



HAL
open science

Understanding the origin of the most isolated endemic reef fish fauna of the Indo-Pacific: Coral reef fishes of Rapa Nui

Erwan Delrieu-Trottin, Laura Brosseau-Acquaviva, Stefano Mona, Valentina Neglia, Emily Giles, Cristian Rapu-Edmunds, Pablo Saenz-Agudelo

► To cite this version:

Erwan Delrieu-Trottin, Laura Brosseau-Acquaviva, Stefano Mona, Valentina Neglia, Emily Giles, et al.. Understanding the origin of the most isolated endemic reef fish fauna of the Indo-Pacific: Coral reef fishes of Rapa Nui. *Journal of Biogeography*, 2019, 46 (4), pp.723-733. 10.1111/jbi.13531 . hal-03923941

HAL Id: hal-03923941

<https://hal.science/hal-03923941v1>

Submitted on 10 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Understanding the origin of the most isolated endemic reef fish fauna of the Indo-Pacific: coral reef fishes of Rapa Nui

Erwan Delrieu-Trottin^{1,*}, Laura Brosseau-Acquaviva¹, Stefano Mona^{2,3},
Valentina Neglia¹, Emily C. Giles¹, Cristian Rapu-Edmunds⁴, and Pablo
Saenz-Agudelo¹

¹Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia, Chile

²Institut de Systématique, Evolution, Biodiversité (ISYEB), UMR 7205 - CNRS, MNHN, UPMC, EPHE, Ecole Pratique des Hautes Etudes, Paris Sorbonne Universités, Paris, France

³EPHE, PSL Research University, Paris, France^{[1][2]}

⁴Mike Rapu Diving Center, Caleta Hanga Roa O'tai, Isla de Pascua, Chile

**Corresponding author, email: erwan.delrieu.trottin@gmail.com*

Abstract

Aim: To understand the origin of the most isolated endemic fish fauna of the Indo-Pacific, Easter Island (Rapa Nui), and to infer divergence times and colonization routes of the endemic coral reef fish fauna from their closest relatives.

Location: Easter Island, Pacific Ocean.

Methods: Samples of ten species were used: six small-range species endemic to Rapa Nui and Motu Motiro Hiva (Salas y Gómez) (*i.e.* small-range endemic species) and four large range species endemic to the southern subtropical Pacific (*i.e.* large-range endemic species). We present phylogenetic reconstruction results based on mitochondrial (1 to 5) and nuclear (1 to 6) loci to place these endemic species in their respective family phylogenies (8). Using these newly calibrated phylogenetic trees, information of species distributions and geological data, we inferred the divergence times from the closest relatives of these ten endemic fishes, compared biogeographical history estimation models to reconstruct their ancestral geographic ranges, colonization routes, and underlying mechanisms of speciation.

Results: The divergence times (*i.e.* divergence times from the closest relatives) of all of the small-range endemics studied were more recent than the age of Rapa Nui and Motu Motiro Hiva; thus, these species can be considered as neoendemics. Biogeographical history estimation models indicated that speciation following a founder-event is the most likely scenario. In contrast, the divergence estimates of the large-range endemic species were highly variable. This being said, the divergence times of all species were more recent than the age of the oldest islands within their distributions.

Main conclusions: Taken together, these results demonstrate that Rapa Nui acts as a cradle of coral reef biodiversity, being responsible for the emergence of small-range endemic fish species, but also a route of dispersion for several large range endemics and as a stepping stone in the diversification of the *Myripristis* and *Pseudolabrus* genera. While no common divergence time was recovered for all of the ten endemic species studied here, the common mechanism of speciation following a founder event was recovered for most of the small-range endemic species.

KEYWORDS endemism, coral reef fishes, neoendemism, phylogeny, Easter Island.

1 | INTRODUCTION

One of the most striking aspects of coral reef fishes is their particularly uneven geographic distributions where species diversity is highest in the Indo Australian Archipelago (IAA) and tends to decline with latitude and longitude (Bellwood & Wainwright, 2002; Mora et al., 2003; Reaka et al., 2008; Bellwood & Meyer, 2009). Of particular interest are endemism hotspots, which are geographic regions that harbor an unusually high number of species that have restricted geographic distributions (Anderson, 1994) and that are found in the periphery of the IAA (Bellwood & Wainwright, 2002; Hughes et al., 2002; Allen, 2008). For instance, despite their relatively low species diversity, the Hawai'ian archipelago, Rapa Nui (Easter Island), the Marquesas Islands, and the Red Sea have an unusually high proportion of endemic species with 14% to 25% of their fish communities being endemics (Randall, 2005, 2007; Randall & Cea, 2011; Delrieu-Trottin et al., 2015). These particular regions constitute centers of endemism in the Indo-Pacific region and have been an important part in a long standing debate regarding the origin and current patterns of the biodiversity of Indo-Pacific coral reefs. From this debate two major hypotheses have emerged to explain the current distribution of coral reef biodiversity. One proposes the IAA as a Centre-of-Origin, a cradle of tropical marine biodiversity where lineages originate, and the other suggests that the IAA is rather a Centre-of-Accumulation where distinct faunas the Pacific and Indian Oceans overlap (Hughes et al., 2002; Connolly et al., 2003; Mora et al., 2003; Reaka et al., 2008). Understanding the evolutionary history of endemic species from peripheral centers of endemism can thus contribute to this ongoing debate.

Endemism by definition should be described within a specific spatial and temporal framework (Anderson, 1994). Spatially this is done by defining the geographic boundaries or ecological characteristics of the area that

encompass a specific species distribution. Temporally, endemic species have long been classified as either old or new endemics (Engler, 1882) and alternative hypothesis have been proposed to explain their current endemic status. One of these hypotheses suggests that endemic species occupy a restricted geographic area because they are young and newly established species (neoendemic species) (Willis, 1922). The alternative hypothesis suggests that endemic species are the remnants of an ancestral widespread species (paleoendemic species) (Stebbins & Major, 1965). Both of these hypotheses imply changes in the demographic history of the species that in turn should leave detectable genetic footprints (signals of expansion or contraction). These hypotheses are thus compatible with classical neutral models of speciation (allopatric, parapatric, and peripatric speciation). However, it has also been suggested that endemic species can arise by ecological speciation via local adaptation ("Gene pool – Ecological niche interaction hypothesis" sensu Stebbins (1980)). Under this hypothesis, the distribution of endemic species is restrained by the availability of specific ecological conditions. Endemics arising from ecological speciation could maintain or expand their range size throughout their evolutionary history as ecological conditions evolve through space and time (Kruckeberg & Rabinovitz, 1985; Anderson, 1994; Pigot et al., 2012). In practical terms, paleoendemic species should be distinguishable from neo or ecological endemics as these should display deep divergence times with signals of contraction. In contrast, neoendemics and ecological endemics are more challenging to tell apart (Cowman et al., 2017), and are not necessarily mutually exclusive. A species could be young and restricted to a specific geographic area because it arose as the product of local adaptation to particular environmental conditions.

Overall and despite the complexity of how endemic species originate, understanding whether endemic species are the product of recent speciation

or ancestral extinctions is of great importance to assess their likelihood of extinction (Myers et al., 2000). For instance, small range size is one of the characteristics that defines endemics and is also used to determine if a species is likely to be threatened or to become extinct (Pinheiro et al., 2017). In addition, regions that have unusual proportions of endemic species (centers of endemism) have been proposed as conservation priorities (Roberts et al 2002) because they can harbor both genetic novelty (neoendemics) and genetic history (paleoendemics). Thus, despite the fact that elucidating how a species originates remains challenging, elucidating their time of origin and their demographic history is a first step towards this end. It can also help to evaluate their risk of extinction and help to design and implement conservation strategies accordingly (Foote et al., 2007).

In general, little information is available regarding the evolutionary processes that have shaped the histories of endemic fish species and thus the distribution of fish fauna in and among centers of endemism. From an evolutionary point of view, these centers have been proposed to be either evolutionary cradles (neoendemism) or graveyards (paleoendemism) of tropical marine biodiversity (Bowen et al., 2013). However, the mechanisms by which these reservoirs of endemic fish species have been generated are not well characterized. Only a handful of studies have analysed divergence times and distribution of coral reef fishes to infer whether areas were hosting neoendemics or paleoendemics (e.g. Hodge et al., 2014, Cowman et al., 2017, Pinheiro et al., 2017). These studies have found that endemic species display a wide range of divergence times and so far support both the paleo and neoendemism hypothesis. However, whether these emerging patterns can be generalized remains an open question as divergence times for most endemic species still remain unknown.

Rapa Nui (Easter Island, Chile) is among the most remote islands on Earth and is the most isolated tropical island of the Pacific Ocean. Located at 27°07'10 S and 109°21'17 W in the Pacific, Rapa Nui lies 3,700 km from South America and 2,100 km east of the Pitcairn Islands. This small triangular island (166 km²) and the islet Motu Motiro Hiva (Salas y Gómez, 0.15 km²), located 400 km further east, are the easternmost islands of Polynesia. Rapa Nui and Motu Motiro Hiva constitute the only two emerged islands of the Easter Chain that also includes dozens of seamounts extending 2,232 km east until the Nazca seamount (23°36'S et 83°30'W) (Clouard & Bonneville, 2005; Ray et al., 2012). Due to a subtropical gyre, the mean current field in the Rapa Nui region is to the east, favoring the transportation of waters from southern Pacific islands to Rapa Nui (Martinez et al., 2009) and more locally from Rapa Nui to Motu Motiro Hiva (Andrade et al., 2014). The two islands are relatively young. Vezzoli & Acocella (2009) date Rapa Nui as being 0.8 My while Clouard & Bonneville (2005) date Rapa Nui as being 2.5 My and Motu Motiro Hiva as being 1.7 My. The fish fauna of these two islands, which mainly consists of coral reef fishes, is notable for its low diversity and high proportion of endemic species. Randall & Cea (2011) have identified a total of 169 fish species (139 shore fishes), this being an order of magnitude lower than that found in the Hawai'ian Islands (1,250 species) or in Indonesia (3,000) (Allen & Erdmann, 2012). 21.7% of the fish species found in Rapa Nui are endemic to this island (Randall & Cea, 2011; Cea, 2016). Overall, little is known regarding the evolutionary history of the island's fish fauna though three of the 30 endemic species of Rapa Nui are estimated to have diverged from their closest relatives 2.6 Mya (Cowman et al., 2017).

This study aims to shed light on the evolutionary history of the endemic reef fish fauna of Rapa Nui. In this context we refer to neoendemic species as those species with divergence times post-dating the origin of the island of

Rapa Nui (this means that most likely they originated on the island during or after its appearance; 2.5 My). We refer to paleoendemics as those species with divergence times that pre-date the formation of Rapa Nui (meaning that most likely they originated elsewhere, colonized the island and went extinct everywhere else). We anticipated that if Rapa Nui constitutes a cradle of coral reef biodiversity, i) divergence times for endemic species should post-date the formation of Rapa Nui; ii) we should find historical biogeographic evidence that supports founder-event speciation models, i.e. rare jump dispersal events at the origins of new genetically isolated lineages. Alternatively, finding divergence times for Rapa Nui endemic species older than the age of this island and no historical biogeographic support for founder-event speciation would indicate that the island represents a refuge of biodiversity that originated elsewhere. In the following sections we test these predictions to resolve the evolutionary histories of ten endemic species from eight different families. We place these species with their presumed sister species within currently available phylogenies constructed using mitochondrial and nuclear loci and reconstruct their biogeographic history. We discuss the potential underlying evolutionary processes that led to present day patterns.

2 | MATERIALS AND METHODS

2.1 | Selection of taxa and specimen collection

Randall & Cea (2011) have reported 30 fish species as endemic to Rapa Nui and 17 fish species as southern subtropical species. The defined Rapa Nui endemics are only present around Rapa Nui and Motu Motiro Hiva and have small-ranges (less than 500 km in linear distance, see Delrieu-Trottin et al., 2014). The southern subtropical species or regional endemics (see Friedlander et al., 2013) have large-ranges (1,000 – 8,000 km in linear distance, see Delrieu-Trottin et al., 2014) and are usually distributed from

Southern Polynesia to Rapa Nui. We referenced all of the presumed sister species of the 47 species of these two groups using species descriptions of endemic species and books of reference (Randall, 2005, 2007; Randall & Cea, 2011). To avoid overestimating divergence dates, it was essential that the phylogenies contained all the presumed sister species of each endemic species. Of the 47 species investigated, complete information was only available for nine species; four large-range endemic species: *Centropyge hotumatua* (Pomacanthidae), *Cheilodactylus plessisi* (Cheilodactylidae) *Myripristis tiki* (Holocentridae), *Pseudolabrus fuentesi* (Labridae) and five small-range endemic species: *Cantherhines rapanui* (Monacanthidae), *Chromis randalli*, *Chrysiptera rapanui* (Pomacentridae), *Coris debueni* (Labridae), and *Kuhlia nutabunda* (Kuhliidae). We added a tenth species, the small range endemic *Chaetodon litus* (Chaetodontidae), albeit no genetic data is available for one of its presumed sister species to obtain preliminary results regarding its evolutionary history (Table 1, Figure 1). Sampling was performed using pole-spears or an anesthetic (clove oil) while SCUBA diving around Rapa Nui in October 2016. Identification of collected specimens was done according to Randall & Cea (2011). Upon collection, a small piece of fin tissue was preserved in 96% EtOH.

2.2 | Laboratory procedures

Whole genomic DNA was extracted from fin tissue preserved in 96% EtOH at ambient temperature. DNA extraction was performed using the GeneJet Genomic DNA purification kit according to the manufacturer's protocols (Thermo Fisher Scientific). For each species, we amplified the same gene fragments using the corresponding primers as used in previously generated phylogenies. It is worth noting that we also generated molecular data for three presumed sister species, *Chrysiptera galba* (presumed sister

species of *C. rapanui*), *Chromis pamae* (presumed sister species of *C. randalli*) and *Coris roseovidis* (presumed sister species of *C. debueni*) to complete the dataset (see Table 2 for details). Fragments were amplified using PCR protocols and sequencing as described by Williams et al. (2012).

2.3 | Phylogeny

All generated sequences were deposited in GenBank (Accession numbers: MK100716- MK100761). Species were placed in the phylogenies of cheilodactylids (Burrige & Smolenski, 2004), holocentrids (Dornburg et al., 2012), kuhlids (Feutry et al., 2013), monacanthids (Santini et al., 2013), pomacanthids (Gaither et al., 2014), and pomacentrids (Frédérich et al., 2013) while several phylogenies were used to compose the dataset of chaetodontids (Cowman & Bellwood (2011) for eight of the nine genes and Gaboriau et al (2017) for COI). For labrids, we used the 12S, 16S, Tmo-4C4, RAG2, S7 and COI genes as described in Aiello et al (2017) and followed the methodology of Wainwright et al (2018): we included only a few representatives of all genera other than that of the target species (Table 2). This strategy allowed us to optimize computation time without compromising the accuracy of divergence date estimates.

We accessed sequence data stored in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) for each coral reef fish family using the R package seqinr (Charif & Lobry, 2007). The endemic species *Kuhlia nutabunda* and *Cheilodactylus plessisi* were already present in the phylogenies of their genera, so we did not add sequences for these species. Sequences were aligned with Clustal W (Thompson et al., 1994) or MAFFT (Katoh & Standley, 2013) and edited using Geneious v. 9.0.5 (<http://www.geneious.com>, Kearse et al. (2012)).

Maximum Likelihood (ML) analyses were performed for all of the families under study using the online version of IQ-TREE (Minh et al., 2013; Nguyen et al., 2015) available at <http://iqtree.cibiv.univie.ac.at> (Trifinopoulos et al., 2016). The best model of evolution of each gene was retrieved with ModelFinder (Kalyaanamoorthy et al., 2017) implemented in IQ-TREE using the Bayesian Information Criterion (BIC) prior to the construction of the ML trees (see Appendix S1 in Supporting Information Table S1). To assess branch support, all IQ-TREE analyses included the ultrafast bootstrap approximation (UFboot) with 1000 replicates (Minh et al., 2013) and the SH-like approximate likelihood ratio test (SH-aLRT) also with 1000 bootstrap replicates (Guindon et al., 2010). The time-calibrated phylogenies (BI) were constructed with the software BEAST2 v. 2.4.6 (Bouckaert et al., 2014) using the fossil calibrations of each phylogeny. We used secondary calibration points from the Actinopterigian super-phylogeny (Near et al., 2013) when studies did not include time-trees (Table S2). In the same way, we used tree priors (Yule model for all the analyses except for pomacentrids (Birth-Death model)) and partitioning (by individual molecular markers for all analysis except for pomacentrids (all genes concatenated)) according to the reference phylogenies while the substitution models were selected according to the ModelFinder results using the SSM v. 1.1.0 package in BEAUti (Bouckaert & Xie, 2017) Table S1). We used a relaxed log normal clock model as the molecular clock. The number of generations was set to reach analysis convergence (Table S2). We assessed the convergence and appropriate burn-in of each analysis using Tracer v. 1.5 (Drummond & Rambaut, 2007). Three independent analyses were run to ensure convergence. We constructed a maximum clade credibility tree using TreeAnnotator 1.8.1 (Drummond & Rambaut, 2007) to get median ages and 95% highest posterior density (HPD) intervals for each node. The 95% HPD represents the smallest interval that contains 95% of the posterior probability

and can be loosely thought of as a Bayesian analog to a confidence interval (Gelman et al., 2004).

2.4 | Phylogeography

We used the R package BioGeoBEARS (Matzke, 2013) in R (2017) to estimate the ancestral geographic ranges and to investigate the biogeographical history of the endemic species. This package includes three of the most common biogeographical history estimation models in a likelihood framework, i.e. the Dispersal-Extinction-Cladogenesis model (DEC; (Ree et al., 2005), DIspersal Vicariance Analysis (DIVA; (Ronquist, 1997) and the BayArea model (Landis et al., 2013). All of these models allow for anagenetic processes (range expansion and/or range contraction), but differ in the range of cladogenetic processes that are permitted (vicariance / sympatry) among/within-areas or single areas. As an example, all models allow for a sympatry event within one area, but only the DEC model allows for sympatry event within a subset of the range of a species. In the same manner, only the DEC and the DIVA models allow for a vicariance event within a subset of the range of a species, and only the DIVA model allows for a vicariance event resulting in the existence of species with multiple areas; but see the supermodel figure of Matzke for more details (2015). Here we assessed the fit of each of the different models using the corrected Akaike information criterion (AICc). The three models were tested including or not a “jump dispersal” parameter J that permits founder-event speciation in the species's history (Paulay & Meyer, 2002; Templeton, 2008). Founder-event speciation is the result of a long-distance colonization event of a small number of individuals founding a population that is then genetically isolated from the ancestral population; these events have been shown to be common in island systems (Matzke, 2014). BioGeoBEARS requires a clade including the endemic species and all of its closely related

species as an input. We used the maximum clade credibility tree from BEAST2 that we obtained for each family and extracted the clade of interest using the R package *ape* v.5.0 (Paradis et al., 2004). The geographic ranges of each species were coded with one or more areas.

Finally, given that the colonization of an island before its existence is technically impossible, and before its emergence is unlikely, we constrained the timing of dispersal as 1.7 Ma and 2.5 Ma for Motu Motiro Hiva and Rapa Nui respectively ('area allowed' parameter in BioGeoBEARS). Thus, this resulted in the removal of those areas in the ancestral range estimation of nodes that were older than the ages of these islands. We chose to implement the oldest age estimate for Rapa Nui to try to balance the fact that BioGeoBEARS cannot take into account the 95% highest posterior density (HPD) intervals of each node. We also constrained the timings of colonization of the Juan Fernandez Islands by *Pseudolabrus fuentesis* (2 Ma, Clouard & Bonneville, 2005) and of the Marquesas Islands by *Cantherhines*, *Kuhlia* and *Myripristis* (5.5 Ma, Clouard & Bonneville, 2005). For each analysis, we set the maximum number of ancestral areas to the maximum number of areas inhabited by a single species.

