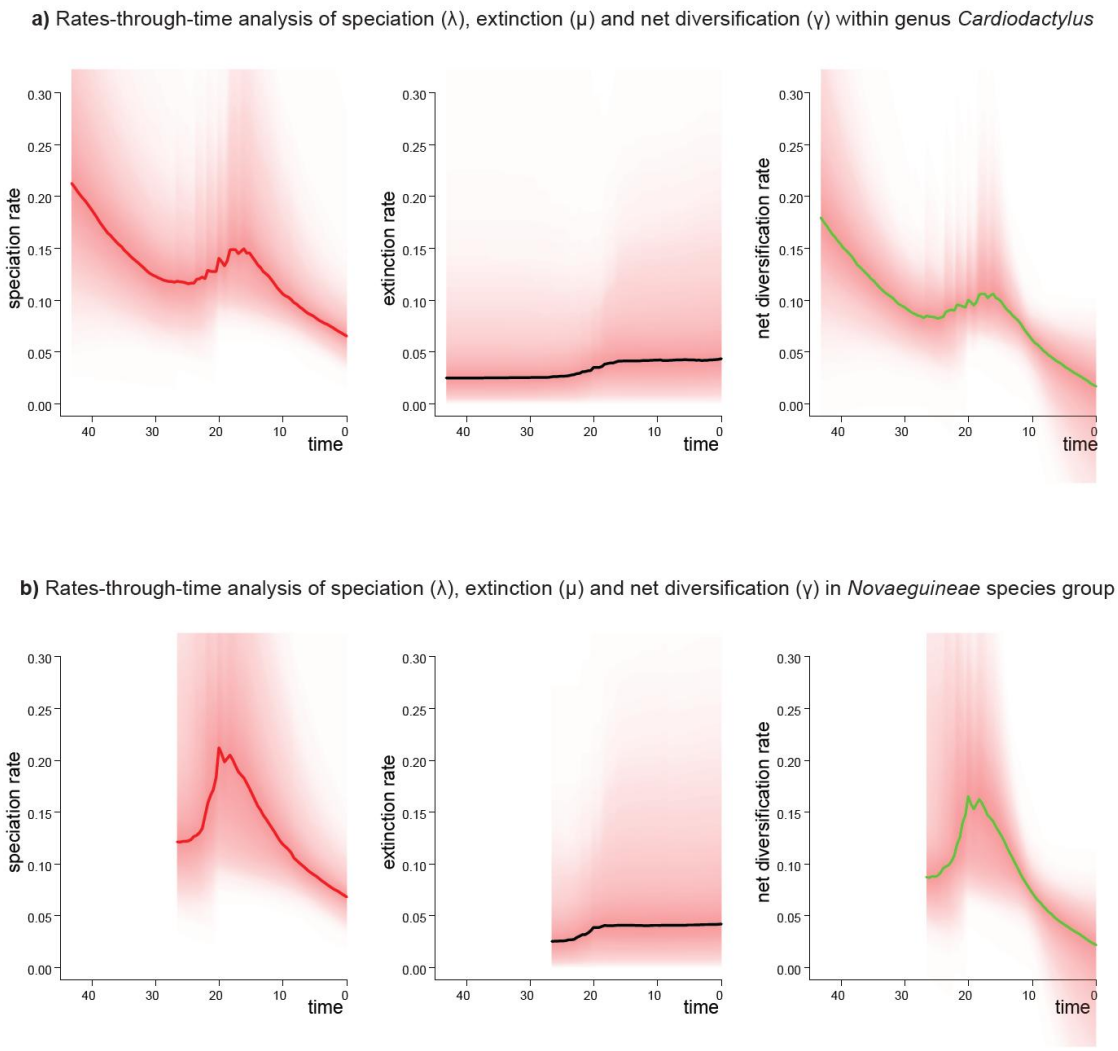
1036 Fig.5.



1037

1038

1039 Fig.6.



1040

1041

Supporting information

Appendix S1. Taxon sampling.

List of all *Cardiodactylus* and outgroups used in this study with current geographical distribution and GenBank accession numbers of each marker. The abbreviations of taxonomic status indicate as following: subfamily Gryllinae (GRYL), subfamily Eneopterinae (ENEO), subfamily Gryllinae (Gry), tribe Lebinthini (Leb), tribe Eneopterini (Ene), tribe Eurepini (Eur), tribe Nisitrini (Nis), tribe Xenogryllini (Xen). NA means missing data. The samples used in the dating analyses to represent each species are show with an asterisk (*).

Species	Taxonomy	Voucher	Molecular code	Geographic area/code	16S	12S	COI	COII	Cytb	18S	28S	EF1a	H3
Acheta domesticus (Linnaeus, 1758)	GRYL/Gry	MNHN-EO-ENSIF3523	Adom	France (Laboratory strain)	AF248698	ADZ97611	JX897403	JX897439	AF248682	AD18SITS1	JX897465	GQ886692	KR900315 0
Agnotecous albifrons Desutter-Grandcolas, 1997	ENEO/Leb	MNHN-EO-ENSIF-2767	AalCT	New Caledonia	JX897353	JX897394	JX897418	JX897446	JX897314	JX897583	JX897490	JX897527	JX897572
Agnotecous albifrons Desutter-Grandcolas, 1997	ENEO/Leb	MNHN-EO-ENSIF-1771	AalGe	New Caledonia	JX897354	JX897396	JX897416	JX897445	JX897316	JX897585	NA	JX897529	JX897574
Agnotecous azurensis Desutter-Grancolas, 2006	ENEO/Leb	MNHN-EO-ENSIF-2777	AazGK	New Caledonia	JX897360	JX897389	JX897426	NA	JX897331	JX897593	JX897472	JX897505	JX897565
Agnotecous azurensis Desutter-Grancolas, 2006	ENEO/Leb	MNHN-EO-ENSIF-2780	AazRB	New Caledonia	JX897358	JX897376	JX897423	JX897453	JX897329	JX897595	JX897475	JX897502	JX897566
Agnotecous meridionalis Desutter-Grandcolas, 2006	ENEO/Leb	MNHN-EO-ENSIF-2772	AmeIDP	New Caledonia	JX897349	JX897401	JX897420	NA	JX897311	JX897579	JX897488	JX897519	JX897553
Agnotecous meridionalis Desutter-Grandcolas, 2006	ENEO/Leb	MNHN-EO-ENSIF-2771	AmePB	New Caledonia	JX897350	JX897402	JX897410	JX897442	JX897313	JX897597	JX897489	JX897520	JX897550
Agnotecous obscurus (Chopard, 1970)	ENEO/Leb	MNHN-EO-ENSIF-2786	AobAo2	New Caledonia	JX897356	JX897398	JX897415	JX897449	JX897319	JX897591	NA	JX897510	NA
Agnotecous obscurus (Chopard, 1970)	ENEO/Leb	MNHN-EO-ENSIF-2785	AobMa	New Caledonia	JX897357	JX897393	JX897412	JX897450	JX897320	JX897587	JX897487	JX897525	JX897576
Agnotecous robustus (Chopard, 1915)	ENEO/Leb	MNHN-EO-ENSIF-2752	AroAo2	New Caledonia	JX897359	JX897375	JX897406	JX897443	JX897333	JX897588	NA	JX897498	JX897555
Agnotecous sarramea Desutter-Grancolas, 1997	ENEO/Leb	MNHN-EO-ENSIF-2764	AsaMA	New Caledonia	JX897372	JX897380	JX897430	JX897456	JX897342	JX897598	JX897471	JX897511	JX897561
Agnotecous occidentalis Desutter-Grandcolas, 2006	ENEO/Leb	MNHN-EO-ENSIF-2765	AspCR2	New Caledonia	JX897362	JX897386	JX897434	JX897461	JX897322	JX897589	NA	JX897512	JX897570
Agnotecous clarus Desutter-Grandcolas, 2006	ENEO/Leb	MNHN-EO-ENSIF-2788	AspTRPP	New Caledonia	JX897347	JX897400	JX897407	JX897451	JX897324	JX897590	JX897492	JX897523	JX897554
Agnotecous yahoue Otte, 1987	ENEO/Leb	MNHN-EO-ENSIF-2766	AyaKo2	New Caledonia	JX897366	JX897387	JX897438	JX897463	JX897334	JX897602	JX897478	JX897500	JX897549
Agnotecous yahoue Otte, 1987	ENEO/Leb	MNHN-EO-ENSIF-2773	AyaMo5	New Caledonia	JX897367	JX897388	JX897437	JX897462	JX897337	NA	JX897479	JX897501	NA
**Cardiodactylus admirabilis* Tan & Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3441	Cadm_C84_Singapore	Singapore/B	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus borneoe Robillard & Gorochov, 2014	ENEO/Leb	RMNH	Cbor_C30_Borneo	Borneo/F	this study	this study	NA	this study	this study	this study	this study	this study	this study
Cardiodactylus borneoe Robillard & Gorochov, 2014	ENEO/Leb	ZIN	Cbor_C95_Sabah	Borneo/F	this study	this study	NA	NA	this study	NA	NA	NA	this study
Cardiodactylus borneoe Robillard & Gorochov, 2014	ENEO/Leb	ZIN	Cbor_C96_Sabah	Borneo/F	this study	this study	NA	this study	this study	this study	this study	this study	this study
Cardiodactylus borneoe Robillard & Gorochov, 2014	ENEO/Leb	ZIN	Cbor_C98_Manukan	Borneo/F	this study	this study	NA	this study	this study	this study	this study	NA	this study
**Cardiodactylus borneoe* Robillard & Gorochov, 2014	ENEO/Leb	ZIN	Cbor_C99_Tioman	Malaysia/B	*this study	*this study	NA	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus busu Otte, 2007	ENEO/Leb	Leng-TR119	Cbus_C146_TritonBay	New Guinea/K	this study	this study	this study	this study	this study	this study	NA	this study	this study

Continued

1055 *Continued*

Species	Taxonomy	Voucher	Molecular code	Geographic area/code	16S	12S	COI	COII	Cytb	18S	28S	EF1a	H3
Cardiodactylus busu Otte, 2007	ENEO/Leb	ZIN	Cbus_C162_Fawi	New Guinea/K	this study	this study	this study	this study	this study	NA	NA	this study	this study
Cardiodactylus busu Otte, 2007	ENEO/Leb	MNHN-ENSIF184	Cbus_C56_Baitabag	New Guinea/K	NA	this study	this study	this study	NA	this study	this study	NA	this study
**Cardiodactylus busu* Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF101	Cbus_C57_Wanang	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus busu Otte, 2007	ENEO/Leb	MNHN-ENSIF109	Cbus_C58_MW200	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus contrarius Gorochov, 2014	ENEO/Leb	ZIN	Ccon_C100_Sulawesi	Sulawesi/H	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus contrarius* Gorochov, 2014	ENEO/Leb	MNHN-EO-ENSIF3653	Ccon_C115_Sulawesi	Sulawesi/H	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus doloduo* Gorochov, 2014	ENEO/Leb	ZIN	Cdol_C36_Sulawesi	Sulawesi/H	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus empagatao Otte, 2007	ENEO/Leb	ZIN	Cemp_C35_Bunaken	Sulawesi/H	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus empagatao* Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF3353	Cemp_C79_Leyte	Philippines/D	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus empagatao Otte, 2007	ENEO/Leb	UPLB MNH	Cemp_C81_Barangay	Philippines/D	this study	this study	this study	this study	this study	this study	NA	this study	this study
Cardiodactylus enkraussi Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF2313	Cenk_C50_Luganville	Vanuatu/L	this study	this study	this study	this study	NA	this study	NA	NA	this study
Cardiodactylus enkraussi Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF2398	Cenk_C74_Peaoru	Vanuatu/L	NA	this study	this study	this study	NA	this study	NA	NA	this study
Cardiodactylus enkraussi Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF2335	Cenk_C75_Peavot	Vanuatu/L	NA	this study	this study	this study	this study	this study	NA	NA	NA
**Cardiodactylus enkraussi* Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF2354	Cenk_C76_Butmas	Vanuatu/L	*this study	*this study	*this study	*this study	NA	*this study	*this study	*this study	*this study
Cardiodactylus enkraussi Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF2366	Cenk_C9_Santo	Vanuatu/L	JF972518	JF972502	KU705561	KU705550	NA	JF972533	NA	NA	KU705595
Cardiodactylus erniae Robillard & Gorochov, 2014	ENEO/Leb	MNHN-EO-ENSIF3492	Cern_C110_Java	Java/G	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus erniae* Robillard & Gorochov, 2014	ENEO/Leb	MNHN-EO-ENSIF3489	Cern_C24_Java	Java/G	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus floresiensis* Robillard, 2014	ENEO/Leb	MZB	Cflo_C31_Flores	Flores island/I	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus floresiensis Robillard, 2014	ENEO/Leb	MZB	Cflo_C48_Flores	Flores island/I	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus fruhstorferi Gorochov & Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF1596	Cfru_C11_Sulawesi	Sulawesi/H	this study	this study	NA	this study	this study	this study	this study	this study	this study
**Cardiodacıylus fruhstorferi* Gorochov & Robillard, 2015	ENEO/Leb	MNHN-ENSIF1123	Cfru_C46_Sulawesi	Sulawesi/H	*this study	*this study	NA	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus guttulus (Matsumura, 1913)	ENEO/Leb	ZIN	Cgut_C101_Vietnam	Vietnam/A	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus guttulus (Matsumura, 1913)	ENEO/Leb	ZIN	Cgut_C103_Vietnam	Vietnam/A	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus guttulus (Matsumura, 1913)	ENEO/Leb	MNHN-ENSIF1193	Cgut_C10_Japan	Japan/A	JF972519	JF972503	KU705562	NA	JF972487	JF972534	KU705580	KU705612	KU705596
Cardiodactylus guttulus (Matsumura, 1913)	ENEO/Leb	ZIN	Cgut_C137_Vietnam	Vietnam/A	this study	this study	NA	this study	this study	this study	this study	this study	this study
**Cardiodactylus guttulus* (Matsumura, 1913)	ENEO/Leb	LibinMa	Cgut_C45_Hainan	Hainan/A	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus guttulus (Matsumura, 1913)	ENEO/Leb	MNHG	Cgut_C86_Ryukyu	Ryukyu island/A	this study	this study	this study	this study	this study	this study	this study	this study	this study

1056 *Continued*

1057 *(Continued)*

Species	Taxonomy	Voucher	Molecular code	Geographic area/code	16S	12S	COI	COII	Cytb	18S	28S	EF1a	H3
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	ZIN	Chad_C125_Bagusa	Bagusa/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	ZIN	Chad_C130_MCyclop	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	ZIN	Chad_C134_Manokwari	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	ZIN	Chad_C139_Fawi	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus haddocki* Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF3541	Chad_C64_MW700	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF3548	Chad_C66_Baitabag	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF3549	Chad_C69_MW200	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF3546	Chad_C89_Wanang	New Guinea/K	this study	this study	this study	this study	this study	this study	NA	this study	this study
Cardiodactylus haddocki Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF3547	Chad_C90_Wanang	New Guinea/K	this study	this study	this study	this study	this study	this study	NA	this study	this study
**Cardiodactylus halmahera* Gorochov & Robillard, 2014	ENEO/Leb	ZIN	Chalm_C38_Halmahera	Moluccas/J	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus javarare Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF91	Cjav_C164_Wanang	New Guinea/K	NA	this study	this study	this study	this study	this study	NA	this study	NA
Cardiodactylus javarare Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF60	Cjav_C165_MW200	New Guinea/K	NA	this study	this study	NA	this study	this study	NA	this study	this study
Cardiodactylus javarare Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF103	Cjav_C63_Wanang	New Guinea/K	this study	this study	this study	NA	this study	this study	this study	this study	this study
**Cardiodactylus javarare* Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF3538	Cjav_C78_Baitabag	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus kondoi Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF1555	Ckon_C41_Luzon	the Philippines/D	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus kondoi* Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF3154	Ckon_C42_Luzon	the Philippines/D	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus kondoi Otte, 2007	ENEO/Leb	UPLB-MNH	Ckon_C51_Negros	the Philippines/D	NA	this study	this study	this study	this study	this study	this study	NA	this study
Cardiodactylus kondoi Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF3160	Ckon_C53_Pollilo	the Philippines/D	this study	this study	this study	NA	this study	this study	this study	this study	this study
**Cardiodactylus kotandora* Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3519	Ckot_C65_Flores	Flores island/I	*this study	*this study	*this study	*this study	*this study	NA	*this study	*this study	*this study
**Cardiodactylus* aff. *lampongsi* Robillard & Gorochov, 2014	ENEO/Leb	ZIN	Clam_C105_Sumatra	Sumatra/C	*this study	*this study	NA	*this study	*this study	NA	*this study	*this study	*this study
**Cardiodactylus lampongsi* Robillard & Gorochov, 2014	ENEO/Leb	MNHN-EO-ENSIF3495	Clam_C117_Sumatra	Sumatra/C	*this study	*this study	*this study	NA	*this study	*this study	*this study	*this study	*this study
**Cardiodacylus lombrinjani* Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3675	Clom_C19_Sumbawa	Sumbawa/I	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodacylus lombrinjani Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3673	Clom_C21_Lombok	Sumbawa/I	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus lucus Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF68	Cluc_C61_MW200	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	NA	this study
Cardiodactylus lucus Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF67	Cluc_C62_Wanang	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	NA	this study
**Cardiodactylus lucus* Dong & Robillard, 2016	ENEO/Leb	MNHN-ENSIF94	Cluc_C68_MW700	New Guinea/K	*this study	*this study	*this study	NA	*this study	*this study	*this study	*this study	*this study

1058 *Continued*

1059 *(Continued)*

Species	Taxonomy	Voucher	Molecular code	Geographic area/code	16S	12S	COI	COII	Cytb	18S	28S	EF1a	H3
**Cardiodactylus maai* Otte, 2007	ENEO/Leb	ZIN	Cmaa_C119_Biak	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus maai Otte, 2007	ENEO/Leb	ZIN	Cmaa_C128_Supiori	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus manus Otte, 2007	ENEO/Leb	ZIN	Cman_C127_Rani	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus manus* Otte, 2007	ENEO/Leb	ZIN	Cman_C37_Manokwari	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus muiri Otte, 2007	ENEO/Leb	MNHN-E0-ENSIF4179	Cmui_C158_Kumawa	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	NA
**Cardiodactylus muiri* Otte, 2007	ENEO/Leb	LEN2014-TR617(MZB)	Cmui_C160_Kumawa	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus muria* Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3481	Cmur_C18_Java	Java/G	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus niugini* Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF113	Cniu_C54_Baitabag	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus niugini Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF1583	Cniu_C55_MW200	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus niugini Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF73	Cniu_C70_Wanang	New Guinea/K	this study	this study	this study	this study	this study	this study	NA	this study	this study
Cardiodactylus nobilis Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF102	Cnob_C59_MW1200	New Guinea/K	this study	this study	this study	NA	this study	this study	this study	this study	this study
**Cardiodactylus nobilis* Dong & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF131	Cnob_C67_MW1200	New Guinea/K	*this study	*this study	*this study	NA	*this study	*this study	*this study	*this study	*this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	ZIN	Cnov_C141_Queensland	Australia/L	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	ZIN > MNHN-EO-ENSIF38	Cnov_C142_NewGeorgia	Solomon Island/L	this study	this study	this study	this study	this study	NA	this study	this study	this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	ZIN	Cnov_C143_Mhagen	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-E0-ENSIF4384	Cnov_C151_TritonBay	New Guinea/K	this study	this study	this study	this study	this study	this study	NA	this study	NA
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF1921	Cnov_C1_Lifou	New Caledonia/L	this study	this study	KU70556 4	KU70555 2	this study	KU70558 8	KU70558 8	KU70561 4	KU70559 7
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MZB	Cnov_C27_Biak	New Guinea/K	this study	this study	this study	this study	this study	NA	NA	NA	NA
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF2030	Cnov_C2_Peavot	Vanuatu/L	JF972521	JF972506	KU70556 3	KU70555 1	JF972490	JF972537	KR90350 0	KU70561 3	KR90315 1
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-ENSIF2038	Cnov_C3_Oubatche	New Caledonia/L	JF972520	JF972504	this study	this study	JF972488	JF972535	this study	this study	this study
**Cardiodactylus novaeguineae* (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF1466	Cnov_C49_Wewec	New Caledonia/L	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF1512	Cnov_C5_NewIreland	New Guinea/K	this study	NA	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF3550	Cnov_C60_Loloata	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	CSIRO	Cnov_C6_Queensland	Australia/L	this study	this study	this study	NA	this study	this study	this study	NA	this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF2032	Cnov_C71_Penaoru	Vanuatu/L	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF1925	Cnov_C72_Luganville	Vanuatu/L	this study	this study	this study	this study	this study	this study	this study	this study	this study

1060 *Continued*

1061

1062 *(Continued)*

Species	Taxonomy	Voucher	Molecular code	Geographic area/code	16S	12S	COI	COII	Cytb	18S	28S	EF1a	H3
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF2020	Cnov_C73_BigBay	Vanuatu/L	this study	this study	this study	NA	this study	this study	NA	NA	NA
Cardiodactylus novaeguineae (Haan, 1842)	ENEO/Leb	MNHN-EO-ENSIF1474	Cnov_C87_Tanna	Vanuatu/L	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus oeroe* Robillard, 2014	ENEO/Leb	RMNH	Coer_C43_Sulawesi	Sulawesi/H	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus palawan Gorochov, 2014	ENEO/Leb	ZIN	Cpal_C106_Palawan	Palawan/E	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus palawan Gorochov, 2014	ENEO/Leb	ZIN	Cpal_C107_Palawan	Palawan/E	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus palawan* Gorochov, 2014	ENEO/Leb	MNHN-EO-ENSIF1506	Cpal_C91_Turtle	Palawan/E	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus palawan Gorochov, 2014	ENEO/Leb	MNHN-EO-ENSIF1468	Cpal_C92_Palawan	Palawan/E	this study	this study	this study	this study	this study	this study	this study	this study	this study
Cardiodactylus palawan Gorochov, 2014	ENEO/Leb	MNHN-EO-ENSIF1497	Cpal_C93_Palawan	Palawan/E	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus pelagus* Otte, 2007	ENEO/Leb	ZIN	Cpel_C109_Borneo	Borneo/F	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus quatei Otte, 2007	ENEO/Leb	ZIN	Cqua_C138_Fawi	New Guinea/K	this study	this study	this study	this study	this study	NA	this study	this study	this study
**Cardiodactylus quatei* Otte, 2007	ENEO/Leb	2014-TR367	Cqua_C156_Lobo	New Guinea/K	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus quatei Otte, 2007	ENEO/Leb	2014-TR437	Cqua_C157_Urisa	New Guinea/K	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus reticulatus* Gorochov, 2014	ENEO/Leb	ZIN	Cret_C33_Tioman	Malaysia/B	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus riga* Otte, 2007	ENEO/Leb	UPLB-MNH	Crig_C77_Burauen	the Philippines/D	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus riga Otte, 2007	ENEO/Leb	MNHN-EO-ENSIF3352	Crig_C83_LeyteS	the Philippines/D	this study	this study	this study	this study	this study	this study	this study	this study	this study
**Cardiodactylus rufidulus* Saussure, 1878	ENEO/Leb	NHM-UK-BM1984-865	Cruf_C44_Guadalcanal	Solomon/L	*this study	*this study	NA	*this study	*this study	NA	*this study	NA	*this study
**Cardiodactylus singapura* Robillard, 2011	ENEO/Leb	MNHN-EO-ENSIF2759	Csin_C12_Bukit	Singapour/B	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus sumba* Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3661	Csum_C20_Sumba	Sumba/I	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus sumba Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3658	Csum_C47_Sumba	Sumba/I	this study	this study	NA	this study	this study	this study	this study	this study	this study
**Cardiodactylus tangkoko* Gorochov, 2014	ENEO/Leb	ZIN	Ctang_C116_Sulawesi	Sulawesi/H	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus tankara* Robillard, 2009	ENEO/Leb	MNHN-ENSIF2416	Ctank_C8_Espiritu	Vanuatu/L	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
**Cardiodactylus tello* Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF160	Ctel_C171_Siberut	Sumatra/C	*this study	*this study	NA	NA	NA	NA	NA	NA	NA
Cardiodactylus thailandia Robillard, 2011	ENEO/Leb	ZIN	Ctha_C111_Thailand	Thailand/A	this study	this study	this study	this study	this study	this study	NA	this study	this study
**Cardiodactylus thailandia* Robillard, 2011	ENEO/Leb	ZIN	Ctha_C112_Thailand	Thailand/A	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study	*this study
Cardiodactylus variegatus Gorochov & Robillard, 2014	ENEO/Leb	ZIN	Cvar_C113_Borneo	Borneo/F	this study	this study	this study	NA	this study	NA	this study	this study	this study
Cardiodactylus variegatus Gorochov & Robillard, 2014	ENEO/Leb	MNHN-EO-ENSIF3510	Cvar_C28_Borneo	Borneo/F	this study	this study	NA	this study	this study	this study	NA	this study	this study
**Cardiodactylus variegatus* Gorochov & Robillard, 2014	ENEO/Leb	RMNH	Cvar_C52_Borneo	Borneo/F	*this study	*this study	*this study	*this study	*this study	*this study	NA	*this study	*this study
Eneoptera guyanensis Chopard, 1931	ENEO/Ene	MNHN-EO-ENSIF2741	Egu	French Guiana	AY905301	AY905272	JX897404	KU705553	AY905331	AY905331	KU705581	JX897495	JX897547
Eurepini sp.	ENEO/Eur	MNHN-EO-ENSIF3155	Eursp	Australia	KR903674	KR903834	KU705565	KU705554	KR903331	KR904028	KR903503	NA	KR903153
Gryllus bimaculatus De Geer, 1773	GRYL/Gry	MNHN-EO-ENSIF3524/3404	Gbi	France (Laboratory strain)	AF248685	AY905292	NA	KU705555	AF248659	AF514509	KR903002	NA	KR903154

1063 *Continued*

1064 *(Continued)*

Species	Taxonomy	Voucher	Molecular code	Geographic area/code	16S	12S	COI	COII	Cytb	18S	28S	EF1a	H3
Pixibinthus sonicus Anso & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF99	L59LeNCRBa	New Caledonia	KU705530	KU708012	KU705577	NA	KU705538	KU705546	KU705592	NA	KU705608
Pixibinthus sonicus Anso & Robillard, 2016	ENEO/Leb	MNHN-EO-ENSIF83	L66LeNCFN2	New Caledonia	KU705532	KU708015	KU705574	NA	KU705540	KU705547	KU705590	NA	KU705605
Lebinthus luae Robillard & Tan, 2013	ENEO/Leb	MNHN-EO-ENSIF2740	LbiS1	Singapore	JF972524	KR904017	KU705567	KU705557	JF972493	KR904199	KR903665	KU705615	KR903321
Lebinthus lifouensis Desutter-Grandcolas, 1997	ENEO/Leb	MNHN-EO-ENSIF1346	Lli	New Caledonia	AY905309	AY905279	KU705566	KU705556	AY905364	AY905336	KU705583	NA	KU705599
Lebinthus nattawa Robillard, 2009	ENEO/Leb	MNHN-EO-ENSIF2564	LnaN	Vanuatu	JF972525	JF972510	KU705568	KU705558	JF972494	JF972541	KU705584	NA	KU705600
Lebinthus sp. Stål, 1877	ENEO/Leb	MNHN-EO-ENSIF117	Lpng1	Papua New Guineae	JF972528	JF972513	KU715289	KU715288	KU705536	JF972544	KU715290	KU715292	KU715291
Lebinthus santoensis Robillard, 2009	ENEO/Leb	MNHN-EO-ENSIF2484	LsaPe	Vanuatu	KU705528	KU708011	KU705569	NA	KU705535	KU705543	KU705585	NA	KU705601
Lebinthus santoensis Robillard, 2009	ENEO/Leb	MNHN-EO-ENSIF2437	LsaV	Vanuatu	JF972527	JF972511	JX897405	JX897441	JF972495	JF972542	JX897467	NA	JX897548
Lebinthus villemantae Robillard, 2010	ENEO/Leb	MNHN-EO-ENSIF2739	Lvil	Sulawesi	JF972526	JF972512	KU705570	KU705559	JF972496	JF972543	KU705586	NA	KU705602
Nisitrus vittatus (Haan, 1842)	ENEO/Nis	MNHN-EO-ENSIF2742	NviS	Singapore	this study	this study	KU705572	NA	this study	AY905340	KR903667	JN887883	JX897546
Nisitrus vittatus (Haan, 1842)	ENEO/Nis	MNHN-EO-ENSIF3134	Nvith	Malaysia	AY905314	AY905284	KU705571	NA	AY905369	AY905340	KU705587	NA	KU705603
Paranisitra longipes Chopard, 1925	ENEO/Nis	MNHN-EO-ENSIF3157	Plo2	the Philippines	KR903827	KR904020	KU715287	NA	NA	KR904202	KR903668	NA	KR903325
Swezwilderia sp.	ENEO/Leb	MNHN-EO-ENSIF2737	Ssp	Fiji	JF972529	JF972514	KU705579	NA	JF972498	JF972545	NA	NA	KR903327
Xenogryllus marmoratus Bolívar, 1890	ENEO/Xen	MNHN-EO-ENSIF3161	Xma2	Japan	KR903830	KR904024	NA	KU705560	KR903491	KR904206	NA	KU705610	KR903329

1065

1066

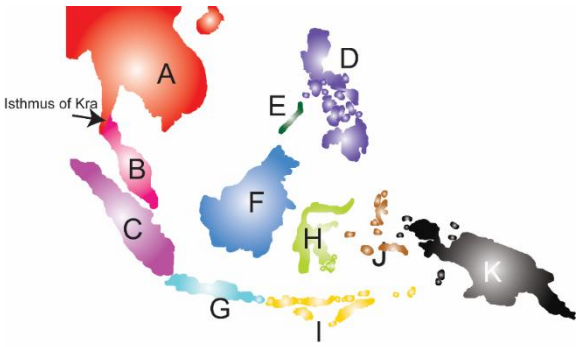
1067

Appendix S2. Time-stratified biogeographical standards implemented in BioGeoBEARS analyses.

1. Definition of geographical areas

The map of SEA and surrounding regions was subdivided into 12 areas for the purpose of the biogeographic analyses: A: Indochina, B: Thai-Malay Peninsula, C: Sumatra, D: Philippines (except Palawan island), E: Palawan, F: Borneo, G: Java, H: Sulawesi, I: Lesser Sunda Islands, J: Moluccas, K: New Guinea, L: Southwest pacific.

Here Indochina corresponds to the Southeast Asia mainland, including the northern Ryukyu Islands (not showed on the simplified map), southeastern China (including Taiwan), Vietnam and northern Thailand. The boundary between Indochina and the Thai-Malay Peninsula corresponds to the Isthmus of Kra which intersects the north and south of Thailand. Southwest pacific is not shown on the map, but it includes the archipelagoes located east from New Guinea.



2. Dispersal rates:

Matrix of scaling factors (between 0 and 1.0) for dispersal rates between areas were constructed according to the geographical position of the areas during a given time slice, interpreting greater distances and/or the extent of geographical barriers (inland river, shallow sea, deep sea) as being inversely proportional to expected rates (Condamine et al., 2013; Toussaint and Balke, 2016). Dispersal rates were set as follows: no barrier between two connected areas: 1.0; minor barrier between two

areas, *e.g*., inland river, shallow sea water: 0.75; large barrier between two areas, *e.g*., deep sea, strait, high mountain: 0.25; long-distance dispersal between two areas: 0.01; absence of one area at a given time: 0. In addition, whenever two areas were separated by one area, dispersal rates were modified as follows: 0.5 when not involving a minor barrier; 0.25 when the area in between includes a minor barrier; and 0.1 when the area in between encompass a major barrier.

3. Time slices and dispersal multiplier matrices:

Collision events and the resumption of subduction that happened during the Cenozoic had a great impact on SEA terrestrial biogeographic patterns and on the current archipelago setting (Hall, 2013). Time-stratified biogeographical models were defined according to three distinct time slices following three major paleogeological events in SEA:

3.1. Time slice 1 (45-25 Ma): Eocene to Mid Oligocene

The India-Asia collision was related to the 45 Ma plate reorganization during the Eocene, resulting in Australia's northward moving. The Makassar Straits became a wide marine barrier and separated west Sulawesi from core SEA. The Reed Bank-Palawan-Mindoro Block separated from the South China margin by the opening of the South China Sea during the Oligocene (Hall, 2002). However, at that period Palawan was still a part of Southeast China Mainland. In the Eocene, the southern part of the Philippines, along with the Halmahera Arc, formed part of an arc system which extended into the Southwest pacific at an equatorial/southern hemisphere position. The northern Philippines, principally Luzon, was connected via the Sulu–Cagayan Arc to northern Borneo and Sabah before the Miocene (Hall, 2002). Despite

marked geological activities (collision and subduction events), oceanic islands in the Wallacean region and New Guinea did not emerge from the sea (Hall, 2013). During this period, some islands or parts of current island in the Southwest pacific began to emerge, which is for example the case of New Caledonia that emerged *ca.* 37 Ma (Grandcolas et al., 2008) or the Fiji islands, which were formed by subduction with the oldest rocks being from a volcanic island-arc of late Eocene (37-34 Ma; Neall and Trewick, 2008).

Dispersal rates between well-connected areas (no barrier) were set to 1.0; this corresponds to dispersal events between Indochina [A] and the Thai-Malay Peninsula [B], Indochina [A] and Palawan [E], Sumatra [C] and Java [G], and Borneo [F] and Java [G]. Dispersal areas between slightly less connected areas (minor barrier) were assigned a rate of 0.75 (for example between the Thai-Malay Peninsula [B] and Palawan [E]; see below for the remaining pairings). Dispersal rates between Palawan [E] and Borneo [F] and the Philippines (minus Palawan) [D] and Borneo [F] were set to 0.5 considering the presence of small continuous fragments of islands between these areas. As explained above whenever two areas were separated by one barrier, lower dispersal rates were used (0.25 and 0.1; see below for all corresponding pairings). Finally, dispersals from the Southwest Pacific to Southeast Asia were considered as long-distance dispersal events (rate of 0.01; see below for all corresponding pairings).

Matrix of scaling factors for dispersal rates in time slice 1 (45-25 Ma)

	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.5	0.25	1	0.5	0.25	0.1	0	0	0	0.01
B		1	1	0.25	0.75	0.75	0.75	0.1	0	0	0	0.01
C			1	0.25	0.75	1	1	0.1	0	0	0	0.01
D				1	0.1	0.5	0.25	0.1	0	0	0	0.01
E					1	0.75	0.5	0.1	0	0	0	0.01
F						1	1	0.25	0	0	0	0.01
G							1	0.25	0	0	0	0.01

H								1	0	0	0	0.01
I									0	0	0	0
J										0	0	0
K											0	0
L												1

1142

1143

1144

1145 **3.2. Time slice 2 (25-15 Ma): Mid Oligocene to Mid Miocene**

1146 Between the Eocene and Early Miocene, the southern and southeast margin of
1147 Sundaland were constituted of volcanic arcs (Hall, 2002). The volcanic arcs located
1148 on the southern and southeast margin of Sundaland were active and these events
1149 created a minor inner barrier within Sundaland. The Palawan Continental Terrane
1150 was isolated from the mainland because of the opening of Southeast China Sea and
1151 it drifted towards Borneo (Walia et al., 2012). New Guinea began to emerge,
1152 probably as small islands along the northern edge, but most of the areas, which is
1153 nowadays a mountain range, were covered by shallow sea (Hall, 2009, 2013).
1154 Oceanic islands emerged from the sea and rift deformation formed some parts of the
1155 Philippines and Sulawesi. In addition, deformation and uplift began in Borneo, Java
1156 and Sulawesi.

1157 Therefore, there are several minor barriers within Sundaland (dispersal rate of
1158 0.75). Long-distance dispersal was assigned to dispersal events between SEA and
1159 New Guinea, as well as between SEA and Southwest pacific. However, the
1160 emergence of continuous small fragments / islands between Sulawesi and New
1161 Guinea probably increased the chances for dispersal events (dispersal rate of 0.75).

1162

1163 Matrix of scaling factors for dispersal rates in time slice 2 (25-15 Ma).

	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.5	0.1	0.25	0.75	0.25	0.1	0	0	0.01	0.01

B		1	0.75	0.25	0.5	0.75	0.25	0.1	0	0	0.01	0.01
C			1	0.25	0.5	0.75	0.75	0.1	0	0	0.01	0.01
D				1	0.25	0.5	0.25	0.5	0	0	0.1	0.01
E					1	0.75	0.25	0.1	0	0	0.01	0.01
F						1	0.75	0.25	0	0	0.01	0.01
G							1	0.1	0	0	0.01	0.01
H								1	0	0	0.75	0.01
I									0	0	0	0.01
J										0	0	0.01
K											1	0.1
L												1

3.3. Time slice 3 (15-0 Ma): Mid Miocene to present-day

The Australian margin drifted from the continent, forming the Bird's Head region of New Guinea, (Sula Spur promontory). By the Mid Miocene, about 15 Ma, the Sula Spur promontory collided with the margin of SEA in Sulawesi, forming the region now known as Wallacea, including the Philippines, the Moluccas, a completed Sulawesi and the Lesser Sunda Islands (Hall, 2009; Spakman and Hall 2010). The geological pattern in SEA did not change so much since these geological events. Thus, dispersal rates within Sundaland were assigned the same values as before (time slice 25-15 Ma). Considering the persistence of strings of small islands involved in the origin of the Wallacean region, dispersal rates between the Philippines and Palawan, the Philippines and Borneo, the Philippines and the Moluccas, and the Moluccas and New Guinea were set to 0.75.

Matrix of scaling factors for dispersal rates in time slice 3 (15 Ma to present day).

	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.5	0.1	0.25	0.75	0.1	0.1	0.01	0.01	0.01	0.01
B		1	0.75	0.1	0.25	0.75	0.25	0.1	0.1	0.01	0.01	0.01
C			1	0.25	0.25	0.75	0.75	0.1	0.25	0.01	0.01	0.01
D				1	0.75	0.75	0.25	0.5	0.25	0.75	0.25	0.01
E					1	0.75	0.25	0.1	0.25	0.1	0.01	0.01
F						1	0.75	0.25	0.5	0.1	0.01	0.01
G							1	0.1	0.75	0.1	0.01	0.01
H								1	0.5	0.75	0.25	0.01
I									1	0.5	0.25	0.01
J										1	0.75	0.25
K											1	0.5

L												1

4. Alternative dispersal multipliers

We implemented alternative sets of dispersal multiplier to test the robustness of our biogeographical analyses and limit the impact of the arbitrariness of the probability attributed to each event. Three more simplified sets of multipliers were tested using the following changes: except the probabilities of 0, 0.01, 0.1 and 1 that remained unchanged, the events with probabilities of 0.25, 0.50 and 0.75 were given homogenous values (either 0.15, 0.50 or 0.75) in the alternative sets of multipliers. The three resulting simplified sets of multipliers are presented below.

Simplified set of dispersal multipliers 1 (0, 0.01, 0.1, 0.25, 1.0) – probabilities of 0.25, 0.5 and 0.75 in the base matrix are given a probability of 0.25.

time slice 1: 45-25 Ma												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.25	0.25	1	0.25	0.25	0.1	0	0	0	0.01
B		1	1	0.25	0.25	0.25	0.25	0.1	0	0	0	0.01
C			1	0.25	0.25	1	1	0.1	0	0	0	0.01
D				1	0.1	0.25	0.25	0.1	0	0	0	0.01
E					1	0.25	0.25	0.1	0	0	0	0.01
F						1	1	0.25	0	0	0	0.01
G							1	0.25	0	0	0	0.01
H								1	0	0	0	0.01
I									0	0	0	0
J										0	0	0
K											0	0

L												1
time slice 2: 25-15 Ma												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.25	0.1	0.25	0.75	0.25	0.1	0	0	0.01	0.01
B		1	0.25	0.25	0.25	0.75	0.25	0.1	0	0	0.01	0.01
C			1	0.25	0.25	0.75	0.75	0.1	0	0	0.01	0.01
D				1	0.25	0.25	0.25	0.25	0	0	0.1	0.01
E					1	0.25	0.25	0.1	0	0	0.01	0.01
F						1	0.25	0.25	0	0	0.01	0.01
G							1	0.1	0	0	0.01	0.01
H								1	0	0	0.25	0.01
I									0	0	0	0.01
J										0	0	0.01
K											1	0.1
L												1
time slice 3: 15 Ma to present												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.25	0.1	0.25	0.25	0.1	0.1	0.01	0.01	0.01	0.01
B		1	0.25	0.1	0.25	0.25	0.25	0.1	0.1	0.01	0.01	0.01
C			1	0.25	0.25	0.25	0.75	0.1	0.25	0.01	0.01	0.01
D				1	0.25	0.25	0.25	0.25	0.25	0.75	0.25	0.01
E					1	0.25	0.25	0.1	0.25	0.1	0.01	0.01
F						1	0.25	0.25	0.25	0.1	0.01	0.01
G							1	0.1	0.25	0.1	0.01	0.01
H								1	0.25	0.25	0.25	0.01
I									1	0.25	0.25	0.01
J										1	0.25	0.25
K											1	0.25
L												1

Simplified set of dispersal multipliers 2 (0, 0.01, 0.1, 0.5, 1.0) – probabilities of 0.25, 0.5 and 0.75 in the base matrix are given a probability of 0.5.

time slice 1: 45-25 Ma												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.5	0.5	1	0.5	0.5	0.1	0	0	0	0.01
B		1	1	0.5	0.5	0.5	0.5	0.1	0	0	0	0.01
C			1	0.5	0.5	1	1	0.1	0	0	0	0.01
D				1	0.1	0.5	0.5	0.1	0	0	0	0.01
E					1	0.5	0.5	0.1	0	0	0	0.01
F						1	1	0.5	0	0	0	0.01
G							1	0.5	0	0	0	0.01
H								1	0	0	0	0.01
I									0	0	0	0
J										0	0	0
K											0	0
L												1
time slice 2: 25-15 Ma												

	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.5	0.1	0.5	0.5	0.5	0.1	0	0	0.01	0.01
B		1	0.5	0.5	0.5	0.5	0.5	0.1	0	0	0.01	0.01
C			1	0.5	0.5	0.5	0.5	0.1	0	0	0.01	0.01
D				1	0.5	0.5	0.5	0.5	0	0	0.1	0.01
E					1	0.5	0.5	0.1	0	0	0.01	0.01
F						1	0.5	0.5	0	0	0.01	0.01
G							1	0.1	0	0	0.01	0.01
H								1	0	0	0.5	0.01
I									0	0	0	0.01
J										0	0	0.01
K											1	0.1
L												1
time slice 3: 15 Ma to present												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.5	0.1	0.5	0.5	0.1	0.1	0.01	0.01	0.01	0.01
B		1	0.5	0.1	0.5	0.5	0.5	0.1	0.1	0.01	0.01	0.01
C			1	0.5	0.5	0.5	0.5	0.1	0.5	0.01	0.01	0.01
D				1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.01
E					1	0.5	0.5	0.1	0.5	0.1	0.01	0.01
F						1	0.5	0.5	0.5	0.1	0.01	0.01
G							1	0.1	0.5	0.1	0.01	0.01
H								1	0.5	0.5	0.5	0.01
I									1	0.5	0.5	0.01
J										1	0.5	0.5
K											1	0.5
L												1

Simplified set of dispersal multipliers 3 (0, 0.01, 0.1, 0.75, 1.0) – probabilities of 0.25, 0.5 and 0.75 in the base matrix are given a probability of 0.75.

time slice 1: 45-25 Ma												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.75	0.75	1	0.75	0.75	0.1	0	0	0	0.01
B		1	1	0.75	0.75	0.75	0.75	0.1	0	0	0	0.01
C			1	0.75	0.75	1	1	0.1	0	0	0	0.01
D				1	0.1	0.75	0.75	0.1	0	0	0	0.01
E					1	0.75	0.75	0.1	0	0	0	0.01
F						1	1	0.75	0	0	0	0.01
G							1	0.75	0	0	0	0.01
H								1	0	0	0	0.01
I									0	0	0	0
J										0	0	0
K											0	0
L												1
time slice 2: 25-15 Ma												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.75	0.1	0.75	0.75	0.75	0.1	0	0	0.01	0.01

B		1	0.75	0.75	0.75	0.75	0.75	0.1	0	0	0.01	0.01
C			1	0.75	0.75	0.75	0.75	0.1	0	0	0.01	0.01
D				1	0.75	0.75	0.75	0.75	0	0	0.1	0.01
E					1	0.75	0.75	0.1	0	0	0.01	0.01
F						1	0.75	0.75	0	0	0.01	0.01
G							1	0.1	0	0	0.01	0.01
H								1	0	0	0.75	0.01
I									0	0	0	0.01
J										0	0	0.01
K											1	0.1
L												1
time slice 3: 15 Ma to present												
	A	B	C	D	E	F	G	H	I	J	K	L
A	1	1	0.75	0.1	0.75	0.75	0.1	0.1	0.01	0.01	0.01	0.01
B		1	0.75	0.1	0.75	0.75	0.75	0.1	0.1	0.01	0.01	0.01
C			1	0.75	0.75	0.75	0.75	0.1	0.75	0.01	0.01	0.01
D				1	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.01
E					1	0.75	0.75	0.1	0.75	0.1	0.01	0.01
F						1	0.75	0.75	0.75	0.1	0.01	0.01
G							1	0.1	0.75	0.1	0.01	0.01
H								1	0.75	0.75	0.75	0.01
I									1	0.75	0.75	0.01
J										1	0.75	0.75
K											1	0.75
L												1

References

Condamine, F.L., Sperling, F.A.H., Kergoat, G.J., 2013. Global biogeographical pattern of swallowtail diversification demonstrates alternative colonization routes in the Northern and Southern hemispheres. J. Biogeogr. 40, 9–23.