3 | RESULTS

3.1 | Phylogenetic analysis

We worked on aligned sequence matrices including 31 to 232 species per phylogeny and consisting of two to nine genes (932 to 7157 base pairs per matrix) (Table 2). The ML analyses consistently returned the same tree topology across all runs with strong bootstrap support at most nodes (Figures S1 - S8). Similarly, the Bayesian analyses produced the same tree topology across all three runs with high posterior probabilities (PP, Figures S9 - S16). For all analyses except those of Labridae and Pomacanthidae, parameters

reached effective sample sizes higher than 200 in the Bayesian inferences. For Labridae and Pomacanthidae, most parameters reached effective sample sizes higher than 200 and only few parameters had ESS within 100 and 200.

For all eight families under study, we found topologies congruent with those of the published reference phylogenies. For eight of the ten endemic species of this study, the Maximum Likelihood analysis and the Bayesian analysis retrieved the same topology. Differences among the ML and BI topologies were observed between *Cantherhines rapanui* and *Coris debueni* and their closest relatives.

Large-range endemic species were found among species-rich clades (*Cheilodactylus plessisi* and *Pseudolabrus fuentesi*) but also among species poor clades (*Centropyge hotumatua* and *Myripristis tiki*) while all small-range endemic species were part of species-rich clades. Several of the species presumed to be sister species based on morphology were confirmed by the molecular analysis (*Chromis randalli* / *C. pamae*; *Chrysiptera rapanui* / *C. galba*, *Kuhlia nutabunda* / *K. sandvichensis*, *Pseudolabrus fuentesi* / *P. gayi*).

Our molecular analysis helped to resolve five species complexes that to date are recognized as polytomies. First, our results clearly show that *Coris debueni* is closely related to *Coris roseoviridis*, a Southern endemic species distributed from the Cook Islands to Pitcairn (Table 1). Second, *Cantherhines rapanui* is part of a four-species complex that was poorly resolved by the ML analysis, yet the BI showed that *C. rapanui* is closely related to *C. longicaudus* (Cook Is. and Society Is.). Third, for *Cheilodactylus plessisi*, both the BI and ML analyses placed it as the closest relative to *C. vittatus* from Hawai'i. Fourth, by adding *Centropyge hotumatua* to the phylogeny of Pomacanthidae, we show that the position of *Centropyge hotumatua* is not within the 'multicolor' complex as previously

thought. Both the ML and BI placed *C. hotumatua* at the base of the ‘multicolor’ and ‘bispinosa’ complexes. The BI indicated that *C. hotumatua* is the closest relative of the Hawai’ian endemic *C. potteri*. Fifth and last, *Myripristis tiki* was not recovered as the closest relative of *M. amaena*. Instead both our BI and ML analyses provided strong support for *M. tiki* as the closest relative of *M. leiognathus*, a species distributed in the Eastern Tropical Pacific (from Gulf of California to Ecuador). Finally, acknowledging that the current available Chaetodontidae phylogeny does not include *Chaetodon smithi*, our results indicate that *Chaetodon litus* is the closest relative to the Hawai’ian endemic *C. miliaris* in a clade that also includes the West Pacific species *C. guentheri*.

3.2 | Dating divergence time between endemic species and their closest relatives

We generated time-trees for the eight families studied here, and these include the first-ever generated time-trees for Cheilodactylidae, Kuhliidae, and Holocentridae. For chaetodontids, labrids, monacanthids, pomacanthids, and pomacentrids, the divergence times that we obtained between species for the internal nodes of the maximum clade credibility trees were similar to those of the available published phylogenies. Overall, we obtained divergence times between endemic species and their closest relatives that ranged from 0.54 Ma (0.07 - 1.3, 95% HPD, *Cantherhines rapanui*) to 9.52 Ma (6.77 - 12.4, 95% HPD, *Myripristis tiki*) (Figure 1, Figures S9 - S16).

We observed that all of the divergence estimates for the small-range endemic species (*Cantherhines rapanui*, *Chaetodon litus*, *Chromis randalli*, *Chrysiptera rapanui*, *Coris debueni*, *Kuhlia nutabunda*) were less than the geological ages of Motu Motiro Hiva (1.7 Ma) and Rapa Nui (2.5 - 0.8 Ma). We observed a wider range of divergence times for the large-range endemic

species and species with divergence times older than Rapa Nui and Motu Motiro Hiva include *Cheilodactylus plessisi* with 2.80 Ma (0.88 - 4.00 % HPD), *Centropyge hotumatua* with 6.29 Ma (4.46 - 8.25 Ma 95% HPD) and *Myripristis tiki* with 9.52 Ma (6.77 - 12.4 Ma 95% HPD). Despite these results, we also found a large-range endemic species with divergence times that were younger than the geological ages of Motu Motiro-Hiva and Rapa Nui (*Pseudolabrus fuentesi* 2.21 Ma (0.88 - 4.21 % HPD)). None of the species under study had divergence times older than the oldest island that it inhabits (Australis: 12 Ma, Gambier Islands: 8 Ma, and Pitcairn: 1 Ma).

3.3 | Phylogeography

The BioGeoBEARS analyses of the maximum clade credibility trees allowed us to estimate the ancestral range and the evolutionary origin of the different endemic species (Figure S17 - S25). Interestingly, the majority (7/9) of the best-fitting models included a founder-event (+J) (BayAreaLike+J for *Myripristis tiki* and *Chrysiptera rapanui*; DIVALike+J for *Cantherhines rapanui*, *Coris debueni*, and *Cheilodactylus plessisi*; DEC+J for *Chromis randalli* and *Centropyge hotumatua*, Table S3).

Overall, BioGeoBEARS analyses showed that the best-fitting model was very often the one including a founder-event speciation process. Four of the small-range endemic species (*Chromis randalli*, *Chrysiptera rapanui*, *Coris debueni* and *Cantherhines rapanui*) were closely related to species distributed in the South Pacific Islands (Australis Islands, Gambier Islands, and Pitcairn) and likely emerged from a founder-event. *Pseudolabrus fuentesi* was closely related to an East Australian species and a Juan-Fernandez endemic species. *Cheilodactylus plessisi*, *Chaetodon litus* and *Centropyge hotumatua* were recovered as part of an antitropical clade with a close relative from Hawai'i. Finally, *Kuhlia nutabunda* was closely related to

a species largely distributed in the Pacific while *Myripristis tiki* likely emerged from a founder-event and was found to be related to an East Tropical Pacific species.

4 | DISCUSSION

This study represents the first attempt to better understand the origin of Rapa Nui endemic reef fishes including all their presumed sister species using newly computed time calibrated phylogenetic data and employing biogeographic analysis. Our results show that the coral reef fishes endemic to Rapa Nui have diverse evolutionary histories and two trends seem to emerge strongly related to the type of range distribution: (1) Species strictly endemic to Rapa Nui (small-range endemic species) have divergence dates with their closest relatives that are less than the geological age of Motu Motiro Hiva and Rapa Nui and thus can be considered as neoendemic species that likely emerged from a founder-event. Additionally, these species have sister species with restricted distributions only in the central south Pacific Islands (Austral Islands, Gambier Islands, Rapa Iti, Pitcairn). (2) Large-range endemic species have divergence times with their closest relatives that differ drastically (up to an order of magnitude), and their sister species are widely distributed among the Hawai'ian Islands, East Pacific, and Central Pacific.

The posterior distribution of divergence times (HPD) between endemic species and their sister species should be compared to the geological ages of the islands where endemic species are found. An HPD smaller or containing the values of the geological age of the island where the endemic species is found should support neoendemism while HPDs encompassing older ages should support paleoendemism. From our phylogenetic reconstruction, we found that most of the endemic species emerged since the Pliocene (5 Ma),

which was a period of high diversification of coral reef fishes in the Indo Australian Archipelago (Cowman & Bellwood, 2011; Alfaro et al., 2007; Klanten et al., 2004; Hodge et al., 2012, 2014). Geological data suggests that Rapa Nui emerged between 2.5 and 0.8 M ago while Motu Motiro Hiva is dated at 1.7 Ma (Clouard & Bonneville, 2005; Vezzoli & Acocella, 2009). Based on these dates, more than half of the species (6 out of 10) are strictly younger than Rapa Nui and most (8 out of 10) have HPDs that encompass the age of Rapa Nui. Additionally, we have found that only *Myripristis tiki* and *Centropyge hotumatua*, two large-range endemics, are older (both the median and HPD bounds) than Rapa Nui. Overall, the estimated divergence times between the small range endemic species and their sister species taken together with the geological data indicate that most of these species are likely neoendemics.

Founder-event speciation was expected for neoendemic species while range contraction was expected for paleoendemic species. Using phylogenies to disentangle the biogeographic histories of species is a complex task (Hedges, 2009), and estimating the internal nodes of the phylogenies generated here proved difficult. Despite this, the analyses indicated that most of the models selected (7/9 analysis) included a founder-event, which have been shown to be of prime importance in other island systems (Matzke, 2014; see Litsios et al., 2014; Pinheiro et al., 2017 and Wainwright et al., 2018 for examples). Overall, our biogeographic analyses show that dispersal and founder-event speciation played a major role in the colonization of Rapa Nui. For the small-range endemics, two speciation schemes can be recognized. First, a southern Pacific species diverged from a largely distributed species and then colonized and diverged in Rapa Nui via a founder event. This scenario is likely for *Coris debueni*, *Chrysiptera rapanui*, and *Chromis randalli*. All three species have a sister species distributed in several islands of the South Pacific (from Australs to Pitcairn), and the

emergence of the Rapa Nui archipelago was very likely followed by colonization and divergence. This scenario is also very likely for *Chaetodon litus* whose presumed sister species, the Southern endemic species *C. smithi* (Randall & Caldwell, 1973), is absent from our dataset. Second, a largely distributed species colonized and diverged in Rapa Nui through a founder event, a scenario that is likely for *Kuhlia nutabunda* and *Cantherhines rapanui*. This scenario differs from the first in that the largely distributed species (*Cantherhines longicaudus* and *Kuhlia sandvicensis*) colonized the southern islands of the Pacific but not Rapa Nui.

Finally, given the young age of Rapa Nui compared to the Austral Islands (20 Ma; Clouard & Bonneville, 2005), Rapa Nui has likely represented a new area for the colonization of the large-range endemic species included in this study (*Myripristis tiki*, *Pseudolabrus fuentesi*, *Centropyge hotumatua*, *Cheilodactylus plessisi*). Rapa Nui could have acted as a stepping stone in the diversification of the *Myripristis* and *Pseudolabrus* genera; *Myripristis tiki* and *Pseudolabrus fuentesi* are possible links between the fauna of the Tropical Eastern Pacific, of Juan Fernandez, and of the Desventuradas islands.

Local abundance and geographical range size are not correlated in marine fishes (Hobbs et al., 2011, Hughes et al., 2014). Indeed, many marine endemic reef fish are abundant (Hobbs et al., 2012; Delrieu-Trottin et al., 2014) and the local population size of small-range endemics can be an order of magnitude greater than the population size of widespread species at the same location (Hobbs, 2010). Similar to what has been found in Hawai'i (DeMartini & Friedlander, 2004; Kosaki et al., 2017) and the Marquesas islands (Delrieu-Trottin et al., 2015), the endemic fishes of Rapa Nui constitute a major component of the total assemblage. According to Friedlander et al. (2013), the small-range endemic species *Chromis randalli*, *Chrysiptera rapanui* and *Chaetodon litus* are among the most abundant

species in Rapa Nui, and here we show that all three have HPDs that are younger or encompassing the geological age of Rapa Nui and Motu Motiro-Hiva. In contrast, we found *Kuhlia nutabunda* to have a similar HPD, yet it is rather rare on the island. Overall, abundance does not seem to be a good proxy to differentiate neoendemism from paleoendemism in Rapa Nui as contrasting levels of abundance are found even among neoendemics.

Taken together, our results indicate that Rapa Nui acts as a cradle of coral reef biodiversity, being responsible of the emergence of small-range endemic fish species. We have found a common speciation mechanism for the small-range endemic species, highlighting the importance of founder events. Such results are of high importance from a conservation point of view: Rapa Nui and Motu Motiro-Hiva constitute centers of endemism in the Indo-Pacific region and are among the most unique marine environments in the Pacific. The creation of a Marine Protected Area in 2018 excluding extractive industries and industrial fishing is an important first step towards the conservation of this center of speciation.

ACKNOWLEDGEMENTS

We thank Rebeca Tepano, Nina, Taveke Olivares Rapu, Liza Garrido Toleado (SERNAPESCA), Ludovic Burns Tuki (Mesa del Mar), and the people of the Rapa Nui Island for their kind and generous support. This study was funded by FONDECYT Postdoctorado fellowship N°3160692 to E. Delrieu-Trottin and FONDECYT Iniciación fellowship N°11140121 to P. Saenz-Agudelo. The authors declare no conflict of interest. All applicable institutional guidelines for the care and use of animals were followed. Specimens were collected under permit No. 724, 8 March 2016 obtained from the Chilean Subsecretary of Fishing. The Universidad Austral de Chile Ethical Care Committee and Biosecurity Protocol approved our use and

handling of animals. Finally, we thank C. Riginos, H. Pinheiro and three anonymous reviewers for providing constructive reviews of earlier versions of the manuscript.

References

- Aiello, B. R., Westneat, M. W., & Hale, M. E. (2017). Mechanosensation is evolutionarily tuned to locomotor mechanics. *Proceedings of the National Academy of Sciences*, *114* (17), 4459–4464. doi: 10.1073/pnas.1616839114
- Alfaro, M. E., Santini, F., & Brock, C. D. (2007). Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (order tetraodontiformes). *Evolution*, *61*(9), 2104–2126. doi: 10.1111/j.1558-5646.2007.00182.x
- Allen, G. R. (2008). Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *18*(5), 541–556. doi: 10.1002/aqc.880
- Allen, G. R., & Erdmann, M. V. (2012). *Reef fishes of the east indies*. Honolulu: University of Hawaii Press.
- Anderson, S. (1994). Area and endemism. *The Quarterly Review of Biology*, *69*(4), 451–471. doi: 10.1086/418743
- Andrade, I., Hormazabal, S., & Correa Ramirez, M. (2014). Time space variability of satellite chlorophylla in the Easter Island Province, south-eastern Pacific Ocean. *Latin American Journal of Aquatic Research*, *42*(4), 871–887. doi: 10.3856/vol42-issue4-fulltext -13
- Bellwood, D. R., & Wainwright, P. C. (2002). The History and Biogeography of Fishes on Coral Reefs. In *Coral reef fishes* (pp. 5–32). doi: 10.1016/B978-012615185-5/50003 -7
- Bellwood, D. R., Klanten, S., Cowman, P. F., Pratchett, M. S., Konow, N., & Van Herwerden, L. (2010). Evolutionary history of the butterflyfishes (f: Chaetodontidae) and the rise of coral feeding fishes. *Journal of Evolutionary Biology*, *23*(2), 335–349. doi: 10.1111/j.1420-9101.2009.01904.x
- Bellwood, D. R., & Meyer, C. P. (2009). Searching for heat in a marine biodiversity hotspot. *Journal of Biogeography*, *36*(4), 569–576. doi: 10.1111/j.1365-2699.2008.02029.x
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D.,... Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, *10*(4). doi: 10.1371/journal.pcbi.1003537
- Bouckaert, R. & Xie, D. (2017). Standard Nucleotide Substitution Models v1.0.1. <http://doi.org/10.5281/zenodo.995740>
- Bowen, B. W., Rocha, L. A., Toonen, R. J., & Karl, S. A. (2013). The origins

- of tropical marine biodiversity. *Trends in Ecology and Evolution*, 28(6), 359–366. doi: 10.1016/j.tree.2013.01.018
- Burridge, C. P. (2004). *Cheilodactylus (Goniistius) francisi*, a new species of morwong (Perciformes: Cirrhitidae) from the Southwest Pacific. *Records of the Australian Museum*, 56(2), 231–234. doi: 10.3853/j.0067-1975.56.2004.1425
- Burridge, C. P., & Smolenski, A. J. (2004). Molecular phylogeny of the Cheilodactylidae and Latridae (Perciformes: Cirrhitidae) with notes on taxonomy and biogeography. *Molecular Phylogenetics and Evolution*, 30(1), 118–127. doi: 10.1016/S1055-7903(03)00157-X
- Cea, A. (2016). *Ika Rapa Nui*. Rapa Nui Press.
- Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In U. Bastolla, M. Porto, H. E. Roman, & M. Vendruscolo (Eds.), *Structural approaches to sequence evolution*, (pp. 207–232). New York: Springer Verlag.
- Clouard, V., & Bonneville, A. (2005). Ages of seamounts, islands, and plateaus on the Pacific plate. *Special Paper 388: Plates, plumes and paradigms*, 388, 71–90. doi: 10.1130/0-8137-2388-4.71
- Connolly, S. R., Bellwood, D. R., & Hughes, T. P. (2003). Indo-Pacific biodiversity of coral reefs: Deviations from a mid-domain model. *Ecology*, 84(8), 2178–2190. doi: 10.1890/02-0254
- Cowman, P. F., & Bellwood, D. R. (2011). Coral reefs as drivers of cladogenesis: Expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *Journal of Evolutionary Biology*, 24(12), 2543–2562. doi: 10.1111/j.1420-9101.2011.02391.x
- Cowman, P. F., Parravicini, V., Kulbicki, M., & Floeter, S. R. (2017). The biogeography of tropical reef fishes: endemism and provinciality through time. *Biological Reviews*, 92(4), 2112–2130. doi: 10.1111/brv.12323
- Delrieu-Trottin, E., Maynard, J., & Planes, S. (2014). Endemic and widespread coral reef fishes have similar mitochondrial genetic diversity. *Proceedings of the Royal Society London B*, 20141068.
- Delrieu-Trottin, E., Williams, J. T., Bacchet, P., Kulbicki, M., Mourier, J., Galzin, R., . . . Planes, S. (2015). Shore fishes of the Marquesas Islands, an updated checklist with new records and new percentage of endemic species. *Check List*, 11(5). doi: 10.15560/11.5.1758
- DeMartini, E., & Friedlander, A. (2004). Spatial patterns of endemism in shallow-water reef fish populations of the Northwestern Hawaiian

- Islands. *Marine Ecology Progress Series*, 271, 281–296. doi: 10.3354/meps271281
- Dornburg, A., Moore, J. A., Webster, R., Warren, D. L., Brandley, M. C., Iglesias, T. L., ... Near, T. J. (2012). Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei: Beryciformes: Holocentridae): Reconciling more than 100 years of taxonomic confusion. *Molecular Phylogenetics and Evolution*, 65(2), 727–738. doi: 10.1016/j.ympev.2012.07.020
- Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1), 214. doi: 10.1186/1471-2148-7-214
- Engler, A. 1882. *VersucheinerEntwick lungsgeschichte der Pflanzenwelt*. Leipzig: Engelmann
- Feutry, P., Castelin, M., Ovenden, J. R., Dettai, A., Robinet, T., Cruaud, C., & Keith, P. (2013). Evolution of Diadromy in Fish: Insights from a Tropical Genus (*Kuhlia* Species). *The American Naturalist*, 181(1), 52– 63.
- Frédérich, B., Sorenson, L., Santini, F., Slater, G. J., & Alfaro, M. E. (2013). Iterative Ecological Radiation and Convergence during the Evolutionary History of Damsel-fishes (Pomacentridae). *The American Naturalist*, 181(1), 94–113.
- Friedlander, A. M., Ballesteros, E., Beets, J., Berkenpas, E., Gaymer, C. F., Gorny, M., & Sala, E. (2013). Effects of isolation and fishing on the marine ecosystems of Easter Island and Salas y Gómez, Chile. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 23(4), 515–531. doi: 10.1002/aqc.2333
- Foote, M., Crampton, J.S., Beu, A.G., Marshall, B.A., Roger, A., Maxwell, P.A., & Matcham, I. (2007) Rise and Fall of Species Occupancy in Cenozoic Fossil Mollusks. *Science*, 318, 1131–1134.
- Gaboriau, T., Leprieur, F., Mouillot, D., & Hubert, N. (2018) Influence of the geography of speciation on current patterns of coral reef fish biodiversity across the Indo-Pacific. *Ecography*, 41, 1295–1306.
- Gaither, M. R., Schultz, J. K., Bellwood, D. R., Pyle, R. L., DiBattista, J. D., Rocha, L. A., & Bowen, B. W. (2014). Evolution of pygmy angelfishes: Recent divergences, introgression, and the usefulness of color in taxonomy. *Molecular Phylogenetics and Evolution*, 74(1), 38–47. doi: 10.1016/j.ympev.2014.01.017
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis*. doi: 10.1007/s13398-014-0173-7.2

- Greenfield, D. W. (1974). *A revision of the squirrelfish genus Myripristis Cuvier (Pisces, Holocentridae)* (No. 19). Natural History Museum, Los Angeles County.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phylml 3.0. *Systematic biology*, *59*(3), 307–321.
- Heads, M. (2009). Inferring biogeographic history from molecular phylogenies. *Biological Journal of the Linnean Society*, *98*(4), 757–774. doi: 10.1111/j.1095-8312.2009.01320.x
- Hobbs, J.-P. A. (2010). Poleward range expansion of a tropical coral reef fish (*Centropyge flavissima*) to Lord Howe Island, Australia. *Marine Biodiversity Records*, *3*, 1–2. doi: 10.1017/S1755267210000990
- Hobbs, J. P. A., Jones, G. P., & Munday, P. L. (2011). Extinction Risk in Endemic Marine Fishes. *Conservation Biology*, *25*(5), 1053–1055. doi: 10.1111/j.1523-1739.2011.01698.x
- Hobbs, J. P. A., Jones, G. P., Munday, P. L., Connolly, S. R., & Srinivasan, M. (2012). Biogeography and the structure of coral reef fish communities on isolated islands. *Journal of Biogeography*, *39*(1), 130–139. doi: 10.1111/j.1365-2699.2011.02576.x
- Hodge, J. R., Read, C. I., van Herwerden, L., & Bellwood, D. R. (2012). The role of peripheral endemism in species diversification: Evidence from the coral reef fish genus *Anampses* (Family: Labridae). *Molecular Phylogenetics and Evolution*, *62*(2), 653–663. doi: 10.1016/j.ympev.2011.11.007
- Hodge, J. R., van Herwerden, L., & Bellwood, D. R. (2014). Temporal evolution of coral reef fishes: Global patterns and disparity in isolated locations. *Journal of Biogeography*, *41*(11), 2115–2127. doi: 10.1111/jbi.12356
- Hughes, T. P., Bellwood, D. R., & Connolly, S. R. (2002). Biodiversity hotspots, centres of endemism, and the conservation of coral reefs. *Ecology Letters*, *5*(6), 775–784. doi: 10.1046/j.1461-0248.2002.00383.x
- Hughes, T.P., Bellwood, D.R., Connolly, S.R., Cornell, H. V., & Karlson, R.H. (2014) Double jeopardy and global extinction risk in corals and reef fishes. *Current Biology*, **24**, 2946–2951.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. doi:

10.1038/nmeth.4285

- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. doi: 10.1093/molbev/mst010
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. doi: 10.1093/bioinformatics/bts199
- Klanten, S. O., van Herwerden, L., Choat, J. H., & Blair, D. (2004). Patterns of lineage diversification in the genus *Naso* (Acanthuridae). *Molecular Phylogenetics and Evolution*, *32*(1), 221–235.
- Kosaki, R. K., Pyle, R. L., Leonard, J. C., Hauk, B. B., Whitton, R. K., & Wagner, D. (2017). 100% endemism in mesophotic reef fish assemblages at Kure Atoll, Hawaiian Islands. *Marine Biodiversity*, *47*(3), 783–784. doi: 10.1007/s12526-016-0510-5
- Kruckeberg AR, Rabinovitz D (1985) Biological aspects of endemism in higher plants. *Annual review of Ecology and Systematics*, *16*, 447–479
- Landis, M. J., Matzke, N. J., Moore, B. R., & Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, *62*(6), 789–804. doi: 10.1093/sysbio/syt040
- Litsios, G., Pearman, P. B., Lanterbecq, D., Tolou, N., & Salamin, N. (2014). The radiation of the clownfishes has two geographical replicates. *Journal of Biogeography*, *41*(11), 2140–2149. doi: 10.1111/jbi.12370
- Martinez, E., Maamaatuaiahutapu, K., & Taillandier, V. (2009). Floating marine debris surface drift: Convergence and accumulation toward the South Pacific subtropical gyre. *Marine Pollution Bulletin*, *58*(9), 1347–1355. doi: 10.1016/j.marpolbul.2009.04.022
- Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.*, *5*(4), 242–248.
- Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, *63*(6), 951–970. doi: 10.1093/sysbio/syu056
- Matzke, N.J. (2015). Stochastic mapping under biogeographical models. Available at: http://phylo.wikidot.com/biogeobears#stochastic_mapping (accessed 11 September 2018).

- Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195.
- Mora, C. (2015). *Ecology of fishes on coral reefs*. Cambridge University Press.
- Mora, C., Chittaro, P. M., Sale, P. F., Kritzer, J. P., Ludsin, S. a., & Africa, S. (2003). Patterns and processes in reef fish diversity. *Nature*, *421*, 933–936. doi: 10.1038/nature01421.1.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Near, T. J., Dornburg, A., Eytan, R. I., Keck, B. P., Smith, W. L., Kuhn, K. L, ... Wainwright, P. C. (2013). Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences*, *110*(31), 12738–12743. doi: 10.5061/dryad.d3mb4
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. doi: 10.1093/molbev/msu300
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. doi: 10.1093/bioinformatics/btg412
- Paulay, G., & Meyer, C. (2002). Diversification in the tropical pacific: comparisons between marine and terrestrial systems and the importance of founder speciation. *Integrative and Comparative Biology*, *42*(5), 922–934.
- Pigot, A.L., Owens, I.P.F., & Orme, C.D.L. (2012) Speciation and Extinction Drive the Appearance of Directional Range Size Evolution in Phylogenies and the Fossil Record. *PLoS Biology*, **10**, e1001260.
- Pinheiro, H.T., Bernardi, G., Simon, T., Joyeux, J.-C., Macieira, R.M., Gasparini, J.L., Rocha, C., & Rocha, L.A. (2017) Island biogeography of marine organisms. *Nature*, **549**, 82.
- Pyle, R. L. (2003). *A systematic treatment of the reef-fish family Pomacanthidae (Pisces: Perciformes)*. University of Hawaii at Manoa.
- R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. doi: [http:// www.R-project.org/](http://www.R-project.org/)
- Randall, J. E. (2005). *Reef and Shore Fishes of the South Pacific : New Caledonia to Tahiti and the Pitcairn* (Vol. 1). Honolulu: University of

- Hawaii Press.
- Randall, J. E. (2007). *Reef fishes of Hawaii*. Honolulu: University of Hawaii Press.
- Randall, J. E., & Caldwell, D. K. (1973). *A new butterflyfish of the genus chaetodon and a new angelfish of the genus centropyge from easter island*. Natural History Museum, Los Angeles County.
- Randall, J. E., & Cea, A. (2011). *Shore Fishes of Easter Island*. University of Hawai'i Press.
- Ray, J. S., Mahoney, J. J., Duncan, R. A., Ray, J., Wessel, P., & Naar, D. F. (2012). Chronology and geochemistry of lavas from the Nazca Ridge and Easter Seamount Chain: An ~30 myr hotspot record. *Journal of Petrology*, 53(7), 1417–1448. doi: 10.1093/petrology/egs021
- Reaka, M. L., Rodgers, P. J., & Kudla, A. U. (2008). Patterns of biodiversity and endemism on Indo-West Pacific coral reefs. *Proceedings of the National Academy of Sciences of the United States of America*, 105 Suppl, 11474–81. doi: 10.1073/pnas.0802594105
- Ree, R. H., Moore, B. R., Webb, C. O., & Donoghue, M. J. (2005). A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, 59(11), 2299. doi: 10.1554/05-172.1
- Roberts, C.M. (2002) Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reefs. *Science*, 295, 1280–1284.
- Ronquist, F. (1997). Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, 46(1), 195–203. doi: 10.1093/sysbio/46.1.195
- Russell, B. C., & Randall, J. E. (1980). The Labrid Fish Genus *Pseudolabrus* from Islands of the Southeastern Pacific, with Description of a New Species from Rapa. *Pacific Science*, 34(4), 433–440.
- Santini, F., Sorenson, L., & Alfaro, M. E. (2013). A new multi-locus timescale reveals the evolutionary basis of diversity patterns in triggerfishes and filefishes (Balistidae, Monacanthidae; Tetraodontiformes). *Molecular Phylogenetics and Evolution*, 69(1), 165–176. doi: 10.1016/j.ympev.2013.05.015
- Siu, G., Bacchet, P., Bernardi, G., Brooks, A., Carlot, J., Claudet, J., ... Galzin, R. (2017). Shore fishes of French polynesia. *Cybium*, 41(3), 1–34.
- Stebbins, G. L. (1980). Rarity of plant species: a synthetic viewpoint. *Rhodora*, 82(829), 77–86.
- Stebbins, G. L., & Major, J. (1965). Endemism and speciation in the California flora. *Ecological Monographs*, 35(1), 1–35.

- Templeton, A. R. (2008). The reality and importance of founder speciation in evolution. *Bioessays*, 30(5), 470–479.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680. doi: 10.1093/nar/22.22.4673
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B.Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44, W232–W235.
- Vezzoli, L., & Acocella, V. (2009). Easter Island, SE Pacific: An end-member type of hotspot volcanism. *GSA Bulletin*, 121(5-6), 869–886.
- Victor, B. C. (2015). How many coral reef fish species are there? Cryptic diversity and the new molecular taxonomy. *Ecology of Fishes on Coral Reefs*. Cambridge University Press, Cambridge, United Kingdom, 76–87.
- Wainwright, P. C., Santini, F., Bellwood, D. R., Ross Robertson, D., Rocha, L. A., & Alfaro, M. E. (2018). Phylogenetics and geography of speciation in New World Halichoeres wrasses. *Molecular Phylogenetics and Evolution*, 121, 35–45. doi: 10.1016/j.ympev.2017.12.028
- Westneat, M. W., & Alfaro, M. E. (2005). Phylogenetic relationships and evolutionary history of the reef fish family Labridae. *Molecular Phylogenetics and Evolution*, 36(2), 370–390. doi: 10.1016/j.ympev.2005.02.001
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York,
- Williams, J. T., Delrieu-Trottin, E., & Planes, S. (2012). A new species of Indo-Pacific fish, *Canthigaster criobe*, with comments on other *Canthigaster* (Tetraodontiformes: Tetraodontidae) at the Gambier Archipelago. *Zootaxa*, 3523(3523), 80–88.
- Willis, J. C. (1922). *Age and Area - A Study in Geographical Distribution and Origin of Species*. Cambridge, UK: Cambridge University Press.

BIOSKETCHES

Erwan Delrieu-Trottin is broadly interested in molecular ecology and the evolution of fishes. Stefano Mona and Pablo Saenz-Agudelo are interested broadly in population genetics. Laura Brosseau-Acquaviva and Cristian Rapu Edmunds are interested in conservation. Valentina Neglia and Emily Giles are interested in the ecology and molecular ecology of marine organisms.

Author contributions: EDT and PSA conceived the project; EDT, VN, CRE, ECG and PSA designed and conducted sampling; EDT and LBA performed molecular experiments; EDT and LBA analysed and interpreted the data; EDT wrote the first draft and all co-authors contributed.

DATA ACCESSIBILITY

These sequence data have been submitted to the GenBank databases under accession numbers MK100716- MK100761. All the .tree files from ML and Bayesian analyses are available on FigShare (<https://figshare.com/s/e09f85988fca07e4b2ba>).

SUPPORTING INFORMATION

Additional Supporting Information can be found online in the supporting information tab for this article.

FIGURE 1 Geographic distribution of the species under study and divergence times from their closest relatives. (a) Geographic distributions of large-range endemic species and of (b) small-range endemic species in this study. The large-range endemics have larger distributions while the small-range endemics are only present around Rapa Nui and Motu Motiro Hiva (MMH) (see Table 1 for detailed distributions). The map was generated using the R package ggplot2 (Wickham, 2009) and drawn with rectangular projection (c) Divergence times of species endemic to Rapa Nui from their closest relatives. Median (dot) and 95 % HPD (lines) are represented. Colors denote the distribution. Vertical dotted lines represent the age of the oldest emerged island of the archipelago (Austral Islands: 12 Ma, Gambier Islands: 8 Ma and Pitcairn: 1 Ma Clouard & Bonneville (2005); Vezzoli & Acocella (2009)). An asterix next to the species name indicates that there was strong statistical support for jump dispersal according to the Biogeographical history estimation models.

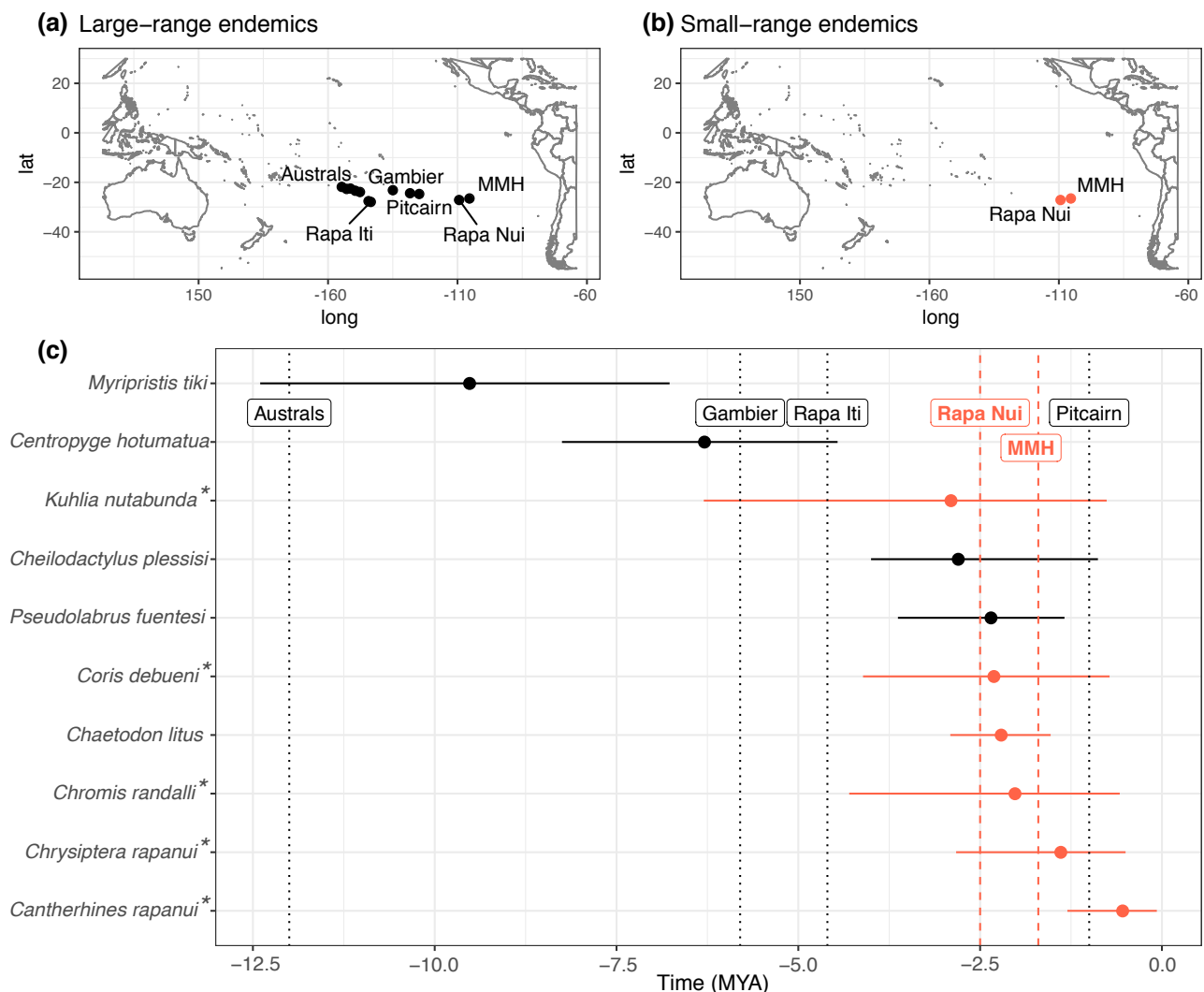


TABLE 1 Distribution of the endemic species (Rapa Nui name) under study and of their presumed sister species based on morphology.
 RN: Rapa Nui, MMH: Motu Motiro Hiva, RI: Rapa Iti

Species & Distribution	Sister species
<i>Cantherhines rapanui</i> (koreva) RN + MMH	Species complex with <i>C. fronticictus</i> (East coast of Africa to the Marshall Is.), <i>C. verecundus</i> (Hawai'ian Is.), <i>C. longicaudus</i> (Cook Is. and Society Is.) (Randall, 2005; Siu et al. 2017)
<i>Chromis randalli</i> (mamata) RN + MMH	<i>Chromis pamae</i> (Austral Is., and RI, Gambier Is., Pitcairn) (Randall, 2005)
<i>Chaetodon litus</i> (tipi tipu 'uri) RN + MMH	<i>C. smithi</i> (RI and Pitcairn) (Randall & Caldwell, 1973)
<i>Coris debueni</i> (teteme) RN + MMH	Species complex with <i>C. caudimaculata</i> (Indian Ocean), <i>C. dorsomaculata</i> (western Pacific), <i>C. roseoviridis</i> (Cook Is. to Pitcairn), <i>C. venusta</i> (Hawai'ian Is.) (Randall & Cea, 2011)
<i>Kuhlia nutabunda</i> (mahore) RN + MMH	<i>K. sandvicensis</i> (Pacific Ocean) (Randall & Cea, 2011)

Chrysiptera rapanui
(mamata)
RN + MMH

Chrysiptera galba (Austral Is., Gambier Is. and Pitcairn) (Randall, 2005)

Centropyge hotumatua
(kototi para)
RN+ MMH + Pitcairn + RI+ Australs

Species complex with *C. multicolor* (Central Pacific), *C. interrupta* (Japan and Hawai'ian Is.), *C. jocular* (Coco & Christmas Is.), *C. nahackyi* (Johnston atoll), *C. debelius* (Mauritius) (Pyle, 2003)

Cheilodactylus plessisi
(ra'ea)
RN + MMH + RI

C. zebra (Japan), *C. francisi* (Lord Howe Is. to New Caledonia) & *C. vittatus* (Hawai'ian Is.) Burridge, 2004)

Myripristis tiki
(marau)
RN + MMI + Pitcairn + RI

M. amaena (Japan to Pitcairn) (Greenfield, 1974)

Pseudolabrus fuentesis
(kotea)
RN + MMH+ Pitcairn + RI+ Australs

Peudolabrus gayi (Juan Fernandez Is. and Desventudaras Is.) (Russell & Randall, 1980)

TABLE 2 Loci used and total number of species included in each phylogeny to place endemic species in their respective phylogenies. In bold are loci that amplified successfully (Accession numbers MK100716- MK100761). In italics are loci that were already published. Phylogenies of Cowman & Bellwood (2011)¹; Gaboriau et al (2017)²; Burrige & Smolenski (2004)³; Dornburg et al. (2012)⁴; Feutry et al. (2013)⁵; Westneat & Alfaro (2005)⁶; Aiello et al. (2017)⁷; Santini et al. (2013)⁸; Gaither et al. (2014)⁹; Frédéricich et al. (2013)¹⁰.