Dong, J., Robillard, T., 2016. The *Cardiodactylus* crickets from Eastern New Guinea, with description of five new species (Orthoptera: Gryllidae: Eneopterinae), in Robillard, T., Legendre, F., Villemant, C., Leponce, M. (Eds.), Insects of Mount Wilhelm, Papua New Guinea. Muséum national d'Histoire naturelle, Paris, pp. 203–258.

Grandcolas, P., Murienne, J., Robillard, T., Desutter-Grandcolas, L., Jourdan, H., Guilbert, E., Deharveng, L., 2008. New Caledonia: a very old Darwinian island? Phil. Trans. R. Soc. B. 363, 3309–3317.

Hall, R., 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SE Pacific: computer-based reconstructions, model and animations. J. Asian Earth Sci. 20, 353–431.

Hall, R., 2009. Southeast Asia's changing palaeogeography. Blumea. 54, 148–161.

Hall, R., 2013. The palaeogeography of Sundaland and Wallacea since the Late Jurassic. J Limnol 72, 1–17.

Neall, V.E., Trewick, S.A., 2008. The age and origin of the Pacific islands: a geological overview. Phil. Trans. R. Soc. B 363, 3293–3308.

Robillard, T., Gorochov, A.V., Poulain, S., Suhardjono, Y.R., 2014. Revision of the cricket genus *Cardiodactylus* (Orthoptera, Eneopterinae, Lebinthini): the species from both sides of the Wallace line, with description of 25 new species. Zootaxa, 3854, 1–104.

Spakman, W., Hall, R., 2010. Surface deformation and slab-mantle interaction during Banda arc subduction rollback. Nat. Geosci. 3, 562–566.

Toussaint, E.F.A., Balke, M., 2016. Historical biogeography of *Polyura* butterflies in the oriental Palaeotropics: trans-archipelagic routes and South Pacific island hopping. J. Biogeogr. 43, 1560–1572.

Walia, M., Knittel, U., Suzuki, S. Chung, S., Pena, R.E., 2012. No Paleozoic metamorphics in Palawan (the Philippines)? Evidence from single grain U-Pb dating of detrital zircons. J. Asian Earth Sci. 52, 134–145.

Appendix S3. Summary of ancestral area estimation with different models and dispersal rate multipliers with time slices (**Table S3a**) and without time slices (**Table S3b**). The gray line highlights the model with the highest statistical support.

Table S3a.

Multipliers sets	max_range_size	Speciation model	LnL	Parameter estimates	
				d	e
(0, 0.01,0.1,0.25,0.5,0.75,1.0)	2	DEC	-128.05	0.024	0.0175
		DIVALIKE	-126.55	0.024	0.0137
		BAYAREALIKE	-140.47	0.036	0.0445
	3	DEC	-131.32	0.016	0.0162
		DIVALIKE	-131.19	0.018	0.0139
		BAYAREALIKE	-143.99	0.019	0.0537
	4	DEC	-131.55	0.015	0.0154
		DIVALIKE	-131.51	0.018	0.0134
		BAYAREALIKE	-145.04	0.016	0.0584
(0, 0.01, 0.1,0.25, 1.0)	2	DEC	-128.99	0.038	0.0168
		DIVALIKE	-127.65	0.039	0.0133
		BAYAREALIKE	-142.26	0.055	0.0421
	3	DEC	-132.57	0.026	0.0157

		DIVALIKE	-132.70	0.029	0.0135
		BAYAREALIKE	-146.60	0.030	0.0538
	4	DEC	-132.86	0.024	0.0150
		DIVALIKE	-132.99	0.028	0.0131
		BAYAREALIKE	-147.61	0.025	0.0594
(0, 0.01, 0.1, 0.5, 1.0)	2	DEC	-130.83	0.022	0.0172
		DIVALIKE	-129.53	0.022	0.0139
		BAYAREALIKE	-143.73	0.032	0.0436
	3	DEC	-134.27	0.015	0.0160
		DIVALIKE	-134.48	0.017	0.0141
		BAYAREALIKE	-147.39	0.017	0.0545
	4	DEC	-134.42	0.014	0.0153
		DIVALIKE	-134.80	0.016	0.0136
		BAYAREALIKE	-148.17	0.014	0.0602
(0, 0.01, 0.1, 0.75, 1.0)	2	DEC	-132.14	0.015	0.0175
		DIVALIKE	-130.88	0.016	0.0143
		BAYAREALIKE	-144.67	0.022	0.0446
	3	DEC	-133.82	0.010	0.0158
		DIVALIKE	-135.75	0.012	0.0144
		BAYAREALIKE	-147.93	0.012	0.0549
	4	DEC	-135.47	0.009	0.0155
		DIVALIKE	-136.10	0.011	0.0140
		BAYAREALIKE	-148.59	0.010	0.0607

Table S3b (without time slices)

Multipliers sets	max_range_size	Speciation model	LnL	Parameter estimates	
				d	e
(0, 0.01,0.1,0.25,0.5,0.75,1.0)	2	DEC	-111.70	0.016	0.0122
		DIVALIKE	-105.83	0.015	0.0066
		BAYAREALIKE	-131.77	0.005	0.0342
	3	DEC	-113.29	0.012	0.0109
		DIVALIKE	-108.70	0.012	0.0056
		BAYAREALIKE	-138.14	0.002	0.0358
	4	DEC	-113.23	0.011	0.0093
		DIVALIKE	-120.49	0.003	0.0065
		BAYAREALIKE	-138.69	0.002	0.0369

Appendix S4. Sampling of species groups for the BAMM analyses.

Because *C. javarere* and *C. kolombangara* show intermediate characters between *Novaeguineae* and *Efordi* species groups (Dong and Robillard, 2016), these two species form a third species group, hereby referred as the *Javarere* species group. Therefore, for the purpose of the BAMM analyses we considered that there are 60 species in *Novaeguineae* species group, 20 in *Efordi* species group and two in *Javarere* species group. In total, there are 35 species of the *Novaeguineae* species

group and one of the *Efordi* species group in SEA. The number of species in each species group and the sampling fraction in BAMM analyses were given in TableS4.

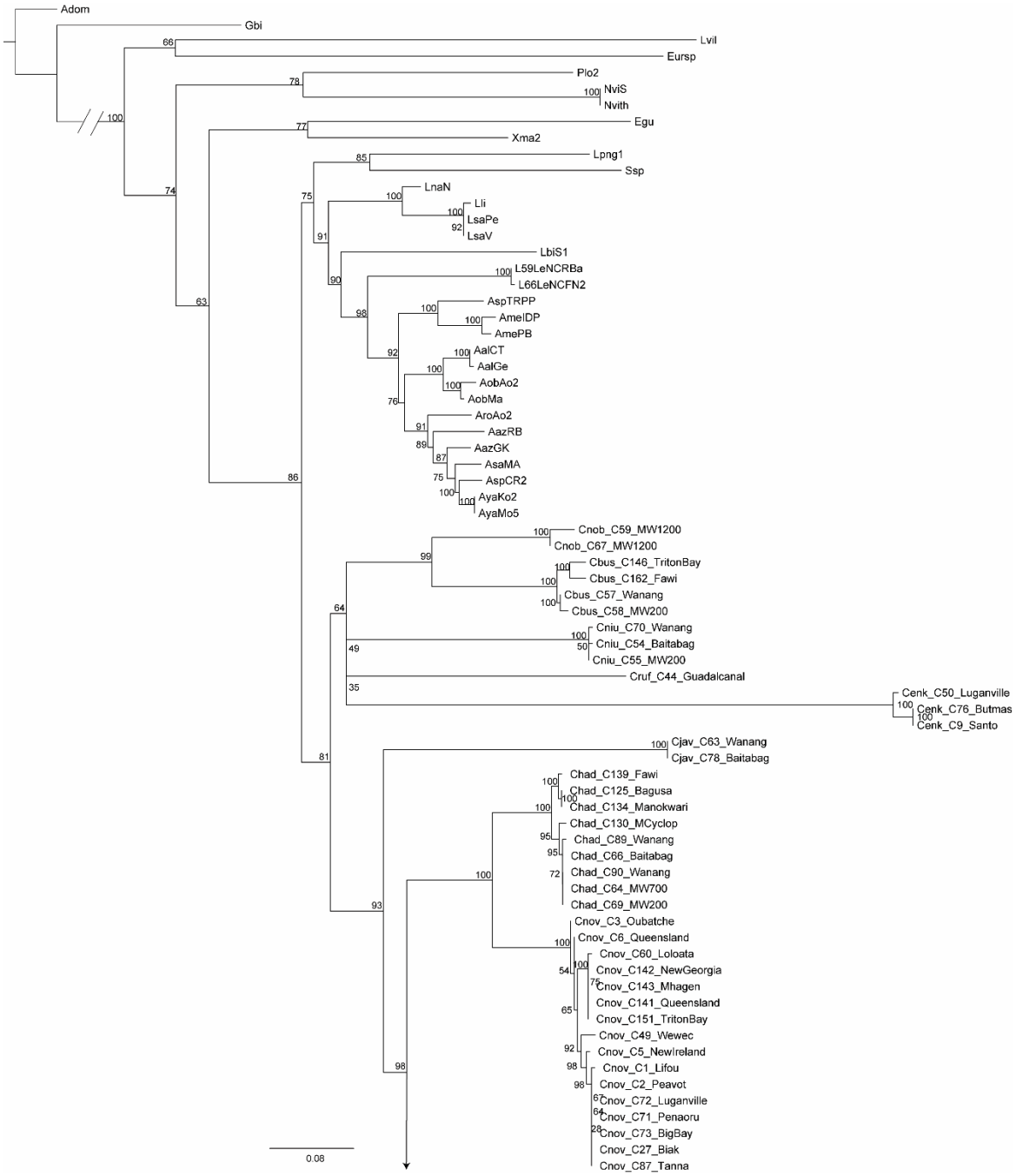
Table S4. Species sampling fraction in BAMM analyses.

Species group	Total numbers	Numbers in SEA	Sampling fraction
Novaeguineae	60/35	35/28	0.58
Efordi	20/5	1/0	0.25
'Javarere'	2/1	0/0	0.5

Note: Numbers (*/*) in the column named "Total numbers" correspond to the total numbers in each species group (left of sign /) and the species numbers in this study (right of sign /), respectively. Numbers (*/*) in the column named "Numbers in SEA" correspond the total numbers in each species group distributed in SEA (left of sign /) and the species numbers in this study in SEA (right of sign /), respectively. "Sampling fraction" corresponds the proportion of the number of species sampled in this study vs total numbers of each species group.

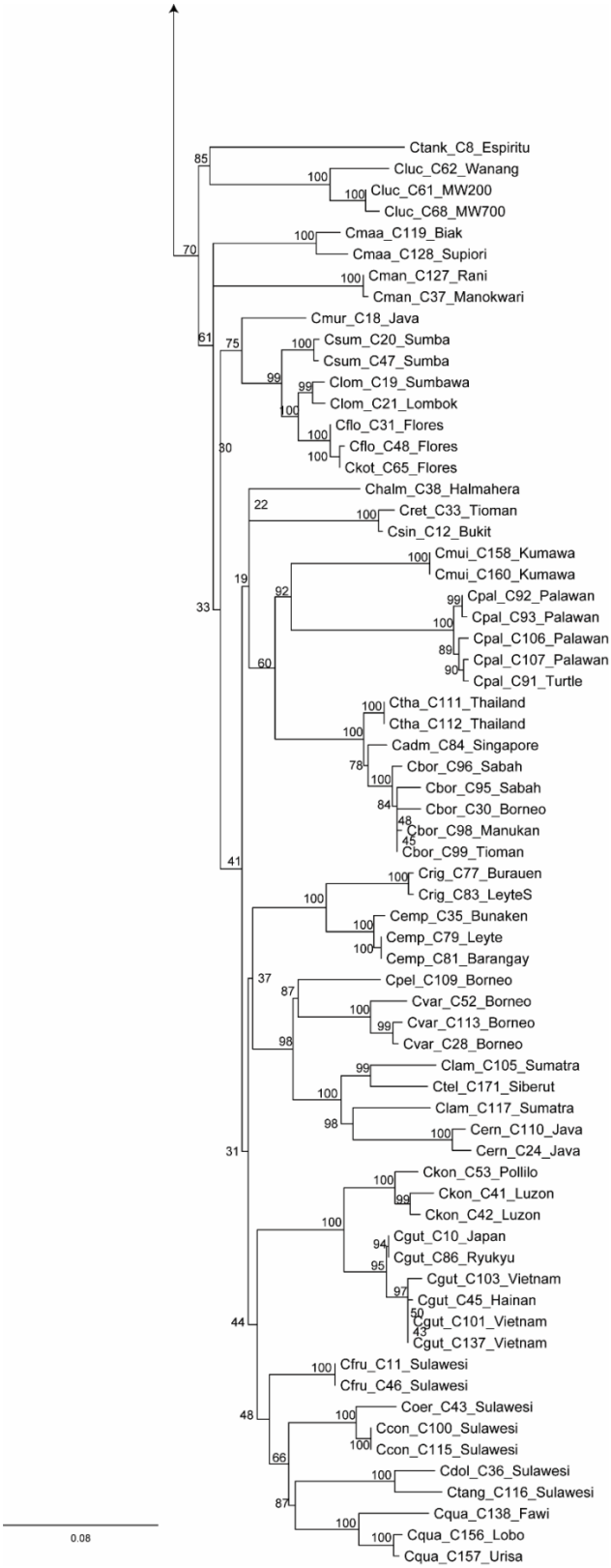
Appendix S5. Gene trees inferred for each marker: 16S (a), 12S (b), COI (c), COII (d), Cytb (e), 18S (f), 28S (g), EF1a (h), H3 (i).

16S (a):

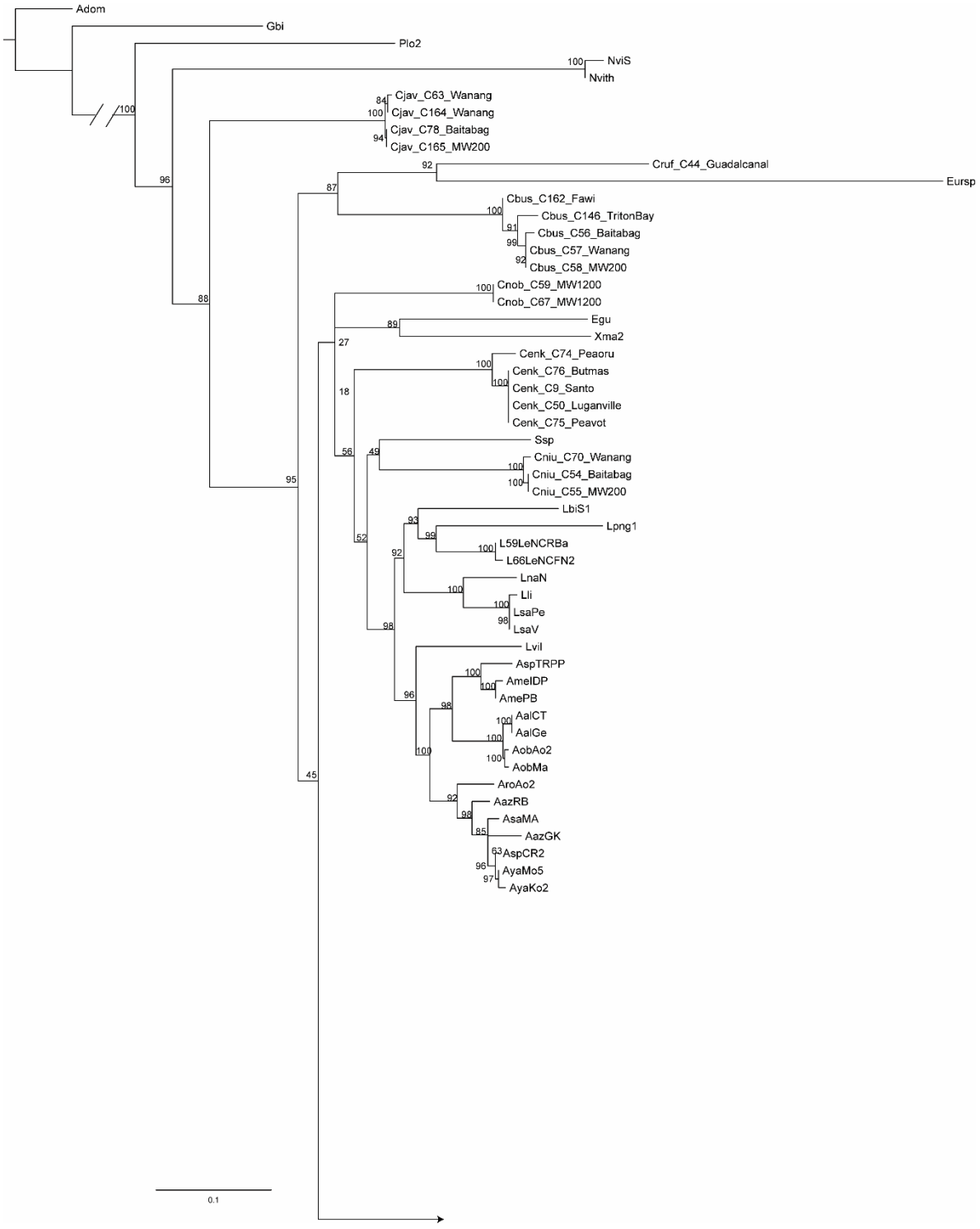


(continued)

16S (a): continued

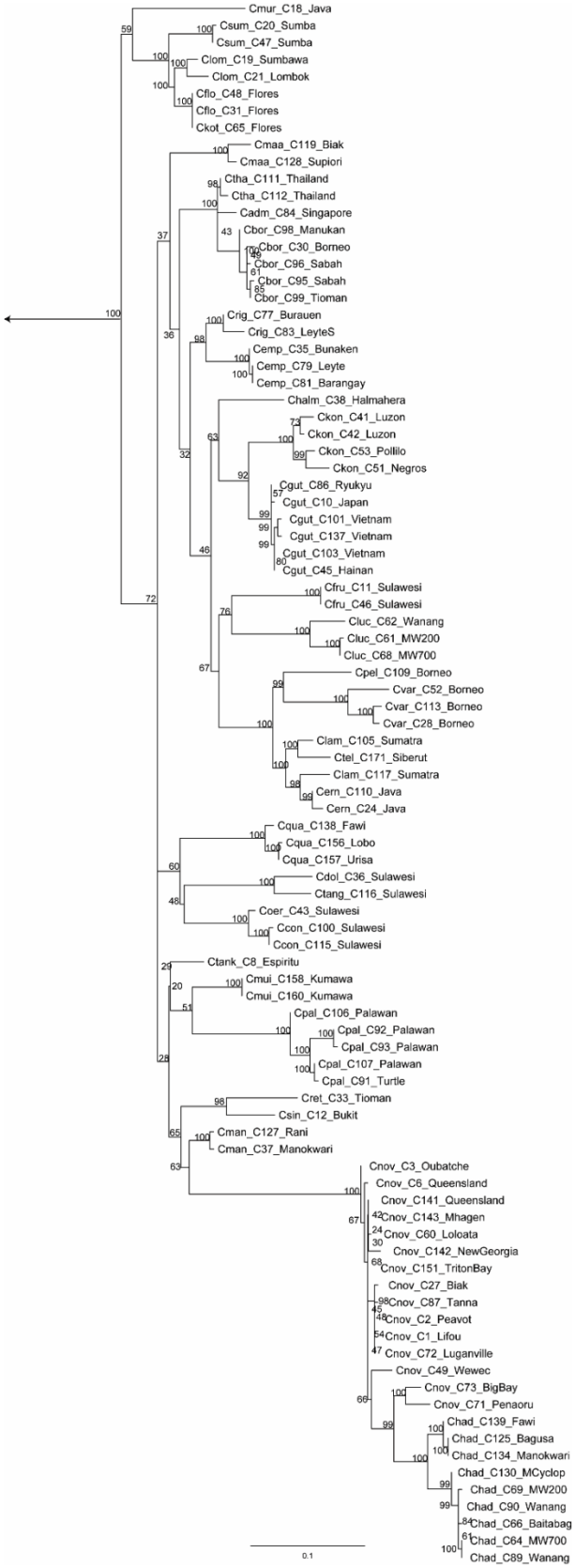


12S (b):