Family	Genes	Nb of sp included
<i>Chaetodon litus</i> ¹	12S, 16S, COI, Cytb, ETS2, ND3, Rag2, S7, TMO4C4	122
<i>Cheilodactylus plessisi</i> ³	<i>COI, Cytb</i>	30
<i>Myripristis tiki</i> ⁴	COI, ENC1, Glyt, Myh6, Ptr, Rag1, Sreb	42
<i>Kuhlia nutabunda</i> ⁵	<i>COI, Rh, 16S, Tmo4c4, IRBP</i>	37
<i>Coris debueni</i> ^{6,7}	12S, 16S, Rag2, COI, Tmo4c4	105
<i>Coris roseoviridis</i> ^{6,7}	12S, 16S, Rag2, COI, Tmo4c4	105
<i>Pseudolabrus fuentesi</i> ^{6,7}	12S, 16S, Rag2, COI, Tmo4c4	105
<i>Cantherhines rapanui</i> ⁸	COI, Cytb, Myh6, Rag1, Rh	94
<i>Centropyge hotumatua</i> ⁹	COI, Cytb, Rag2, S7, Tmo4C4	121
<i>Chromis randalli</i> ¹⁰	COI, Cytb, 12S, 16S, Bmp4, Nd3, Rag1, Rag2	163
<i>Chrysiptera rapanui</i> ¹⁰	COI, Cytb, 12S, 16S, Bmp4, Nd3, Rag1, Rag2	163
<i>Chrysiptera galba</i> ¹⁰	COI, Cytb, 12S, 16S, Bmp4, Nd3, Rag1, Rag2	163
<i>Chromis pamae</i> ¹⁰	COI, Cytb, 12S, 16S, Bmp4, Nd3, Rag1, Rag2	163

Journal of Biogeography

SUPPORTING INFORMATION

**Understanding the origin of the most isolated endemic reef fish
fauna of the Indo-Pacific: coral reef fishes of Rapa Nui**

Erwan DELRIEU-TROTTIN^{1,*}, Laura BROSSEAU-ACQUAVIVA¹, Stefano MONA^{2,3}, Valentina NEGLIA¹, Emily GILES¹, Cristian RAPU-EDMUNDS⁴, and Pablo SAENZ-AGUDELO¹

¹*Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia, Chile*

²*Institut de Systematique, Evolution, Biodiversite (ISYEB), UMR 7205 - CNRS, MNHN, UPMC, EPHE, Ecole Pratique des Hautes Etudes, Paris Sorbonne Universites, Paris, France*

³*EPHE, PSL Research University, Paris, France*

⁴*Mike Rapu Diving Center, Caleta Hanga Roa O'tai, Isla de Pascua, Chile*

**Corresponding author, email: erwan.delrieu.trottin@gmail.com*

Appendix S1 Supplementary tables to provide informations regarding the Maximum Likelihood (ML) and Bayesian inferences analyses (Tables [S1](#) and [S2](#)) and the BioGeoBEARS (Table [S3](#)). ML trees (Figures [S1](#) - [S8](#)) and time trees (Figures [S9](#) - [S16](#)) to place the 10 endemic species in their respective phylogenies. Estimates of ancestral ranges for every species under study (Figures [S17](#) - [S25](#)).

List of Tables

S1	The best model of evolution of each gene as selected retrieved using the Bayesian Information Criterion (BIC) in ModelFinder implemented in IQ-TREE	4
S2	Age constraints used for age estimation in BEAST using fossil calibrations of each phylogeny of reference (see main text). We used secondary calibration points from the Actinopterigian super-phylogeny (Near et al. 2013) when studies did not include time-trees.	6
S3	AICc results from the BioGeoBEARS model comparisons. The first three models with best AICc are displayed. Analysis not computed for <i>Chaetodon litus</i> as its phylogeny is missing a presumed sister species, namely <i>Chaetodon litus</i>	7

Table S1: The best model of evolution of each gene as selected retrieved using the Bayesian Information Criterion (BIC) in ModelFinder implemented in IQ-TREE

	Gene	Substitution models	Invariable sites	Gamma
Labridae	12S	GTR+I+G	0.353431	0.693969
	16S	TIM2+I+G	0.410218	0.576535
	RAG2	TIM2e+I+G	0.293299	1.15982
	Tmo4c4	TIM2e+I+G	0.495671	1.24214
	COI	TPM2+I+G	0.620058	1.49506
Kulhiidae	16S	K2P+I+G	0.660725	0.672442
	COI	HKY+G	-	0.116777
	IRBP	K2P+G	-	0.162922
	RH	HKY+I	0.890448	-
	TMO4C4	K2P	-	-
Holocentridae	COI	TVM+G	-	0.125531
	ENC1	TNe+G	-	0.182762
	Glyt	K2P+G	-	0.278142
	MYH6	TNe+I+G	0.60426	0.725441
	PTR	TNe+I+G	0.678241	0.908126
	RAG1	K2P+G	-	0.25112
	SREB	HKY+I	0.867407	-
Cheilodactylidae	COI	TN+I+G	0.637556	1.35744
	Cytb	HKY+I+G	0.0564855	0.207658
Chaetodontidae	12S	GTR+I+G	0.389997	0.747254
	16S	TIM2+I+G	0.448689	0.523588
	COI	TIM2+I+G	0.572826	1.29948
	CYTB	TN+I+G	0.387876	0.803337
	ETS2	K2P+G	-	0.75847
	ND3	TN+I+G	0.427064	0.834432
	RAG2	TNe+I+G	0.344559	0.878103
	S7	HKY+I+G	0.090315	4.14272
	TMO4C4	K2P+I+G	0.418497	0.798892
Monacanthidae	COI	TN+I+G	0.556395	0.780737
	CytB	TIM2+I+G	0.470359	0.771422
	myh6	TIMe+I+G	0.455102	0.726666
	RAG1	TIM2+I+G	0.381691	0.844701
	Rh	HKY+I+G	0.59121	0.910396
Pomacanthidae	COI	TPM2+I+G	0.608247	1.53093
	Cytb	HKY+I+G	0.516902	1.08296
	Rag2	K2P+I+G	0.476705	0.762621

	S7	K2P+G	-	1.21254
	Tmo4c4	K2P+I	0.782091	-
Pomacentridae	16S	TIM2+I+G	0.381917	0.732404
	Bmp4	TPM2+I+G	0.485169	0.897711
	Cytb	TN+I+G	0.500194	0.656067
	ND3	TN+I+G	0.601689	1.08341
	Rag1	GTR+I+G	0.458773	0.798158
	12S	TIM3+I+G	0.417452	0.904553
	COI	TIM+I+G	0.562566	0.908294
	Rag2	K2P+I+G	0.375754	0.829911

Table S2: Age constraints used for age estimation in BEAST using fossil calibrations of each phylogeny of reference (see main text). We used secondary calibration points from the Actinopterigian super-phylogeny (Near et al. 2013) when studies did not include time-trees.

Family	MCMC length (millions)	Age constrains
Chaetodontidae	40	Root (Scat vs Chaetodontidae-Pomacanthidae): 50 Ma (Lognormal: M = 1.5; S = 1.5; Offset = 50); Crown Chaetodontidae: 30 Ma (Lognormal: M = 1.0; S = 1.0; Offset = 30); Crown <i>Chaetodon</i> : 16 Ma (Normal: M = 1.0; S = 8.0; Offset = 16)
Cheilodactylidae	15	Secondary calibration using Near et al. 2013 phylogeny. Crown Cheilodactylidae : 25 Ma (Lognormal: M=0.5; S=1.25; Offsset = 25.0)
Holocentridae	30	Secondary calibration with Near et al. 2013. Crown Holocentridae : 50 Ma (Lognormal: M=0.672; S=0.8; Offset=50.0)
Kuhliidae	50	Secondary calibration using Near et al. 2013 phylogeny. Crown Kuhliidae: 25 Ma (Lognormal: M=1.0; S=1.25; Offset = 25.0); Divergence between <i>Kuhlia malo</i> , <i>salelea</i> , <i>marginata</i> clades and <i>Kuhlia rupestris</i> & <i>sauvagii</i> : 5,2 Ma (Exponential: M=1.6; Offset=5.2)
Labridae	80	Root (crown Labridae): 61.5 (Lognormal: M=0,8; S=0,9; Offset 61.5); Hypsigenyines: 50 Ma (Exponential: M=11; offset 50); Labridae (- Hypsigenyines): 50 Ma (Exponential: M = 12; Offset = 50); Pseudodax - Bodianus: 14 Ma (Exponential: M = 9.0; Offset = 14)
Monacanthidae	50	Root: 59 Ma (Exponential: M = 6.5; Offset=59.0); crown balistoids: 41 Ma (Exponential: M = 3.0; Offset = 41); crown molids: 22 Ma (Exponential: M = 6.5; Offset = 22.0)
Pomacanthidae	30	Root: 50 Ma (Exponential: M = 1.0; Offset = 50); Crown Pomacanthids: 38 Ma (Normal: M = 38; S=2.0; Offset = 0.0)
Pomacentridae	60	Root: 74 Ma (Exponential: M = 8.0; Offset = 74); crown pomacentrids (Exponential: M = 12.0; Offset = 50); crown <i>Chromis</i> (Exponential: M = 11.0; Offset = 6.5)

Table S3: AICc results from the BioGeoBEARS model comparisons. The first three models with best AICc are displayed. Analysis not computed for *Chaetodon litus* as its phylogeny is missing a presumed sister species, namely *Chaetodon litus*

Family	Species	AICc
Chaetodontidae	<i>Chaetodon litus</i>	-
Cheilodactylidae	<i>Cheilodactylus plessisi</i>	DIVALike+J : 72.71 DEC+J : 73.9 BayAreaLike+J : 75.36
Holocentridae	<i>Myripristis tiki</i>	BayAreaLike + J : 111.1 BayAreaLike : 123.2 DEC+ J : 129.9
Kuhliidae	<i>Kuhlia nutabunda</i>	BayAreaLike : 57.59 DEC : 60.06 DIVALike : 60.08
Labridae	<i>Pseudolabrus fuentesi</i>	DIVALike : 63.31 DEC : 64.95 DIVALike+J : 65.36
Labridae	<i>Coris debueni</i>	DIVALike+J : 49.05 DEC+J : 50.67 BayAreaLike+J : 51.03
Monacanthidae	<i>Cantherhines rapanui</i>	DIVALike + J : 93.41 DEC + J : 94.77 BayAreaLike+J : 95.13
Pomacanthidae	<i>Centropyge hotumatua</i>	DEC+J : 96.07 DEC : 96.98 BayAreaLike+J : 97.95
Pomacentridae	<i>Chrysiptera rapanui</i>	BayAreaLike+J : 61.07 DIVALike : 66.37 DIVALike + J : 68.35
Pomacentridae	<i>Chromis randalli</i>	DEC+J : 54.86 DIVALike+J : 55.2 BayAreaLike+J : 56.51

List of Figures

S1	Maximum-Likelihood tree for Chaetodontidae. Numbers on nodes denote ultrafast bootstrap approximation (UFboot) values with 1000 replicates and the SH-like approximate likelihood ratio test (SH-aLRT) also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.	11
S2	Maximum-Likelihood tree for Cheilodactylidae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90. . . .	12
S3	Maximum-Likelihood tree for Holocentridae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.	13
S4	Maximum-Likelihood tree for Kuhliidae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.	14
S5	Maximum-Likelihood tree for Labridae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.	15
S6	Maximum-Likelihood tree for Monacanthidae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90. . . .	16
S7	Maximum-Likelihood tree for Pomacanthidae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90. . . .	17
S8	Maximum-Likelihood tree for Pomacentridae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90. . . .	18
S9	Timetree for the Chaetodontidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	20
S10	Timetree for the Cheilodactylidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	21
S11	Timetree for the Holocentridae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	22
S12	Timetree for the Kuhliidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	23
S13	Timetree of of Labridae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	25
S14	Timetree for the Monacanthidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	26
S15	Timetree for the Pomacanthidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	28
S16	Timetree for the Pomacentridae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.	29

S17	Estimates of ancestral ranges for <i>Cheilodactylus</i> spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	30
S18	Estimates of ancestral ranges for <i>Chromis</i> spp using DEC + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	31
S19	Estimates of ancestral ranges for <i>Chrysispera</i> spp using BAYAAREA + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	32
S20	Estimates of ancestral ranges for <i>Myripristis</i> spp using BAYAREA + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	33
S21	Estimates of ancestral ranges for <i>Kuhlia</i> spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	34
S22	Estimates of ancestral ranges for <i>Coris</i> spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	35
S23	Estimates of ancestral ranges for <i>Pseudolabrus</i> spp using DIVALIKE model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	36
S24	Estimates of ancestral ranges for <i>Cantherhines</i> spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.	37

S25 Estimates of ancestral ranges for *Centropyge* spp using DEC + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis. 38

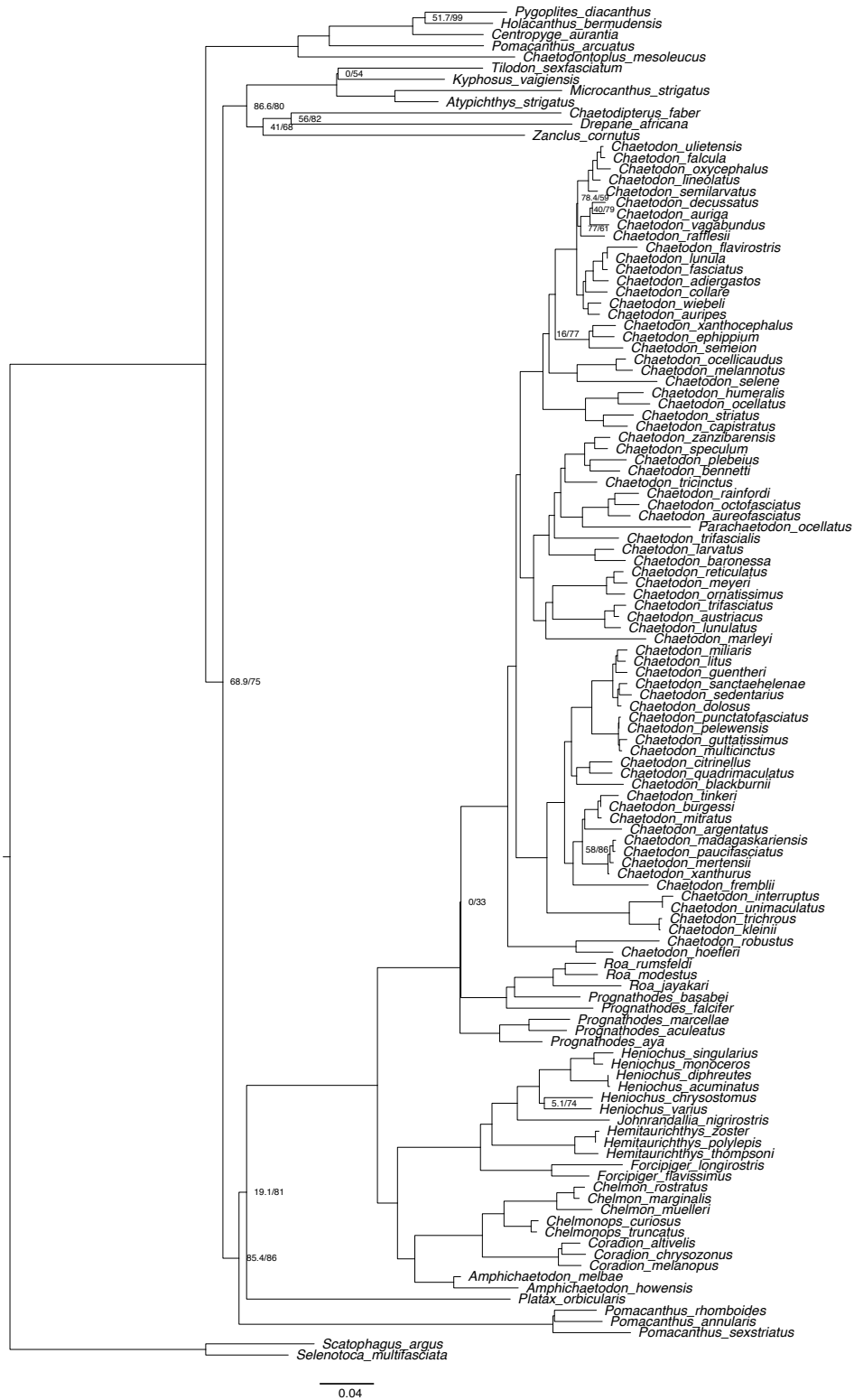


Figure S1: Maximum-Likelihood tree for Chaetodontidae. Numbers on nodes denote ultrafast bootstrap approximation (UFboot) values with 1000 replicates and the SH-like approximate likelihood ratio test (SH-aLRT) also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.

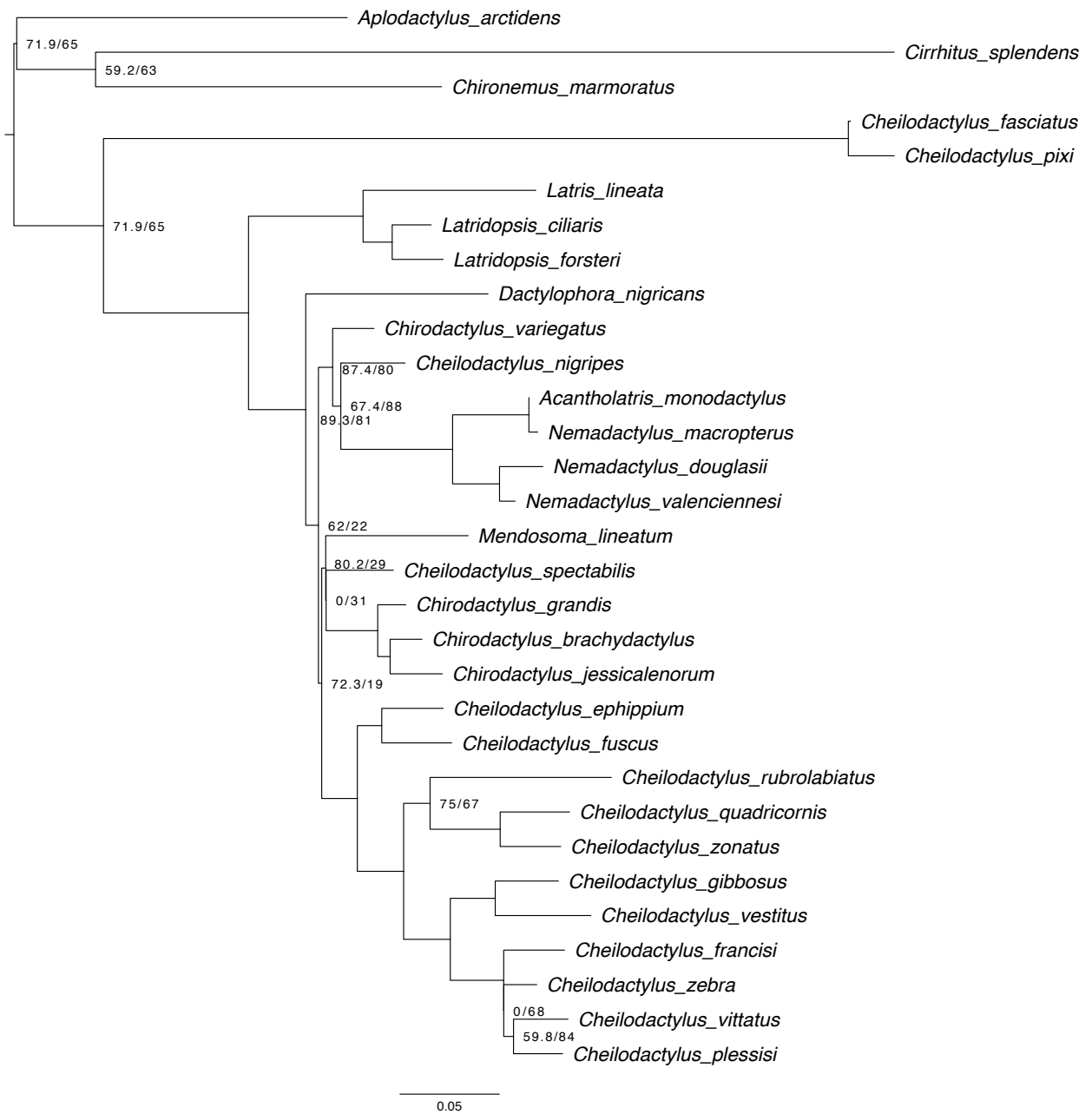


Figure S2: Maximum-Likelihood tree for Cheilodactylidae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.

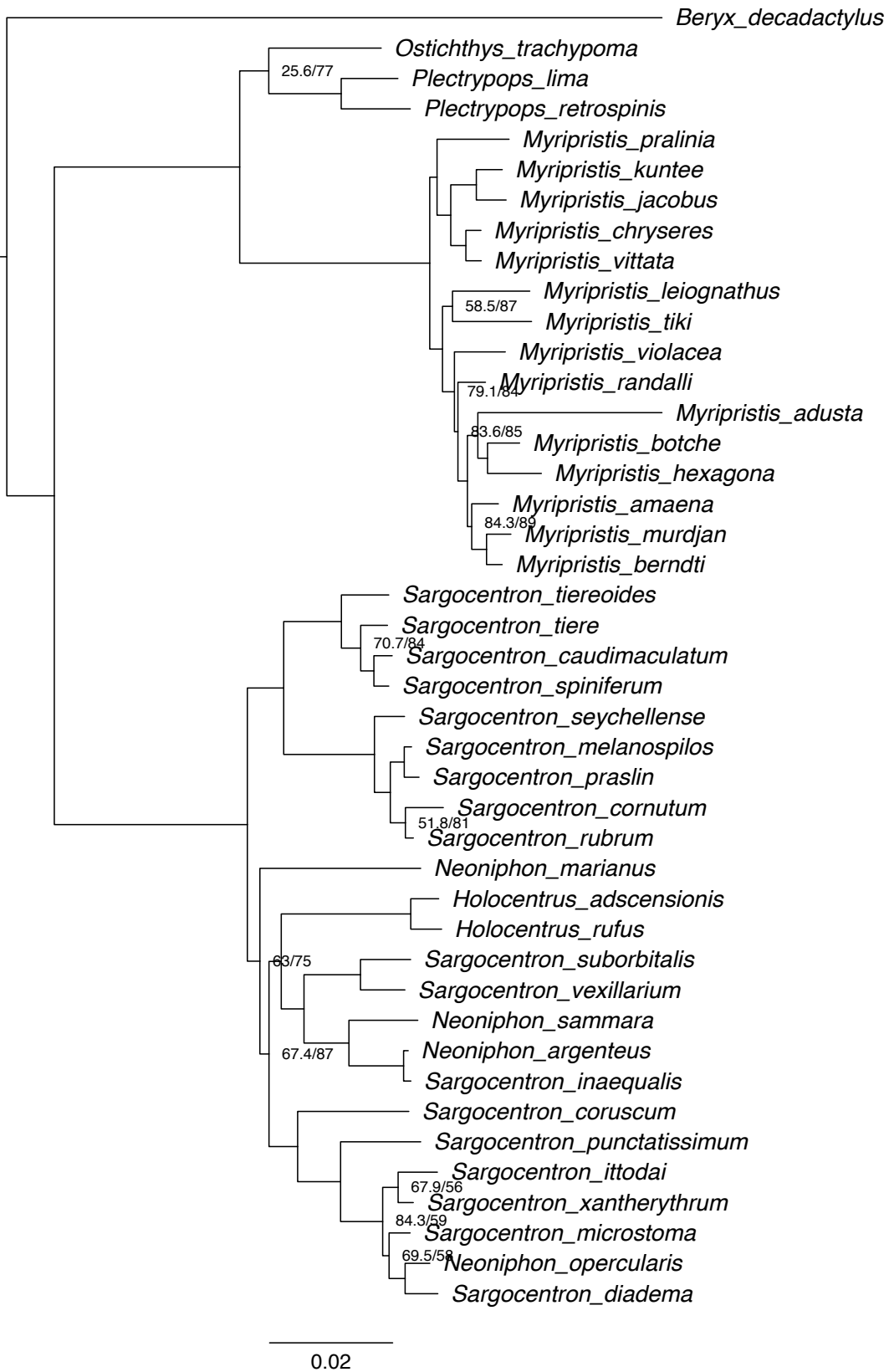


Figure S3: Maximum-Likelihood tree for Holocentridae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.

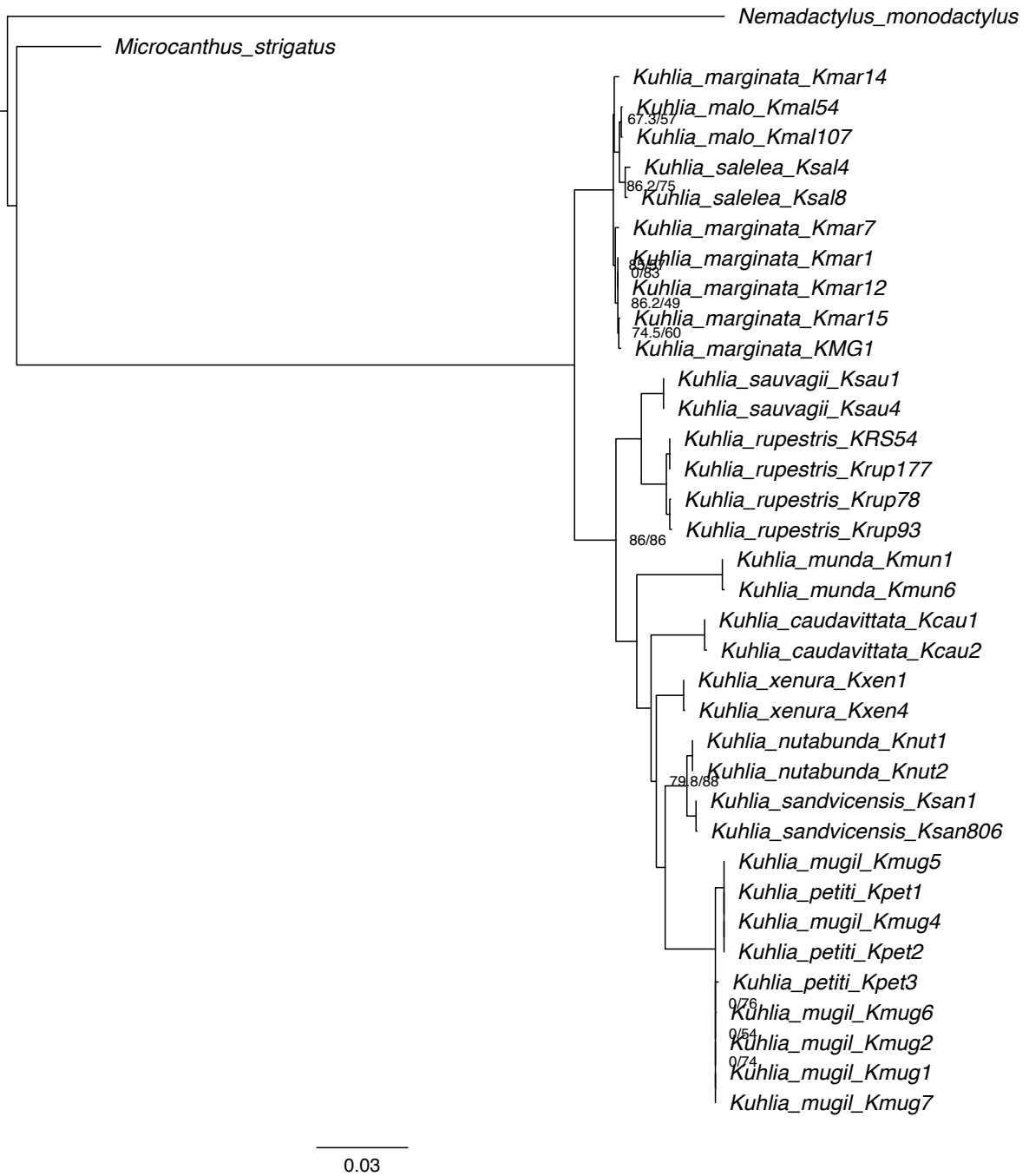
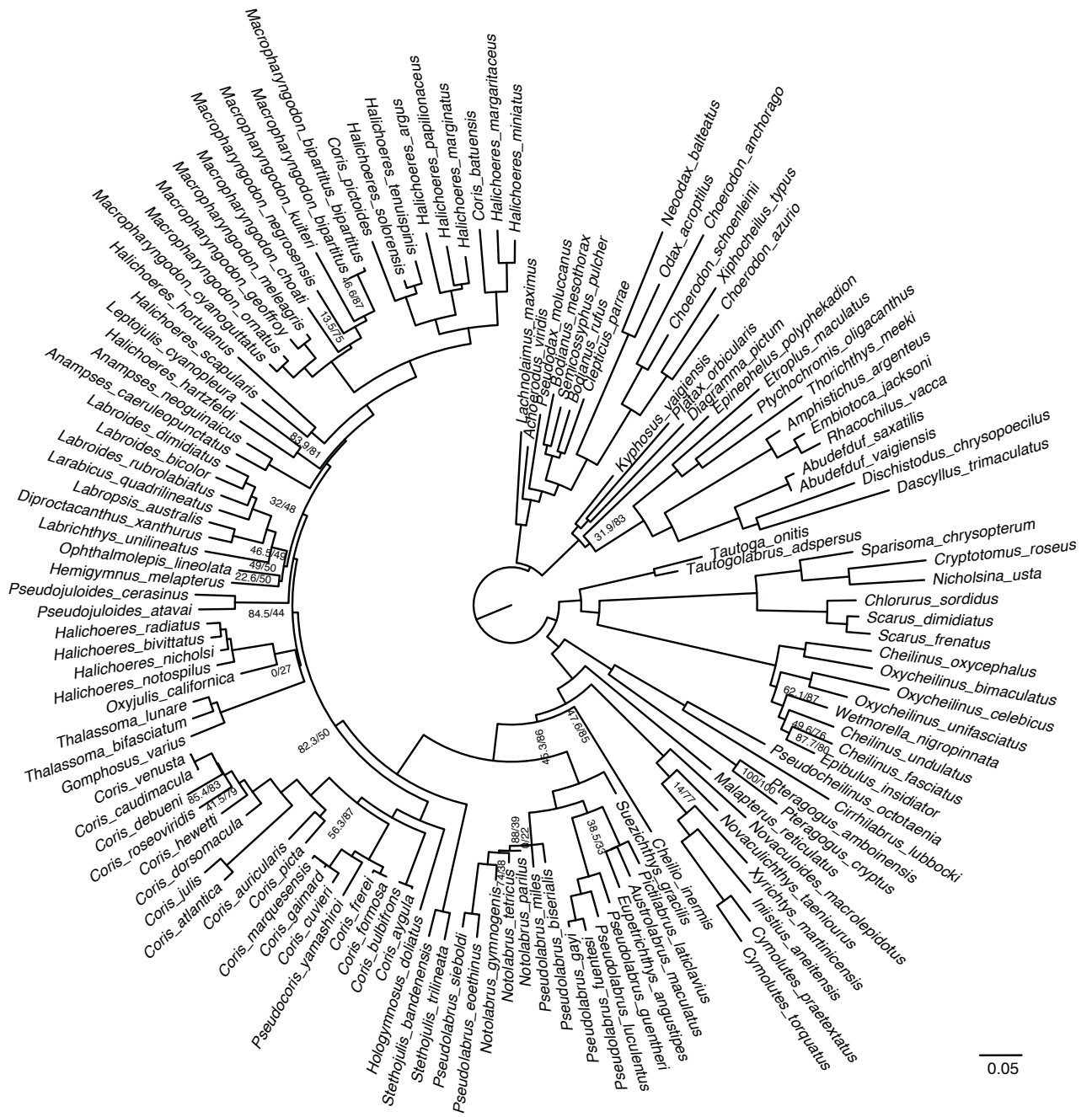


Figure S4: Maximum-Likelihood tree for Kuhliidae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.



0.05

Figure S5: Maximum-Likelihood tree for Labridae. Numbers on nodes denote UFboot values with 1000 replicates and the SH-aLRT also with 1000 bootstrap replicates. UFboot and SH-aLRT displayed only for nodes with values below 90.

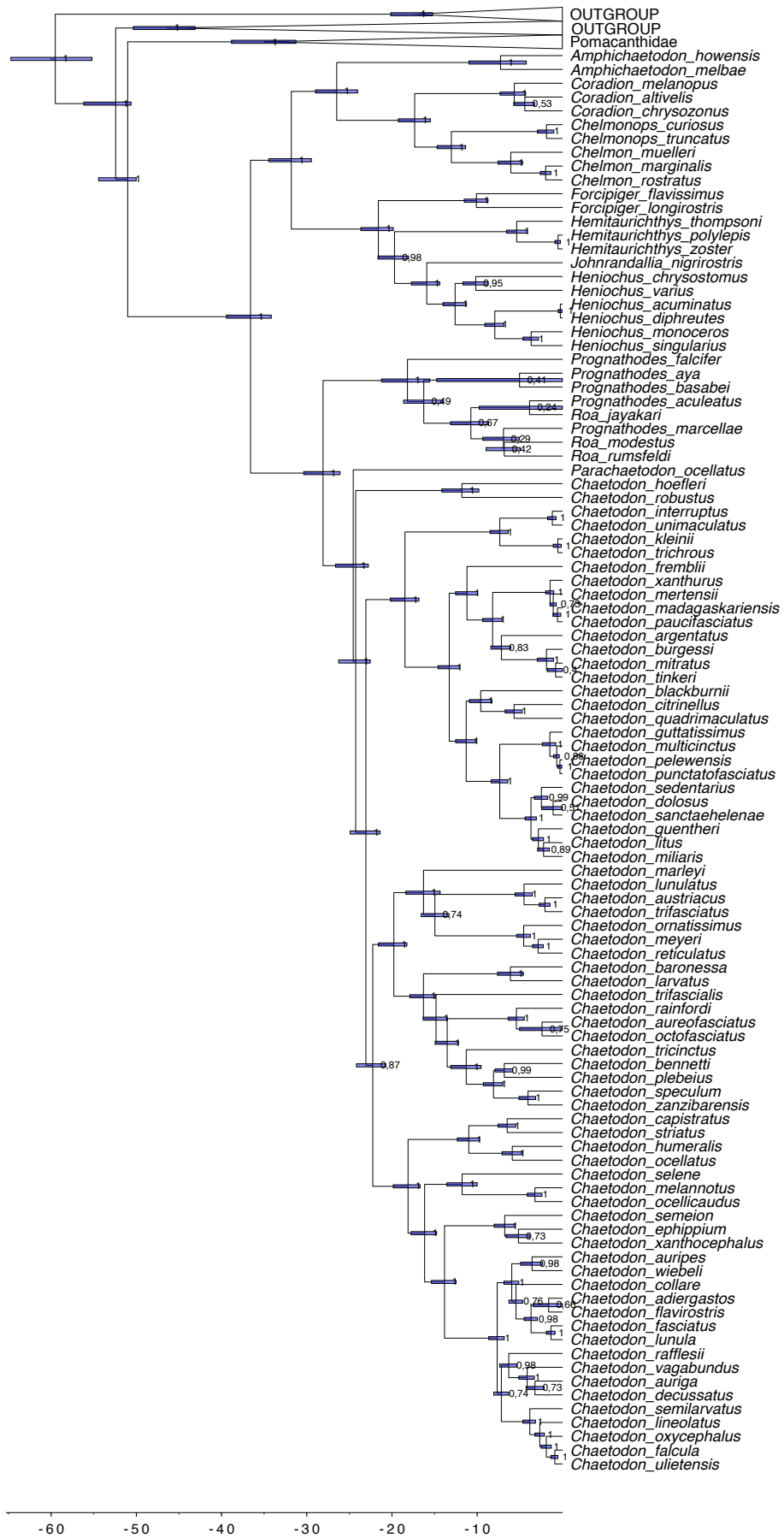


Figure S9: Timetree for the Chaetodontidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

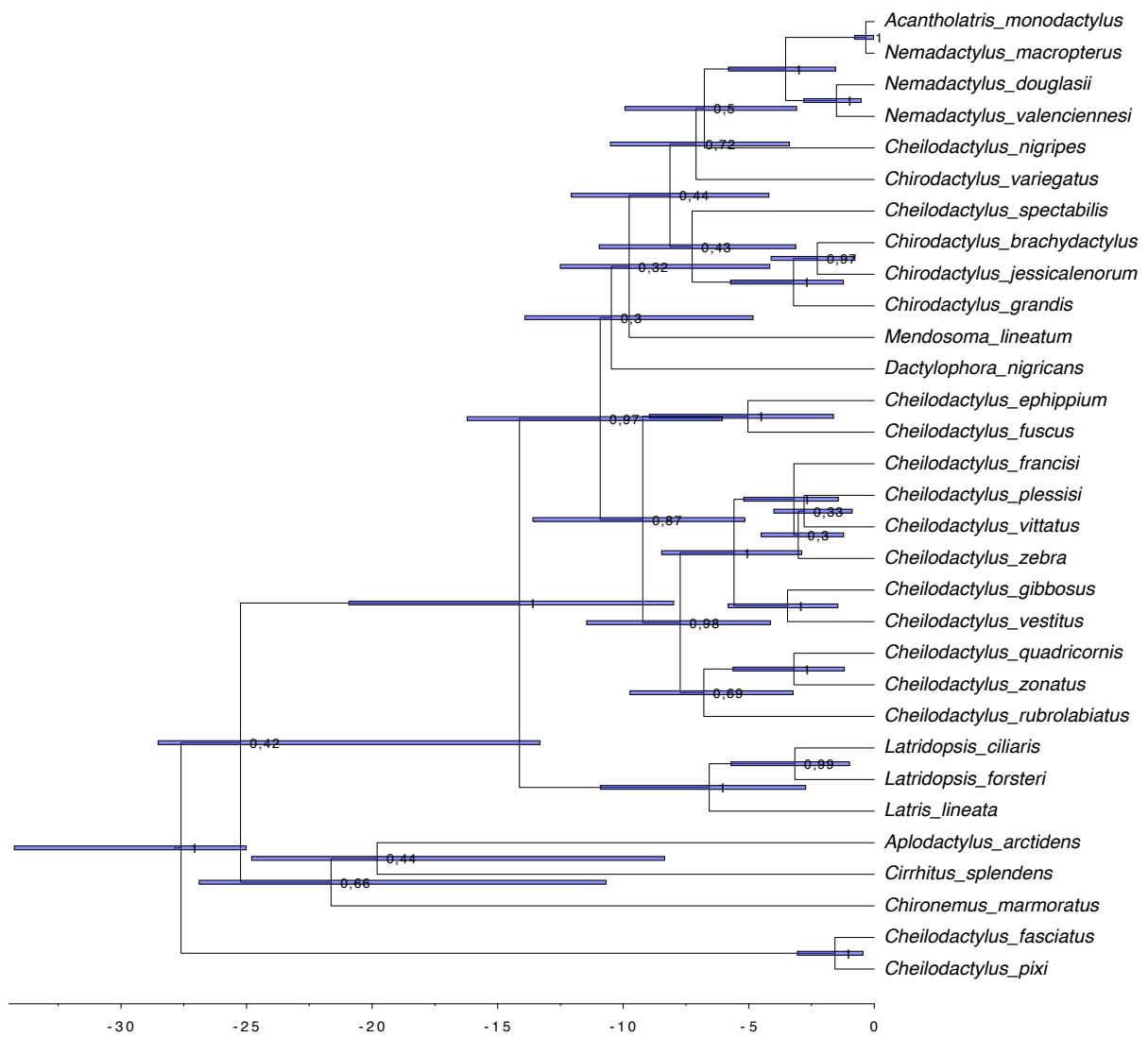


Figure S10: Timetree for the Cheilodactylidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