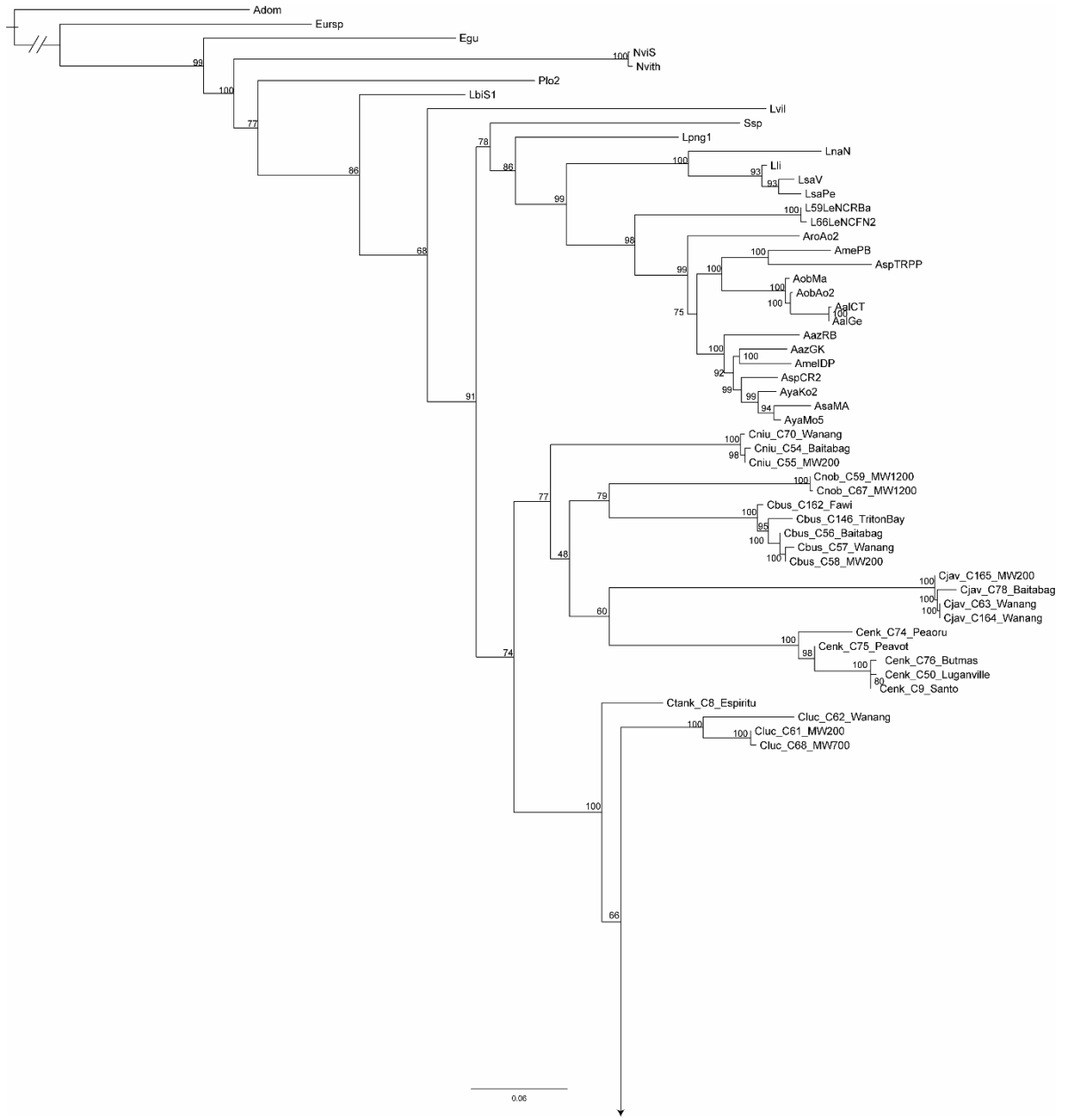


(continued)

12S (b): continued

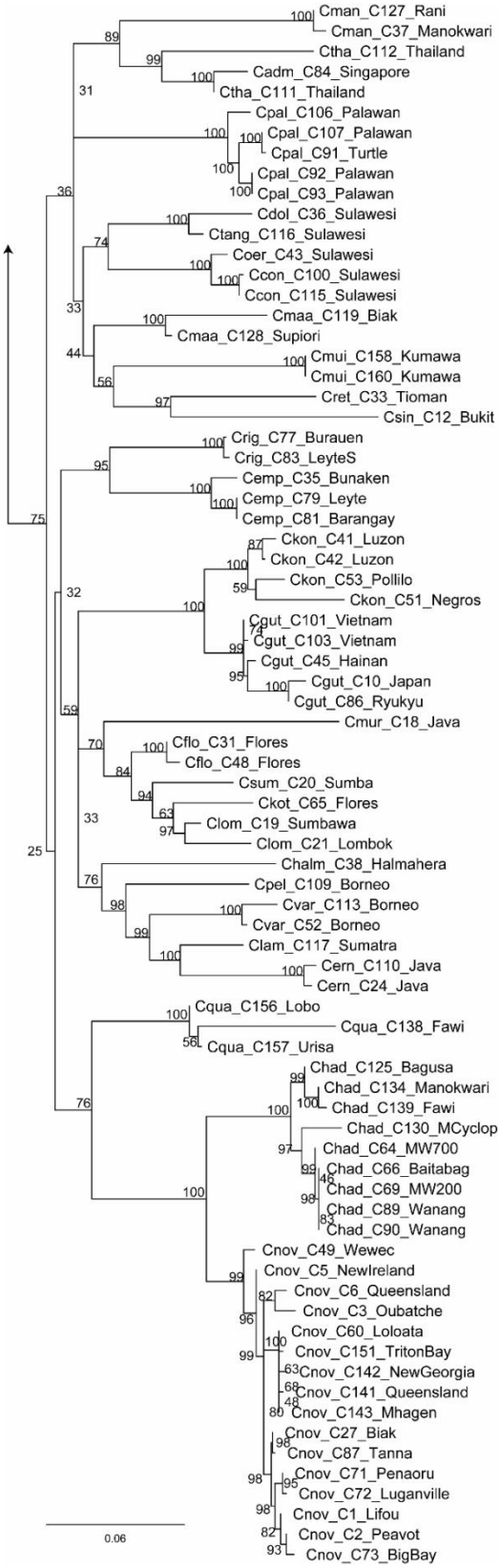


COI (c):

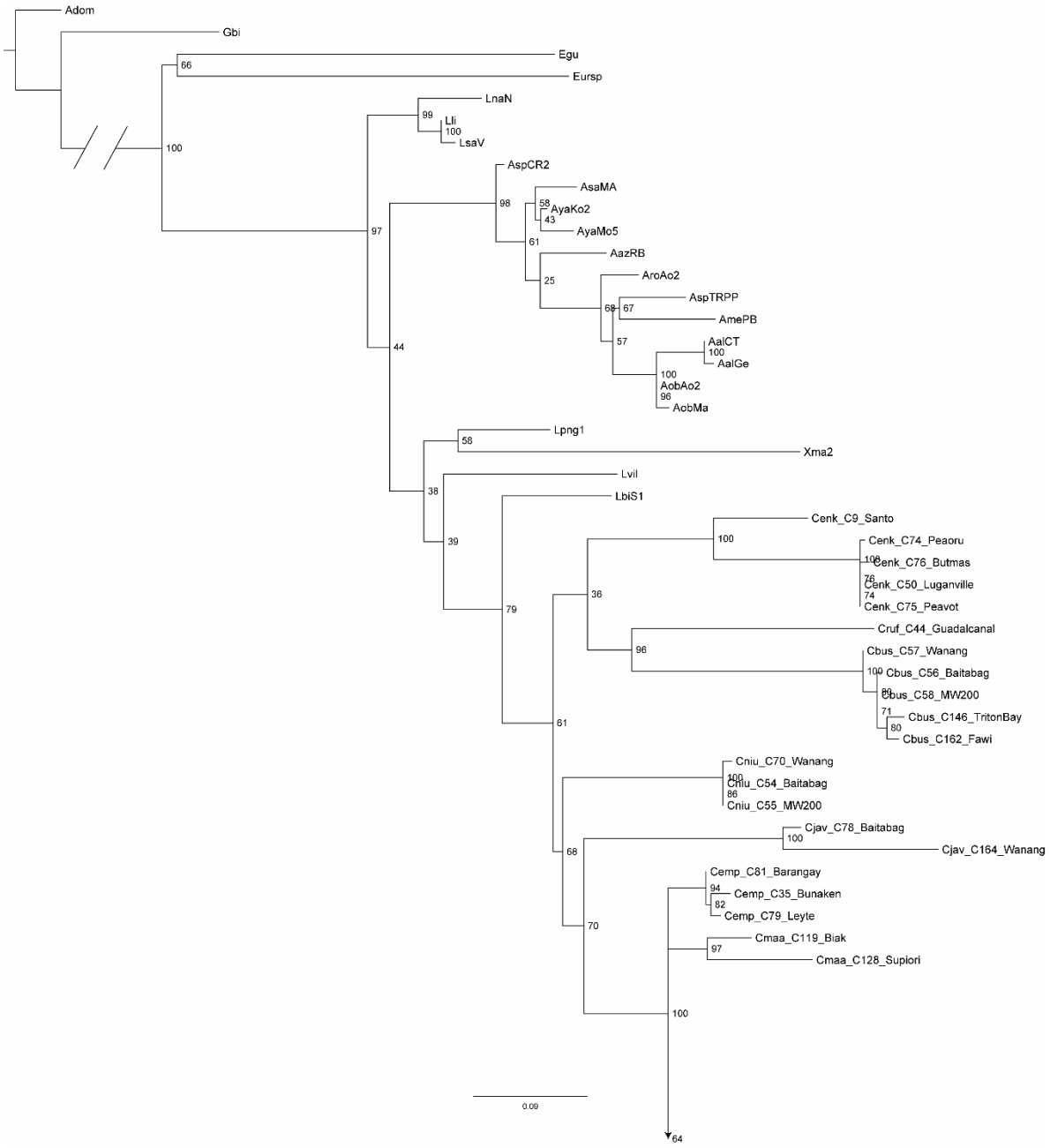


(continued)

COI (c): continued

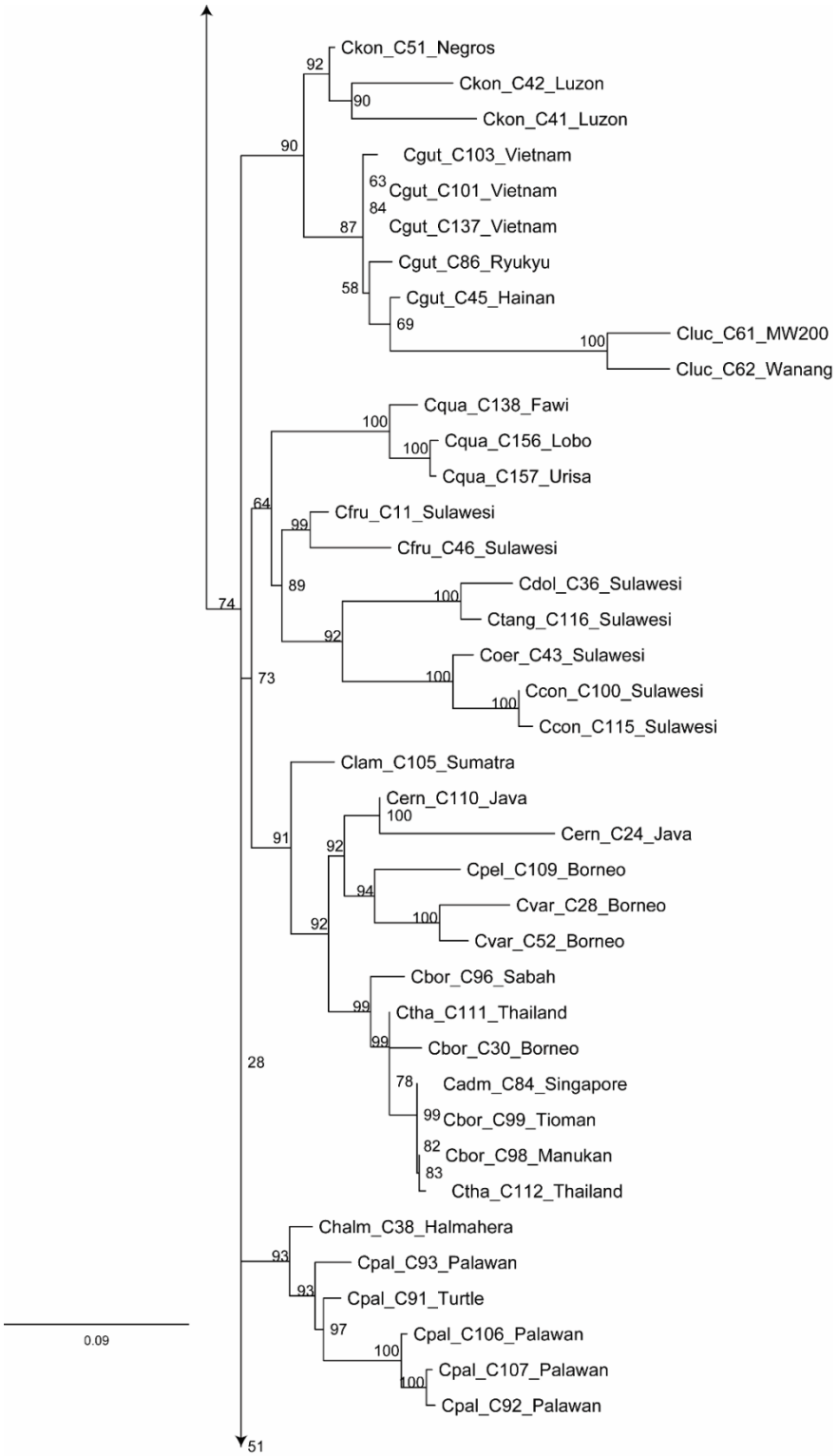


COII (d):



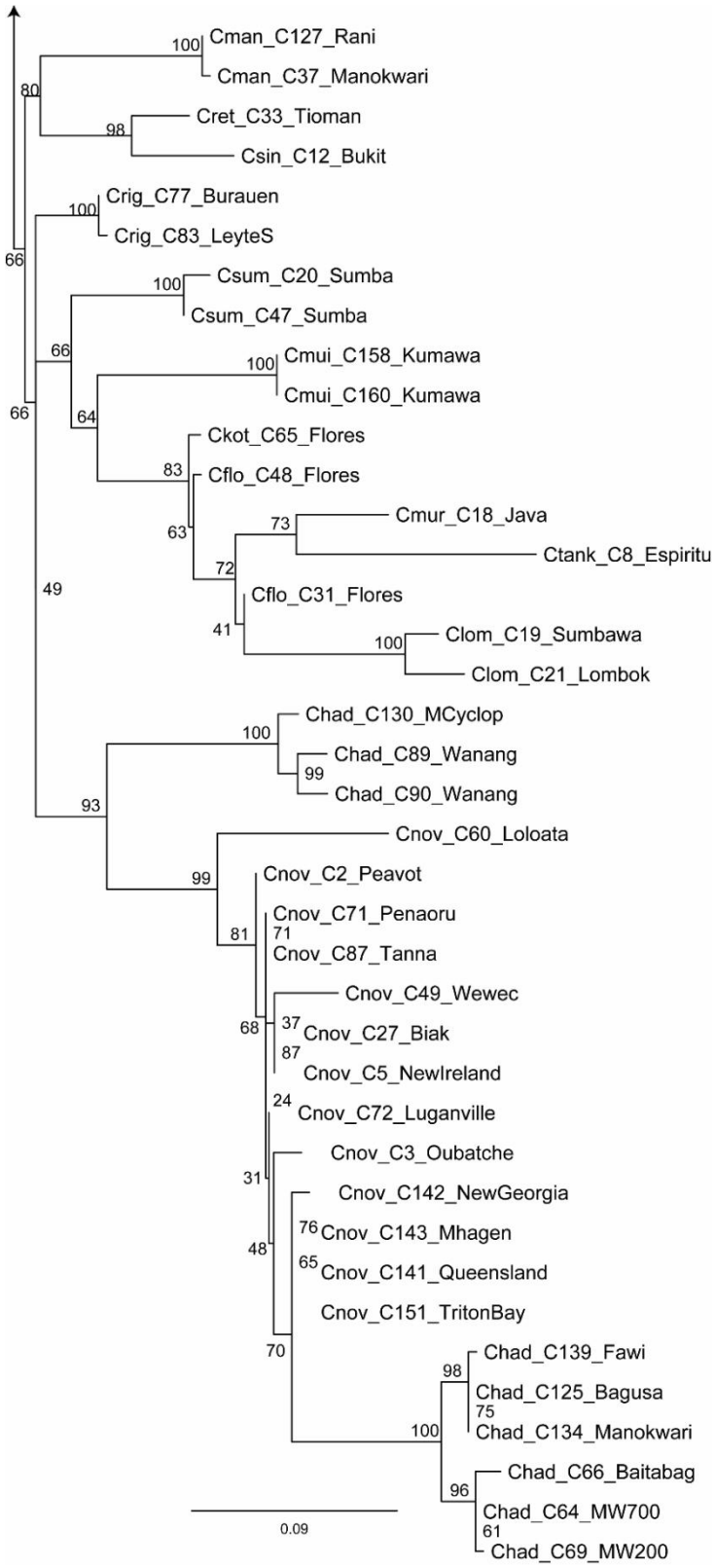
(continued)

COII (d): continued

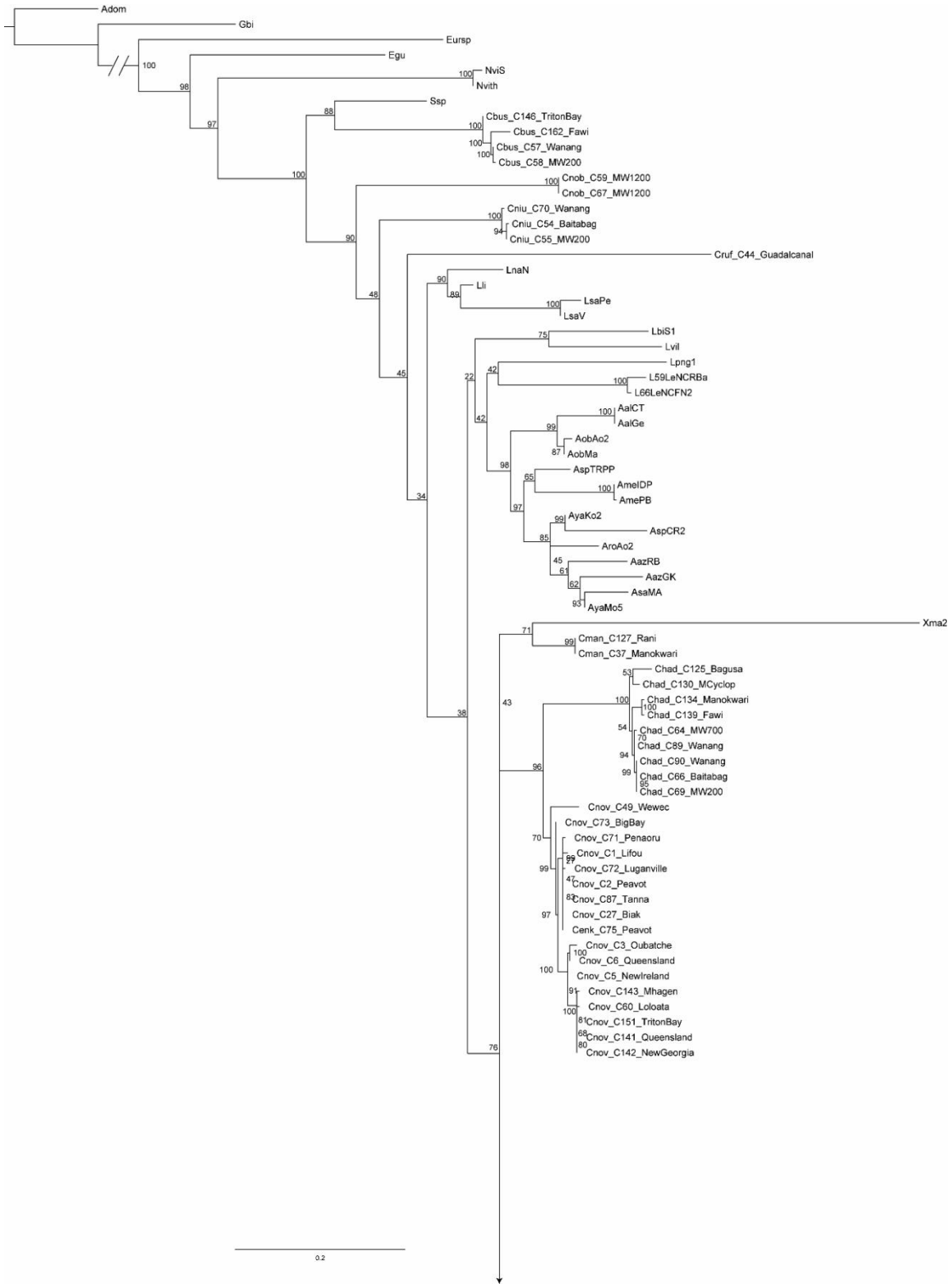


(continued)

COII (d): continued

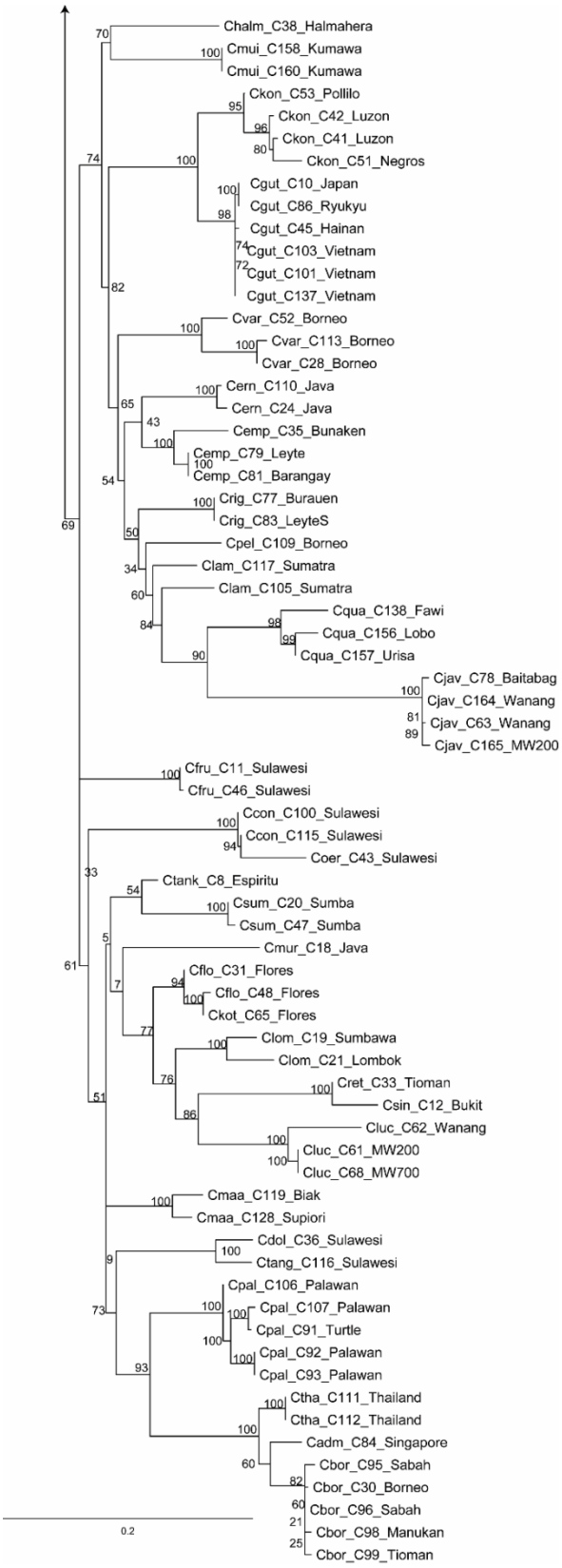


Cytb (e):

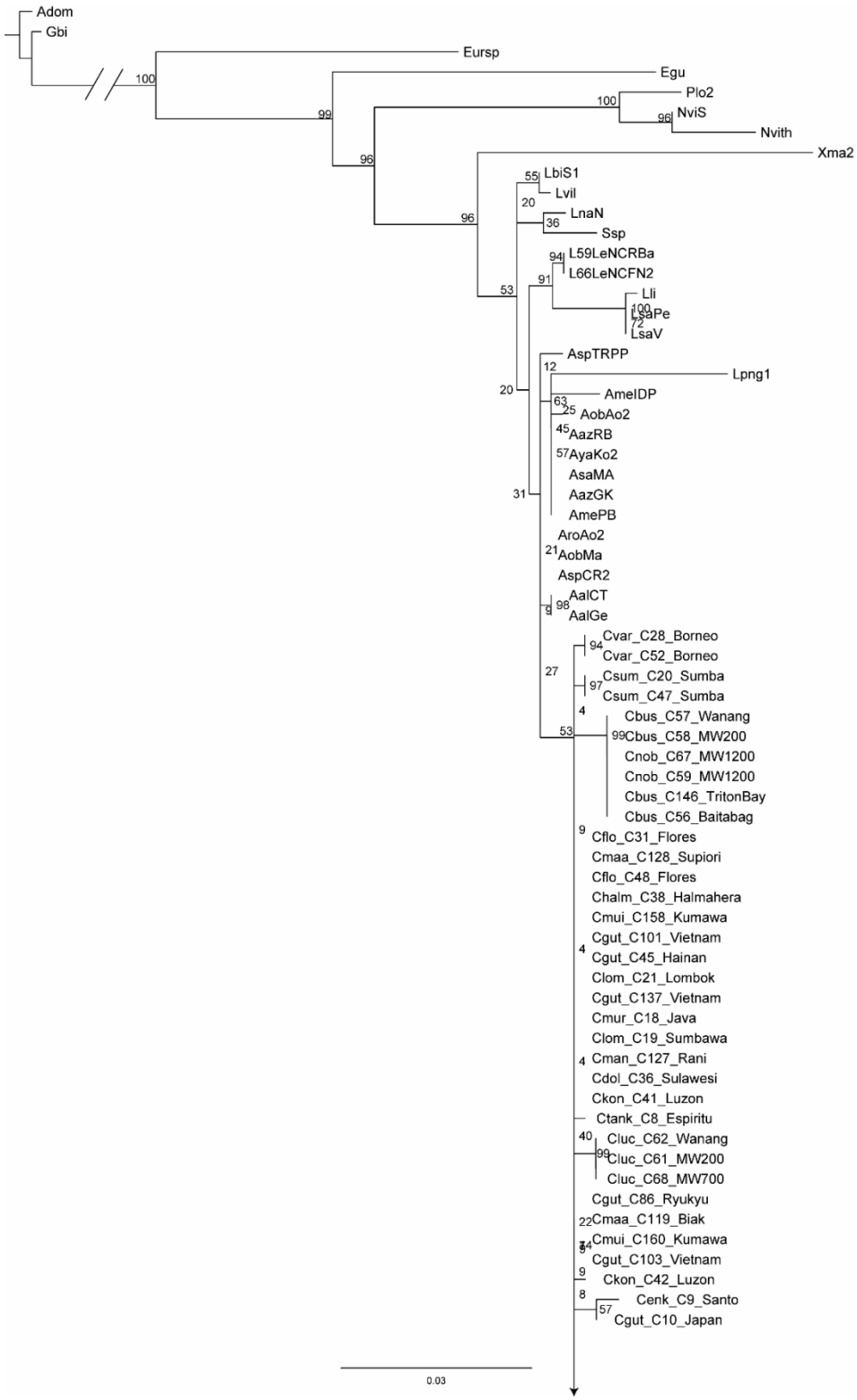


(continued)

Cytb (e): continued

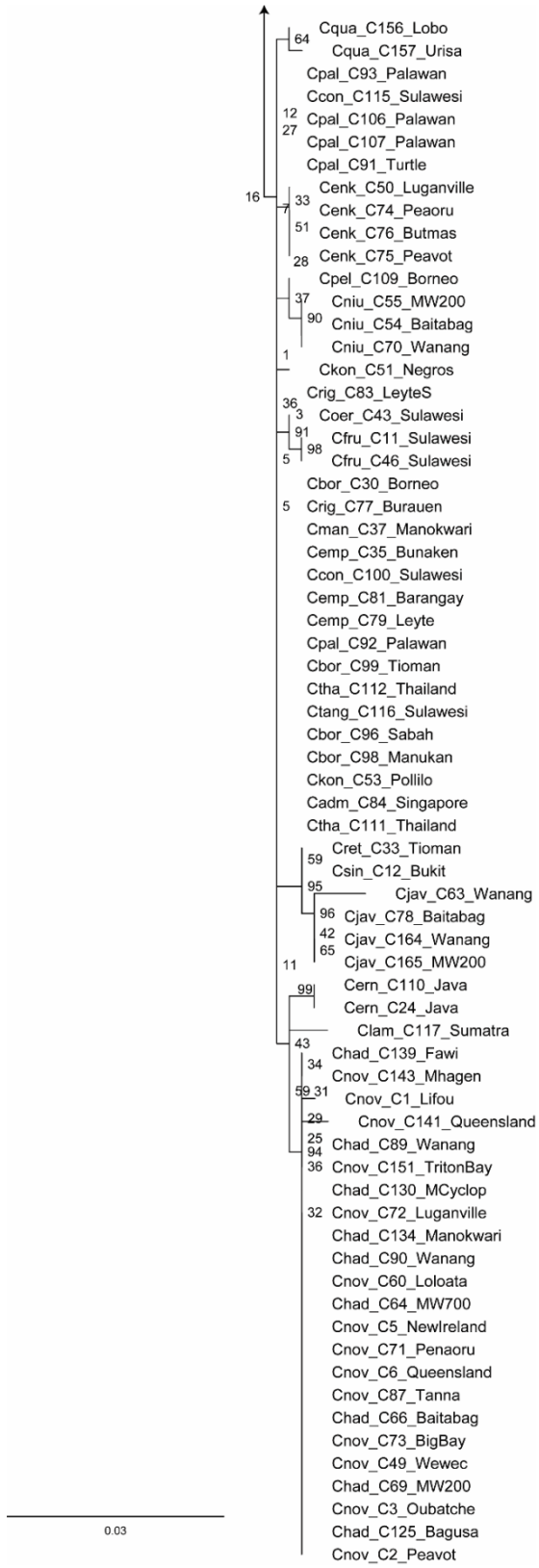


18S (f):



(continued)

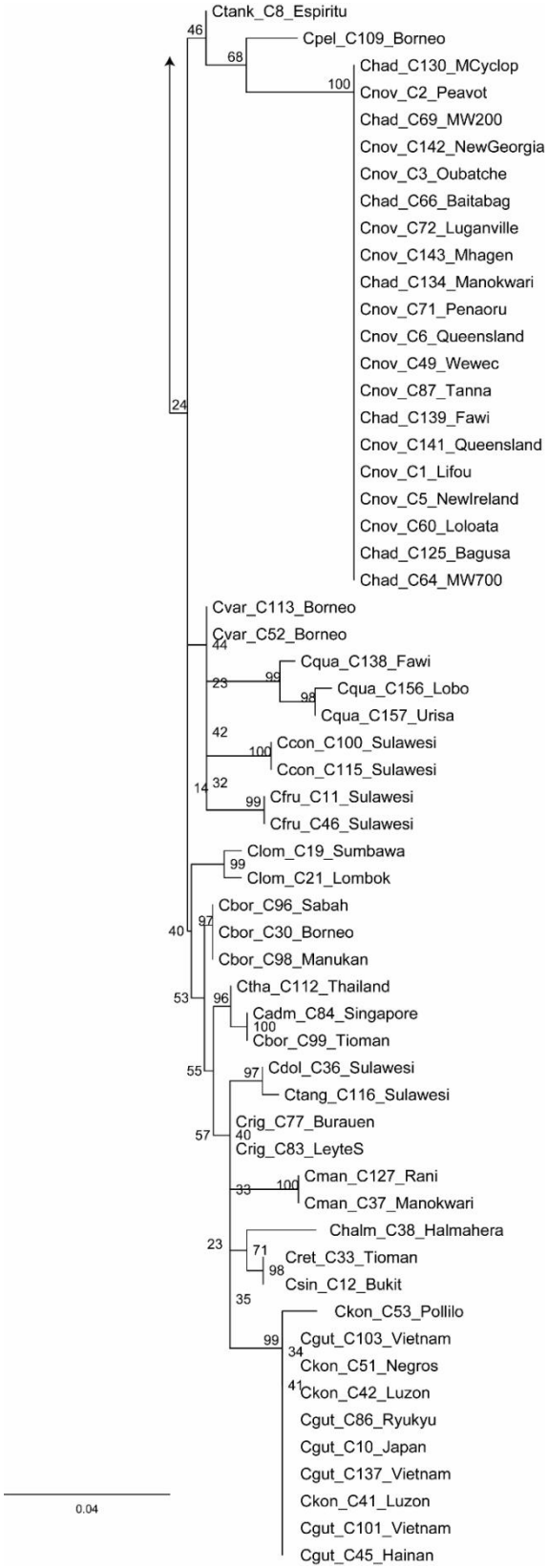
18S (f): continued



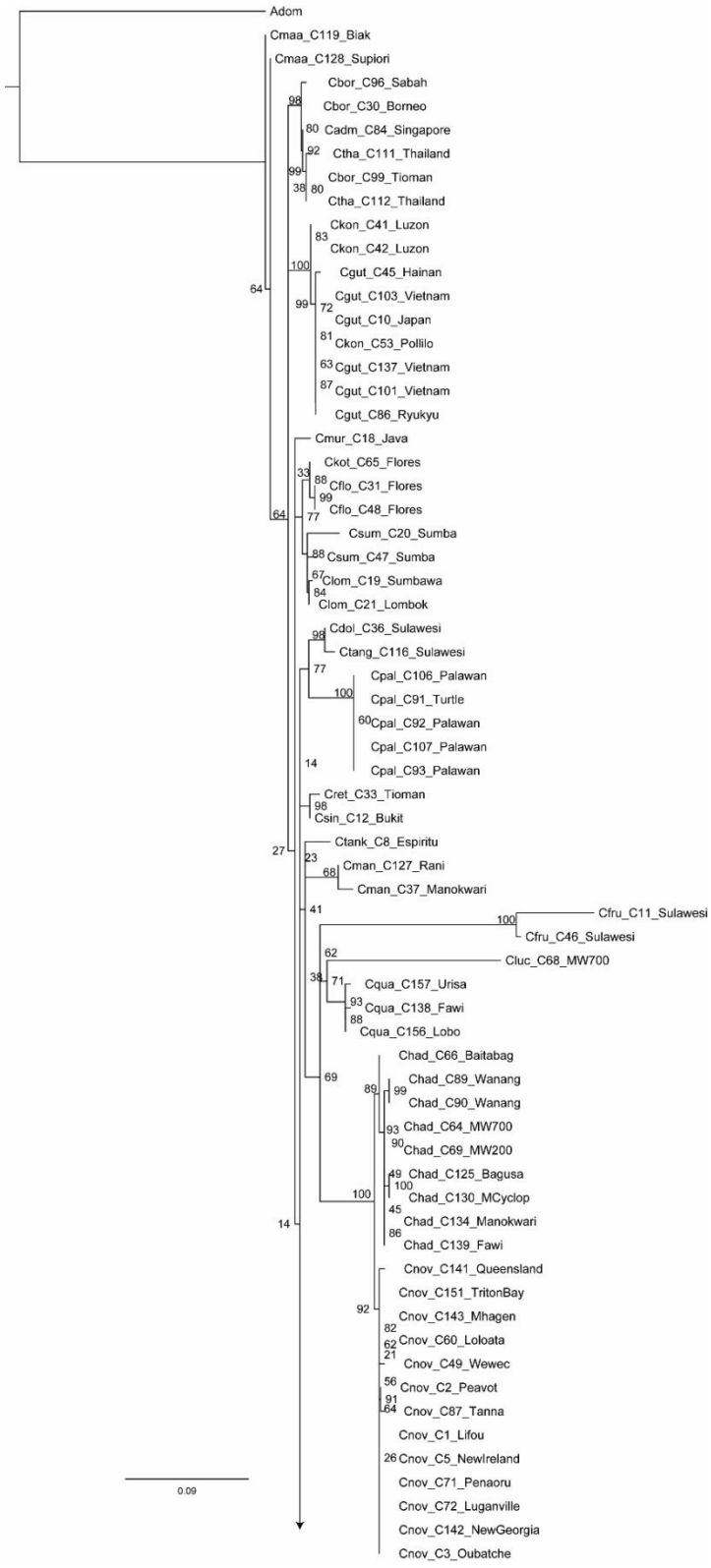
28S (g):

(continued)

28S (g): continued

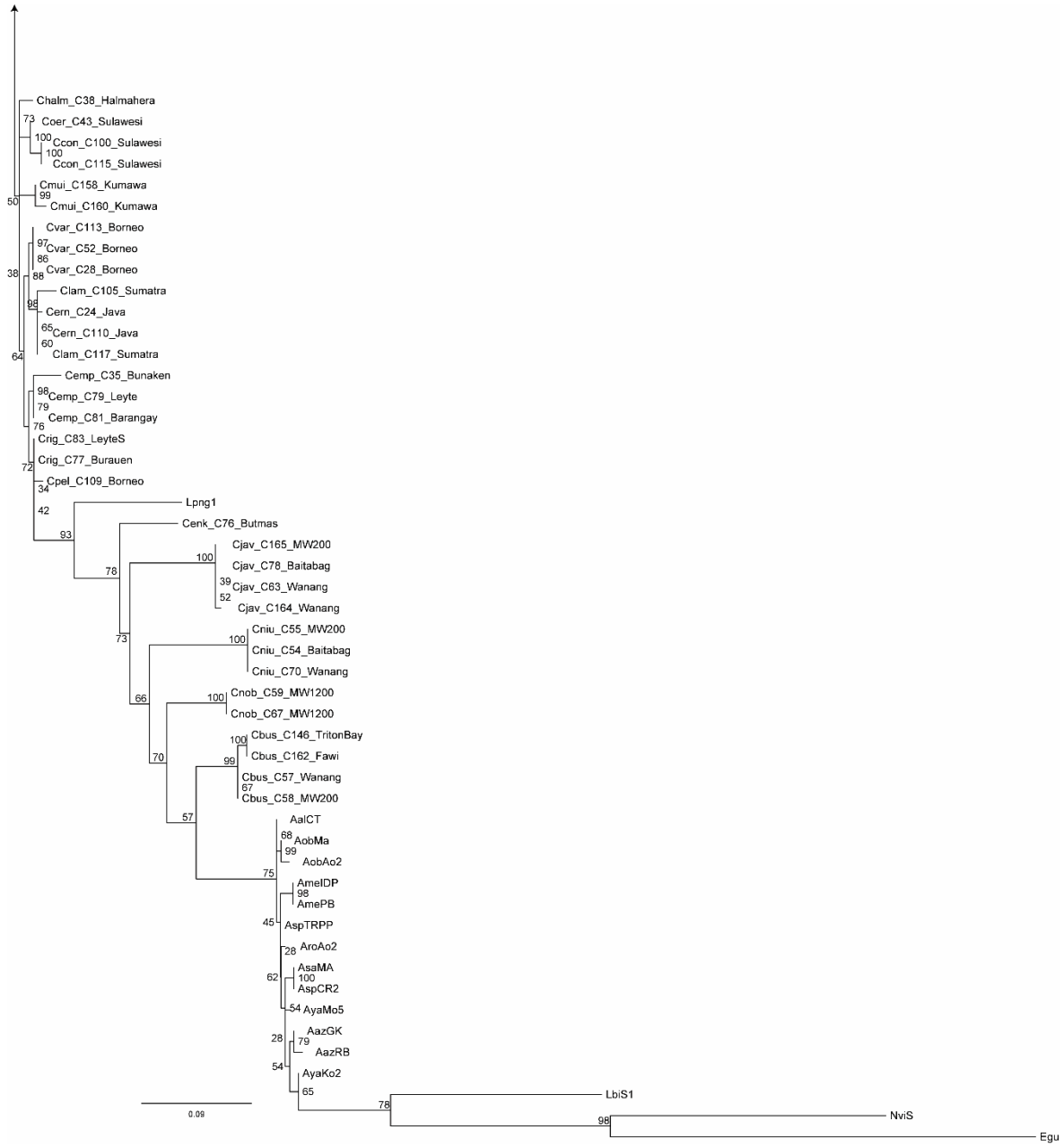


EF1a (h):