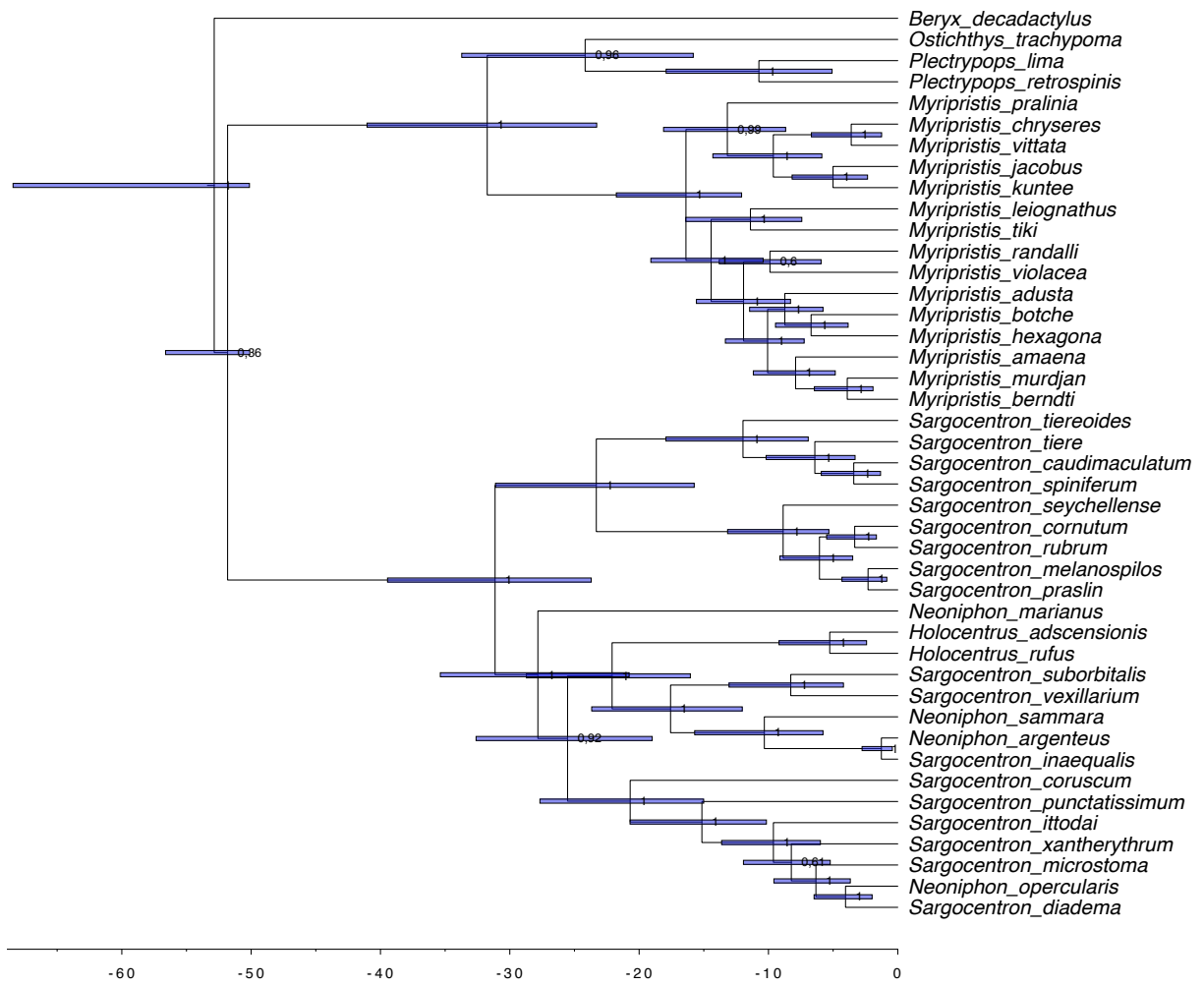


Figure S11: Timetree for the Holocentridae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

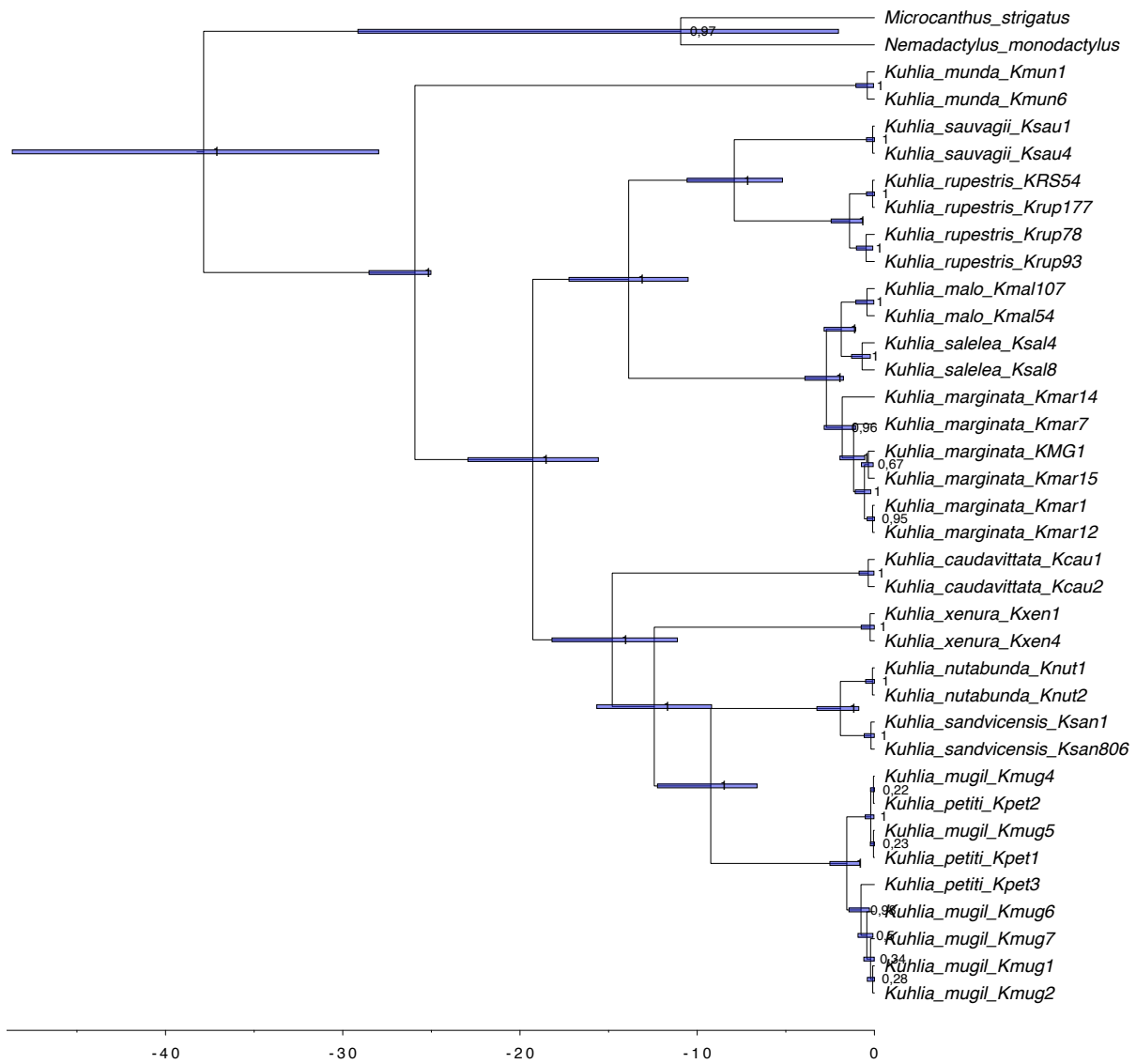


Figure S12: Timetree for the Kuhliidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

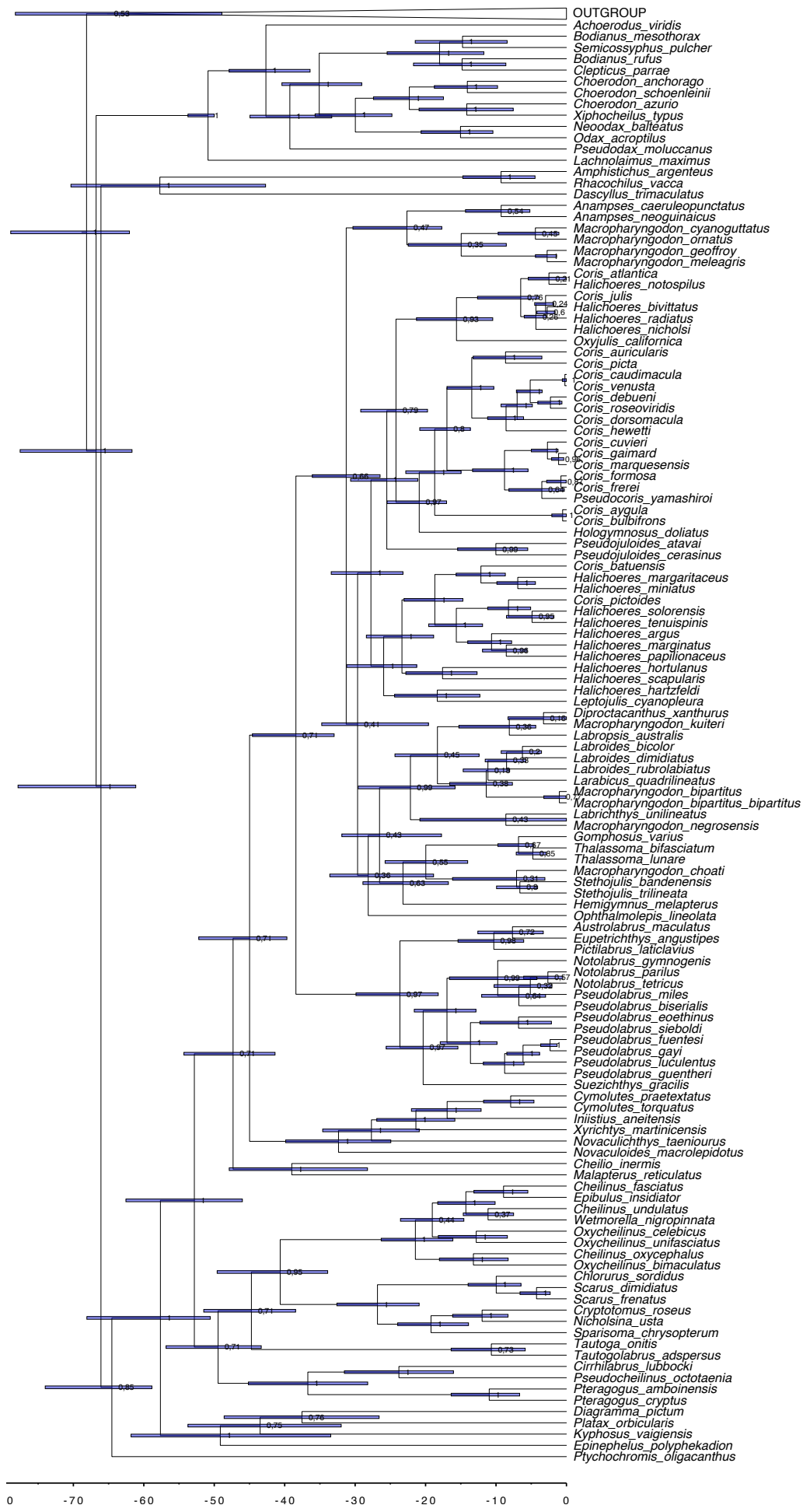


Figure S13: Timetree of of Labridae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

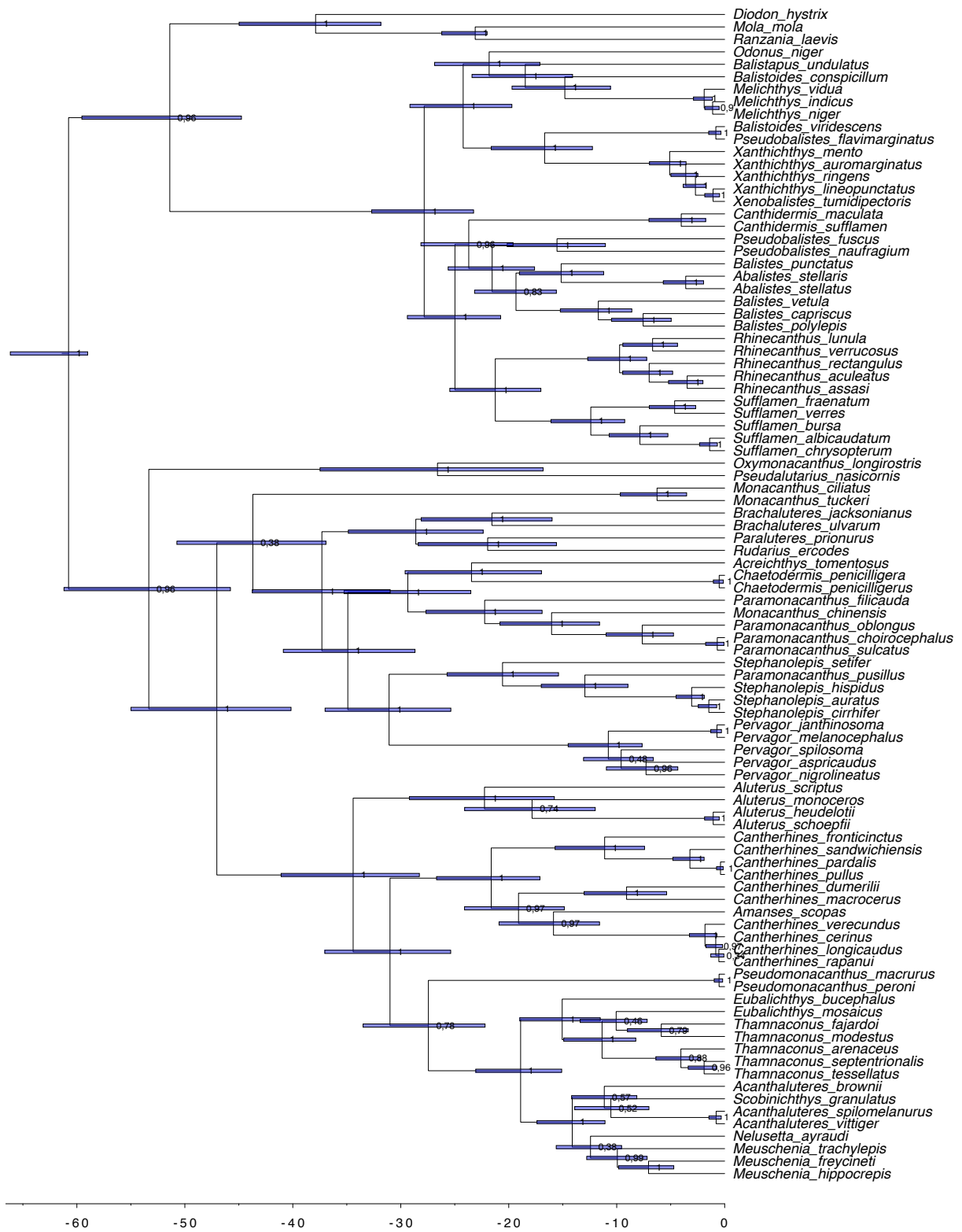


Figure S14: Timetree for the Monacanthidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

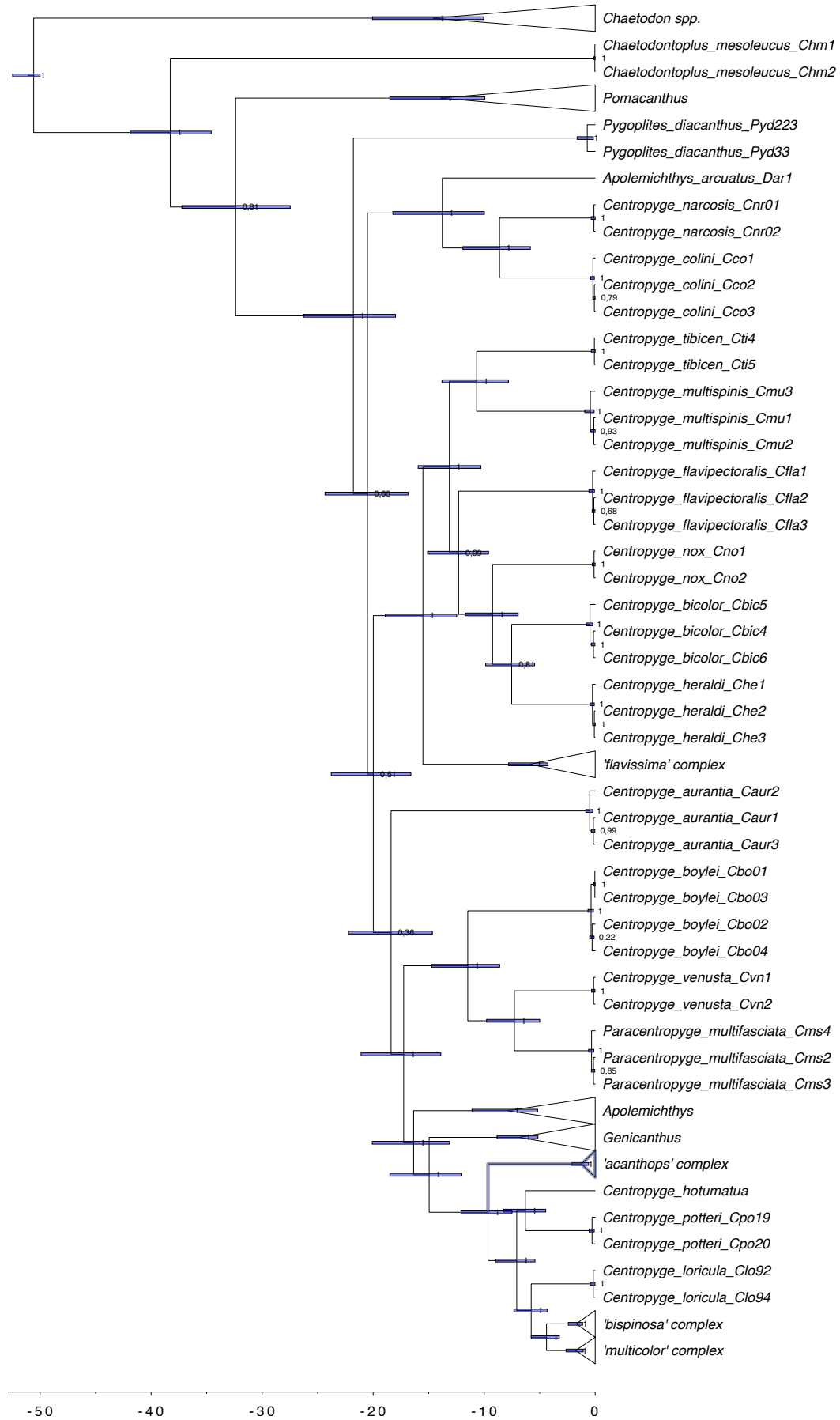


Figure S15: Timetree for the Pomacanthidae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

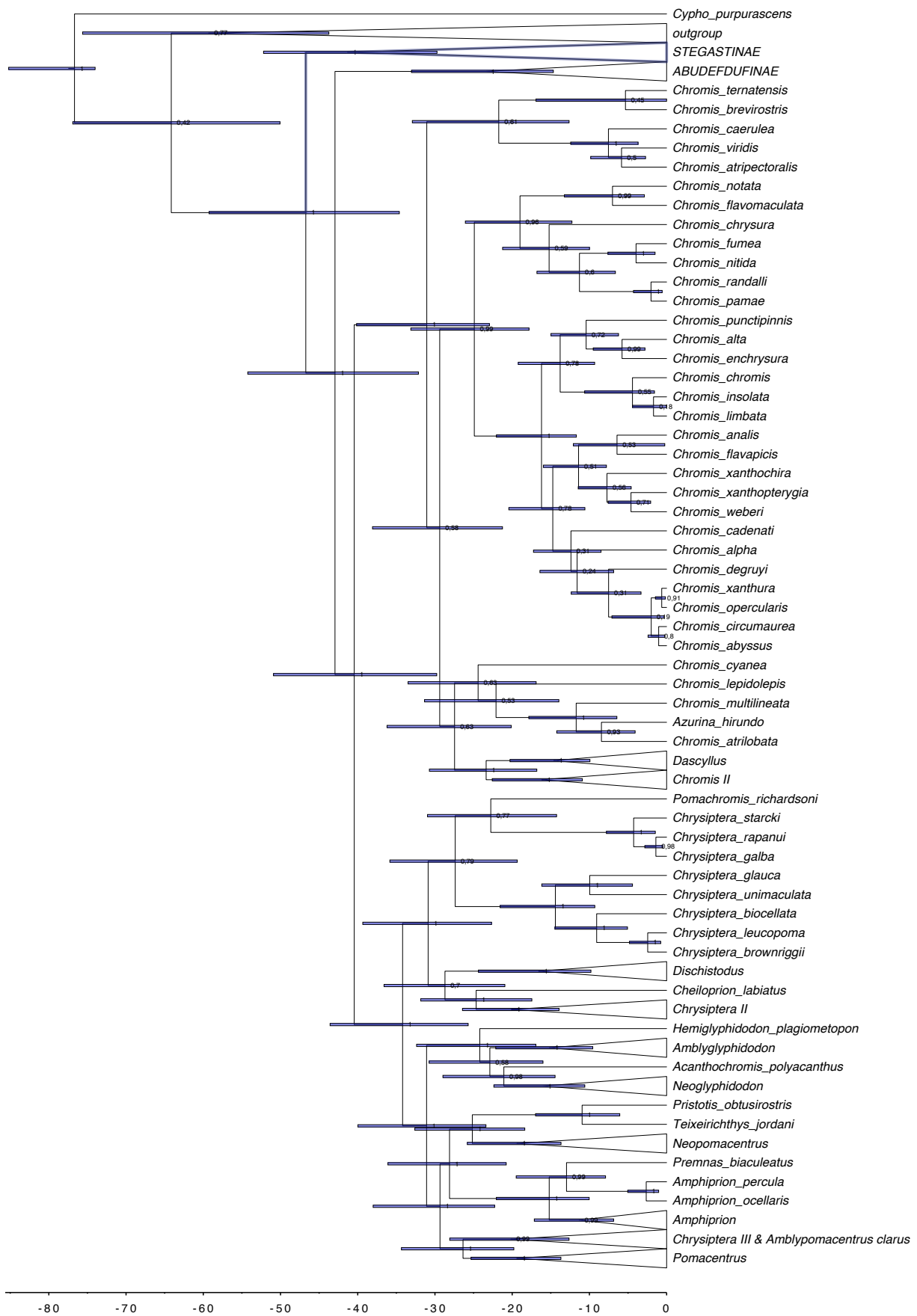


Figure S16: Timetree for the Pomacentridae based on the Bayesian analysis performed in BEAST2 V 2.4.6. Values next to nodes indicate PP support. Horizontal bars indicate 95% HPD.