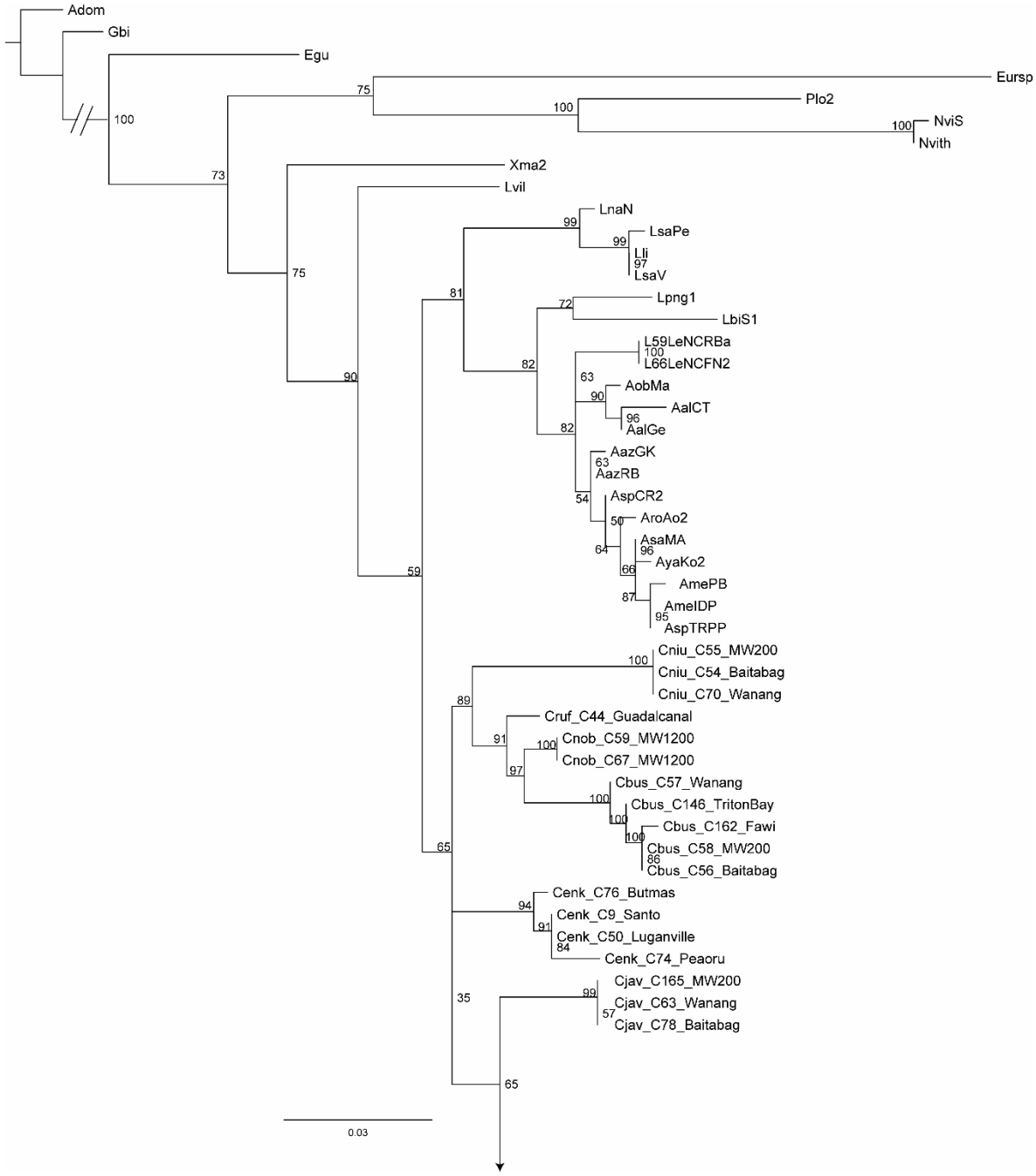


(continued)

EF1a (h): continued

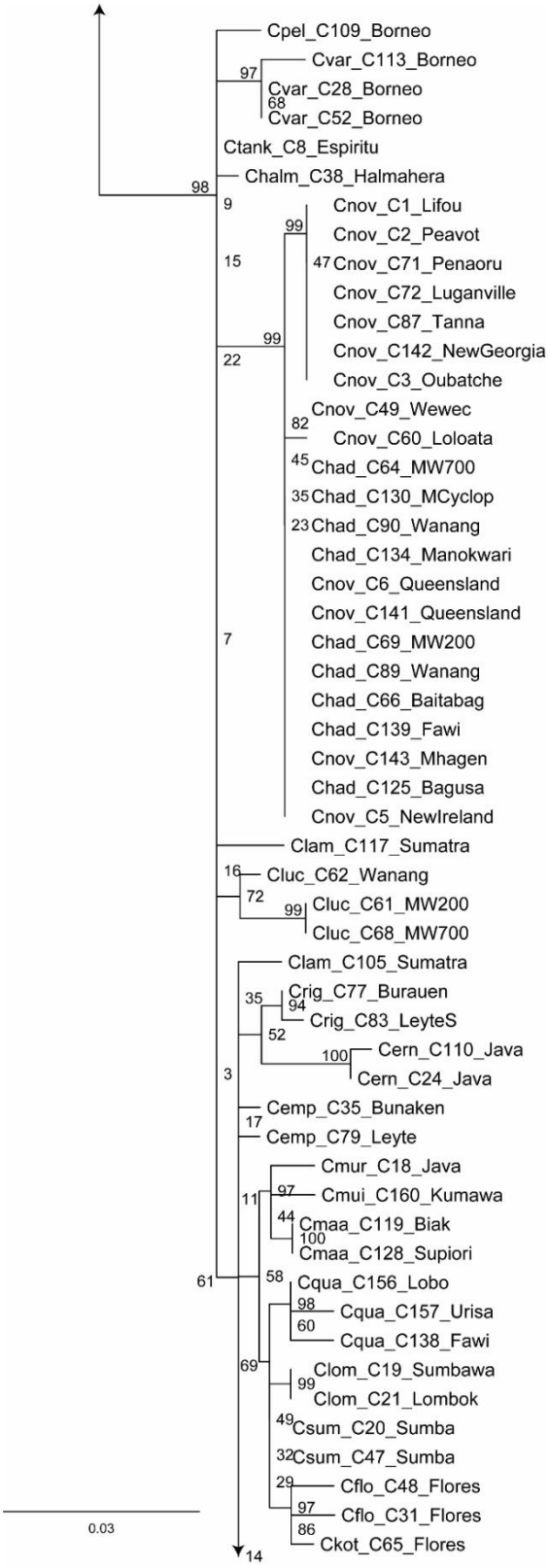


H3 (i):



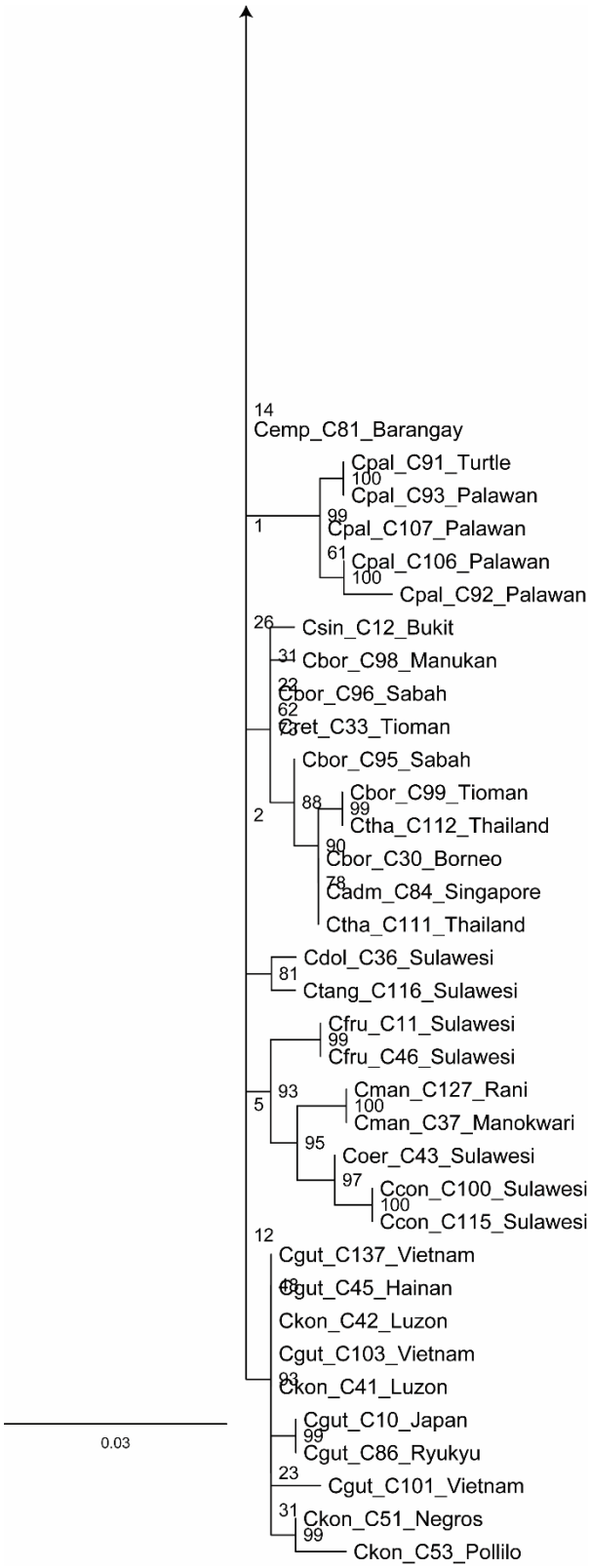
(continued)

H3 (i): continued



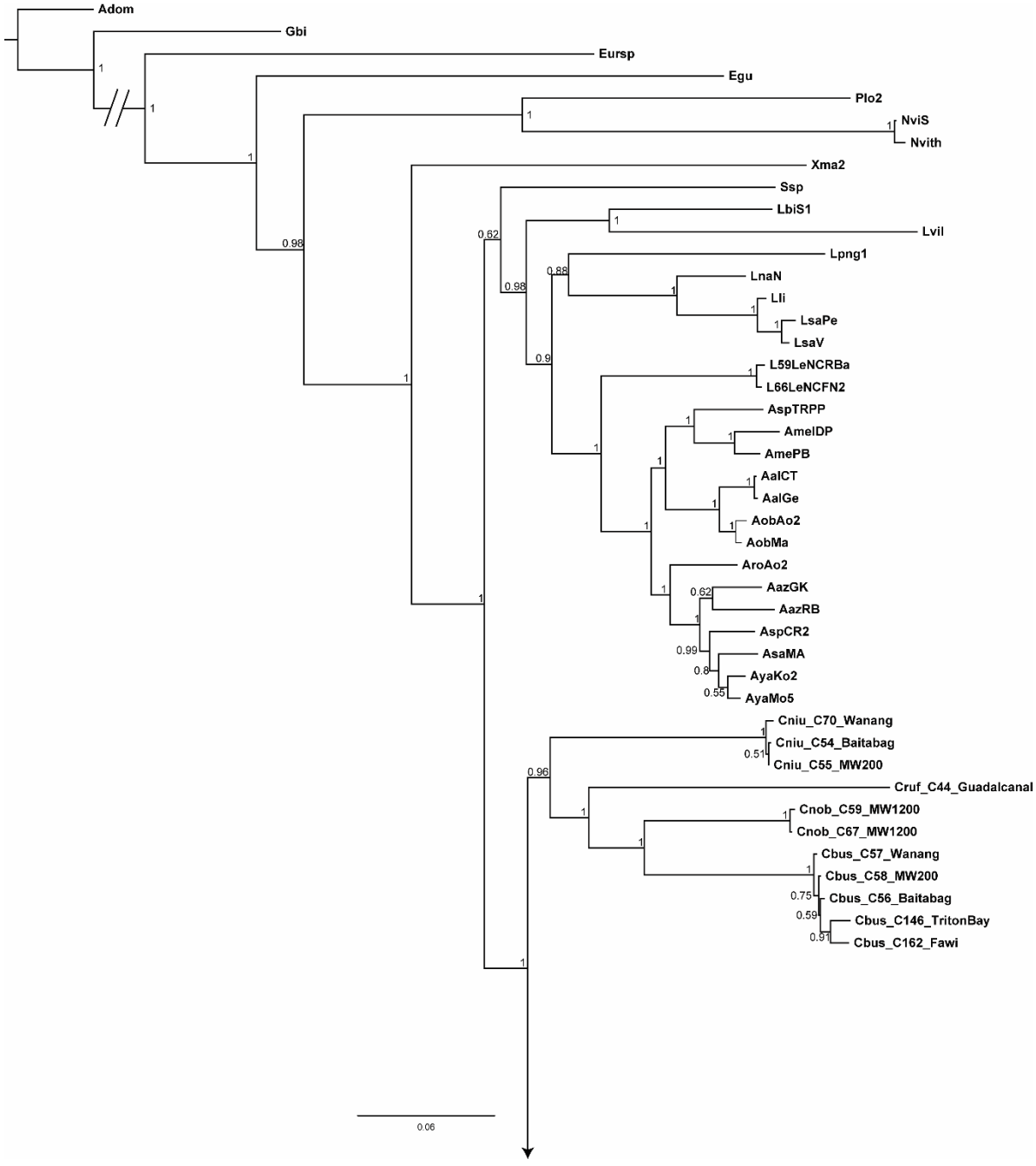
(continued)

H3 (i): continued



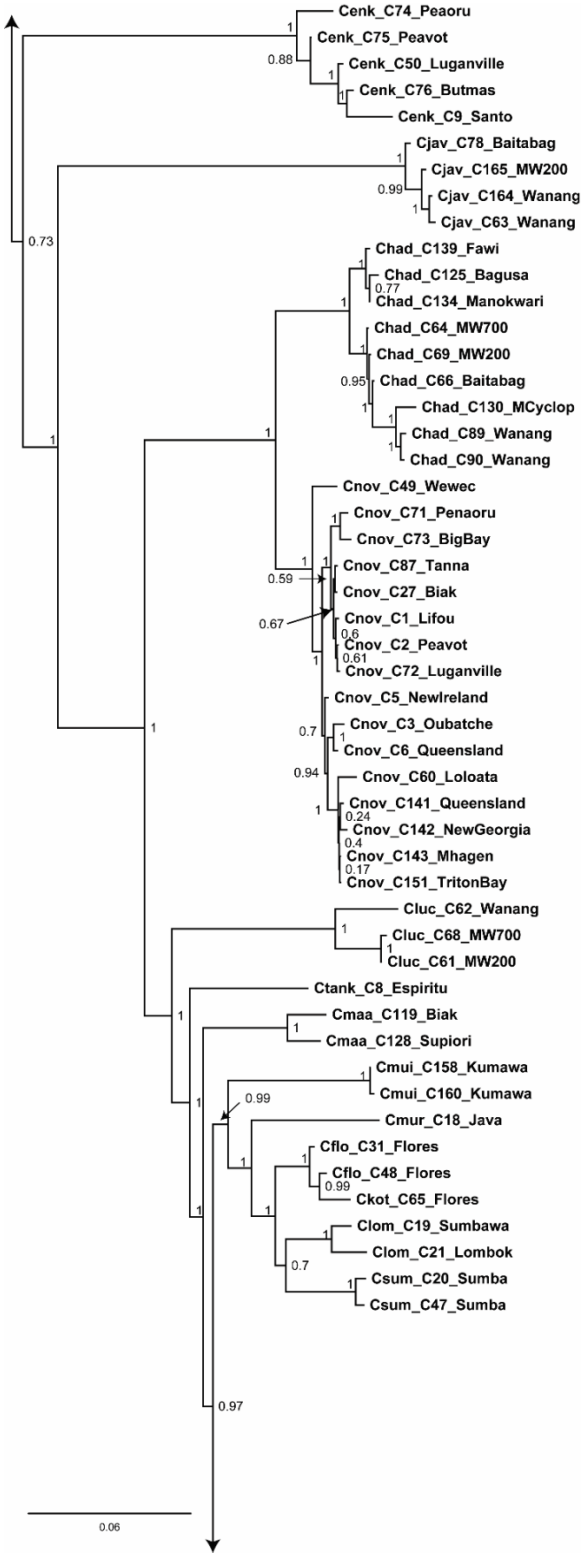
Appendix S6. Original output results of Bayesian inference (BI) and maximum likelihood (ML) analyses for the concatenated dataset.

BI maximum consensus tree (concatenated dataset):



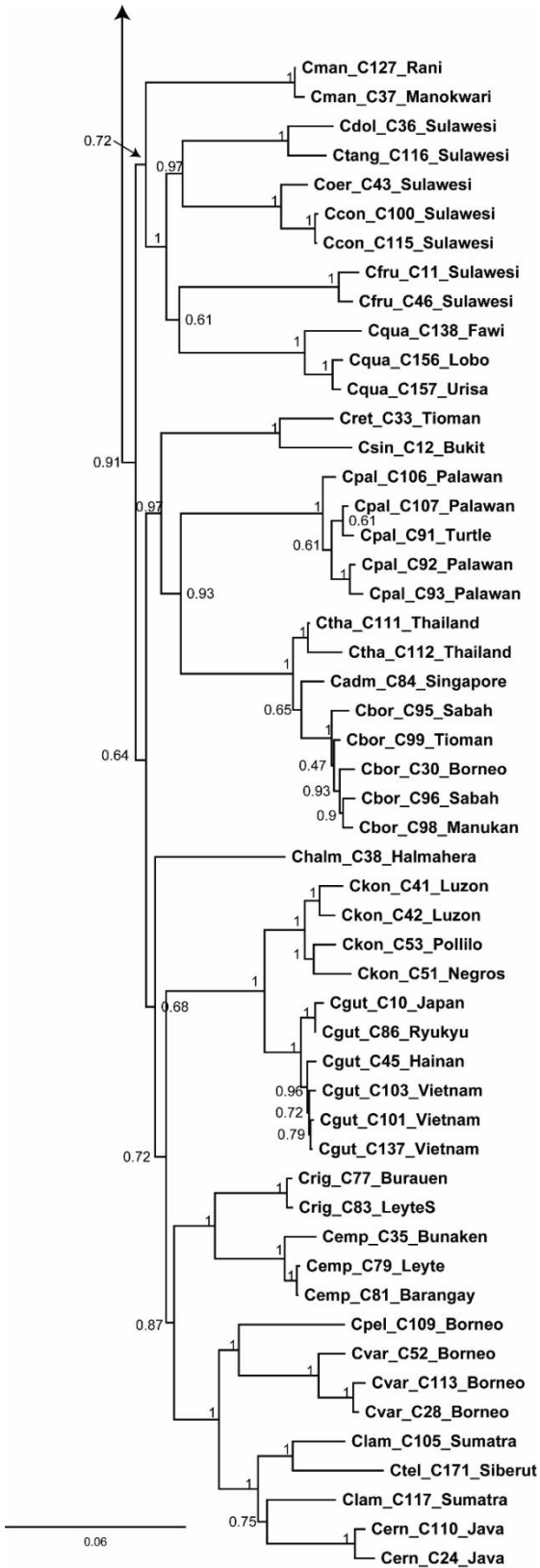
(continued)

BI maximum consensus tree (concatenated dataset): continued

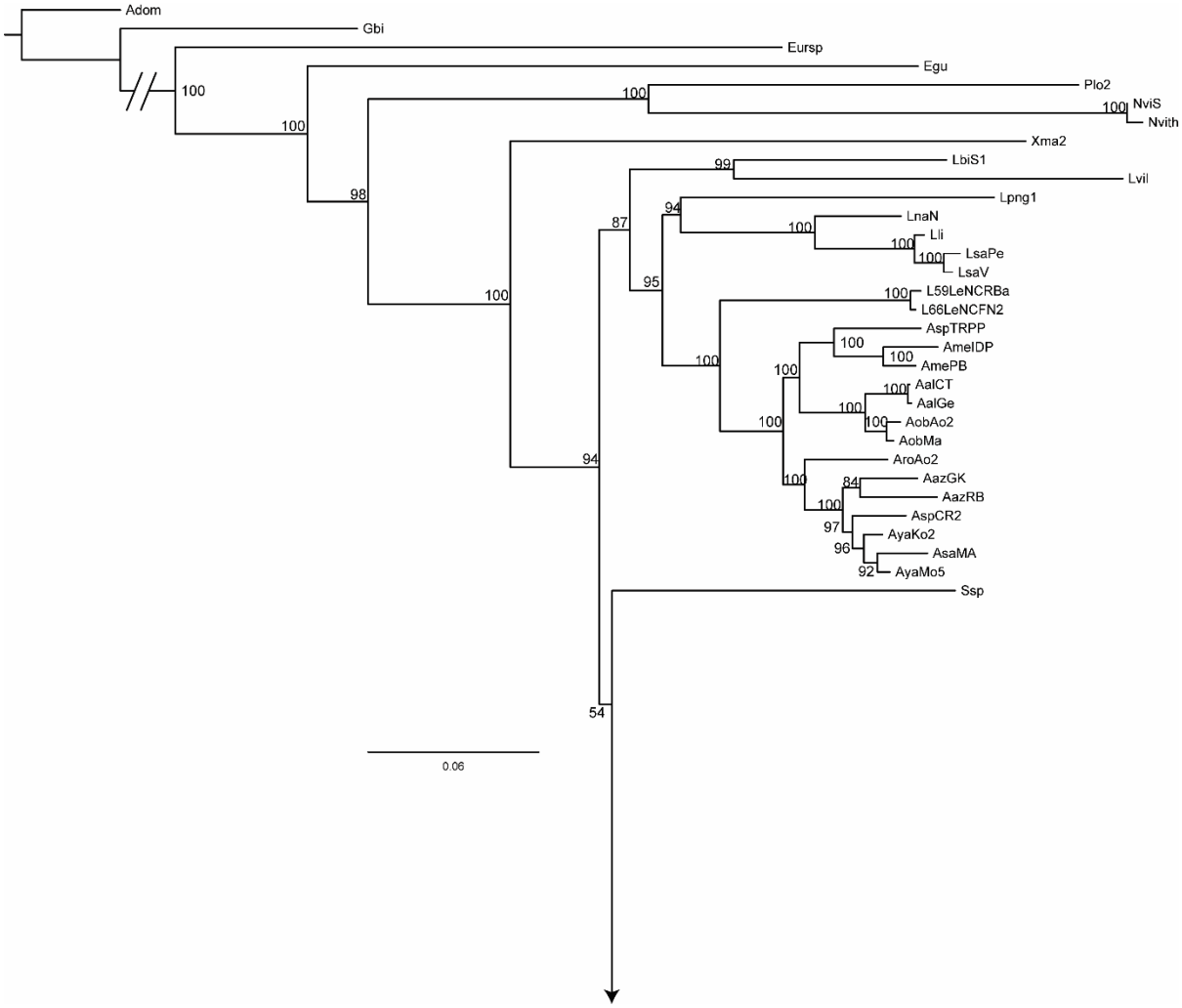


(continued)