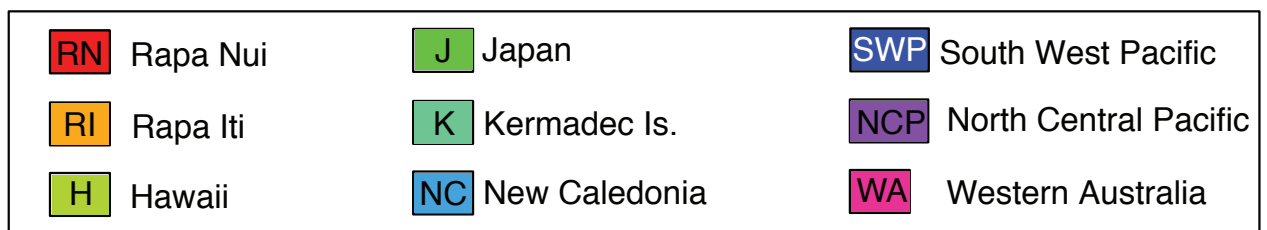
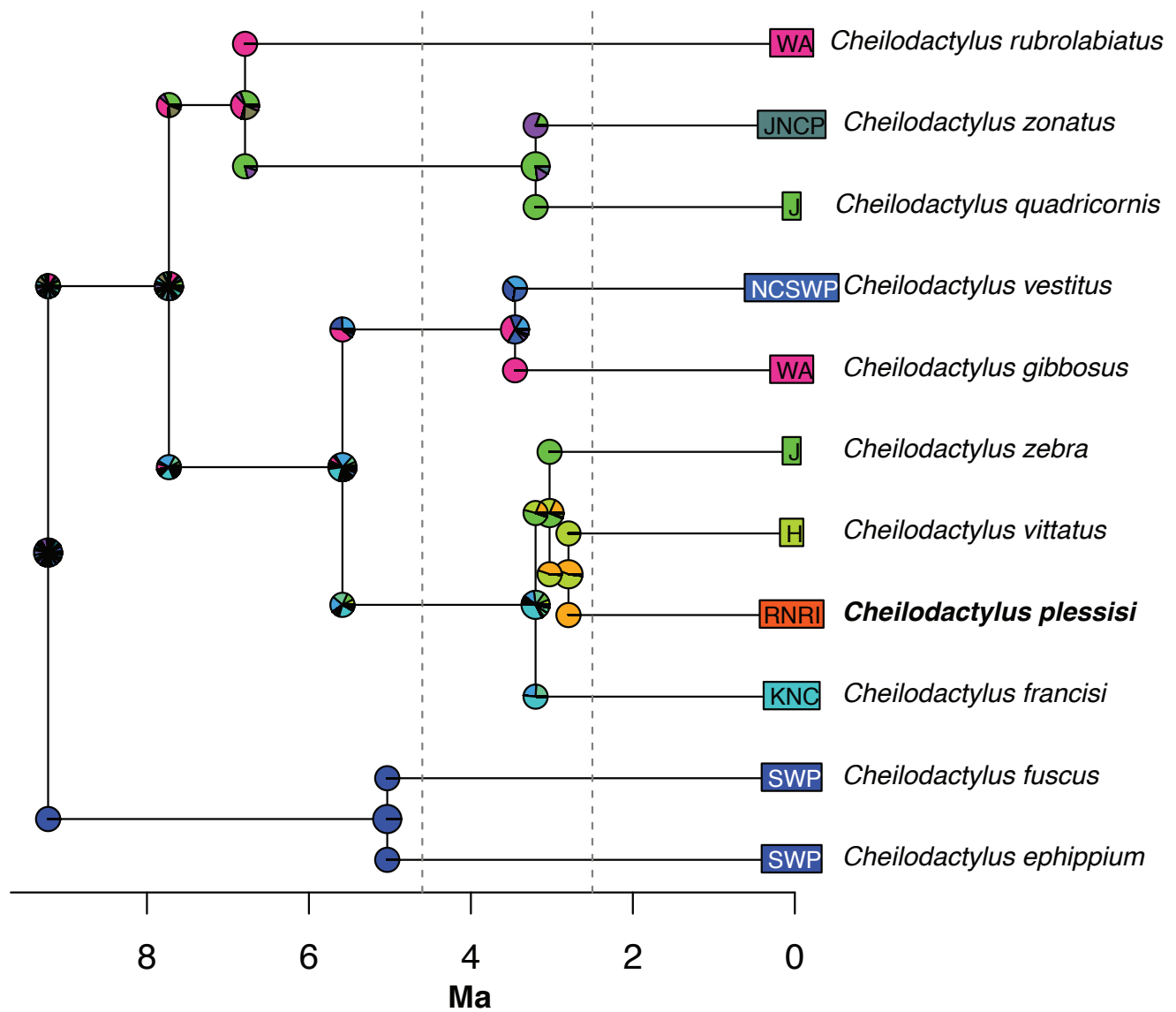


Figure S17: Estimates of ancestral ranges for *Cheilodactylus* spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.

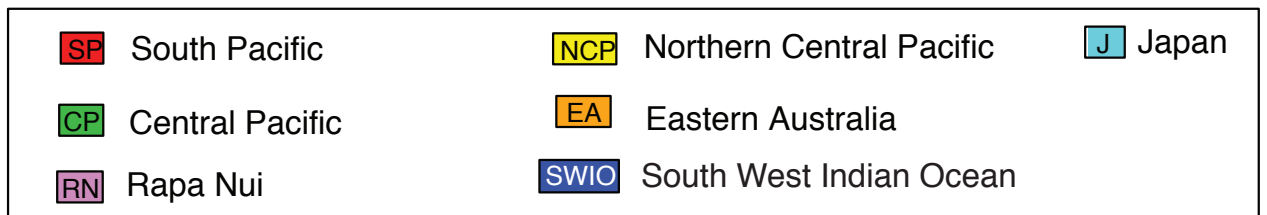
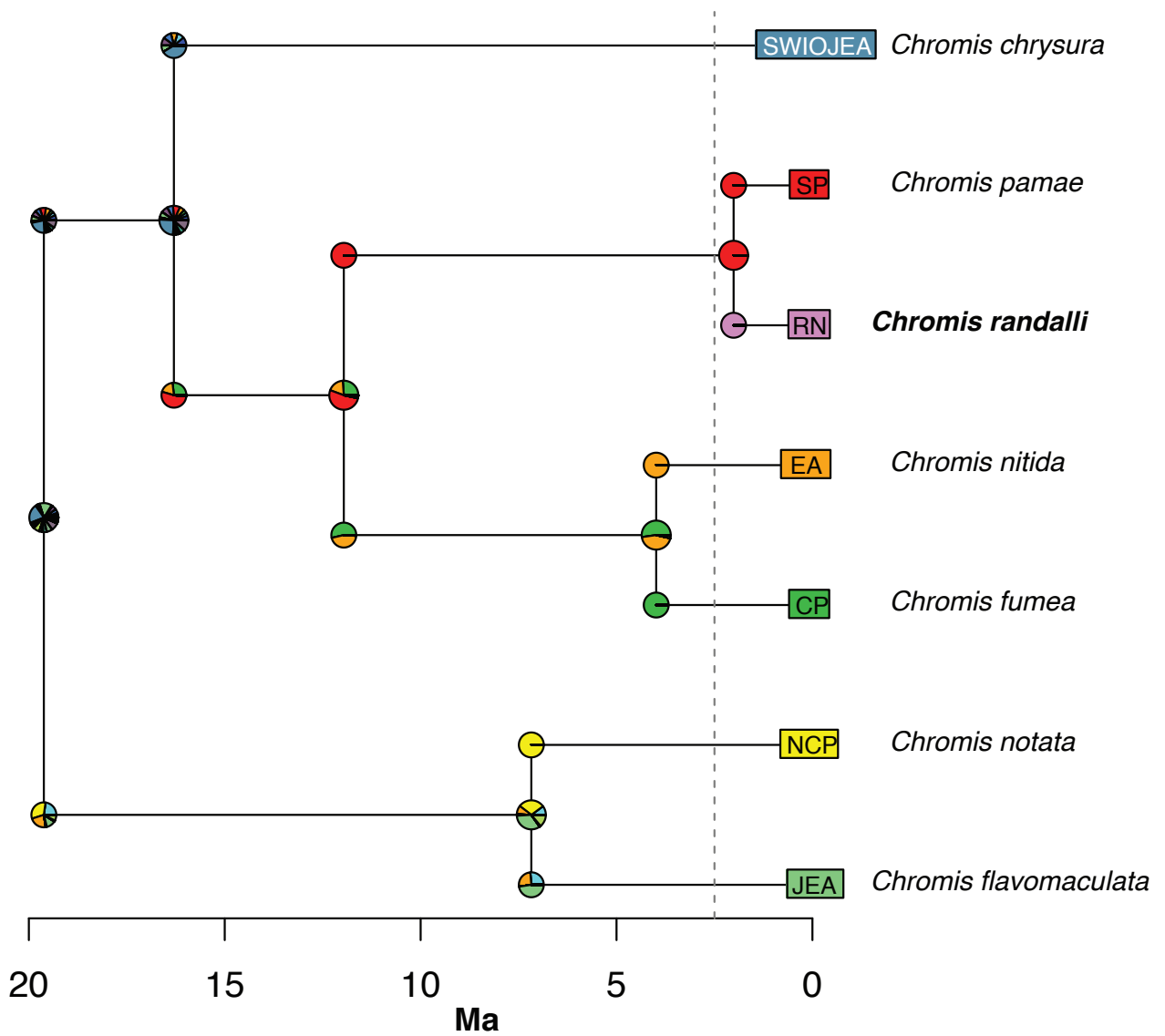


Figure S18: Estimates of ancestral ranges for *Chromis* spp using DEC + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.

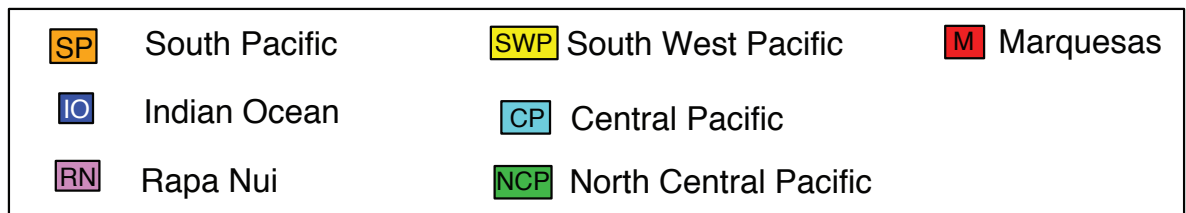
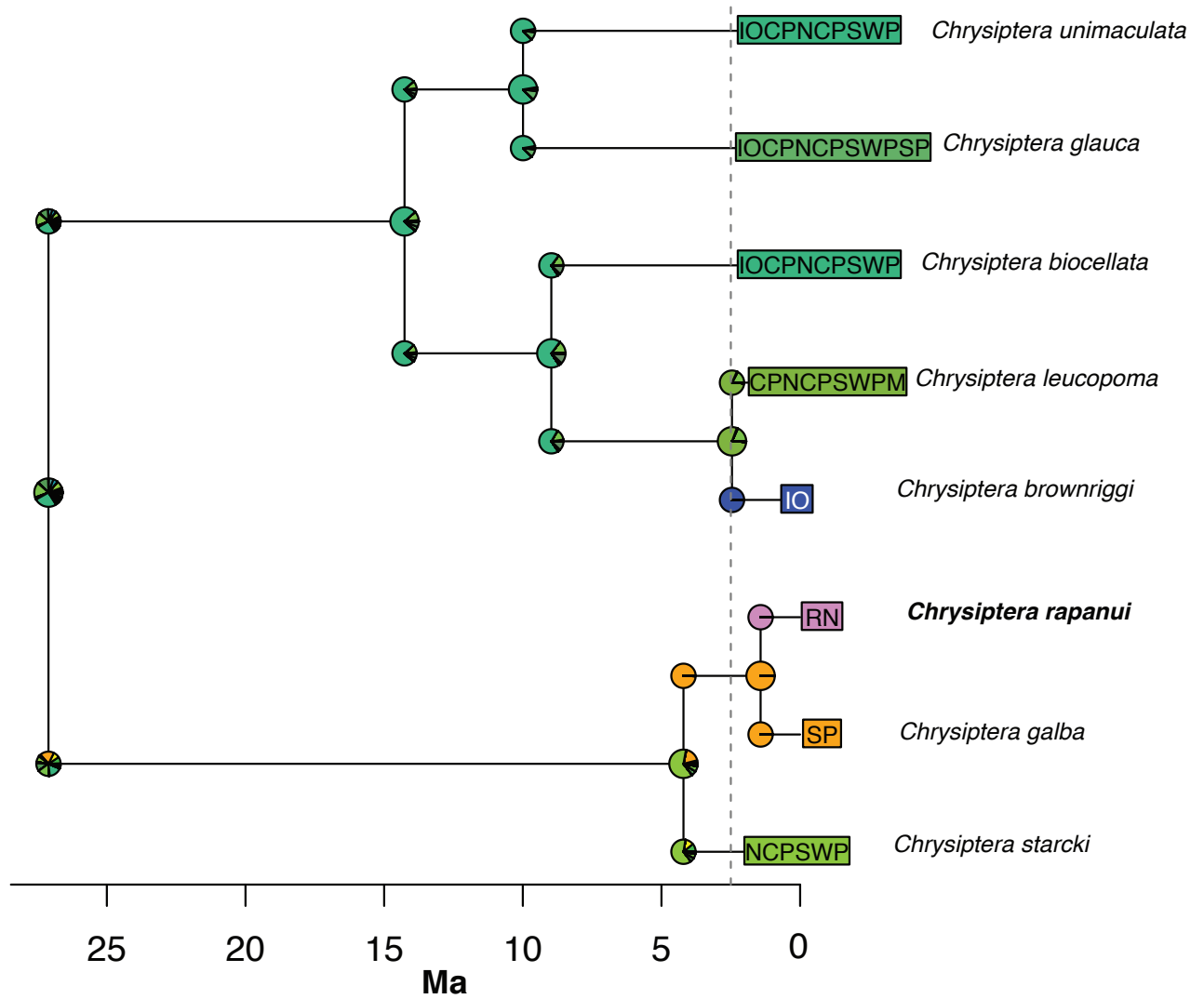


Figure S19: Estimates of ancestral ranges for *Chrysiptera* spp using BAYAEREA + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.

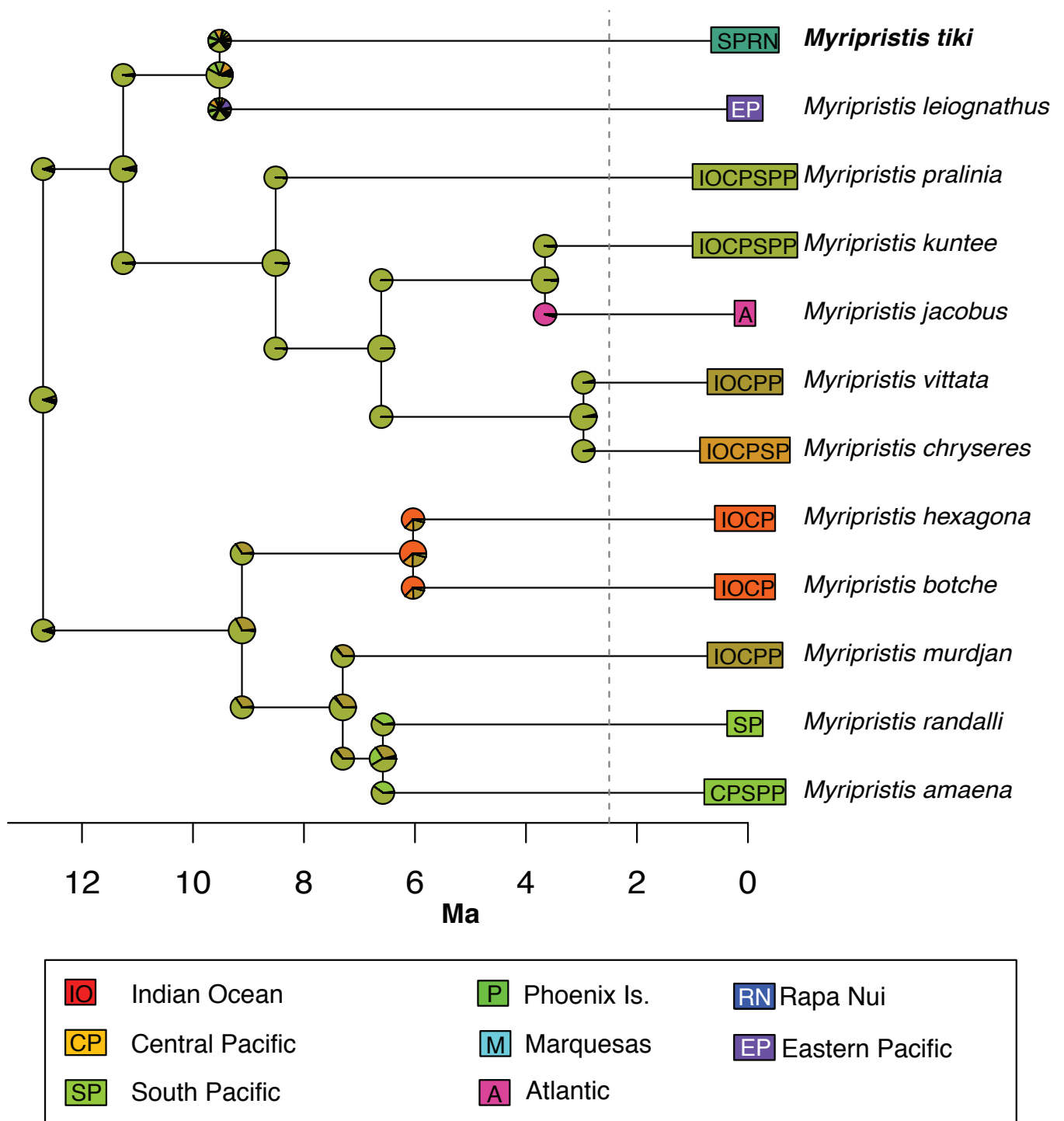


Figure S20: Estimates of ancestral ranges for *Myripristis* spp using BAYAREA + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.