BI maximum consensus tree (concatenated dataset): continued

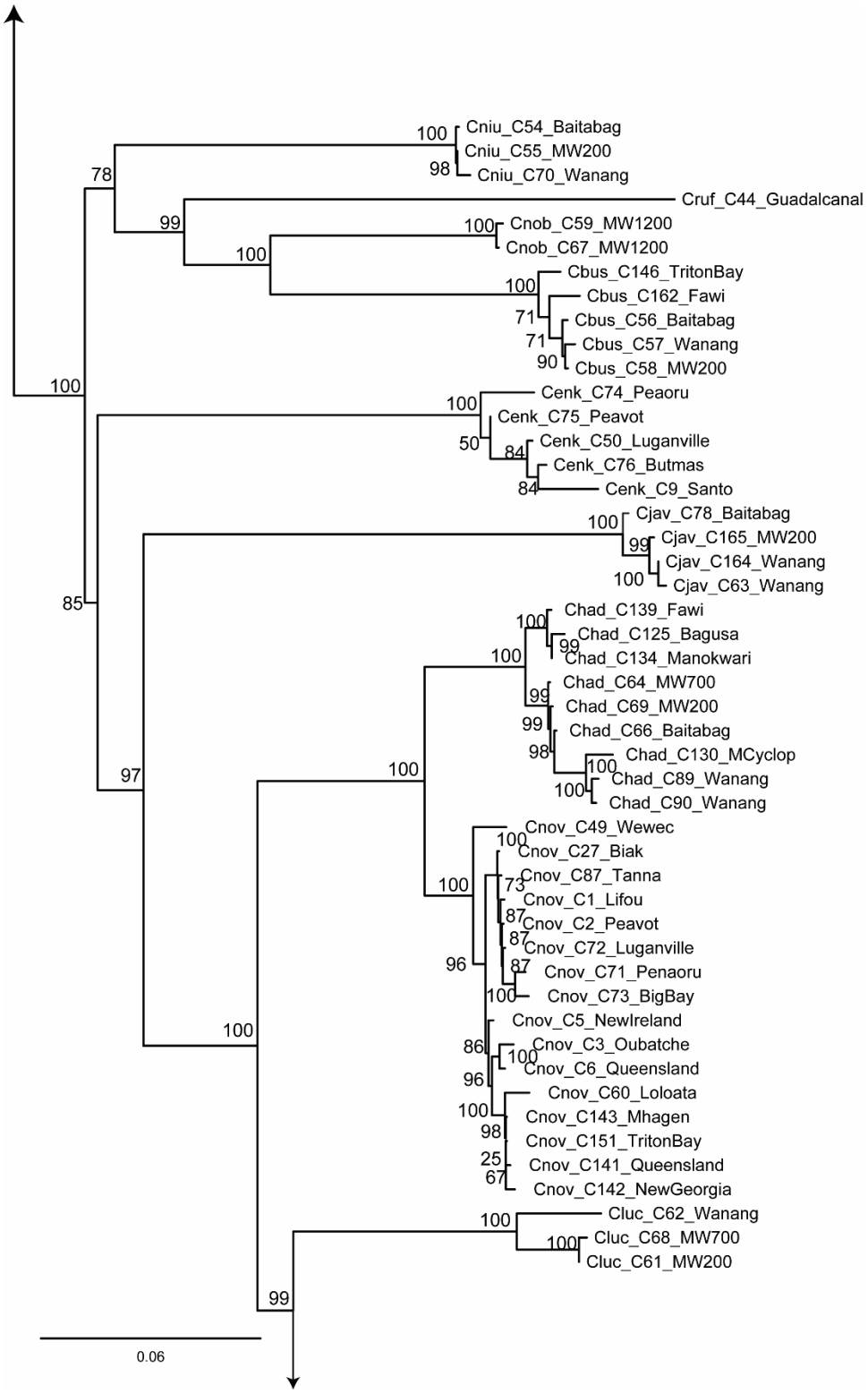


Best ML tree, RAxML (concatenated dataset):



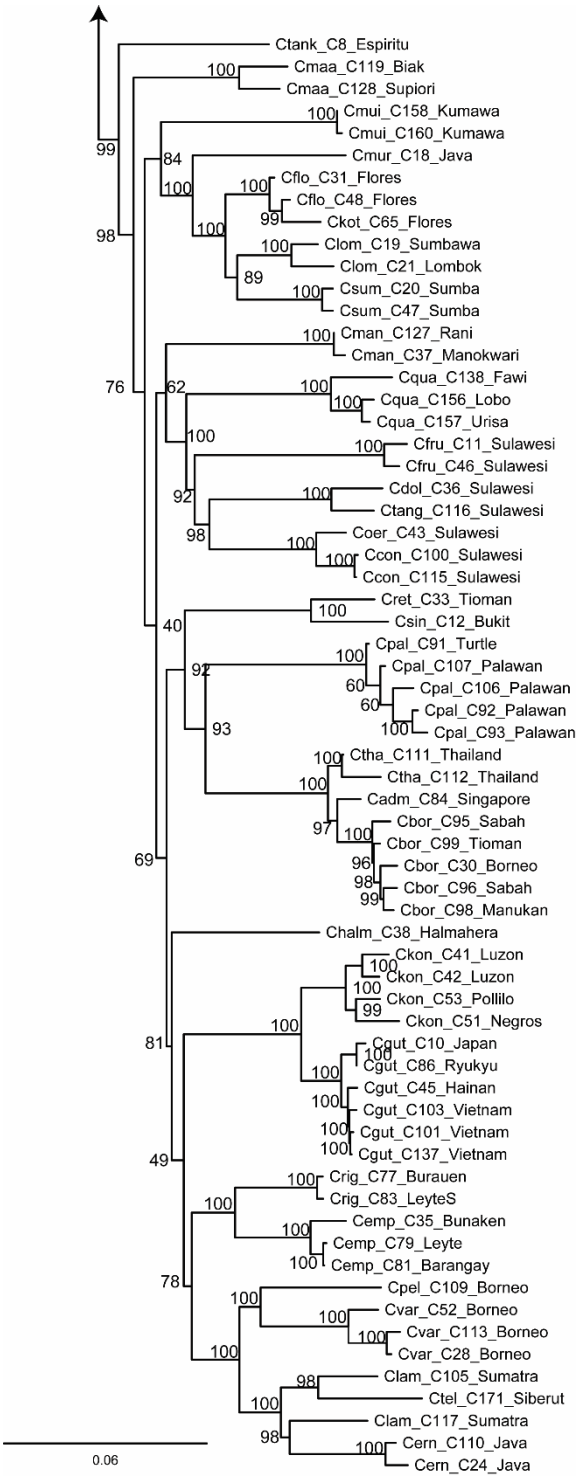
(continued)

Best ML tree, RAxML (concatenated dataset): continued

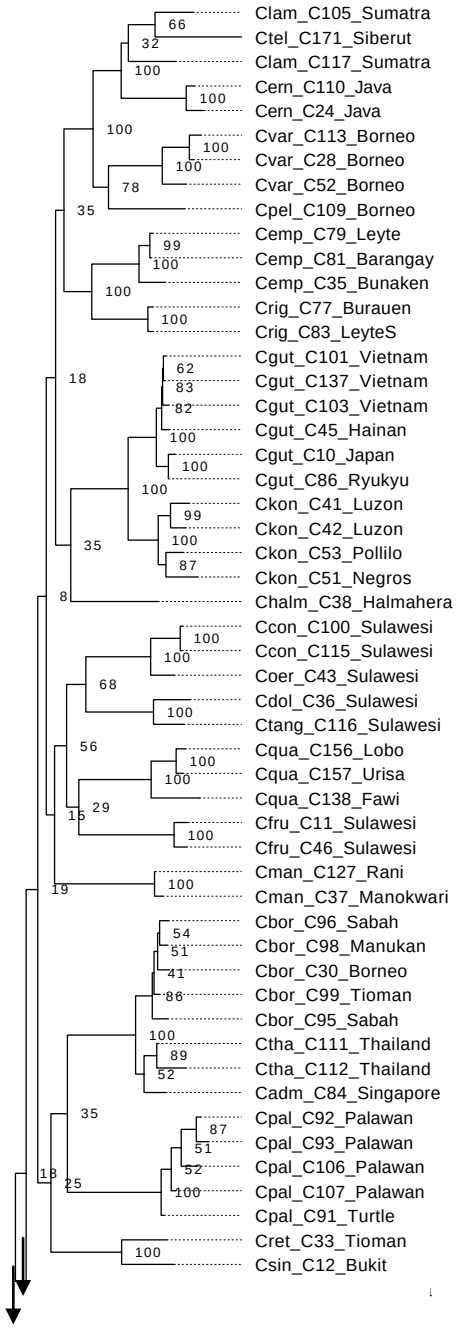


(continued)

Best ML tree, RAxML (concatenated dataset): continued

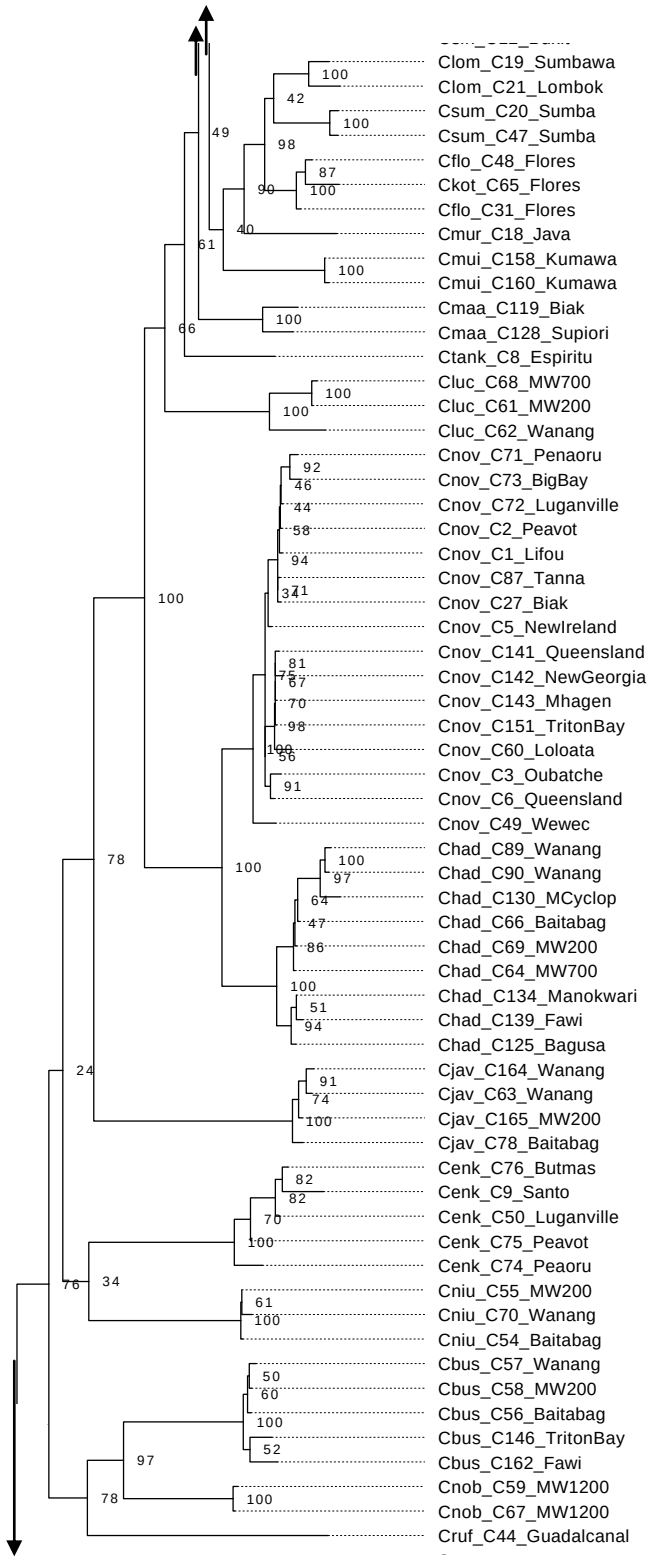


Best ML tree, IQ-TREE:



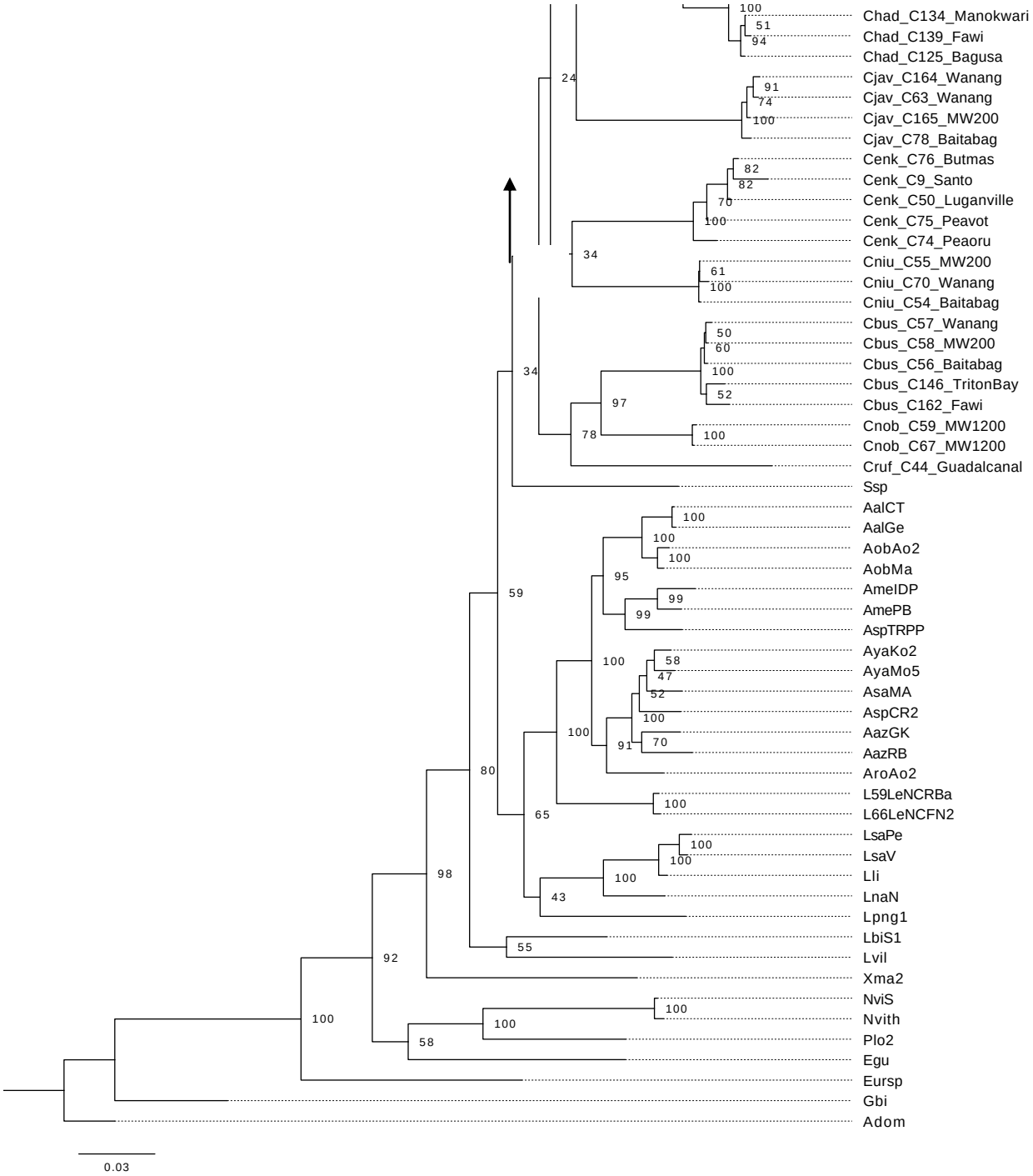
(continued)

Best ML tree, IQ-TREE (concatenated dataset): continued



(continued)

Best ML tree, IQ-TREE (concatenated dataset): continued



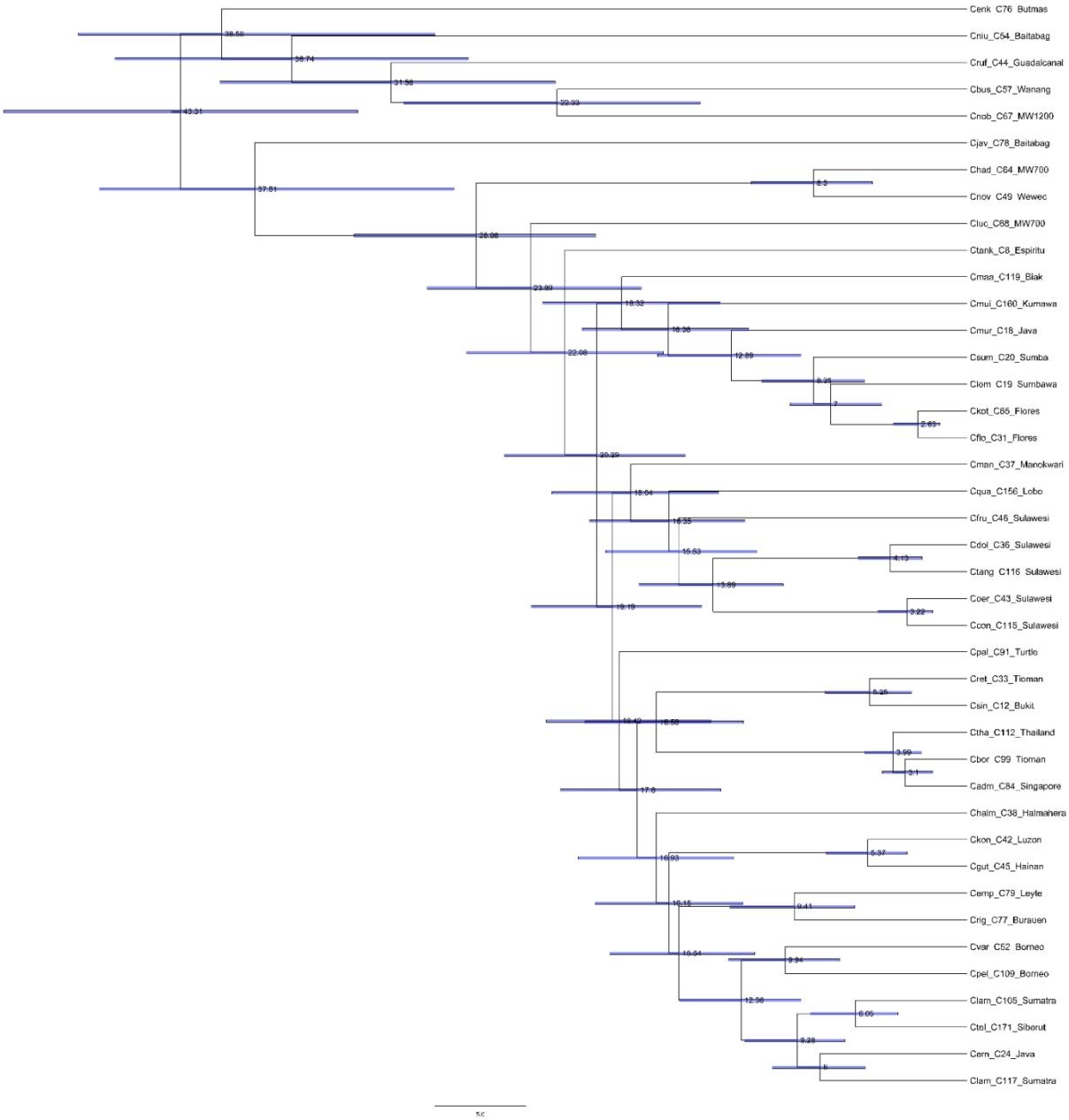
1438

1439

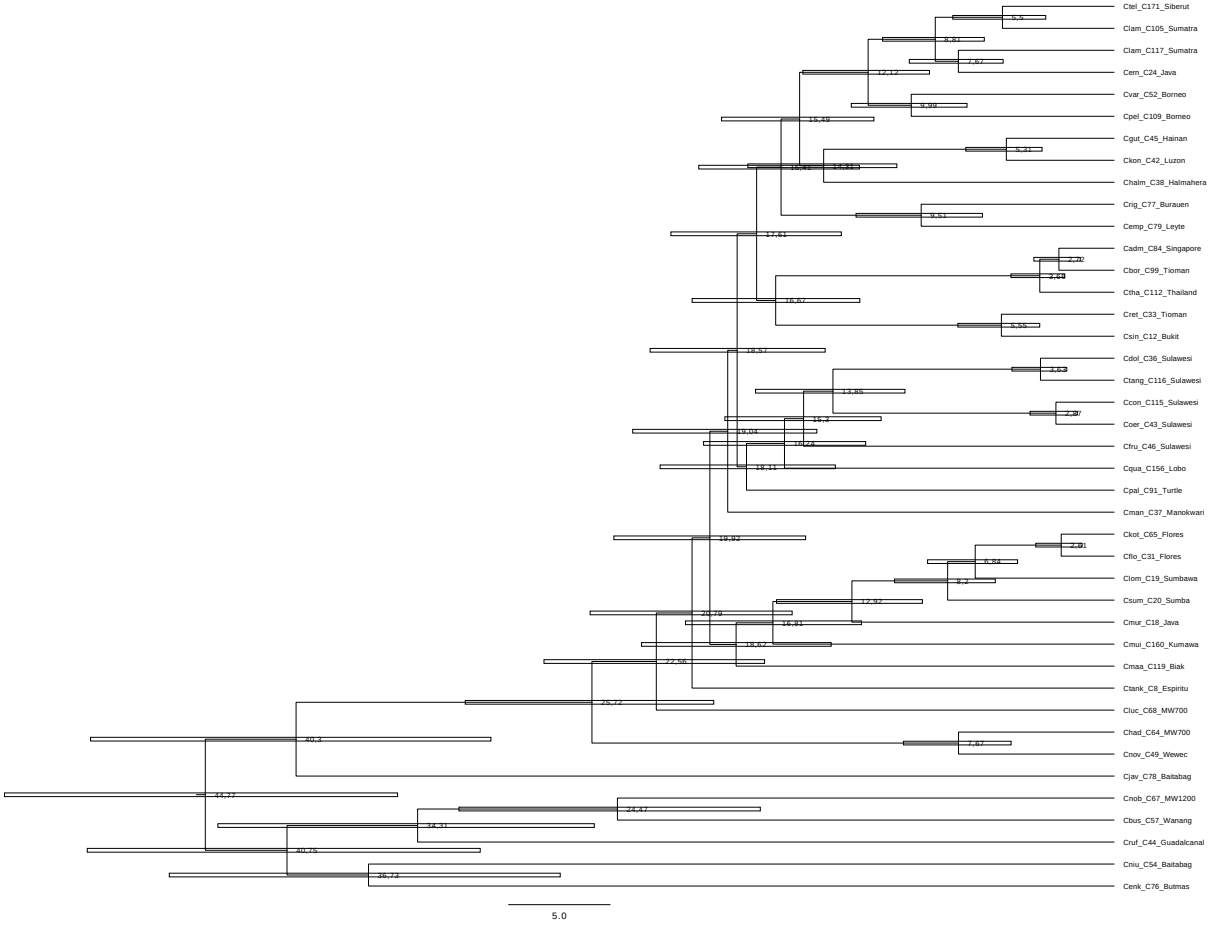
1440

Appendix S7. Original outputs for dating analyses.

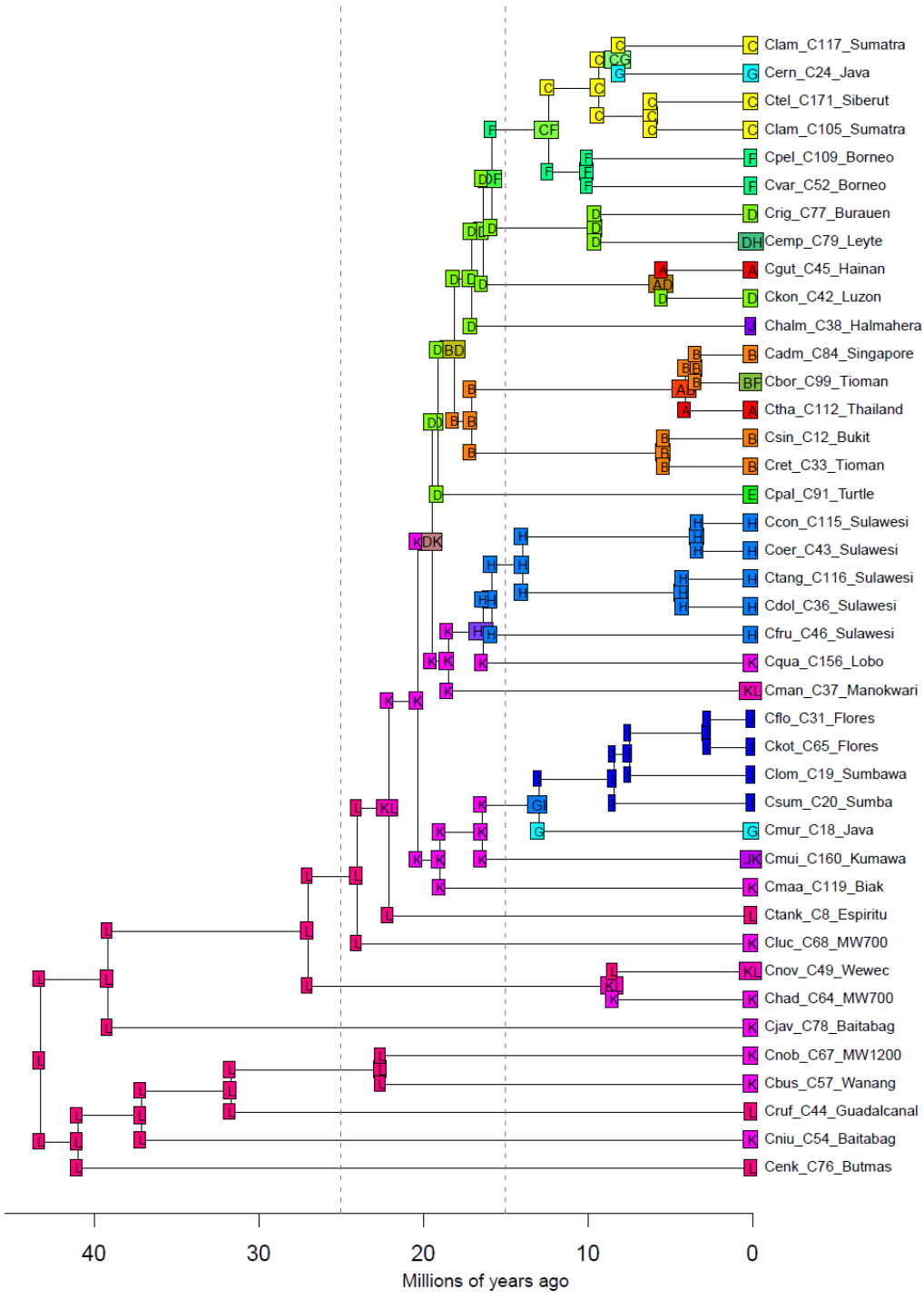
Dated phylogeny resulting from BEAST analyses relying on the *unlinked branch lengths* option of PartitionFinder:

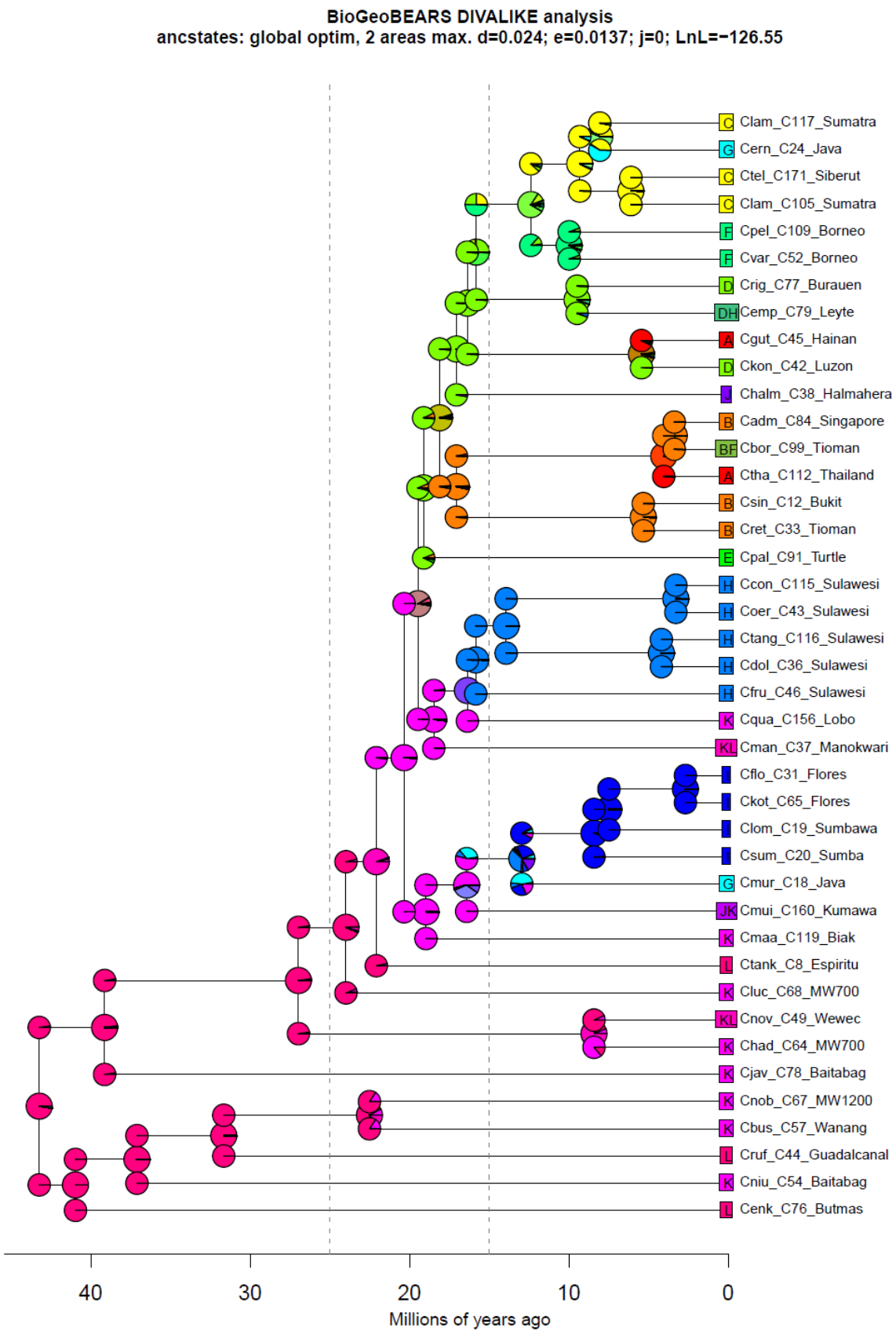


Dated phylogeny resulting from BEAST analyses relying on the *linked branch lengths* option of PartitionFinder:



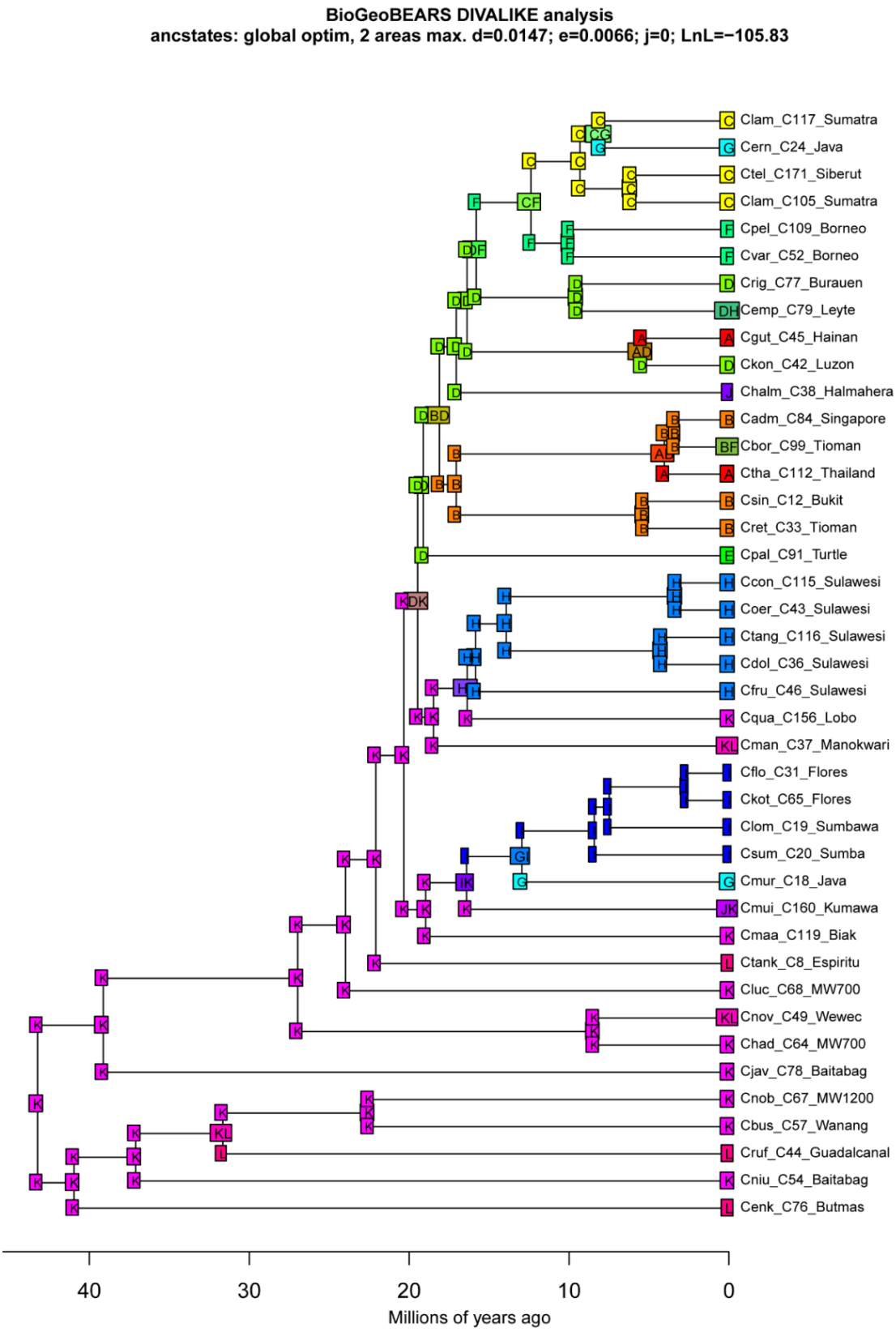
Appendix S8. Results of ancestral area estimation in BioGeoBEARS based on the most complex set of dispersal rate multipliers with three time slices, two maximum ancestral areas and DIVALIKE speciation model.

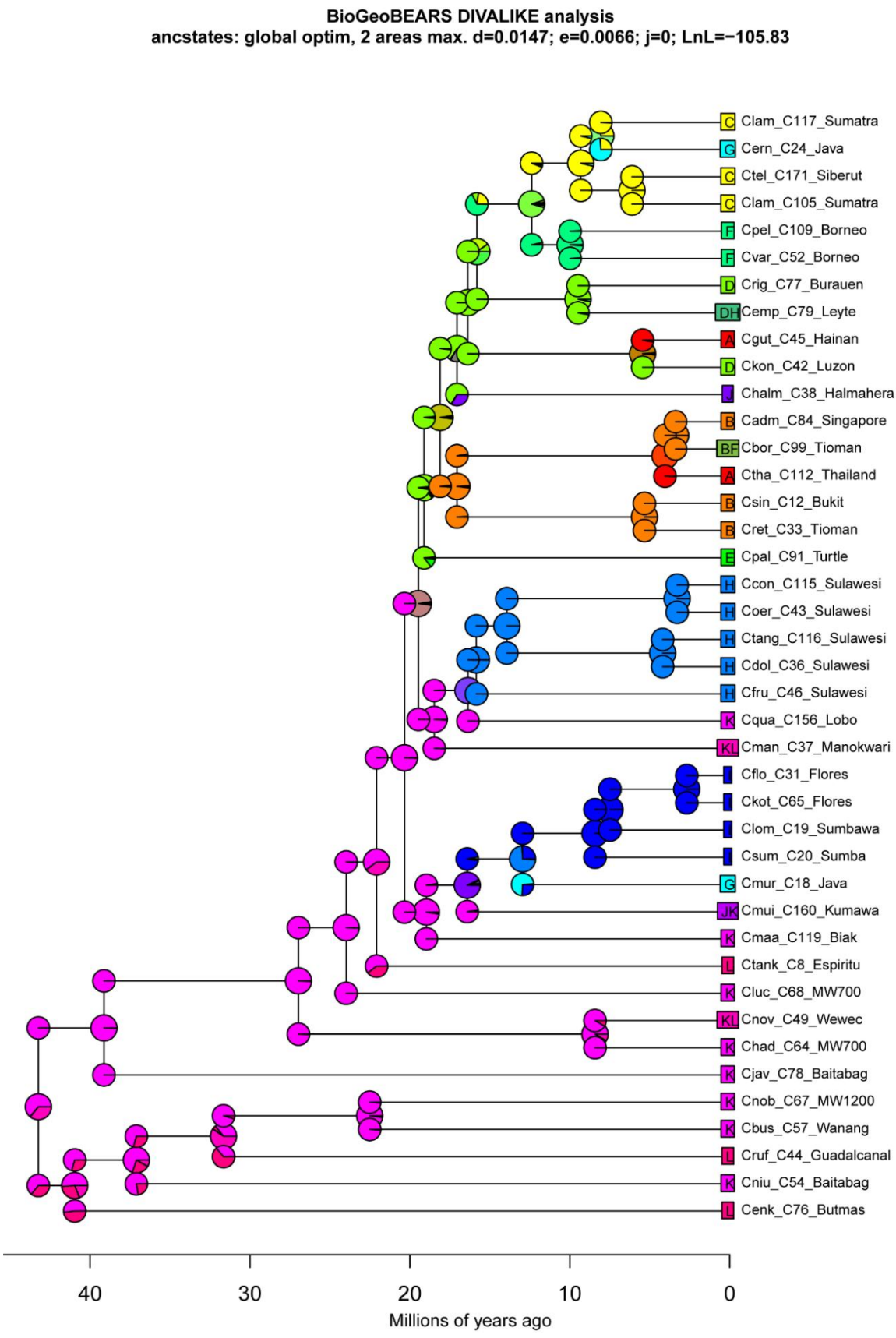



BioGeoBEARS DIVALIKE analysis
ancstates: global optim, 2 areas max. d=0.024; e=0.0137; j=0; LnL=−126.55
C Clam_C117_Sumatra
G Cern_C24_Java
C Ctel_C171_Siberut
C Clam_C105_Sumatra
F Cpel_C109_Borneo
F Cvar_C52_Borneo
D Crig_C77_Burauen
DH Cemp_C79_Leyte
A Cgut_C45_Hainan
D Ckon_C42_Luzon
J Chalm_C38_Halmahera
B Cadm_C84_Singapore
BF Cbor_C99_Tioman
A Ctha_C112_Thailand
B Csin_C12_Bukit
B Cret_C33_Tioman
E Cpal_C91_Turtle
H Ccon_C115_Sulawesi
H Coer_C43_Sulawesi
H Ctang_C116_Sulawesi
H Cdol_C36_Sulawesi
H Cfru_C46_Sulawesi
K Cqua_C156_Lobo
KL Cman_C37_Manokwari
I Cflo_C31_Flores
I Ckot_C65_Flores
I Clom_C19_Sumbawa
I Csum_C20_Sumba
G Cmur_C18_Java
JK Cmui_C160_Kumawa
K Cmaa_C119_Biak
L Ctank_C8_Espiritu
K Cluc_C68_MW700
KL Cnov_C49_Wewec
K Chad_C64_MW700
K Cjav_C78_Baitabag
K Cnob_C67_MW1200
K Cbus_C57_Wanang
L Cruf_C44_Guadalcanal
K Cniu_C54_Baitabag
L Cenk_C76_Butmas
40
30
20
10
0
Millions of years ago

1468

Appendix S9. Results of ancestral area estimation in BioGeoBEARS based on the most complex set of dispersal rate multipliers without time slices, two maximum ancestral areas and DIVALIKE speciation model.




BioGeoBEARS DIVALIKE analysis
ancstates: global optim, 2 areas max. d=0.0147; e=0.0066; j=0; LnL=−105.83
C Clam_C117_Sumatra
G Cern_C24_Java
C Ctel_C171_Siberut
C Clam_C105_Sumatra
F Cpel_C109_Borneo
F Cvar_C52_Borneo
D Crig_C77_Burauen
DH Cemp_C79_Leyte
A Cgut_C45_Hainan
D Ckon_C42_Luzon
J Chalm_C38_Halmahera
B Cadm_C84_Singapore
BF Cbor_C99_Tioman
A Ctha_C112_Thailand
B Csin_C12_Bukit
B Cret_C33_Tioman
E Cpal_C91_Turtle
H Ccon_C115_Sulawesi
H Coer_C43_Sulawesi
H Ctang_C116_Sulawesi
H Cdol_C36_Sulawesi
H Cfru_C46_Sulawesi
K Cqua_C156_Lobo
KL Cman_C37_Manokwari
I Cflo_C31_Flores
I Ckot_C65_Flores
I Clom_C19_Sumbawa
I Csum_C20_Sumba
G Cmur_C18_Java
JK Cmui_C160_Kumawa
K Cmaa_C119_Biak
L Ctank_C8_Espiritu
K Cluc_C68_MW700
KL Cnov_C49_Wewec
K Chad_C64_MW700
K Cjav_C78_Baitabag
K Cnob_C67_MW1200
K Cbus_C57_Wanang
L Cruf_C44_Guadalcanal
K Cniu_C54_Baitabag
L Cenk_C76_Butmas
40
30
20
10
0
Millions of years ago

1473

Appendix S10. Summary of diversification rate under BAMM analysis with gradient of prior values ranging from 0.1 to1.0. The gray line highlights the model with the highest statistical support.

poissonRatePrior	effectiveSize (postburn$N_shifts)	effectiveSize (postburn$logLik)	speciation rate	extinction rate
0.1	751.8863	706.4032	0.1242214	0.06180068
0.2	516.7889	646.3655	0.1201581	0.0575648
0.3	590.2721	404.8521	0.1207902	0.05841862
0.4	737.7893	105.6895	0.1222049	0.05950429
0.5	1088.348	970.286	0.1201945	0.05787543
0.6	612.6263	570.5625	0.1205574	0.05654984
0.7	743.6326	658.2969	0.120278	0.05735786
0.8	569.0992	483.0941	0.118358	0.0551307
0.9	590.9629	510.6484	0.119574	0.05630469
1.0	622.4329	630.815	0.118121	0.05380159