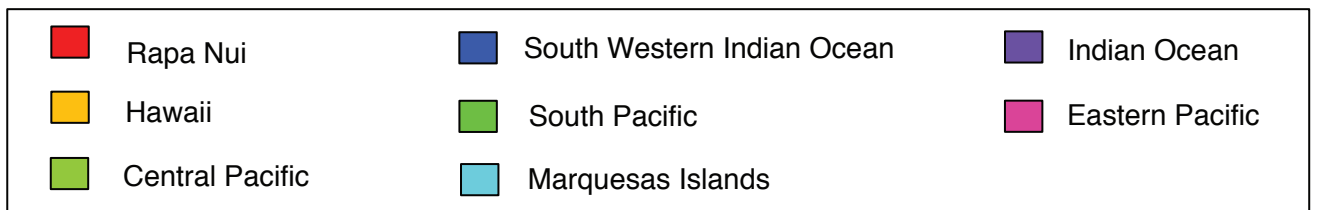
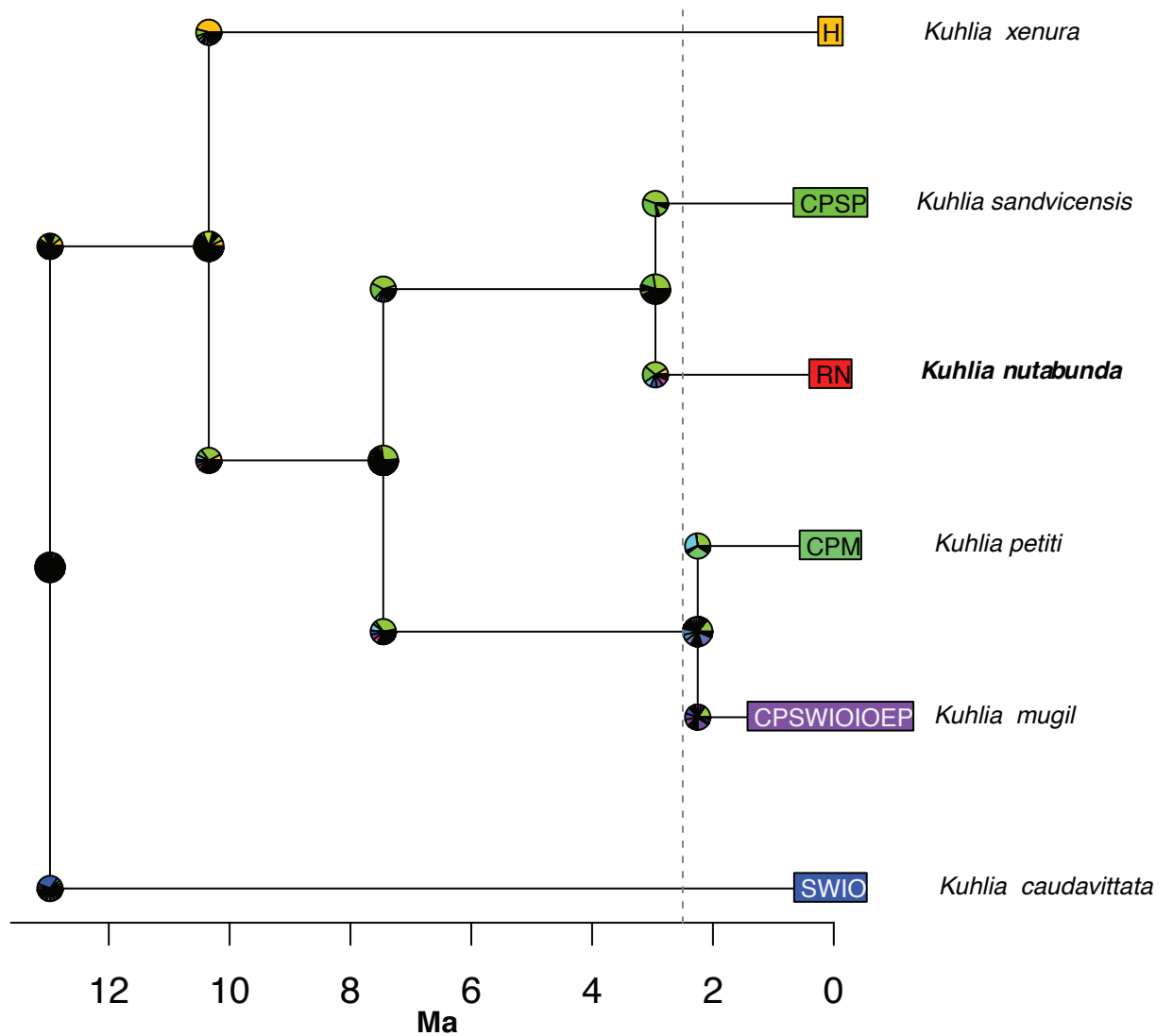
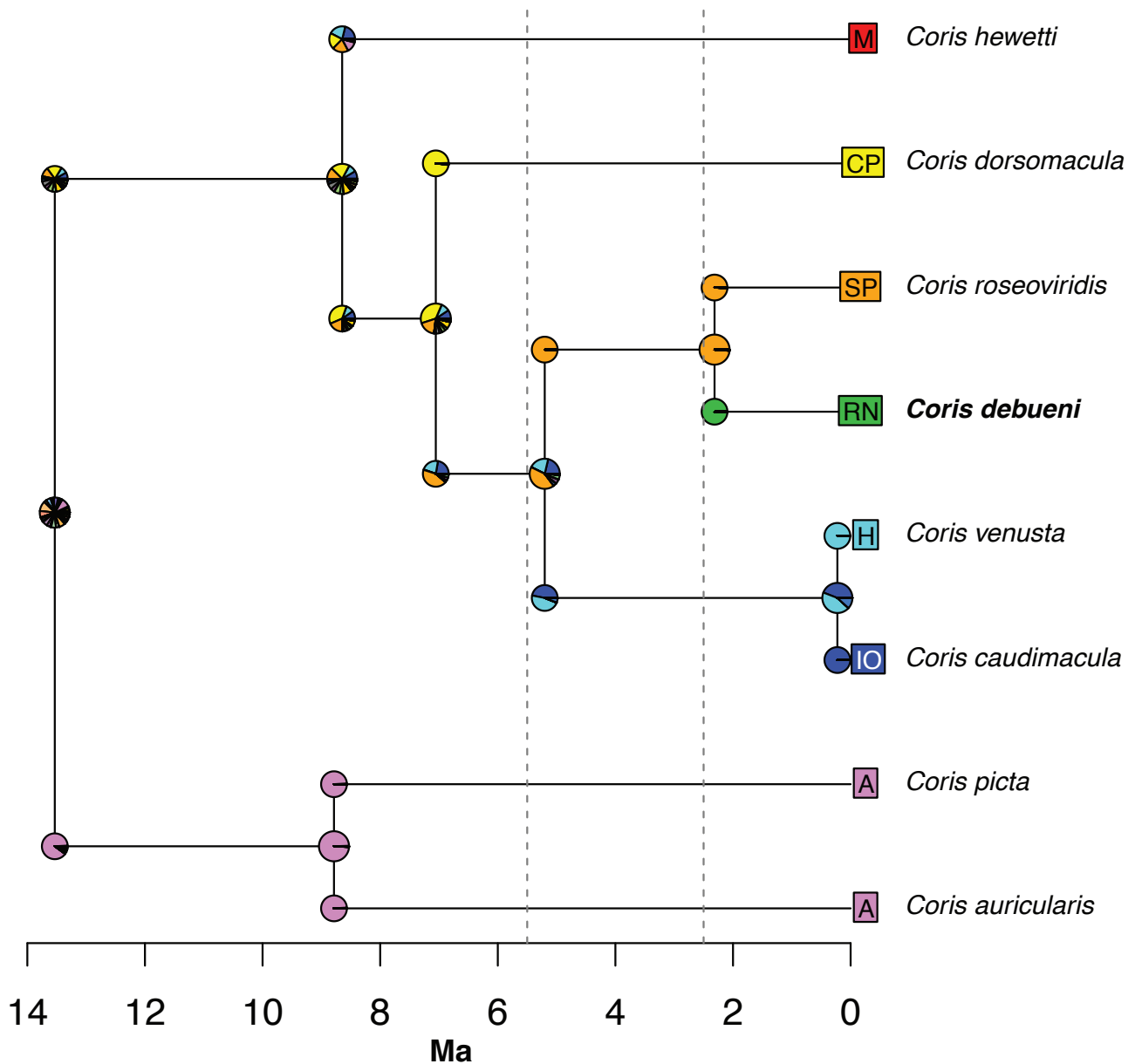


Figure S21: Estimates of ancestral ranges for *Kuhlia* spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.



M	Marquesas Is.	H	Hawai'ian Is.	IO	Indian Ocean
CP	Central Pacific	A	Atlantic Ocean		
SP	South Pacific	RN	Rapa Nui		

Figure S22: Estimates of ancestral ranges for *Coris* spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.

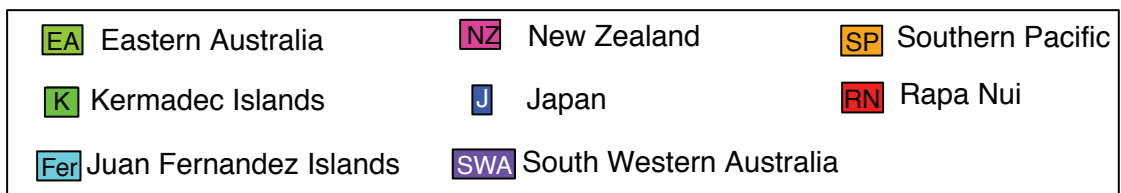
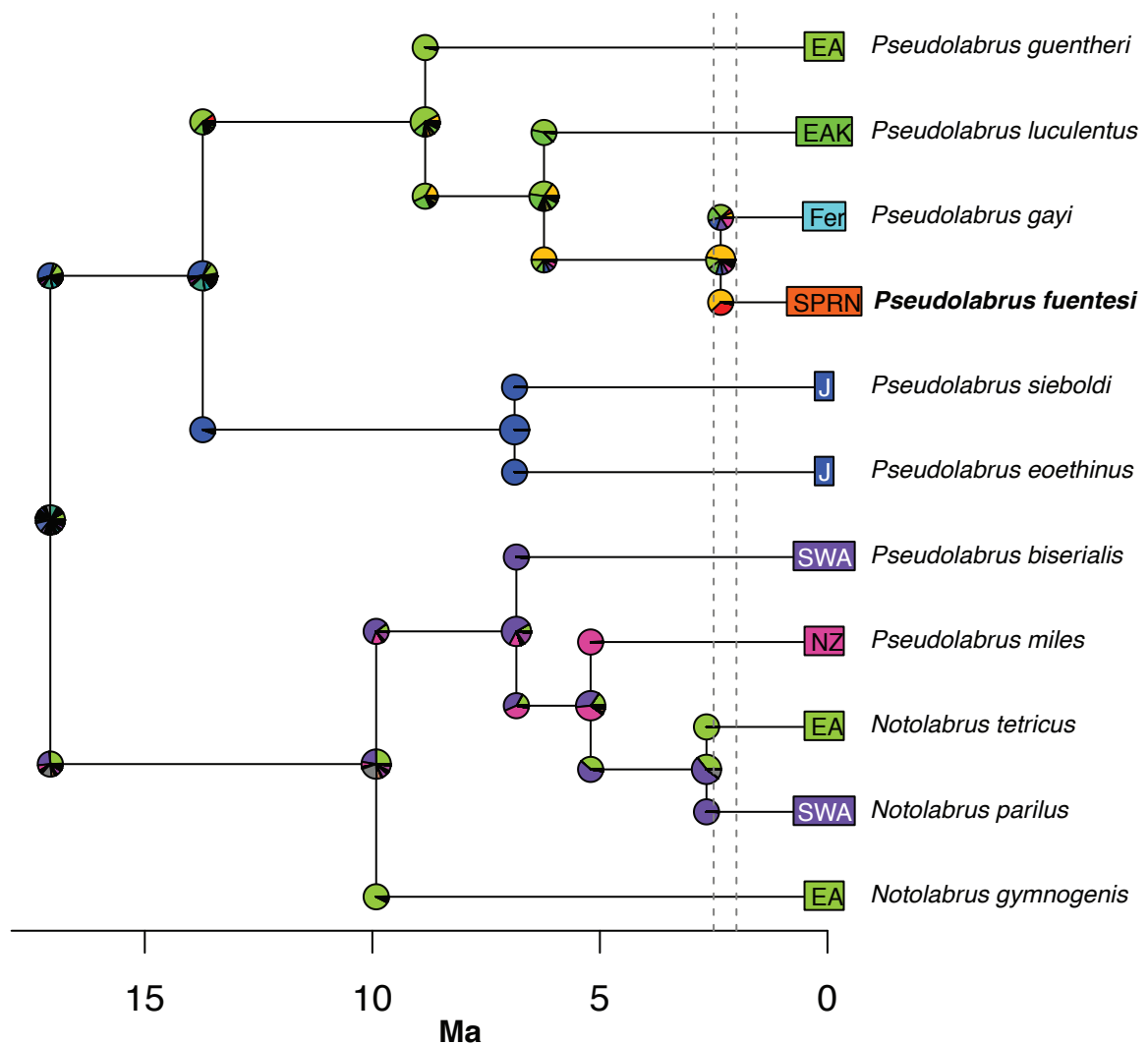


Figure S23: Estimates of ancestral ranges for *Pseudolabrus* spp using DIVALIKE model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.

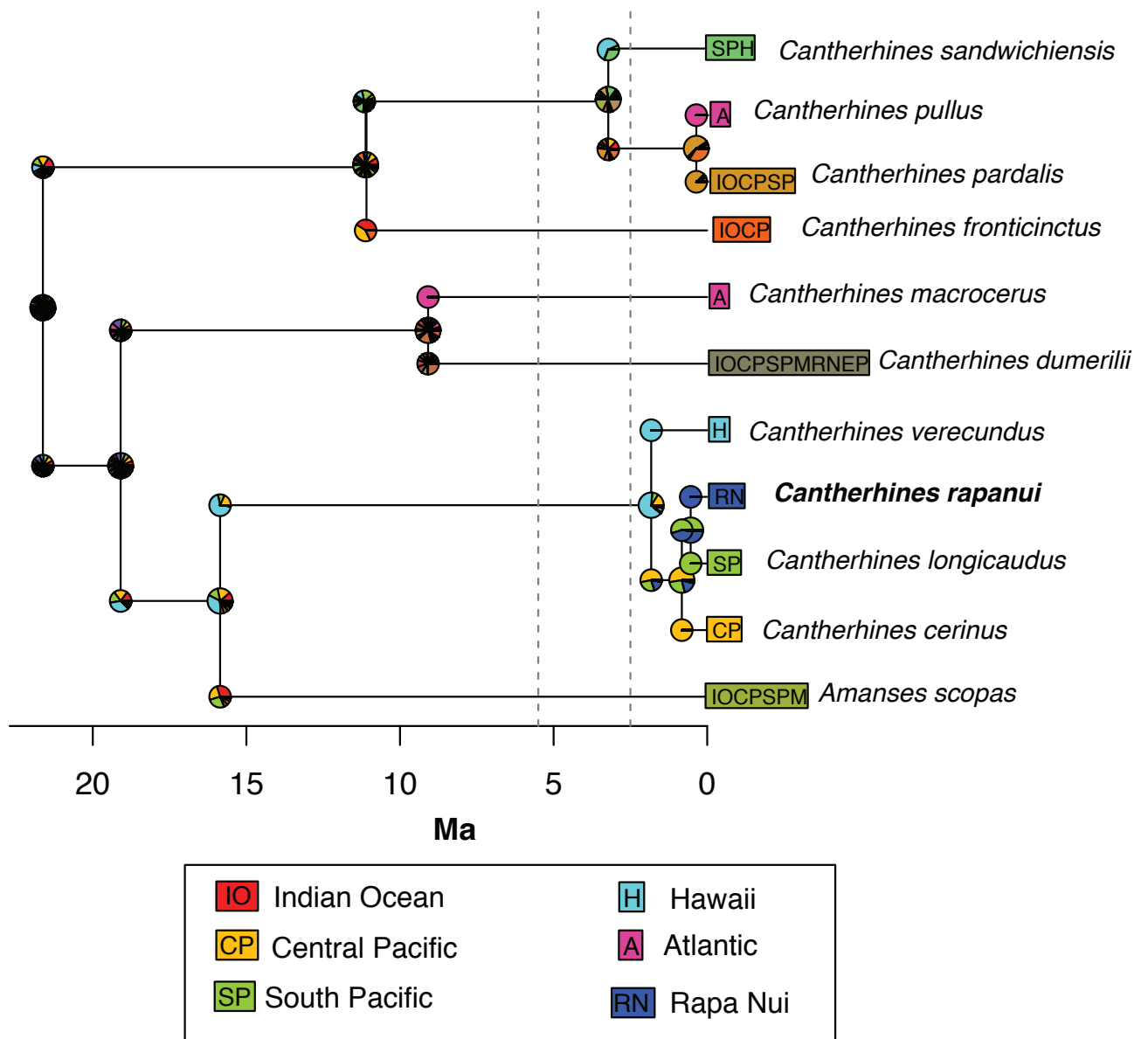


Figure S24: Estimates of ancestral ranges for *Cantherhines* spp using DIVALIKE + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.

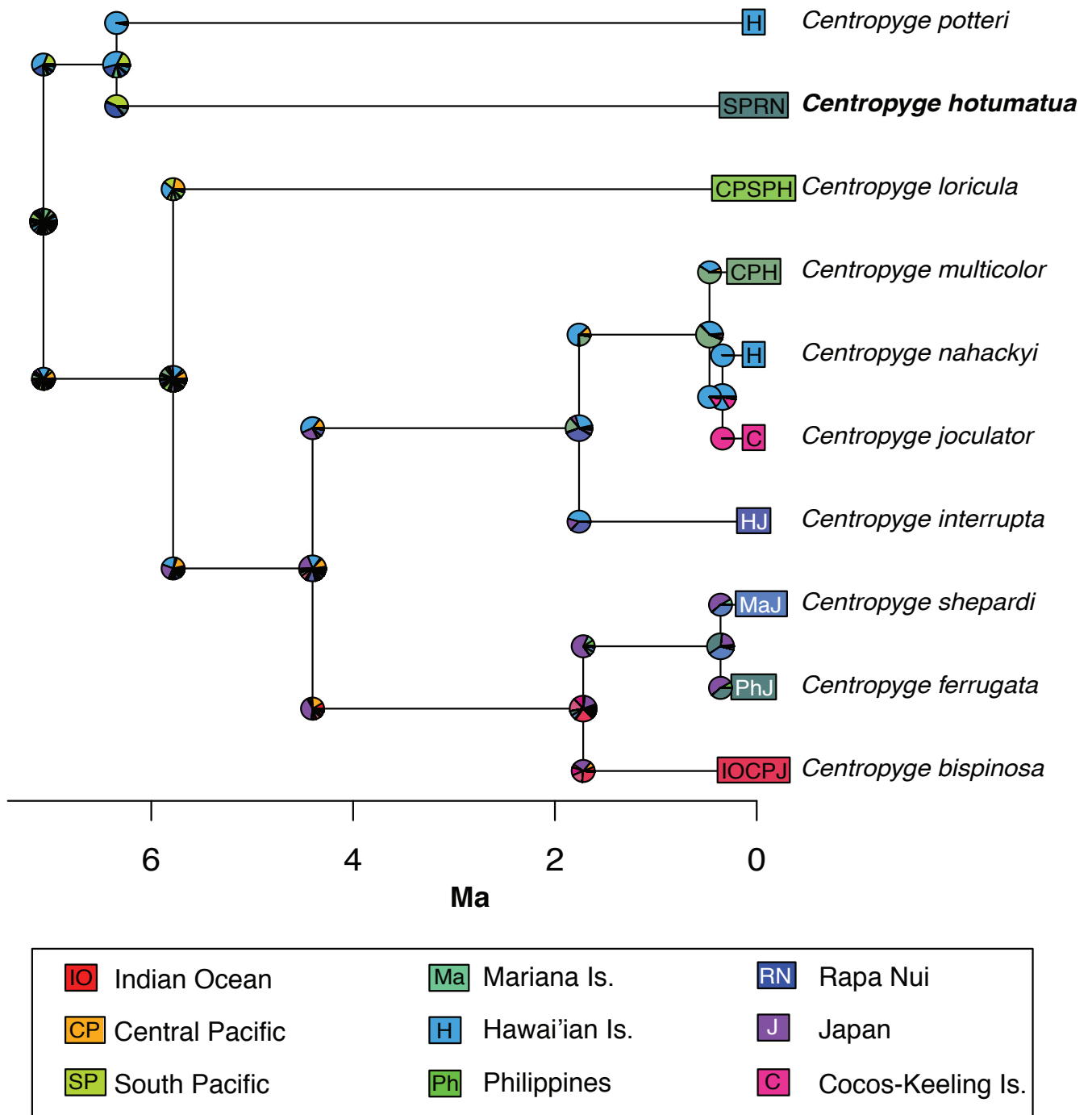


Figure S25: Estimates of ancestral ranges for *Centropyge* spp using DEC + J model. Pie charts at each node denote the probability of all possible ancestral states. The colors of widespread ranges are mixtures of the colors used for the single areas. High uncertainty is shown when multiple pies are present. In bold the species of interest. The time scale is calibrated in millions of years before present. Corners represent the states instantaneously after cladogenesis.